HARVARD
Journal on Legislation

Volume 44, Number 2 Summer 2007

Articles

Recent Developments

ARTICLE

REMEDYING ELECTION WRONGS

STEVEN F. HUEFNER*

Many matters of U.S. election administration have attracted significant popular, political, and scholarly attention in recent years. Largely slighted, however, has been the matter of how the various state election systems respond when an election outcome is unsettled or contested. Moreover, some recent electoral reforms, such as widespread provisional balloting and increased use of no-fault absentee voting, actually may increase the frequency with which contested elections occur. This Article explores the complex issues that arise in remedying a failed election, and urges states to refine and clarify their remedial standards and procedures for resolving an election dispute.

One unmistakable impact of the incredibly close 2000 presidential race and the dramatic litigation over its outcome is that the American public now pays substantially more attention to how states conduct their elections. Much of this attention has focused—properly—on adopting reforms to avoid the kinds of problems that famously plagued Florida in 2000, whether in matters of ballot design, voting technologies, or recount procedures.[1] As a result, most states have strengthened their voting processes in a number of important ways to reduce the risks of election difficulties.[2] Meanwhile, Congress has enacted the Help America Vote Act ("HAVA"),[3] which encourages states to update their voting systems, standardize their voting registration requirements, and otherwise improve their election processes.[4] Addi-

* Associate Professor of Law, Michael E. Moritz College of Law, The Ohio State University, and Senior Fellow, Election Law @ Moritz. J.D., Columbia Law School, 1991; A.B., Harvard University, 1986. The author is grateful for the suggestions of Ruth Colker, Terri Enns, Ned Foley, Rick Hasen, and Dan Tokaji, and for the research assistance of Aaron Applebaum, Emmy Ashmus, Andrew Brasse, Damien Kitte, Joshua Moser, Henry Phillips-Gary, Eric White, and Stephen Wolfson.

[1] Summaries and analyses of the 2000 election are voluminous. *See generally, e.g.*, RICHARD A. POSNER, BREAKING THE DEADLOCK: THE 2000 ELECTION, THE CONSTITUTION, AND THE COURTS (2001); THE VOTE: BUSH, GORE, AND THE SUPREME COURT (Cass R. Sunstein & Richard A. Epstein eds., 2001); Symposium, *Recounting Election 2000*, 13 STAN. L. & POL'Y REV. 1 (2002); Gillian Peele, *The Legacy of Bush v. Gore*, 1 ELECTION L.J. 263 (2002).

[2] For a state-by-state list of recent election reforms, see ELECTIONLINE.ORG, ELECTION REFORM: WHAT'S CHANGED, WHAT HASN'T AND WHY 2000–2006, at 39–72 (2006), *available at* http://www.electionline.org/Portals/1/Publications/2006.annual.report.Final.pdf.

[3] Pub. L. No. 107-252, 116 Stat. 1666 (codified in scattered sections of 42 U.S.C.).

[4] *See, e.g.*, 42 U.S.C. § 15481 (Supp. 2006) (voting systems requirements); 42 U.S.C. § 15302 (Supp. 2006) (replacement of punch card or lever voting machines); 42 U.S.C. § 15483 (Supp. 2006) (computerized statewide voter registration list requirements and requirements for voters who register by mail); 42 U.S.C. § 15482 (Supp. 2006) (provisional voting and voting information requirements).

tional congressional action in this area has been proposed,[5] and many state-level reforms will continue to occur.

Largely absent, however, has been meaningful discussion of how to improve the way that state election systems respond when elections still go awry, as they inevitably will.[6] That is, notwithstanding the array of recent reforms that should help to reduce the frequency of election miscues, post-election controversies nonetheless are sure to recur. For instance, the bare fact of an incredibly close election, such as the 2004 Washington gubernatorial election,[7] can highlight the imperfections that exist in most election processes and potentially create an election crisis, even without the presence of more systematic failures. Meanwhile, systematic failures also are likely to continue to occur on occasion, whether in the form of the nonuniform standards employed in Florida for counting punch cards in the 2000 election,[8] the incredibly long lines in some Ohio precincts in 2004,[9] or the apparent flaw in the ballot layout for the thirteenth congressional district in Sarasota County, Florida, in 2006.[10] Any number of other possible scenarios may similarly create a pressing need for an answer to the question of what to do when an election goes wrong or its outcome is for some reason in doubt.[11]

Furthermore, some recent election reforms actually may have the unintended result of increasing the number of occasions when an election outcome can be meaningfully contested. For instance, increased use of both absentee and provisional ballots[12] may have the effect of increasing the "margin of litigation."[13] Although in any particular election a number of

---

[5] Some proposals in the 109th Congress included H.R. 5913, 109th Cong. (2d Sess. 2006) (proposing to require government photo identification in order to vote); H.R. 3058, 109th Cong. § 83 (1st Sess. 2005) (proposing to fund the Election Assistance Commission to carry out election reform portions of HAVA); H.R. 550, 109th Cong. (1st Sess. 2005) (proposing to require paper audit trails for electronic voting equipment); S. 17, 109th Cong. (1st Sess. 2005) (proposing to amend HAVA in several respects); S. Con. Res. 53, 109th Cong. (1st Sess. 2005) (proposing to express the "sense of Congress that any effort to impose photo ID requirements for voting should be rejected").

[6] *See infra* notes 136–37 and accompanying text.

[7] *See* SAM REED, WASH. SEC'Y OF STATE, 2004 GOVERNOR'S RACE (2004), http://www.secstate.wa.gov/elections/2004gov_race.aspx (providing complete election results from the Washington gubernatorial race); *see also* Symposium, *Where's My Vote? Lessons Learned from Washington State's Gubernatorial Election*, 29 SEATTLE U. L. REV. 313 (2005).

[8] *See supra* note 1 and accompanying text.

[9] *See Waiting Was the Hardest Part*, COLUMBUS DISPATCH, Nov. 3, 2004, at 1A.

[10] *See* Lloyd Dunkelberger, *Computer Code at Heart of Congressional Election Dispute*, GAINESVILLE SUN (Fla.), Dec. 21, 2006, *available at* http://www.gainesville.com/apps/pbcs.dll/article?AID=/20061221/LOCAL/612210369/-1/today (describing several proffered explanations for abnormally high rate of undervoting in Sarasota County); *infra* note 46 and accompanying text.

[11] *See infra* Part I.A.

[12] Provisional ballots are ballots offered to voters who cannot satisfy poll workers that they are eligible to vote. Issues concerning the eligibility of these voters can then be resolved after the election, and all provisional ballots determined to have been cast by eligible voters then can be processed and included in the vote totals.

[13] *See* Daniel Tokaji, *Are Election Reforms Increasing the Margin of Litigation?*, ELEC-

other factors will affect how large or small the margin of litigation is, there is merit to the claim that some recent election reforms, by expanding the number of votes that may remain at issue after election day, may actually have the unintended effect of increasing the likelihood of contested elections.[14] For example, some election reforms expanding the use of provisional ballots may result in a large number of provisional ballots whose eligibility for inclusion may be contested, and will not be determined until after the election.[15] Meanwhile, other reforms leading to the increased use of absentee ballots may mean that allegations of absentee fraud will call a greater number of ballots into question.[16]

This Article explores what remedies are and should be available whenever some systemic election failure has occurred or whenever an election is within the margin of litigation. It describes the range of potential election problems, as well as their possible solutions, and creates a framework for identifying the most appropriate options for addressing particular problems. In doing so, this Article calls upon states to think more carefully about how to remedy an election failure.[17]

TION LAW @ MORITZ WEEKLY COMMENT, June 21, 2005, http://moritzlaw.osu.edu/electionlaw/comments/2005/050621.php. The phrase "margin of litigation," which gained widespread use during the 2004 presidential election, describes an election outcome close enough to draw post-election legal action. *See id.* Gratifyingly, the 2006 congressional election did not produce a dramatic increase in post-election litigation, despite prospects that it might. *See* Edward B. Foley, *Will the Election System Function Properly This Year?*, ELECTION LAW @ MORITZ WEEKLY COMMENT, Sept. 5, 2006, http://moritzlaw.osu.edu/electionlaw/comments/2006/060905.php. In fact, the only major election to end up in court was the race for Florida's thirteenth congressional district. *See* ELECTIONLINE.ORG, BRIEFING: THE 2006 ELECTION 1–3 (2006), *available at* http://www.electionline.org/Portals/1/Publications/EB15.briefing.pdf.

[14] *See* Richard L. Hasen, *Beyond the Margin of Litigation: Reforming U.S. Election Administration to Avoid Electoral Meltdown*, 62 WASH. & LEE L. REV. 937, 957–59 (2005) (cataloguing a dramatic increase in the number of election cases since 2000). However, this increase cannot be wholly attributed to the underlying facts necessary to support such litigation becoming more commonplace. At least some of this increase (and perhaps much of it) may instead reflect candidates' and the public's increased awareness of the potential of post-election litigation that *Bush v. Gore* precipitated. *See id.*

[15] Well-intentioned efforts to provide all eligible voters the right to vote by guaranteeing any person who appears at a polling place at least a provisional ballot have dramatically increased the number of ballots cast provisionally. *See* ELECTIONLINE.ORG, *supra* note 2, at 32–33.

[16] *See id.* at 28–29 (reporting that in 2000, twenty-nine states required an excuse to vote absentee, while only twenty-three required such an excuse in 2006); *see also* JOHN C. FORTIER, ABSENTEE AND EARLY VOTING: TRENDS, PROMISES, AND PERILS 54–57 (2006); Edward B. Foley, *Will the Expansion in Absentee Voting Yield an Increase in Abuse?*, ELECTION LAW @ MORITZ WEEKLY COMMENT, Mar. 28, 2006, http://moritzlaw.osu.edu/electionlaw/comments/2006/060328.php; CTR. FOR DEMOCRACY AND ELECTION MGMT., BUILDING CONFIDENCE IN U.S. ELECTIONS—REPORT OF THE COMMISSION ON FEDERAL ELECTION REFORM 2005, at 35–36 (2005), *available at* http://www.american.edu/ia/cfer/report/full_report.pdf.

[17] The term "election failure" describes any election that does not satisfactorily produce an answer to the choice on which the electorate is voting. It does not necessarily describe a design flaw, however, as even a perfectly designed system could generate a tie vote. In any event, it is unrealistic to expect perfection in the design of our election systems; even

The general issue of how to remedy an election failure in turn generates a number of specific related questions. Among these are (1) when to invalidate some portion of votes; (2) when, if ever, to use statistical adjustments to correct vote totals for demonstrated errors; (3) in what circumstances to compromise the anonymity of the polling booth by compelling voters to reveal their ballot choices; (4) whether to adjust election rules once an election is underway; and (5) when to take the dramatic steps of either postponing an election or invalidating a completed election and holding a new one. Although many state courts have had to deal with these questions on an ad hoc basis over the years, these questions have not received much scholarly or popular attention.[18] They also admit of few easy answers and require balancing a number of fundamental values, such as fairness, balloting access, election integrity, public trust, accuracy, and accountability.

Furthermore, a number of still deeper issues are lurking beneath the questions of how to remedy an election failure. For example, how much reliability and certainty do or should we expect from our election processes? What costs are we willing to pay for this certainty? How much public money are we willing to spend? How much personal convenience or privacy are we willing to sacrifice? How long are we willing to wait for the final determination of an election? Who should bear the burdens of post-election uncertainty and of resolving that uncertainty? What is the appropriate role of the judiciary in resolving election problems? When should courts eschew intervention, and when should they act aggressively to protect democratic processes?

In raising and reflecting on these and other questions, this Article proceeds in four Parts. Part I briefly describes several prototypes of election miscues and the range of existing remedial tools available to respond

---

in a well-designed system the reality is that some implementation error is almost always present. While such error usually is harmless, it inevitably gets intense scrutiny when an election is close.

[18] Some of these issues were addressed by a younger Kenneth Starr, *see* Kenneth W. Starr, *Federal Judicial Invalidation as a Remedy for Irregularities in State Elections*, 49 N.Y.U. L. REV. 1092 (1974), and by a 1975 article containing a section on "Postelection Remedies," *see Developments in the Law—Elections*, 88 HARV. L. REV. 1114, 1298–1339 (1975). Since then, the legal literature has not focused much sustained attention on these issues, with the exception of the variety of articles analyzing *Bush v. Gore* and the specific problems in Florida in the 2000 election. *See, e.g.*, Hugh M. Lee, *An Analysis of State and Federal Remedies for Election Fraud: Learning from Florida's Presidential Election Debacle*, 63 U. PITT. L. REV. 159 (2001); Daniel Tokaji, *First Amendment Equal Protection: On Discretion, Inequality and Participation*, 101 MICH. L. REV. 2409, 2487–95 (2003); Louise Weinberg, *When Courts Decide Elections: The Constitutionality of* Bush v. Gore, 82 B.U. L. REV. 609 (2002); *see also* R. Michael Alvarez, Betsy Sinclair & Richard L. Hasen, *How Much is Enough? The "Ballot Order Effect" and the Use of Social Science Research in Election Law Disputes*, 5 ELECTION L.J. 40 (2006) (urging caution in applying generalized social science data to resolve particular election contests); *Developments in the Law—Voting and Democracy*, 119 HARV. L. REV. 1127, 1155–65, 1188–1200 (2006) (summarizing ways to adjust election outcomes for illegal votes, and describing frequent judicial reluctance to remedy election administration problems).

to these unsettled election outcomes. Part II identifies several competing values and priorities relevant to assessing the strengths and weaknesses of election remedies. As this Part notes, these sometimes conflicting values will often require trade-offs as states work to promote and prioritize these values independently of any particular election controversy. To highlight these values and trade-offs, Part III discusses several prominent election failures and discusses the appropriate responses. Finally, Part IV urges states to reconstitute the remedies available for election problems, capitalizing on the judicial branch's strength in fact-finding while sparing the judiciary from making ad hoc policy choices with obvious partisan implications. This Part argues first that state election codes should provide a clearer articulation of which specific remedies are appropriate for particular types of election failure. It then identifies several additional issues that a state election contest statute should address with clarity. It also recommends that nonjudicial forums, such as administrative or legislative tribunals, be empowered to settle some election disputes, and that citizens be encouraged to adjust their expectations about our election processes—and about the appropriate remedies for failures in these processes—in recognition of the practical impossibility of developing a flawless system.

I. EXISTING REMEDIAL SCHEMES FOR COMMON ELECTION MISCUES

At the outset, it is worth noting that whenever an election outcome is in doubt, the typical first step is to retabulate the ballots. This process is an integral part of routine election administration,[19] and most states have fairly detailed statutory provisions governing when and how deputized election workers are to count and recount the election returns.[20] Especially since the 2000 election, many states have taken a closer look at these provisions of their code.[21]

In some respects, many of these statutes are holdovers from the days of paper ballots.[22] Indeed, as states increasingly rely on electronic voting,

---

[19] The right to a recount is a statutory right that did not exist at common law. *See, e.g.*, Abbene v. Bd. of Election Comm'rs of Revere, 202 N.E.2d 827, 829 (Mass. 1964); Eldredge v. Nickerson, 78 N.E. 461, 462 (Mass. 1906); *In re* Hearst, 76 N.E. 28, 29 (N.Y. 1905); Coe v. State Election Bd., 221 P.2d 774, 776 (Okla. 1950).

[20] For background regarding state laws concerning recounts, see Daniel Tokaji, *The Paperless Chase: Electronic Voting and Democratic Values*, 73 FORDHAM L. REV. 1711 app. B at 1817–36 (2005). During the 2000 election, some states' recount provisions became famous for identifying how many corners needed to be detached in order for a hanging chad to be included in a recount. *See, e.g.*, Joseph Crawford, *Michigan's Recount—No Dimples, Dapples or Disarray*, GRAND RAPIDS PRESS (Mich.), Dec. 22, 2000, at A12 (describing Michigan's requirement that at least two corners be detached).

[21] Questions receiving the most attention concern whether to require a paper trail to accompany electronic voting and whether a paper trail should be the official ballot in the event of a recount. *See* Peter Katel, *Voting Controversies*, 16 CQ RESEARCHER 745, 750 (2006), *available at* http://library.cqpress.com/cqresearcher/getpdf.php?type=color&file=cqr20060915C.pdf.

[22] Many updated recount statutes explicitly contemplate electronic voting. *See, e.g.*,

recount mechanisms may come to look more and more anachronistic, although the creation of paper audit trails may keep the practice of recounting ballots alive in some jurisdictions.[23] However, in other jurisdictions recounts are now accomplished at the simple touch of a button.[24] Recounts serve only a very narrow purpose in these jurisdictions, given that the original count occurred through an electronic tabulation process that is almost certain to generate an identical result when repeated. Ultimately, in many close elections the real fight therefore is not over whether to conduct a recount, but rather over *which* ballots to count.[25]

Accordingly, in addition to recount provisions, typical state election codes also include contest provisions that establish a judicial procedure to resolve those elections that remain unsettled even after an administrative recount, or whose final results are otherwise disputed. These provisions authorize the judicial branch (or occasionally some other tribunal) to resolve what otherwise traditionally would have been deemed nonjusticiable political questions.[26] Unfortunately, however, these election contest provisions often provide courts with little substantive guidance for determining whether a remediable election failure in fact has occurred, and if so, how to remedy it.[27] Instead, the focus of typical contest statutes is on the procedures for bringing a contest action.[28] Many courts adjudicating election problems therefore have had to develop their own standards for deciding if an actionable failure has occurred and how to resolve it. The unsurprising result has been a variety of judicially developed tests for when

---

FLA. STAT. ANN. § 102.141(6)(a)–(c) (West 2006) (differentiating between procedures for touchscreen voting and optical scan voting); 10 ILL. COMP. STAT. ANN. § 5/22-9.1 (West 2006); OKLA. STAT. ANN. tit. 26, § 8-114 (West 2005).

[23] Some states that have chosen to use electronic voting machines with paper audit trails provide that for purposes of a recount, the official ballot is the paper audit trail. *See, e.g.*, CAL. ELEC. CODE § 19253(b)(1) (West 2005); 10 ILL. COMP. STAT. § 5/24C-2 (2003); OHIO REV. CODE ANN. § 3506.18(A) (West 2006). In the event of a recount, these paper audit trails may be incredibly difficult to count, perhaps more difficult than manually recounting punch cards or other types of previously used paper ballots (as when the audit trail is recorded sequentially on a continuous paper scroll, which is harder to handle and read than a stack of individual punch cards or ballots).

[24] *See, e.g.*, N.M. Stat. Ann. § 1-1-6 (West 2006) (defining "recheck" as method of recounting electronic ballots by regenerating printout of electronic record for comparison with original printout).

[25] *See infra* Part I.B.2.

[26] *See, e.g.*, Johnson v. Stevenson, 170 F.2d 108, 110 (5th Cir. 1948); Griffin v. Buzard, 342 P.2d 201, 202 (Ariz. 1959); McPherson v. Flynn, 397 So. 2d 665, 668 (Fla. 1981); Missouri *ex rel.* Bouchard v. Grady, 86 S.W.3d 121, 123–24 (Mo. Ct. App. 2002); *cf.* Whitley v. Cranford, 119 S.W.3d 28, 31–32 (Ark. 2003) (ordering new elections solely on authority of state constitution's provision requiring "free and equal" elections). *But see* State *ex rel.* Nicely v. Wildey, 197 N.E. 844, 847 (Ind. 1935) (holding that "courts have inherent power to protect the sovereign people, and those who are candidates for office or claiming title to or rights in an office from fraud or unlawfulness"). *See generally* 26 AM. JUR. 2D *Elections* § 412 (1996) (describing election contest as "purely a constitutional or statutory proceeding").

[27] *See Developments in the Law—Elections*, *supra* note 18, at 1311.

[28] *See id.* at 1306–07.

courts will uphold, invalidate, call for a rerunning of, or themselves declare the winners of, a contested election.[29]

Disputed election outcomes can arise from any number of causes. A few examples, followed by an identification of typical judicial responses to them, will help to set the stage for a discussion of the challenges of seeking to remedy these election miscues.

A. Typical Kinds of Election Miscues

The primary causes of election failures can be divided into two categories: fraud and mistake. Voting fraud of course is a long-standing plague of democratic elections.[30] Fraud involves a deliberate attempt to manipulate the system unfairly, usually by candidates or their supporters. In contrast, mistake involves an unintentional disturbance or distortion of the election processes, usually caused by those administering the election. Many instances of both mistake and fraud may have little or no impact on the validity of the election outcome, however, and may even go undetected.

In addition to these two primary causes of election failures, there are at least two other possibilities: improper conduct by candidates or their supporters that does not fit the ordinary definition of voting fraud but nevertheless may provide grounds for questioning the integrity of the election (such as campaign spending in excess of agreed-upon limits); and "acts of God," such as hurricanes or other natural disasters, terrorist attacks, massive power failures, or other events outside the control of candidates or election officials that significantly disrupt the ability of voters to cast their votes or to have them properly counted.

1. Fraud

Voting fraud can be committed by dishonest candidates who clearly have a motive to commit it if they can find an opportunity to do so. It also can be committed by polling judges or other elections officials, who typically have much greater opportunity, provided they have a motive. Fraud can also be committed by isolated individuals or organized groups among the electorate, whose motives and opportunities may both be more attenuated.

Although a variety of media reports and other anecdotal accounts have convinced much of the American public that our elections today are fre-

---

[29] *See infra* Part I.B.

[30] *See generally* TRACY CAMPBELL, DELIVER THE VOTE (2005) (providing a provocative account of the range of election fraud that the United States has experienced throughout its history); John Fund, *How to Steal an Election*, CITY J. (2004), *available at* http://www.city-journal.org/html/14_4_urbanities-election.html.

quently tainted by voter fraud,[31] demonstrated cases of actual fraud are relatively uncommon, given the frequency with which Americans vote and the number of races involved.[32] Nevertheless, concerns about fraud understandably shape many features of our election system,[33] and credible allegations of voting fraud must be taken seriously.

Moreover, maintaining a sound mechanism for responding to credible allegations of fraud ought in principle to reduce the instances of fraud. Such a mechanism can include both civil and criminal penalties, as discussed below.[34] However, where sanctions alone have failed to deter election fraud, some additional remedy designed to rectify election fraud may often be necessary, at least if the fraud calls into question the outcome of the election.

Voting fraud that may call election outcomes into question can be grouped into several categories. One type of fraud is the manipulation of the number of raw votes cast, as in stuffing the ballot box. The modern day equivalent of ballot-box stuffing is tampering with the electronic counts on the voting equipment. Special access to the equipment is generally required to perpetrate this type of fraud.

Alternatively, the raw vote can be manipulated through voting by individuals who are not eligible to vote. Perpetrators of this brand of fraud may have fraudulently registered, may vote on behalf of dead people, or may vote multiple times. This type of fraud requires no special access to voting equipment.

Similarly, absentee ballot fraud can be accomplished without any special access. Fraudulent absentee balloting may frequently be used as one vehicle for accomplishing voting by ineligible individuals, because it is often harder to detect than in-person voting by ineligible individuals.[35] But absentee ballot fraud also encompasses voting by *eligible* voters who allow

---

[31] *See, e.g.*, JOHN FUND, STEALING ELECTIONS 1–9 (2004); Richard L. Hasen, *Beyond the Margin of Litigation: Reforming U.S. Election Administration to Avoid Electoral Meltdown*, 62 WASH. & LEE L. REV. 937, 942 (2005); Stephanie Philips, *The Risks of Computerized Election Fraud: When Will Congress Rectify a 38-Year Old Problem?*, 57 ALA. L. REV. 1123, 1123–50 (2006).

[32] *See* Eric Lipton & Ian Urbina, *In 5-Year Effort, Scant Evidence of Voter Fraud*, N.Y. TIMES, Apr. 12, 2007, at A1; DAVID CALLAHAN & LORI MINNITE, DEMOS, SECURING THE VOTE: AN ANALYSIS OF ELECTION FRAUD 4 (2003); Spencer Overton, *Voter Identification*, 105 MICH. L. REV. 631, 644–50 (2007); Tova A. Wang, *Competing Values or False Choices: Coming to Consensus on the Election Reform Debate in Washington State and the Country*, 29 SEATTLE U. L. REV. 353, 361–62 (2005).

[33] Not only are many features intended to reduce the possibility of fraud, but the possibility that fraud may have occurred often gives losing candidates a basis for contesting an election. *See* Edward B. Foley, *The Legitimacy of Imperfect Elections: Optimality, Not Perfection, Should Be The Goal of Election Administration*, *in* MAKING EVERY VOTE COUNT: FEDERAL ELECTION LEGISLATION IN THE STATES 97, 105 (Andrew Rachlin ed., 2006).

[34] *See infra* Part I.B.4.

[35] *See* FORTIER, *supra* note 16, at 52–55; John C. Fortier & Norman J. Ornstein, *The Absentee Ballot and the Secret Ballot: Challenges for Election Reform*, 36 U. MICH. J.L. REFORM 483, 512–13 (2003); Wang, *supra* note 32, at 389–90.

a third party to cast or influence their vote,[36] a practice equally antithetical to free elections. Accordingly, a number of restrictions on absentee voting processes are designed to guard against these problems.[37] However, these restrictions are difficult to police, and often become a primary source of controversy in election contests. In fact, absentee ballot fraud is one of the most common causes of election failures.[38]

Another category of voting fraud that can be accomplished without special access to the mechanics of the election process is preelection deception of voters (or potential voters) in ways that may affect who votes or how they vote. For instance, in the 2004 and 2006 elections, several reports circulated of voters receiving leaflets or phone calls announcing an incorrect voting day or location.[39]

A final category of fraud is after-the-fact distortion of the raw vote, either through outright false reporting of precinct tallies or through the intentional alteration, destruction, damage, or loss of physical ballots or memory cards.[40] Only those with official access to the ballots are likely to be in a position to accomplish this type of election fraud.

2. *Mistake*

Mistakes by election officials also can easily throw an election into question whenever those mistakes cannot be corrected before election day or cannot be remedied by provisional voting or a recount process. For instance, in one recent election-day blunder, the outcome of an Ohio school levy, apparently having failed by one vote, was cast in doubt when election officials realized that two voters each had innocently voted twice, first by absentee ballot and then again in person on election day after an election judge mistakenly advised them that their absentee ballots had not arrived at the board of elections.[41] Before anyone determined this advice was mistaken, both the absentee votes and the election day votes of these two voters had already been added to the final tally.[42]

---

[36] *See, e.g.*, Qualkinbush v. Skubisz, 826 N.E.2d 1181, 1187–91, 1206–07 (Ill. App. Ct. 2004) (describing candidate's unlawful efforts to assist absent voters in casting their ballots).

[37] *See* FORTIER, *supra* note 16, at 58.

[38] *See* FUND, *supra* note 31, at 145; FORTIER, *supra* note 16, at 53.

[39] *See* Jeff E. Schapiro, *FBI Looks Into Voter Intimidation*, RICHMOND TIMES-DISPATCH (Va.), Nov. 7, 2006, at A6; Jerry Seper, *Early Charges of Vote Fraud Suggest a Raft of Challenges; Parties, Interest Groups Claim Bungling, Dirty Tricks*, WASH. TIMES, Nov. 3, 2004, at A11.

[40] *See In re* Gen. Election for Dist. Justice, 670 A.2d 629, 633 (Pa. 1996) (describing ballot tampering that occurred after election day).

[41] *See* Holly Zachariah, *London Schools Contest One-Vote Election Loss in Court: District Asks to Have Double Votes Tossed Out*, COLUMBUS DISPATCH, Nov. 25, 2004, at D8.

[42] *See id.*

Absentee ballot problems account for a large portion of mistake-based allegations of election failures.[43] But an irreversible mistake could also be as straightforward as the loss of a precinct's paper ballots or, the functional equivalent, the failure of an an electronic voting machine (and its backup count) because of irretrievable damage or unreadability.[44] Mistakes could also include errors in who is allowed to vote, errors (including miscommunications) in voting instructions, errors in providing appropriate accommodations for voters with disabilities, other errors related to polling place operations, and confusing, misleading, or defective ballots or equipment.[45]

This last type of mistake is exemplified by the screen layout of the electronic voting machines in 2006 in Sarasota County, Florida, which may have contributed to a surprisingly high number of undervotes in the congressional race in that county.[46] It is also exemplified by the design of the butterfly ballot used in 2000 in Palm Beach County, Florida, which had the wholly unanticipated but undeniable effect of inflating the number of votes officially cast for Patrick Buchanan well above the number of voters who had intended to mark their ballot for Buchanan.[47] Notably, although this aspect of the Florida 2000 election may have provided Al Gore with the best moral argument that he should have won the election, the ballot design impact was quickly dismissed as not legally remediable.[48]

---

[43] *See, e.g.*, Miller v. Picacho Elementary Sch. Dist. No. 33, 877 P.2d 277, 279 (Ariz. 1994) (setting election aside because of irregularities associated with absentee ballots); Womack v. Foster, 8 S.W.3d 854, 871–76 (Ark. 2000) (invalidating several hundred absentee ballots with various types of irregularities); Boyd v. Tishomingo County Democratic Executive Comm., 912 So. 2d 124, 131–32 (Miss. 2005) (discussing several types of irregular absentee ballots and invalidating some).

[44] *See, e.g.*, Bauer v. Souto, 896 A.2d 90, 94 (Conn. 2006) (finding that a voting machine malfunctioned and failed to record votes for one candidate); LaCaze v. Johnson, 310 So. 2d 86, 87 (La. 1974) (finding similarly that a voting machine malfunctioned).

[45] *See, e.g.*, Foulkes v. Hays, 537 P.2d 777, 779 (Wash. 1975) (finding that election officials failed to preserve and safeguard ballots between canvassing and recount).

[46] *Newspaper Links Age, 'Undervotes,'* ORLANDO SENTINEL, Jan. 3, 2007, at C5 (citing data showing a higher undervote where the median age was over sixty-five and stating "[s]everal experts have said the trend supports the theory that poor ballot design made the District 13 race hard to see on Sarasota County's touch-screen machines").

[47] *See* Jonathan N. Wand et al., *The Butterfly Did It: The Aberrant Vote for Buchanan in Palm Beach County, Florida*, 95 AM. POL. SCI. REV. 793, 795, 802–03 (2001); *see also* Jon Sawyer, *Party Spin Doctors Battle for Public Opinion: Both Camps Distort Truth in Florida*, ST. LOUIS POST-DISPATCH, Nov. 15, 2000, at A15 (describing Buchanan himself as "sure" that most of the 3407 ballots cast for him in Palm Beach County were intended for Gore); Don Van Natta Jr., *Counting the Vote: The Ballot; Gore Lawyers Focus on Ballot in Palm Beach County*, N.Y. TIMES, Nov. 16, 2000, at A29 (describing *New York Times*' analysis of Palm Beach voting patterns).

[48] *See* Fladel v. Palm Beach County Canvassing Bd., 772 So. 2d 1240, 1242 (Fla. 2000) (holding that claim of confusing ballot design did not state cause of action for noncompliance with election statute); *see also* David Von Drehle et al., *In Florida, Drawing The Battle Lines: Big Guns Assembled as Recount Began*, WASH. POST, Jan. 29, 2001, at A1. *But see* Stephen J. Mulroy, *Substantial Noncompliance and Reasonable Doubt: How the Florida Courts Got It Wrong in the Butterfly Ballot Case*, 14 STAN. L. & POL'Y REV. 203, 204 (2003); Stephen J. Mulroy, *Right Without a Remedy? The "Butterfly Ballot" Case*

A more frequent ballot preparation mistake involves improper ballots or equipment that omit or mislabel a race or a candidate. For instance, in 1996 the voting machine in one New Mexico precinct listed the wrong candidates' names for two races, which were decided by margins of eleven and ninety-eight voters, respectively.[49] Because some sixty-six voters had used the defective machine, the outcome of both races conceivably could have been different had the machine properly named the candidates.[50]

Post-election controversies related to errors in the casting of ballots by voters with disabilities have often involved the question of whether some voters received improper assistance, either at the polls or in filling out an absentee ballot. For instance, one recent local election in Mississippi turned in part on the validity of a score of absentee ballots, all cast by disabled voters and all witnessed by the same person.[51] Because some of the ballots were not properly completed or were otherwise suspect, the trial court invalidated all of them.[52] In another Mississippi case, the court vacated the results of an election because poll workers had provided assistance to many more voters than the "blind, disabled or illiterate" voters for whom the election code permitted voting assistance.[53] As jurisdictions continue to implement new technologies that reduce the need for disabled voters to receive voting assistance, the instances of allegations of improper assistance may decline. Meanwhile, however, failures to provide accessible equipment for disabled voters could continue to trigger post-election contests in close races.[54]

---

and Court Ordered Federal Election "Revotes," 10 GEO. MASON L. REV. 215, 216–17 (2001) (arguing that a judicial remedy should have been available).

[49] *See* Gunaji v. Macias, 31 P.3d 1008, 1010 (N.M. 2001).

[50] *See id.* at 1010, 1017–18 (holding that the proper remedy was not to call a new election but to throw out all the votes from the precinct using the faulty machine); *see also* Whitley v. Cranford, 119 S.W.3d 28, 31–32 (Ark. 2003) (upholding lower court's decision to void election for justice of the peace because 183 ballots omitted that race and margin of victory was only 55 votes); Lakes v. Estridge, 172 S.W.2d 454, 456 (Ky. 1943) (invalidating school board election when ballots in three precincts omitted several candidates' names, notwithstanding election officials' handwritten corrections to ballots); Ferguson v. Rohde, 449 S.W.2d 758, 760 (Ky. Ct. App. 1970) (finding that balloting equipment omitted a candidate's name).

[51] Campbell v. Whittington, 733 So. 2d 820, 825–26 (Miss. 1999).

[52] *Id.*; *see also* Dugan v. Vlach, 237 N.W.2d 104, 106 (Neb. 1975) (refusing to count disabled voter's ballot that did not meet statutory criteria).

[53] O'Neal v. Simpson, 350 So. 2d 998, 1008 (Miss. 1977); *see also* Brooks v. Crum, 216 S.E.2d 220, 223–24, 227–28 (W. Va. 1975) (rejecting votes from entire precinct in which election workers provided voting assistance to large numbers of voters who were not eligible to receive assistance).

[54] *See, e.g.*, Douglas Hadden, *Voters Handicapped*, PAWTUCKET TIMES (R.I.), Nov. 15, 2006, *available at* http://www.zwire.com/site/news.cfm?newsid=17472755&BRD=1713&PAG=461&dept_id=24491&rfi=6 (describing widespread malfunctioning of AutoMark ballot readers for visually impaired voters).

3. Nonfraudulent Misconduct

Particularly given the increasing number of substantive constraints on the election process (including HAVA, state voter registration processes, the ever-changing landscape of campaign finance regulation, and more), candidates or their supporters may violate election laws in ways other than outright voting fraud. Like actual voting fraud, however, these violations may in some circumstances undermine the reliability of the election outcome. For instance, when a whole group of absentee ballots is filled out at a single get-out-the-vote rally, suspicions may arise about how accurately these ballots reflect the will of the individual voters.[55]

Wyoming's contest statute offers one possible method of addressing some of these concerns by expressly providing that a victorious candidate's violation of any one of a variety of prohibited election activities (such as electioneering at the polls) constitutes grounds for contesting the candidate's victory.[56] In contrast, an Arizona court concluded that a candidate's electioneering in a polling place in violation of the state election code did not constitute grounds for a new election.[57]

In a recent case, the Maryland Court of Appeals rejected a challenge predicated on the victor's failure to file a required campaign finance report.[58] Although Maryland law prohibits a person from running for or assuming public office if the person has not filed the required reports,[59] the court held that the matter was ripe before the election and should have been resolved then.[60] However, other types of campaign finance violations might not be ripe until after an election and arguably could have a real (if not quantifiable) impact on the outcome. This would be the case if, for example, in the final days of a campaign a candidate who has agreed to accept public funding nonetheless violates the spending limits upon which the public funding is predicated.

4. Extrinsic Events or "Acts of God"

A final category of problems that may cast the validity of election results into doubt are circumstances that involve "acts of God," which are dramatic events outside the control of election administrators or candidates. For instance, the devastating Hurricane Katrina of 2005 and the

---

[55] *See* FORTIER, *supra* note 16, at 56; Foley, *supra* note 33, at 107.

[56] *See* WYO. STAT. ANN. § 22-17-101(v) (2006) (referencing the criminal provisions in WYO. STAT. ANN. §§ 22-26-101 to 22-26-121 (2006), which prohibit such misdemeanor conduct as electioneering within 100 feet of a polling place or possessing alcoholic beverages at a polling place, as well as more serious felonies such as bribery or tampering with ballot boxes and machines).

[57] *See* Fish v. Redeker, 411 P.2d 40, 43 (Ariz. Ct. App. 1966).

[58] *See* Ross v. State Bd. of Elections, 876 A.2d 692, 706 (Md. 2005).

[59] *See* MD. CODE ANN., ELEC. LAW § 13-332 (West 2006).

[60] *See Ross*, 876 A.2d at 706.

terrorist attacks of September 11, 2001, each caused major disruptions to democratic processes. Hurricane Katrina's long-term dislocation of hundreds of thousands of people from New Orleans and surrounding communities dramatically altered local elections months later,[61] while the attacks on the World Trade Center interrupted a New York City primary election already underway on the day of the attacks.[62] In both cases, the effect was not to call concluded elections into doubt, but to require an adjustment of when and how the elections would occur. However, it is not difficult to hypothesize that similar occurrences in the days or hours after the close of polling could upset a concluded election.

B. Existing Remedial Approaches

How do our current state election systems attempt to fix election failures? Florida's highly publicized efforts to settle its 2000 presidential election demonstrated that, in theory, the touchstone for resolving election difficulties is the "intent of the voters."[63] As Alexander Hamilton famously said, "the people should choose whom they please to govern them."[64] But when fraud, mistake, or other election distortions or disruptions have occurred, how to ascertain whom the voters have chosen is not always so clear. Although most state election codes include provisions that create processes by which aggrieved candidates or voters may seek judicial review of election outcomes, these provisions often provide courts little substantive guidance.[65] As a result, courts asked to adjudicate election controversies must often draw upon common law and equitable principles to fashion appropriate remedies for particular circumstances not specifically addressed in their state's applicable election code.[66] In fashioning such remedies, courts typically grant forms of relief that can be

---

[61] *See National Briefing South: Louisiana: New Orleans Election Postponed*, N.Y. TIMES, Dec. 13, 2005, at A31 (describing Louisiana Governor's indefinite suspension of New Orleans elections); Gary Rivlin, *New Orleans Election in Doubt*, N.Y. TIMES, Dec. 3, 2005, at A15 (describing Louisiana Secretary of State's recommendation that New Orleans' February 2006 election be postponed as long as eight months because of Katrina's disruptive impact).

[62] *See Primary Elections are Cancelled*, N.Y. L.J., Sept. 12, 2001, at 3.

[63] Gore v. Harris, 772 So. 2d 1243, 1253 (Fla. 2000), *rev'd sub nom.* Bush v. Gore, 531 U.S. 98 (2000); Palm Beach County Canvassing Bd. v. Harris, 772 So. 2d 1273, 1282 (Fla. 2000). However, in *Bush v. Gore*, the U.S. Supreme Court was sharply critical of the "intent of the voter" standard when employed without some uniform, objective criteria for determining the will of particular voters whose ballots are ambiguous. 531 U.S. at 104–06.

[64] 2 ELLIOT'S DEBATES 257 (2d ed. 1859), *available at* http://memory.loc.gov/cgi-bin/ampage.

[65] *See infra* notes 132–35 and accompanying text; *see also Developments in the Law—Elections*, *supra* note 18, at 1311 (describing election contest statutes as providing "little guidance as to the grounds that are cognizable").

[66] *See, e.g.*, Gunaji v. Macias, 31 P.3d 1008, 1014–15 (N.M. 2001) (deciding to throw out all votes from precinct using machine that listed wrong candidates' names rather than call new election); State *ex rel.* Olson v. Bakken, 329 N.W.2d 575, 580 (N.D. 1983).

grouped into a handful of categories: recounts; adjustment of vote totals; new elections; fines and penalties; and injunctions concerning some aspect of an ongoing or future election.

1. Recounts

Recount procedures are at the heart of most states' statutory codes for handling close election contests.[67] Administrative recounts, which are routine in close elections, can be of two basic varieties. First, many states provide for an automatic recount if initial election results are within a sufficiently close margin, usually measured in terms of a percentage of the total vote.[68] Second, many states allow administrative recounts upon request, while often requiring the requesting party to cover the costs.[69]

Automatic recounts are perhaps best viewed as part of ordinary election processes, rather than as mechanisms for resolving contested elections. That is, automatic recounts occur simply as part of determining official election returns. They are designed primarily to identify and correct mistakes in initial vote tabulations, and are an implicit acknowledgment that our election processes are not error-free. Of course, recounts themselves are not guaranteed to be error-free, although they may conceivably be conducted more carefully and deliberately than initial counts.

Requested recounts, on the other hand, often play a more remedial function. This kind of recount may seek to call into question the official returns, or to focus attention on some disputed portion of votes, perhaps in tandem with allegations of fraud or mistake. But in most states, it is difficult to challenge the legality of a particular ballot or set of ballots in either a requested or an automatic recount proceeding. Instead, these issues usually are addressed in contest actions, in which courts or other tribunals are authorized to adjudicate allegations of voting irregularities. For example, more than fifty years ago, the Kentucky Supreme Court explained that when a candidate seeks a recount to correct for ballots tainted by irregularities, a contest action is the proper forum.[70] The court observed that "[the] increasing efforts . . . to litigate election irregularities in recount proceedings" perhaps were a result of the court's "failing to mark clearly and distinctly the dividing line between a recount proceeding and a contest suit."[71]

---

[67] *See supra* notes 19–25 and accompanying text.

[68] *See* Tokaji, *supra* note 20, at 1817–36. A few states even conduct some form of recount in every race, as an audit or check on their processes. *See id.*

[69] *See id.*

[70] *See* Hogg v. Howard, 242 S.W.2d 626, 628 (Ky. 1951).

[71] *Id.*; *see also* Kearns v. Edwards, 28 A. 723, 724 (N.J. 1894) (holding that ministerial recount process is inappropriate forum for determining legality of ballots); *cf.* Carlson v. Oconto County Bd. of Canvassers, 623 N.W.2d 195, 197 (Wis. Ct. App. 2000) (finding that contest statute is the "exclusive remedy" for election fraud or irregularity).

Nevertheless, the dividing line between recounts and contests still remains less than clear in some jurisdictions. Moreover, in many contest actions, what the contestant seeks is effectively a judicially conducted or supervised recount.[72] For instance, fifteen years before the Supreme Court in *Bush v. Gore* invalidated recounts conducted under standards that varied among counties,[73] a congressional candidate in Indiana asked the state courts supervising a recount to impose uniform standards for conducting the recount across fifteen counties.[74] In other states, a timely demand for a recount is a prerequisite to filing a contest action,[75] and in any event a recount may serve to set the stage for an election contest.[76]

2. *Adjustments to Vote Totals and Election Outcomes*

As previously noted, the issue underlying many election recount proceedings is often *which* votes to count. Election officials may have some authority to disqualify ballots that they determine have been cast fraudulently or in error, but courts (or administrative tribunals) presented with an election contest frequently must determine the validity of some subset of votes allegedly tainted by mistake or fraud. Such tribunals may often face difficult issues in making these determinations.

Whether disputed votes are in fact valid can depend on a number of factual and legal questions. One set of questions involves the eligibility of the voters who cast the disputed ballots. For example, were they eligible to register and to vote in their district?[77] Did they properly complete the registration process? Did they properly establish their identities when they voted?[78] Did they properly complete absentee ballots?[79] Are there ballots that cannot be attributed to voters who signed the poll books?[80] Another set of questions focuses on the sufficiency of ballots or voting equipment.

---

[72] *See, e.g.*, Hendon v. N.C. State Bd. of Elections, 633 F.Supp. 454 (W.D.N.C. 1986) (describing a contest action in which plaintiffs requested a court-ordered recount that would include "split ticket" votes cast for a specific candidate on a ballot that was also marked as a straight-ticket ballot for the opposite party).

[73] 531 U.S. 98, 110 (2000).

[74] *See* McIntyre v. Fallahay, 766 F.2d 1078, 1080 (7th Cir. 1985) (explaining that the federal court to which the proceedings had been removed no longer had a justiciable issue once the House of Representatives had adjudicated the election).

[75] *See, e.g.*, Miller v. County Comm'n of Boone County, 539 S.E.2d 770, 777 (W. Va. 2000) (interpreting West Virginia election law, W. VA. CODE ANN. §§ 3-6-9, 3-7-6 (West 1999), as requiring contestant who seeks to challenge validity of ballots to request recount first).

[76] *See, e.g.*, ALA. CODE § 17-16-21(d) (West 2007) (describing a recount that changes vote totals sufficiently to alter an election outcome as "grounds for an election contest" rather than cause for automatic adjustment of certified winner).

[77] *See, e.g.*, Waltman v. Rowell, 913 So. 2d 1083, 1086 (Ala. 2005).

[78] *See, e.g.*, Jones v. Jessup, 615 S.E.2d 529, 531 (Ga. 2005).

[79] *See, e.g.*, Cochran v. Grubbs, 913 So. 2d 446, 448 (Ala. 2003).

[80] *See, e.g.*, Joan Mazzolini, *Thousands Voted Illegally*, PLAIN DEALER (Cleveland), Dec. 5, 2006, at A1.

Does a particular ballot meet the statutory requirements for clearly discernible voter intent?[81] Did the ballot design allow voters to make a free choice among all candidates?[82] Were the ballots properly secured, both before and after voting, to ensure the reliability of the vote?[83]

Furthermore, once courts (or other election tribunals) identify a set of invalid votes, they then must face the often equally difficult legal issue of deciding what relief to grant. One of two principal kinds of relief available in a contest action, discussed immediately below, is to adjust the vote totals by discounting the votes found to be invalid. The second principal kind of relief, discussed in the subsequent section, is to void the entire election, leaving the contested office vacant until a new election is held.[84]

The most compelling case for adjusting election results arises when the specific tainted votes, rather than just the total number of affected votes, can be identified. If the particular invalid votes are known, then a court can subtract those votes from the official tally and declare as the winner the candidate who has the most remaining votes.[85] For instance, in a recent judicial election in Arkansas, the outcome was reversed when 518 invalid absentee ballots were specifically identified and removed from the official count because the ballots were not obtained or submitted in compliance with absentee balloting requirements.[86] In 2004, when the two voters in the aforementioned Ohio election (decided by a one-vote margin) were found to have voted both in person and by absentee ballot,[87] their subsequent voluntary testimony about how they had voted allowed the court to conclude that no adjustment was necessary because the outcome would not change.[88]

---

[81] *See, e.g.*, Big Spring v. Jore, 109 P.3d 219, 220 (Mont. 2005) (clarifying legal standard of ballots invalidated by overvotes); *In re* Election for Sch. Comm. Representative for Dist. 3 in Portland, No. CV-04-695, 2004 WL 3196881, at *3 (Me. Super. Dec. 3, 2004).

[82] *See, e.g.*, Whitley v. Cranford, 119 S.W.3d 28, 30 (Ark. 2003) (ordering a new election when the margin was 55 votes and 183 ballots improperly omitted the race from ballot).

[83] *See, e.g.*, *In re* Petition to Contest the Gen. Election for Dist. Justice in Judicial Dist. 36-3-03, 695 A.2d 476, 479 (Pa. Commw. Ct. 1997) (describing ballot tampering after voting had concluded).

[84] *See infra* Part I.B.3.

[85] If discounting the illegal votes will not change the outcome, courts generally will not adjust the official results of the election, and instead will affirm the outcome based on the original tally. *See* 26 AM. JUR. 2D *Elections* § 438 (1996). An alternative framework would be to adjust the totals anyway, even when the outcome would remain unchanged, if only for the sake of greater historical accuracy. However, to obtain judicial relief a contestant is ordinarily required to prove sufficient illegal votes to call the outcome into question. *See id.* When that predicate is not met, courts effectively have no authority to alter the election results, even if some illegal voting can be proven.

[86] *See* Womack v. Foster, 8 S.W.3d 854, 863, 875–76 (Ark. 2000).

[87] *See* Zachariah, *supra* note 41 and accompanying text.

[88] *See* Randy Ludlow, *Double Votes Moot in Levy Loss; Couple Says They Split on London Schools Tax*, COLUMBUS DISPATCH, Dec. 17, 2004, at 1A. In some instances of voter fraud, courts have compelled the voters to disclose for whom they voted, creating an exception to the sanctity of the secret ballot. *See infra* notes 150–53 and accompanying text. Yet voter testimony in such instances may not always be sufficiently reliable, as the

However, when the tainted votes cannot be specifically identified, as is often the case, the proper remedy is less clear. Nevertheless, when the number of tainted votes exceeds the margin of victory, most courts have a comparatively easy time resolving the contest if the ostensible victor of the election is demonstrated to be the intended beneficiary of vote fraud. In such a circumstance, courts will often presume that all of the illegal votes favored the victor (even if the votes cannot be specifically identified) and will deduct that number of votes from the official tally, thereby reversing the outcome.[89]

Courts face a more difficult question when the total number of votes demonstrated to have been tainted by fraud or misconduct is less than the margin of victory but when the fraud may have extended to additional votes. In some such instances, courts have merely voided the election,[90] or have even concluded that the fraud was sufficiently minor to let the election stand.[91] In other cases, courts have reversed the election outcome, thereby emphasizing that even attempted election fraud can be fatal to a candidate's chances.[92] Otherwise, an unscrupulous candidate, thinking that "the worst that will happen is a new election," might not be sufficiently deterred from committing fraud.[93]

For instance, in another Florida contest shortly before *Bush v. Gore*, the courts threw out the results of the 1997 Miami mayoral race.[94] The trial court concluded that a large number of absentee ballots favoring the apparent victor had been cast fraudulently.[95] Although the trial court had called for a new election, the appellate court concluded that it should send a stronger message to discourage voting fraud by disqualifying all absentee ballots.[96] Without the absentee ballots, the runner-up became the winner.[97]

---

voters themselves may choose to testify falsely in order to produce the outcome they desire, and because they have voted secretly, their testimony may not be amenable to independent verification.

[89] *See* Qualkinbush v. Skubisz, 826 N.E.2d 1181, 1207 (Ill. App. Ct. 2004). In some states, candidates proven to have committed fraud are disqualified from office. *See, e.g.*, HAW. REV. STAT. § 19-4 (2006).

[90] *See, e.g.*, Pabey v. Pastrick, 816 N.E.2d 1138, 1154 (Ind. 2004) (ordering special election to remedy pervasive fraud that rendered it impossible to know true winner of regular election); *see also infra* Part I.B.3.

[91] *See, e.g.*, Nugent v. Phelps, 816 So. 2d 349, 357 (La. Ct. App. 2002); Rogers v. Holder, 636 So. 2d 645, 650, 652 (Miss. 1994) (holding that it would be "imprudent" to void all absentee ballots when only twelve of eighty-five were proven illegal).

[92] *See, e.g.*, Bolden v. Potter, 452 So. 2d 564, 567 (Fla. 1984); Ellis v. Meeks, 957 S.W.2d 213, 217 (Ky. 1997).

[93] *In re* Protest of Election Returns and Absentee Ballots in the Nov. 4, 1997 Election for Miami, Fla., 707 So. 2d 1170, 1174 (Fla. Dist. Ct. App. 1998).

[94] *Id.*

[95] *See id.* at 1172.

[96] *See id.* at 1174.

[97] *See id.* A less radical historical approach, where the party affiliation of those casting an invalid ballot is known, was to attribute the vote to that party's candidate. *See, e.g.*, Talbott v. Thompson, 182 N.E. 784, 789 (Ill. 1932).

Similarly, on occasion courts have simply rejected all votes from a particular precinct where voting irregularities occurred.[98] As with invalidating all absentee ballots, the effect of this adjustment is to disenfranchise an entire subset of voters, some (or perhaps many) of whose ballots are not invalid. Despite this consequence, when courts have been unable to ascertain which candidates benefited from the illegal votes, they sometimes have preferred this remedy over requiring a new election.[99] The alternative of holding a new election effectively amounts to disenfranchising *everyone* who voted, although it also provides everyone with a second opportunity to exercise their franchise.

Another option, especially for nonfraudulent voting irregularities that have affected an identifiable group of voters, is the remedy of "proportional deduction."[100] A court using this remedy makes an educated guess about who benefited from the invalid votes, usually by ascribing to the group of invalid votes the same distribution of votes for specific candidates as is found across all the ballots cast in the entire precinct where the invalid ballots were cast.[101] For instance, if ten felons are found to have cast invalid ballots in a precinct that voted sixty percent for Candidate *A* and forty percent for Candidate *B*, then the felons' ballots would be treated as though six ballots favored Candidate *A* and four ballots favored Candidate *B*.[102]

While proportional deduction has often been used to resolve election contests, at least one court has criticized its use. A Washington trial court adjudicating that state's 2004 gubernatorial election contest recognized that proportional deduction may be methodologically unsound because it usually assumes, without adequate foundation, that the invalid votes in a particular precinct are representative of all votes in the precinct.[103] That is, the ten felons in the example above are assumed as a class to be representative of the rest of the voters in that precinct, when in fact the ten felons instead may disproportionately favor one of the candidates. Though it might be possible to develop more sophisticated proxies that would take into account a number of additional demographic factors to predict for

---

[98] *See, e.g.*, Application of Bonsanto, 409 A.2d 290 (N.J. App. 1979); Burkett v. Francesconi, 23 A.2d 780, 782–83 (N.J. 1942); Vigil v. Garcia, 87 P. 543, 545–46 (Colo. 1906).

[99] *See Vigil*, 87 P. at 545–46.

[100] *See Developments in the Law—Voting and Democracy*, *supra* note 18, at 1156.

[101] *See, e.g.*, Canales v. City of Alviso, 474 P.2d 417, 422–23 (Cal. 1970) (describing proportional deduction as an appropriate standard when no evidence for allocating votes exists); McNabb v. Hamilton, 181 N.E. 646, 647 (Ill. 1932) (apportioning disputed ballots in same proportions as overall vote); *In re* Durkin, 700 N.E.2d 1089, 1095 (Ill. App. Ct. 1998) (approving use of proportional deduction and rejecting use of party affiliation as methods for allocating invalid votes).

[102] *See* Huggins v. Superior Court in and for County of Navajo, 788 P.2d 81, 85–86 (Ariz. 1990).

[103] *See* Transcript of Oral Decision at 15–17, Borders v. King County, No. 05-2-00027-3 (Wash. Super. Ct. June 6, 2005) [hereinafter Borders Transcript] (terming this problem the "ecological fallacy").

whom invalid votes were actually cast, for now proportional deduction remains a fairly crude tool.[104]

3. New Elections

The remaining alternative for remedying an unreliable election outcome, whether it occurs as a result of fraud or mistake, is to invalidate the election and hold a new one. Courts have resorted to this remedy in a variety of circumstances, and many courts have wrestled with whether and when they have authority to order a new election.[105]

Although examples of ordering a new election at the federal level are rare,[106] the remedy is not uncommon in local elections. For instance, in an Alabama mayoral race in 1984, one of four voting machines was conclusively shown to have failed to register any votes cast for one of the four candidates. The Alabama Supreme Court ordered the rerunning of the entire election, overruling the trial court's remedy of new balloting only for those voters who voted on the defective machine.[107]

But whether and when the remedy of a new election is available is often unclear. Relying on different provisions of Louisiana's election contest statute, the Louisiana Supreme Court divided over whether the statute permitted a new election as a remedy for widespread fraud when the fraud was not specifically shown to have affected the result.[108] Two dissenting justices agreed with the trial court that a new election was appropriate because the proven fraud made it "impossible to determine the result" of the election, but the majority interpreted the contest statute as allowing a new election only if fraud is proven to have occurred in numbers "sufficient to change the result."[109]

---

[104] In addition, proportional deduction has received "virtually no academic commentary." *Developments in the Law—Voting and Democracy*, *supra* note 18, at 1156.

[105] *See, e.g.*, Pabey v. Pastrick, 816 N.E.2d 1138, 1141 (Ind. 2004) (concluding the court has authority to order new election when candidate's misconduct makes it impossible to determine election's outcome).

[106] The 1974 election for one of New Hampshire's U.S. Senators produced a two-vote margin of victory. Nine months later, after a protracted election contest, the U.S. Senate declared the seat vacant and New Hampshire held a new election. *See* ANNE M. BUTLER & WENDY WOLFF, U.S. SENATE ELECTION, EXPULSION AND CENSURE CASES FROM 1793 TO 1990, S. DOC. NO. 103-33, at 421–25 (1st Sess. 1995).

[107] *See Ex parte* Vines v. Allen, 456 So. 2d 26, 28 (Ala. 1984); *see also* Bauer v. Souto, 896 A.2d 90, 99 (Conn. 2006) (ordering a new election after a lever machine was proven not to have recorded votes); Whitley v. Cranford, 119 S.W.3d 28, 30 (Ark. 2003) (ordering a new election when margin was 55 votes in an election where 183 ballots improperly omitted the race).

[108] *See* Savage v. Edwards, 722 So. 2d 1004, 1004–08 (La. 1998).

[109] *Id.* The same conceptual disagreement is reflected in the Indiana Supreme Court's reversal of a trial court's refusal to order a new election in the face of widespread absentee ballot fraud. *See Pabey*, 816 N.E.2d at 1151. Although the number of absentee ballots cast vastly exceeded the margin of victory, the trial court had interpreted the contest statute to require proof to a "mathematical certainty" of enough illegal votes to alter the outcome. *Id.* at 1149. The Indiana Supreme Court instead concluded that a new election was required

This interpretive disagreement arose under Louisiana's contest statute even though in many other respects the Louisiana statute provides more guidance than a typical state election code. For instance, Louisiana's contest statute makes explicit that in some circumstances the appropriate remedy for an election failure may be a "restricted election,"[110] which allows some subset of voters to vote again.[111] Courts in several other jurisdictions have on occasion ordered just such a remedy, but without clear statutory authority to do so.[112]

Yet even when courts believe they have the authority to call for a new election, they are quite reluctant to do so, and for good reason.[113] Courts must be satisfied both that an election failure has occurred and that the failure is significant enough to taint the outcome. The margin of victory usually plays a determinative role in answering this latter question. For instance, in 2004 some Louisiana polling places opened late.[114] A Louisiana appellate court ordered a new election for one race in which the margin of victory was less than ten votes,[115] but upheld the original results of another contest in which the margin of victory was over 9000 votes.[116] Accordingly, if the margin of victory is slim, it is much more likely that a court will conclude that an election failure renders the outcome unreliable. Nevertheless, even large margins can become suspect if a sufficiently serious failure occurs.

Concerns about protecting ballot secrecy can also increase the need to call for new elections.[117] For instance, in a 1999 Louisiana sheriff's race, the court ordered a new election after determining that the three-vote margin of victory had included the votes of five invalid absentee ballots.[118] Although the court was able to identify the five voters who cast these defective absentee ballots, it was not prepared to order them to disclose their votes, and concluded that "because of the constitutional guarantee to secrecy of the ballot . . . , it is impossible to determine the result of this runoff election."[119]

---

because the absentee ballot fraud "substantially undermin[ed] the reliability of the election and the trustworthiness of its outcome." *Id.* at 1150–51.

[110] LA. REV. STAT. ANN. § 18:1432(A) (2007).

[111] *See* Jenkins v. Williamson-Butler, 883 So. 2d 537, 539–40 (La. 2004).

[112] *See, e.g.*, Gunaji v. Macias, 31 P.3d 1008, 1012, 1016–18 (N.M. 2001) (creating an equitable remedy of partial revote, in contrast to code requirement of disregarding the entire precinct); State *ex rel.* Olson v. Bakken, 329 N.W.2d 575, 579–82 (N.D. 1983) (approving the equitable remedy of partial special election for an identified set of voters whose votes were not counted).

[113] *See infra* Part IV.A.

[114] *See Jenkins*, 883 So. 2d at 540.

[115] *Id.* at 541.

[116] Hester v. McKeithen, 882 So. 2d 1291, 1294 (La. 2004).

[117] *See, e.g.*, Hester v. Kamykowski, 150 N.E.2d 196, 200–01 (Ill. 1958) (ordering a new election because see-through ballots printed on low-quality paper compromised the secrecy of the election).

[118] Adkins v. Huckaby, 755 So. 2d 206, 208 (La. 2000).

[119] *Id.* at 222. This is in contrast to cases in which courts have compelled witnesses

4. Criminal Penalties, Fines, and Damage Awards

Criminal punishment for the wrongdoer constitutes an entirely different class of remedies for fraud and other forms of election misconduct. Many state codes classify some types of election fraud as felonies and other types of election misconduct as misdemeanors, and establish corresponding punishments. Such remedies do nothing to correct whatever flawed outcome has resulted from the misconduct, however, but serve instead to deter future election problems. The mere threat of criminal punishment ideally reduces the incidence of election fraud and some types of election misconduct. In addition, election officials who engage in fraud, misconduct, or neglect of duty typically can be removed from their positions.[120]

In some circumstances, a qualified elector whose voting rights have been violated may bring a civil action for damages,[121] and in recent years some commentators have encouraged additional use of damage actions to encourage better election processes.[122] However, it is often difficult for a particular voter to show sufficient injury-in-fact to support a civil action for violation of voting rights.[123]

5. Other Injunctive Relief

Enjoining some aspect of the conduct of an election, either prior to or during the election, is a final tool appropriate for remedying certain types of election failures.[124] This sort of injunctive relief may apply to

---

who voted fraudulently to disclose for whom they voted, in order to enable the court to adjust the vote totals accordingly. *See infra* note 151 and accompanying text. It also is in contrast to circumstances in which voters have voluntarily disclosed for whom they voted in order to resolve an election dispute. *See supra* note 88, *infra* note 150, and accompanying text.

[120] *See, e.g.*, OHIO REV. CODE ANN. § 3501.27(A) (West 2006) (stating that "[n]o person who has been convicted of a felony or any violation of the election laws . . . shall serve as an election officer"); *id.* § 3501.22(A) (providing that the election officers of the precinct "may be summarily removed from office at any time by the board for neglect of duty, malfeasance, or misconduct in office or for any other good and sufficient reason").

[121] *See* Gage v. Monescalchi, 793 N.Y.S.2d 235, 235 (N.Y. App. Div. 2005). *But see* Hutchinson v. Miller, 797 F.2d 1279, 1280 (4th Cir. 1986) (holding that federal courts should not entertain claims for damages caused by election irregularities).

[122] For instance, attorney Hugh Lee, who participated in the litigation of the 2000 Florida election, has urged greater use of 42 U.S.C. § 1983 (2006). *See* Lee, *supra* note 18; *see also* SAMUEL ISSACHAROFF ET AL., THE LAW OF DEMOCRACY: LEGAL STRUCTURES OF THE POLITICAL PROCESS 1073–86 (rev. 2d ed. 2002) (discussing damages remedy for defective elections).

[123] *See* Santana v. Registrars of Voters of Worcester, 502 N.E.2d 132, 135 (Mass. 1986) (finding no basis for damages for deprivation of voting rights absent financial loss or physical or emotional injury).

[124] Injunctions can also occasionally cover conduct after an election, as with an injunction against counting ballots pending the resolution of a controversy about their eligibility or against a candidate taking office. *See, e.g.*, Tate-Smith v. Cupples, 134 S.W.3d 535, 537 (Ark. 2003) (enjoining victor from taking office until contest was resolved); James M. Fischer, *Preliminarily Enjoining Elections: A Tale of Two Ninth Circuit Panels*, 41 SAN

polling places, poll workers, elector qualifications, and candidates and their supporters. Such injunctive relief is most often employed to compel performance of obligations that are already clearly established.

Courts are understandably more reluctant to order an injunction on the eve or in the middle of an election than well before it.[125] Nevertheless, many examples exist of judicial intervention in an election already underway. When equipment malfunctions or polling places do not open on time, courts may order polls to remain open longer than normal[126] or require the use of alternative methods of voting.[127] When poll workers are not properly performing their duties, courts may direct them to do so.[128]

Injunctive relief also sometimes includes postponing an election,[129] as New York courts and its Governor did on September 11, 2001.[130] This tends to be a rare and serious event, given that candidates, parties, election officials, and the public all have prepared for a race to occur on a certain date, and have rationed their funds and energy in anticipation of that date. The additional costs of rerunning an election are likely to be substantial. On the other hand, these costs are often preferable to going ahead with an election that is likely to result in an indeterminate, unreliable, or unacceptable outcome. For example, postponing the New York election

---

DIEGO L. REV. 1647, 1648–49 (2004); Suzy Loftus, *Punch-Card Ballots, Residual Votes and the Systematic Disenfranchisement of Minority Voters: A Look at the Decision to Allow the California Recall Election to Proceed*, 39 U.S.F. L. REV. 763, 782–85 (2005).

[125] *See Developments in the Law—Voting and Democracy*, *supra* note 18, at 1188–1200 (discussing judicial reluctance to interfere with ongoing elections).

[126] *See, e.g.*, Diana Marrero & Deborah Barry, *Voting Problems Widespread*, GANNETT NEWS SERV., Nov. 8, 2006, *available at* http://www.deseretmorningnews.com/dn/view/0,1249,650205323,00.html (describing Indiana court order that polls "remain open nearly three hours past the regular closing time to make up for late openings"); William Presecky, *Kane County Judge Backs Counting Overtime Ballots: Democrats Lose Appeal About Elgin Township*, CHI. TRIB., Nov. 29, 2006, at M4 (describing a court order to keep polls open an extra ninety minutes because "numerous problems prevented voting from starting on time"). Most states already require that all voters in line at the time polls are scheduled to close be allowed to vote, however. *See, e.g.*, IDAHO CODE ANN. § 34-2422(1) (2007); ME. REV. STAT. ANN. tit. 21-A, § 626(2)(A) (2006); NEB. REV. STAT. § 32-908(3) (2006); NEV. REV. STAT. § 293.305(1); N.M. STAT. ANN. § 3-8-45(A) (2007); OHIO REV. CODE ANN. § 3501.32(A) (2007); *see also* Barry H. Weinberg & Lyn Utrecht, *Problems in America's Polling Places: How They Can Be Stopped*, 11 TEMP. POL. & CIV. RTS. L. REV. 401, 430 (2002) ("Most states provide that any voter in line at the time of poll closing is entitled to vote.").

[127] *See Waiting Was the Hardest Part*, COLUMBUS DISPATCH, Nov. 3, 2004, at 1A (describing judicial order that election workers provide paper ballots to voters waiting in long lines).

[128] *See, e.g.*, Warren Richey, *GOP Slips at Foley Scandal's Epicenter*, CHRISTIAN SCI. MONITOR, Oct. 27, 2006, at USA2, (referencing a judge's order that poll workers not explain to voters why disgraced Florida Congressman Mark Foley's name was on the ballot instead of the name of the actual candidate Joe Negron).

[129] *See* Southwest Voter Registration Educ. Project v. Shelley, 344 F.3d 882, 912, *rev'd on reh'g en banc*, 344 F.3d 914 (9th Cir. 2003); Gilmore v. Green County Democratic Party Exec. Comm., 368 F.2d 328, 329 (5th Cir. 1966); *cf.* Chisom v. Roemer, 853 F.2d 1186, 1192 (5th Cir. 1988) (vacating trial court's order enjoining judicial election).

[130] *See Primary Elections Are Cancelled*, N.Y. L.J., Sept. 12, 2001, at 3.

already underway on September 11, 2001, was obviously the right decision, given how completely disrupted New Yorkers' lives were that day.[131]

* * *

Given the preceding range of available remedies, most states' existing election contest statutes do little to constrain a court's choice when some election irregularity is proven. A typical contest statute provides only that the finder of fact is to determine who received the majority of votes, or otherwise declare the result.[132] Alternative versions specify that the court may order a new election for "irregularities of sufficient magnitude to cast doubt on the validity,"[133] or provide no express identification of possible remedies.[134] As one Indiana trial court recently noted, "Indiana election law provides little insight into the appropriate remedy available in this proceeding. Case authority [interpreting such statutes] on election contests provides virtual[ly] no guidance for circumstances where widespread misconduct has impacted the absentee ballots cast in an election."[135]

Furthermore, with the dramatic exception of the scholarship on the 2000 presidential contest, academic commentary also has focused scant attention on the comparative strengths of these post-election remedies, or the appropriate circumstances for their use. Understandably, articles about *Bush v. Gore* and the 2000 Florida presidential election have focused primarily on issues of federalism, equal protection, and presidential elections, rather than broader questions about the general appropriateness of election remedies. The overwhelming bulk of legal scholarship concerning the role of courts in policing democratic elections has focused on how to conduct these elections, rather than on ways to correct errors after they have occurred.[136] Likewise, the issue of post-election remedies does

---

[131] *See generally* Diane Cardwell, *Questions Face Elections Board Before Primary*, N.Y. TIMES, Sept. 20, 2001, at A20; Jerry H. Goldfeder, *Could Terrorists Derail a Presidential Election?*, 32 FORDHAM URB. L.J. 523, 525–27 (2005); Clyde Haberman, *Reaffirming Democracy, Here and Now*, N.Y. TIMES, Sept. 22, 2001, at A9.

[132] *See, e.g.*, COLO. REV. STAT. § 1-11-216 (2006); N.J. STAT. ANN. § 19:29-8 (West 2006); MISS. CODE ANN. § 23-15-951 (West 2006); WYO. STAT. ANN. § 22-17-08 (2006); N.D. CENT. CODE § 16.1-16-08 (2006) (prohibiting voiding election unless the contestee "connived" to produce illegal votes or number of illegal votes is greater than margin of victory).

[133] MO. REV. STAT. § 115.593 (2006).

[134] *See, e.g.*, HAW. REV. STAT. § 11-174.5 (2006); KAN. STAT. ANN. § 25-1448 (2006). The Texas statute provides a particularly succinct statement: a tribunal may "declare the outcome" if it "can ascertain the true outcome" but "shall declare the election void if it cannot ascertain the true outcome." TEX. ELEC. CODE ANN. § 221.012 (West 2006).

[135] Pabey v. Pastrick, 816 N.E.2d 1138, 1140 (Ind. 2004) (quoting Pabey v. Pastrick, No. 45D10-0305-MI-007, at 99 (Ind. Super. Ct. 2003)).

[136] One of the two principal election law casebooks devotes one chapter to describing "Remedial Possibilities for Defective Elections." *See* ISSACHAROFF ET AL., *supra* note 122, at 1039–88. The other principal election law casebook contains only isolated, brief refer-

not appear to be on the reform agendas of any of the prominent policy organizations most actively studying the election process.[137] Accordingly, a more systematic evaluation of the appropriate remedies for election failures is necessary. The remaining Parts of this Article begin that effort.

II. FUNDAMENTAL VALUES UNDERLYING THE CHOICE OF APPROPRIATE ELECTORAL REMEDIES

Having set out a sample of election problems and the kinds of remedies typically available to resolve them, this Article now considers several interrelated and sometimes conflicting values of our election system that should constrain or shape the aforementioned remedies. These values include (1) fairness and legitimacy; (2) voter anonymity; (3) accuracy and transparency; (4) promptness and finality; and (5) efficiency and cost.[138] Moreover, to the extent that the judicial branch is responsible for implementing the remedies for election failures, the desirability of these remedies must also be assessed in terms of their implications for separation of powers.

Indeed, even before thinking about how various judicial remedies are shaped by or measure up against each of these values, it bears emphasis that it is highly preferable to avoid judicial involvement in elections altogether, especially after voting has begun.[139] It is therefore crucial to continue to refine other election processes in order to reduce the need for election-day or post-election remedies. Any after-the-fact solutions risk at the very least being perceived as upsetting normal democratic processes, and thereby undermining the legitimacy of election outcomes. Accordingly, all election "reforms" should be evaluated in terms of their impact on the margin of litigation.[140]

Efforts to minimize the likelihood of post-election litigation are especially important in developing the standards and procedures governing

---

ences to the topic of remedies. *See* DANIEL H. LOWENSTEIN & RICHARD L. HASEN, ELECTION LAW (2d ed. 2001).

[137] These organizations include The National Research Commission on Elections and Voting, *see* http://elections.ssrc.org; electionline.org, *see* www.electionline.org; The Century Foundation, *see* http://www.reformelections.org; The Brennan Center for Justice, *see* http://www.brennancenter.org/subpage.asp?key=38&projkey=76; and The United States Election Assistance Commission, *see* www.eac.gov. Similarly, the National Conference of State Legislatures has not focused its efforts on encouraging and tracking state reforms in this area. *See* NCSL ELECTION REFORMS TASK FORCE, THE STATES TACKLE ELECTION REFORM, Mar. 24, 2003, http://www.ncsl.org/programs/legismgt/elect/taskfc/electaskfc.htm.

[138] These are by no means the only values important to our election processes, but they are the values most directly implicated in the choice of how to remedy election failures. Other values, such as accessibility to the ballot or the security and integrity of the voting process, may have greater salience in evaluating other components of an election system. *See, e.g.*, Wang, *supra* note 32, at 354; Tokaji, *supra* note 20, at 1774.

[139] *See* Hasen, *supra* note 31, at 995–99; *see generally* JEFFREY ROSEN, THE MOST DEMOCRATIC BRANCH: HOW THE COURTS SERVE AMERICA (2006).

[140] *See supra* note 13 and accompanying text.

absentee and provisional ballots. To date, a large fraction of election contests have involved absentee ballots.[141] Now that federal law has mandated the use of provisional ballots,[142] the dramatic increase in the number and proportion of provisional ballots cast may result in substantially more post-election challenges involving these kinds of ballots as well.[143] For disputes involving either type of ballot, the best reforms will reduce the causes of contested elections rather than merely regularize the manner in which electoral systems deal with the controversies once they have already developed. States should thus strive both to ensure that the availability of absentee ballots does not increase opportunities for fraud and to reduce the need for provisional ballots (while still guaranteeing a provisional ballot to any voter who is not allowed to vote a regular ballot and ensuring that all voters are provided with convenient voting options). But where best efforts fail to prevent a post-election controversy, the available remedies should seek to honor the following fundamental, though potentially conflicting, values.

A. Fairness and Legitimacy

As with the election process itself, remedial processes need to be both fair and perceived as fair.[144] Although fairness can be evaluated in a variety of ways, one measure of fairness is the degree to which a system treats candidates equally (as opposed to favoring one candidate or one type of candidate over another). For instance, a system that always awards an unresolved election to the incumbent or to the incumbent's party would not be accepted as fair, even though it would have the virtue of producing

---

[141] *See supra* notes 35–38, 43, and accompanying text.

[142] *See* HAVA, § 302, 42 U.S.C. § 15482 (Supp. 2006).

[143] Such challenges often revolve around the question of *which* provisional ballots to count, an issue that may loom large in any election close enough to trigger a recount. The standard for determining this issue has been the subject of intense scrutiny during the past several years. *See, e.g.*, Gerald M. Feige, *Refining the Vote: Suggested Amendments to the Help America Vote Act's Provisional Balloting Standards*, 110 PENN ST. L. REV. 449 (2005); Edward B. Foley, *The Promise and Problems of Provisional Voting*, 73 GEO. WASH. L. REV. 1193 (2005); David C. Kimball, Martha Kropf & Lindsay Battles, *Helping America Vote? Election Administration, Partisanship, and Provisional Voting in the 2004 Election*, 5 ELECTION L.J. 447 (2006); *Two Steps Forward, One Step Back, and a Side Step: Asian Americans and the Federal Help American Vote Act*, 10 ASIAN PAC. AM. L.J. 31, 58 (2005). Administrative and legislative guidelines addressing this issue are already plentiful. *See, e.g.*, MICH. COMP. LAWS ANN. § 168.813 (West 2005) (stating that a provisional ballot should only be tabulated if a voter's valid voter registration is located or if an elector's identity is verified with acceptable identification such as a driver's license and verification of current address).

[144] The typical mantra of healthy elections is that they be "free and fair." For instance, the U.S. State Department has published a series of one-page primers on the fundamentals of democracy, including a discussion of the key principles of democratic elections titled "Free and Fair Elections." U.S. DEP'T OF STATE, PRINCIPLES OF DEMOCRACY: FREE AND FAIR ELECTIONS (2005), *available at* http://usinfo.state.gov/products/pubs/principles/election.htm.

a prompt resolution of the election. More generally, a remedial system with any built-in bias that favors incumbents, the majority party, or candidates with friends or relatives on the boards of elections, could not be considered fair. Avoidance of such bias is therefore a key consideration in structuring a sound remedial system.

Another measure of fairness is the degree of equality with which a system treats votes. This measure was at the heart of the U.S. Supreme Court's opinion in *Bush v. Gore*, which invalidated the 2000 Florida recount under the Equal Protection Clause because the recount standards varied from county to county.[145] Greater uniformity in recount and contest procedures would reduce the potential for certain votes to disproportionately affect election outcomes. The extent to which *Bush v. Gore* ultimately will be read to require that election contest procedures treat votes equally remains to be seen,[146] but this goal of equal treatment of votes is laudable, whether motivated by equal protection concerns or simply an interest in fundamental fairness.

Ensuring that the public perceives and accepts a remedy as fair is equally important to the legitimacy of an election remedy. Public acceptance of the process through which the system resolves election failures ultimately involves a complex mix of overlapping factors that sometimes are in tension with each other. For instance, the public not only must believe that the remedial system treats both candidates and voters equally, but also must be confident that the system will protect anonymous voting while allowing an accurate accounting of the outcome.

B. Voter Anonymity

Although it was not always the case, anonymity of voting is a fundamental principle of American democracy today.[147] Indeed, in some states, secret voting is protected in the state constitution.[148] Anonymity protects against fraud, coercion, bribery, and other forms of corruption that might otherwise occur through vote selling.[149]

Anonymity does, however, come with its own price: it makes auditing election returns much more difficult, complicating the exclusion of improper votes and the inclusion of improperly excluded votes. Indeed,

---

[145] 531 U.S. 98, 110 (2000).

[146] *See* Edward B. Foley, *The Future of* Bush v. Gore, 68 OHIO ST. L.J. (forthcoming 2007).

[147] Secret ballots became the norm in the United States after the Populists made them part of their platform in the late nineteenth century. *See, e.g.*, *In re* Hearst, 76 N.E. 28, 29–30 (N.Y. 1905) (describing New York legislature's adoption of secret ballot); *see also* Fortier & Ornstein, *supra* note 35, at 487–92 (describing the origins of the "Australian" or secret ballot and its adoption in the U.S.).

[148] *See, e.g.*, ARIZ. CONST. art. VII, § 1; CAL. CONST., art. 2, § 7; NEB. CONST. art. VI, § 6; UTAH CONST. art. 4, § 8; WASH. CONST., art VI, § 6; WIS. CONST. art. III, § 3.

[149] *See* Fortier & Ornstein, *supra* note 35, at 489–90.

courts on occasion have concluded that resolving the outcome of an election required stripping some voters of their ordinary anonymity.[150] In most cases, however, courts have taken this step only with respect to voters who themselves engaged in fraudulent voting behavior.[151] Some states prohibit disclosure of voter identity entirely.[152] In contrast, until it was repealed in 2002, a provision in the Arkansas Constitution required each ballot to be numbered in such a way as to permit tracing of an individual ballot to an individual voter.[153] Judgments about the need for ballot secrecy thus critically affect the remedies available for election failures.

C. Accuracy and Transparency

Because democratic legitimacy depends on a system in which votes determine political representatives and policy choices, any healthy democracy must have a mechanism for accurate and reliable voting.[154] Similarly, the core function of a system of remedies for election failures is to restore accuracy and reliability where it has been compromised. But other concerns, including ballot secrecy and the need for a prompt, final resolution of election outcomes, may limit the effective restoration of reliability to a failed election.

As a corollary, when allegations of irregularities have raised doubts about the reliability of an election, the public must be able to observe and partake in the remedial process by which those irregularities are investigated and resolved. Transparency of process, therefore, can become even more important in a failed election than in a routine one. When a court or other tribunal acts to remedy an election irregularity, the public must be

---

[150] *See, e.g.*, *In re* General Election for Dist. Justice, 670 A.2d 629, 638–39 (Pa. 1966) (allowing five voters whose valid ballots had been tampered with to reveal their votes voluntarily).

[151] *See, e.g.*, Mahaffey v. Barnhill, 855 P.2d 847, 850 (Colo. 1993) (reaffirming the principle that citizens who had cast invalid votes could be compelled to testify as to how they had voted, but only if they had not voted in good faith); *see also* TEX. ELEC. CODE ANN. § 221.009 (Vernon 2003) (providing that voters who cast invalid votes can be compelled to disclose their votes, but also providing that the tribunal need not compel such disclosure, even if the number of invalid votes is sufficient to create doubt about the outcome).

[152] *See, e.g.*, McCavitt v. Registrars of Voters, 434 N.E.2d 620, 630 (Mass. 1982).

[153] *See* ARK. CONST. amend. L, § 3 (repealed 2002); *see also* Womack v. Foster, 8 S.W.3d 854, 868 (Ark. 2000) (observing that framers of this constitutional provision "chose to continue to subordinate the secrecy of the ballot to the purity of the election"). Similarly, in Britain votes are recorded with unique identifiers precisely to permit tracing in the event of a dispute about the legality of a particular vote. *See* ELECTORAL COMMISSION, BALLOT SECRECY FACTSHEET (Dec. 29, 2006), http://www.electoralcommission.org.uk/templates/search/document.cfm/6127. Another report claims that the use of touch screen voting can create a similar tracing capability. *See* MARK MILLER, SECRET BALLOT COMPROMISED IN GEORGIA!, http://www.countthevote.org/no_secret_ballot.htm (last visited Feb. 21, 2007).

[154] A variety of research has discussed the importance of accurate vote tabulation and the relative accuracy of different voting mechanisms. *See, e.g.*, Tokaji, *supra* note 20, at 1717–41.

able both to understand why the election failed and to accept how it will be fixed. The public must also have confidence that it will be fixed fairly and not arbitrarily.[155]

These concerns may make proportional adjustments less desirable,[156] particularly where the demographic or statistical analyses underlying them are relatively inaccessible to the public, or where the primary message such adjustments convey to the public is uncertainty regarding the election's true outcome. On the other hand, when in fact an election's true outcome is indeterminable, the electoral process should acknowledge that reality and the associated trade-offs between holding a new election versus finding some neutral way of picking a winner.

Election irregularities that do not call into question the ultimate outcome present a more mundane issue. Should the public still be entitled to an accurate accounting of its preferences? That is, might more than just the ultimate outcome of the election, such as the strength of the voters' preference (or the size of a victor's mandate), also be worthy of an accurate tally? Yet while people may often speak of "counting every vote," the public is primarily concerned with simply counting enough votes to be confident in the outcome. As a practical matter, remedying election failures that do not alter the outcome is a luxury that society cannot afford, either in time or expense.[157]

D. Promptness and Finality

In addition to expecting fair elections with accurate results, the public also appropriately expects to have these election results shortly before the office becomes vacant, rather than long before (or after) the vacancy. Were temporal proximity not a priority, elections could be held well ahead of time (much as in a monarchy the heir apparent may be selected well in advance), which would provide ample opportunity for recounts, contests, revotes, and the careful resolution of any issues that arise in an election. Yet because the issues facing politicians are constantly evolving, and because politicians may frequently change their stripes, holding elections roughly contemporaneously with when the victors will take office increases political accountability.

For these reasons, as well as because of the administrative difficulties of leaving an office vacant, American elections generally are scheduled to occur as close to the commencement of the term of office as is practical.[158] It therefore can become problematic if election outcomes are con-

---

[155] *See supra* Part II.A.

[156] *See supra* notes 100–104 and accompanying text.

[157] However, acts of fraud that do not call the outcome of an election into question should still be addressed through imposition of civil liability or criminal prosecution to discourage similar acts that may affect the outcome of future elections.

[158] *See* Foley, *supra* note 33, at 104.

tested for months or years after election day. Instead, election processes generally, and remedial options in particular, should be designed to promote the prompt resolution of election outcomes.

In an important sense, a speedy determination of an election protects the integrity of the process.[159] In the aftermath of the 2000 election, a number of scholars and public officials agreed that the country would be ill-served by a protracted legal battle over the presidential election.[160] In extreme cases, protracted election disputes have become moot when the terms of office have expired before the litigation has concluded.[161]

In addition, it is important that representatives serve with full authority and respect, rather than with unresolved questions about their legitimacy. It is therefore preferable that an election contest be final before the term of service begins. Accordingly, as a practical matter, it may be more important for a contested election to be resolved conclusively than that it be resolved perfectly. With unlimited time, we might investigate an election problem more thoroughly and also be more willing to rerun it. But usually we must move on, knowing that the next election will shortly supersede whatever imperfections occurred in the most recent election.

These interests in promptness and finality suggest that we may choose to sacrifice some absolute certainty about an election outcome for the expediency of the result. They also suggest that the remedy of ordering a new election ought to be disfavored compared to conclusively determining the winner of a completed election, whenever this approach can produce a fair and acceptable outcome.

E. Efficiency and Cost

A reality of the American system of election administration is that it is run on a shoestring budget,[162] particularly compared to the scale of the undertaking and the financial resources otherwise involved in our elections.[163] For instance, we minimize costs by relying on millions of essen-

---

[159] *See In re* 2003 Election for Jackson Twp. Supervisor, 840 A.2d 1044, 1046 (Pa. Commw. Ct. 2003) ("The integrity of the election process requires immediate resolution of disputes . . . ."); Smith v. King, 716 N.E.2d 963, 969–70 (Ind. Ct. App. 1999) ("[E]lection contest procedures . . . manifest a clear legislative intent that election contests be resolved expeditiously.").

[160] *See* R. W. Apple, Jr., *Bush Sues to Halt Recount in Florida: The Limits of Patience*, N.Y. TIMES, Nov. 12, 2000, at A1 (reporting results of interviews with a variety of scholars and public officials).

[161] *See, e.g.*, Gunaji v. Macias, 31 P.3d 1008, 1011 (N.M. 2001) (noting terms of office had expired by time court resolved underlying remedial issue).

[162] *See* NAT. COMM'N ON FED. ELECTION REFORM (THE CARTER-FORD COMM'N), TO ASSURE PRIDE AND CONFIDENCE IN THE ELECTORAL PROCESS 68 (2001).

[163] In addition, it may be that part of the price we pay for our First Amendment freedoms, especially as interpreted to permit unlimited political expenditures, is a frenzied, no holds barred election atmosphere. This atmosphere may further unsettle election outcomes, as losing candidates in close races (and their supporters) may be more likely to challenge election procedures after having spent so much on their campaigns.

tially volunteer laborers to run polling place operations.[164] Although these public-spirited poll workers typically receive a tiny payment for their service and generally are required to (but in fact may not always) participate in a brief training program, they nevertheless frequently lack the experience or background knowledge to respond appropriately and consistently to issues that arise during the election.[165] In addition, they may themselves commit errors that taint an election. A substantial number of mistakes that give rise to election contests therefore might be prevented if we restructured the manner in which we administer elections. We might also be able to prevent more instances of fraud if we invested more money in the security of our election processes.

Cost considerations also play a role in our selection of voting technology. We accept a certain degree of inaccuracy and unreliability, knowing that more accurate, more reliable methods could be developed, but perhaps at a heavy financial cost. Where the accuracy or reliability of a particular election outcome is unsatisfactory, an election contest holds the promise of relief.

Election contests, however, exact their own cost, both in public (and private) dollars and in democratic legitimacy. These costs must be evaluated in relation to the costs of the rest of the election process. In a sense, the decision is between paying for prevention and paying for a cure. But the possibility that states could invest substantially more in preventative measures does not mean that the existing system is inefficient. In the grand scheme, the number of election contests that could be prevented by substantial increases in election administration funding may be comparatively small, resulting in relatively insignificant savings in the costs of election contests.

Furthermore, it is worth noting that both the speed and the accuracy with which we resolve election failures are themselves dependent on the resources that we devote to our remedial processes. For instance, relying on ordinary courts to adjudicate election contests may be more efficient than using election courts with special expertise, yet may produce less accurate or less timely results.

---

[164] *See* Jim Drinkard, *Scarcity of Poll Workers Persists*, USA TODAY, Oct. 26, 2004, *available at* http://www.usatoday.com/news/politicselections/nation/president/2004-10-26-poll-workers_x.htm?csp=19_wxia.

[165] See ADVANCEMENT PROJECT, PLIGHT OF THE POLL WORKER: EFFORTS TO IMPROVE TRAINING AND SUPPORT FOR POLL WORKERS IN OHIO, PENNSYLVANIA, MARYLAND, FLORIDA, AND MICHIGAN 1, 1–2 (2006), *available at* www.projectvote.org/fileadmin/ProjectVote/Publications/Plight_of_the_Poll_Worker-Advancement_Project.pdf.

F. Separation of Powers

As previously noted,[166] absent statutorily conferred authority to adjudicate an election contest, courts traditionally have treated allegations that a particular election outcome was invalid as a nonjusticiable political question.[167] Courts are justifiably wary of interfering in the outcome of the political process, both to protect themselves and to protect the democratically elected branches.[168] To the maximum extent possible, the people, not the courts, should choose their representatives. Accordingly, in those instances in which an election outcome remains unreliable even after a recount, the least intrusive remedy that a court could order would simply be a new election. This remedy would spare the court the difficult burden of identifying the winning candidate and instead would return to the voters the responsibility of determining the election outcome.

Yet rerunning an election imposes significant burdens on other components of our democracy.[169] Conducting elections is hugely expensive, and rerunning an election duplicates this expense, not only for the public treasury and the election officials involved, but to a large extent for the candidates and their campaigns as well. Candidates may have exhausted their resources and be unable to produce an effective second campaign. Rerunning elections may systematically favor certain types of candidates, such as incumbents or established candidates with greater financial resources, who may be in a better position to continue campaigning than challengers or political newcomers. Furthermore, a new election can never be run on a clean slate, but will always be colored by the perceived outcome of the election it superseded. New elections may also be an inconvenience for the voters, and almost certainly will mean that a different set of voters, with different information, will be deciding the election. Moreover, there can be no guarantee that the new election will itself be free from additional problems, including fraud. In the long term, rerunning elections might lead to disillusionment or apathy, even if in the short term

---

[166] *See supra* note 26 and accompanying text.

[167] At the same time, courts have historically adjudicated voters' damage actions alleging that they had been deprived of their personal rights to vote. *See, e.g.*, Memphis Community School Dist. v. Stachura, 477 U.S. 299, 312 n.14 (1986) (collecting cases). But while courts were comfortable deciding that a voter had in fact been disenfranchised (and awarding relief to the voter), they were not comfortable relying on the same facts to examine the validity of the underlying election outcome, absent an election contest brought under a statutory cause of action. *See, e.g.*, *Johnson v. Stevenson*, 170 F.2d 108, 111 (5th Cir. 1948) (distinguishing voters' individual rights, protected under principles of federal law, from candidates' right to political nomination, protected exclusively under state statute); *cf. United States v. Bathgate*, 246 U.S. 220, 226–27 (1918) (distinguishing personal right to vote from political, nonjusticiable, public right to fair election).

[168] *See, e.g.*, Hutchinson v. Miller, 797 F.2d 1279, 1280 (4th Cir. 1986) ("The legitimacy of democratic politics would be compromised if the results of elections were regularly to be rehashed in federal court.").

[169] *See* Huggins v. Superior Court in and for County of Navajo, 788 P.2d 81, 84 (Ariz. 1990) (identifying a range of problems in ordering a new election).

they excite interest in the particular contest. Frequent new elections also would undercut democratic stability by calling into question the security and efficiency of the voting mechanics.

Thus, new elections are generally used only as a last resort, notwithstanding their attractiveness as a means of keeping courts out of the business of selecting election winners.[170] Instead, the preferred remedy in an election contest is a judicial determination of which candidate won. Yet this approach inevitably entangles the courts in the political process and thereby creates uncomfortable pressures on the judicial branch. It therefore is advisable to limit the judiciary's discretionary judgments in election contests as much as possible.

Alternatively, states could remove these issues entirely from the judicial branch. Many states repose the authority to judge the outcome of some of their elections in legislative or administrative bodies.[171] But those bodies often are not politically neutral or free from self-interest, and in many instances may reach outcomes through votes that fall closely along party lines. Nevertheless, as discussed below,[172] this may sometimes be an acceptable, even preferable, approach.

III. Remedial Values Applied: Some Recent Examples

The values discussed in Part II should limit and guide the selection of an appropriate remedy for an election failure. Yet largely unsettled is how we should prioritize these values when they conflict. As the beginnings of an effort both to prioritize these interrelated but sometimes conflicting values and to establish a framework for resolving election contests, this Part applies these values to several prominent examples of recent election problems.

A. Inaccurate Voting Because of Ballot Design Problems

A number of recent election controversies have involved problems in ballot design, including the 2001 mayoral race in Compton, California;[173] the 2000 presidential election in Palm Beach County, Florida;[174] and the 2006 race for Florida's 13th congressional seat in Sarasota County.[175] Even though these design flaws arguably impeded each election's ability to

---

[170] However, in some states the only available remedies are either to validate the entire election or to find it void, in which case a new election is required to fill the seat. *See, e.g.*, Becker v. Pfeifer, 588 N.W.2d 913, 918 (S.D. 1999) (explaining that a court adjudicating an election contest must either uphold the entire election or declare it void).

[171] *See* Iowa Const. art. 4 § 5; Colo. Rev. Stat. Ann. §§ 1-11-205, 1-11-208 (West 2007); Mass. Code Ann. § 23-15-923 (West 2007).

[172] *See infra* notes 292–295and accompanying text.

[173] See *infra* notes 176–181 and accompanying text.

[174] *See supra* notes 47–48 and accompanying text.

[175] *See supra* note 46 and accompanying text.

determine voters' true preferences, the value of both finality and judicial restraint—and the impact of both of these factors on the public's acceptance of election outcomes—caution against permitting these and similar problems to give rise to court adjustments or invalidations of election results.

It was inappropriate, for example, for the trial court to reverse the outcome of the 2001 Compton mayoral race on the basis of a flaw in the ballot layout.[176] There, the incumbent alleged that an error in determining where his name was positioned on the ballot had cost him the election.[177] The court relied on an expert witness's statistical analysis concerning the "primacy effect," a theory that the candidate listed first on the ballot is thereby at an advantage.[178] In this case, the expert claimed that the primacy effect caused the first-listed candidate, on average, to receive 3.32% more votes than candidates further down the ballot.[179] On this basis, the trial court switched 306 votes from the apparently victorious challenger, who had been listed first on the ballot, to the incumbent.[180] With only a 281 vote margin, the switch of these 306 votes reversed the outcome.[181]

Of course, nothing about any one of these 306 votes suggested that it was cast in error or contrary to a voter's true preference. Further, even if the primacy effect (assuming that the primacy effect accurately described what happened in this election, which itself is an uncertain proposition[182]) means that an election outcome may turn on a random choice about which candidate is listed first, there is no basis in law for disenfranchising those voters whose sole reason for preferring one candidate might happen to be the candidate's position on the ballot. The trial court's action, however, caused this very disenfranchisement. A court might just as well have decided that bad weather on election day kept a disproportionate number of Democratic voters from the polls (as some studies have suggested typically occurs),[183] and on that basis reversed a Republican candidate's narrow victory.

Because such problematic design flaws are too vague and standardless to permit objective evaluation, this level of judicial refereeing of elec-

---

[176] *See* Bradley v. Perrodin, 131 Cal. Rptr. 2d 402, 405 (Cal. Ct. App. 2003).

[177] *See id.* at 406.

[178] For a thorough review of the primacy or "ballot order" effect, see Alvarez et al., *supra* note 18.

[179] *Bradley*, 131 Cal. Rptr. 2d at 406 n.2.

[180] *Id.* at 406.

[181] *Id.*

[182] *See* Alvarez et al., *supra* note 18, at 46–52 (presenting statistical analyses that discount the primacy effect hypothesis).

[183] *See, e.g.*, Brad T. Gomez, Thomas G. Hansford, & George A. Krause, *The Republicans Should Pray for Rain: Weather Turnout and Voting in U.S. Presidential Elections* 69 J. POL. (forthcoming 2007), *available sub nom. The Effect of Bad Weather on Voter Turnout and Partisan Vote Share in U.S. Presidential Elections*, 1948–2000, *available at* http://people.cas.sc.edu/gomezbt/WeatherPaper_v2.pdf (concluding that rain reduces turnout by one percent per inch of rain).

tions after they have occurred can only lead to increased public cynicism about the legitimacy of those elections. Instead, we must either tolerate these structural or design flaws as part of what will always be an imperfect process (just as close election outcomes may sometimes turn on the weather, at least as long as voting remains limited to one day and must ordinarily occur in person), or else identify and correct them before voting begins. For instance, to the extent that the primacy effect is a real factor in election outcomes, the proper remedy is to defeat or randomize this advantage before the election by requiring ballot order rotation across precincts (as a number of states already do),[184] or to recognize and accept that whatever method is used to determine ballot order (such as listing incumbents first or allowing the Secretary of State to decide which party's candidates to list first) may thereby confer an advantage.

Similarly, Palm Beach County's infamous butterfly ballot in the 2000 presidential election did not justify judicial correction after the fact. Even though this design flaw, in contrast to the primacy effect, may well have led some voters to mark their ballots for one candidate when in fact they meant to pick another candidate, nothing about this flawed ballot design necessarily precluded or prevented any voter from accurately registering the voter's true preference.[185] It therefore is impossible to know just how many voters in fact miscast their ballots. Furthermore, all candidates and the public had an opportunity to critique the ballot design before the election took place.[186] In these circumstances, any statistical adjustment of the votes would lack transparency and objectivity,[187] and therefore would lead inevitably to accusations of judicial favoritism. Throwing out the election entirely would be a more neutral remedy than adjusting the returns, and therefore preferable in this regard. But requiring the extraordinary remedy of a new election to fix a flaw that could have been fixed ex ante, and did not necessarily preclude any voter from casting a true vote, is simply asking too much of our system.

The same analysis applies to the flawed touch screen layout of the Sarasota County ballot in 2006, the design of which seems the most likely explanation for the astonishingly high number of undervotes in the race for Florida's 13th congressional district.[188] Although it appears that the placement of this congressional race at the top of a second ballot screen,

---

[184] *See, e.g.*, ARIZ. REV. STAT. ANN. § 16-464 (2007); KAN. STAT. ANN. § 25-2115 (2007); MONT. CODE ANN. § 13-12-205 (West 2007); WYO. STAT. ANN. § 22-6-122 (2007).

[185] In other words, every voter using the butterfly ballot had an opportunity to overcome whatever potential confusion existed in the ballot design and to cast a ballot that accurately registered the voter's preference.

[186] *See NewsHour with Jim Lehrer: Florida Recount* (PBS television broadcast Nov. 28, 2000), *available at* http://www.pbs.org/newshour/bb/election/july-dec00/f l_11-28.html.

[187] Transparency and objectivity would almost certainly be lacking in the sense that the public would have difficulty understanding how a contest tribunal neutrally settled upon how many votes to adjust. *See supra* Part II.C.

[188] *See supra* note 46.

which featured additional contests more prominently below it, may have led some voters to overlook this race, nothing about this placement precluded any voter from casting a vote in the congressional race.[189] In addition, this design flaw also could have been identified and fixed ex ante.[190]

In short, courts simply should not void elections or reverse validly cast votes because of imperfections in the way that a ballot was designed, as long as the imperfections have not precluded any voters from communicating their true preferences.[191] Otherwise, public cynicism about the legitimacy of an election would only increase as a result. Instead, the onus should be on candidates and election administrators to identify and fix ballot design flaws ahead of time.

B. Lost Votes Because of Balloting Failures

In contrast to the problems of poor ballot design, other kinds of voting problems necessarily preclude voters from meaningfully registering their preferences. Such failures include ballots that omit particular candidates or races;[192] the loss or destruction of some subset of marked ballots or voting machine memory cards;[193] and the failure to timely mail absentee ballots to voters who have properly requested them.[194] Each of these failures precludes some voters from communicating their choice, resulting in the loss or exclusion of their votes from the official tally. When the number of lost votes exceeds the margin of victory in a contested race, this type of failure thus often merits a judicial response.

Ordinarily, some form of new election will be the most appropriate solution for lost votes that could have determined the election, despite the burdens of this remedy. This approach obviously promotes accuracy and legitimacy and minimizes separation of powers concerns, but sacrifices promptness, efficiency, and lower costs. Fortunately, in many cases such

---

[189] A design that makes it more likely that some subset of voters will overlook the race is categorically different from a ballot design that omits a race entirely, for instance, thereby precluding or making it impossible for voters using that ballot to register their preference.

[190] For instance, Florida, like most states, allows candidates and the public to inspect ballots and to observe the testing of voting equipment prior to the election. *See* FLA. STAT. § 101.5612 (2006).

[191] Accordingly, calling for a new election because the ballot paper was too thin to permit secret voting, *see* Hester v. Kamykowski, 150 N.E.2d 196, 200–01 (Ill. 1958), would be overreaching without some additional showing that the thin paper in fact altered some voters' ability to register their vote fairly.

[192] *See, e.g.*, Gunaji v. Macias, 31 P.3d 1008, 1010 (N.M. 2001).

[193] *See More than 4500 North Carolina Votes Lost Because of Mistake in Voting Machine Capacity*, USA TODAY, Nov. 4, 2004, *available at* http://www.usatoday.com/news/politics elections/vote2004/2004-11-04-votes-lost_x.htm.

[194] *Cf.* Lisa Abraham, *Ballot Postage Problem Licked; Post Office Will Deliver Absentee Votes Anyway*, AKRON BEACON J., Oct. 31, 2006, at A1 (describing potential problem of insufficient postage compounded by lack of time to repost, attributed in part to tardy mailing of absentee ballots).

inefficiencies and costs could be reduced by limiting the new election to just those voters whose votes are known to have been lost.[195]

It also would be sensible to identify some statistical threshold beyond which a new election would not be in order because the probability that the lost votes would overcome the margin of victory is too remote. For instance, if 100 votes have been lost from a precinct whose voters are historically split roughly evenly between the two major parties, but the margin of victory is currently 90 votes, then barring some unusual demographic features of the lost 100 votes, there may simply be too little likelihood that more than 90 of these lost votes would have been for the trailing candidate, despite the hypothetical possibility of such an outcome. The problem of identifying a threshold to trigger a new election is discussed in more depth below,[196] but until legislatures provide specific guidelines, it is sufficient for now to note that courts may need to employ their discretionary authority in deciding when the mathematical possibility of a different outcome is truly a realistic possibility as well.

The line between defects that wholly preclude a voter from casting a meaningful vote, and defects that merely increase the chances that some voters may fail to cast a vote as intended, may not always be clear. Nevertheless, this distinction is helpful in analyzing the appropriateness of remedies for election defects. For instance, consider the difficulty of characterizing the defects of punch card voting systems, which are problematic insofar as votes may be lost or go uncounted if voters have not punched the chad cleanly out of the punch cards. To some, this may be a problem of insufficiently vigilant voters, who are not necessarily precluded from casting accurate votes (especially if voters are reliably instructed to fully detach their punched-out chad). Others may view it as a systemic defect that will inevitably result in the loss of some portion of votes, and which could be avoided through an alternative voting system. Indeed, this latter view may explain why in the run-up to the 2003 California gubernatorial recall election, a Ninth Circuit panel ordered the election postponed until California could replace its punch card equipment.[197]

Of course, a decision to enjoin an impending election because of concerns about a design flaw is entirely different from a decision to void or adjust the outcome of a completed election. However, once courts have decided ahead of time that a voting system or ballot is appropriate notwithstanding its known design flaws (as the Ninth Circuit did in its en banc rehearing of *Shelley*),[198] that determination should preclude any post-election contest over previously approved election processes. The alternative would be too destabilizing.

---

[195] *See infra* notes 233–234 and accompanying text.

[196] *See infra* Part IV.B.4.

[197] *See* SW. Voter Registration Educ. Project v. Shelley, 344 F.3d 882, *rev'd after reh'g en banc*, 344 F.3d 914 (9th Cir. 2003).

[198] *See* SW. Voter Registration Educ. Project v. Shelley, 344 F.3d 914 (9th Cir. 2003).

C. Discouraged Votes Because of Long Lines or Late Poll Openings

Votes can also effectively be lost if voters are discouraged from getting to the voting booth in the first place, as when lines at polling places are long,[199] or when polling places do not open on schedule.[200] Although both of these kinds of election failures have occurred repeatedly in recent years, unfortunately no good method exists for recovering this sort of lost vote after the fact. It generally would be a practical impossibility to identify accurately those voters who would otherwise have voted absent these difficulties, or to make any other after-the-fact determination to "add" lost votes to the tallies. At the same time, in most circumstances it would unduly compromise the values of finality and efficiency to void the entire election on the basis that voting was not sufficiently convenient for some voters. Such a response might be appropriate only in the most egregious circumstances, perhaps if an alarming portion of voters has been affected or if there is compelling evidence that these burdens have deliberately been imposed on only certain types of voters.

Instead, even the more extreme occurrences of these problems should ordinarily give rise only to injunctive relief on election day, to maximize the possibility that inconvenienced voters can find an alternative way to cast their ballots that day. As previously discussed,[201] injunctive relief can include extending polling place hours or making paper ballots available as an alternative to voting machines. Of course, problems of this sort call for greater pre-election preparations as well, and some less serious instances may already be manageable through such existing arrangements as the typical state requirement that all voters who are in line at the official poll closing time be allowed to vote.[202]

D. Accidental Inclusions of Unlawful Votes

The converse of omitting lost votes in election tallies is including improper votes. Election officials often do not realize that invalid votes have been included until after these votes are commingled with valid votes and it is no longer possible to isolate and exclude only the invalid votes. Under these circumstances, the appropriate remedy depends in part on whether the invalid votes have been included accidentally, as discussed in this subsection, or fraudulently, as discussed in the next subsection.[203]

---

[199] *See* George Merritt & Katy Human, *Voting Problems Overwhelm City*, DENVER POST, Nov. 7, 2006, *available at* http://www.denverpost.com/economy/ci_4620304; *Waiting Was the Hardest Part*, COLUMBUS DISPATCH, Nov. 3, 2004, at 1A.

[200] *See supra* note 126.

[201] *See supra* Part I.B.5.

[202] *See supra* note 126.

[203] *See infra* Part III.E.

The 2004 Washington governor's race provides a recent example of accidental inclusion of improper votes. At the conclusion of the recount process, the final margin was 129 votes, out of almost three million votes cast.[204] In the subsequent judicial contest proceedings, the trial court found that almost 1700 unlawful votes were included in the returns.[205] The vast majority of these were votes cast by ineligible felons, although almost 200 were provisional votes that should not have been counted.[206] The court also found other scattered types of illegal voting, including a few cases in which voting apparently had occurred on behalf of deceased persons, as well as the apparent loss of some valid absentee ballots.[207] However, the court found no evidence of fraud by or on behalf of the candidates.[208]

The court rejected petitioners' requested remedy of proportional deduction, which would have entailed the deduction of a portion of the 1700 unlawful votes from the candidates' respective totals according to the proportions of the vote in the precincts in which the illegal votes were cast.[209] Instead, the court ultimately upheld the election results notwithstanding the presence of the unlawful votes.[210] Guiding the court was a provision in Washington's election code that permits a court to invalidate an election result on the basis of illegal votes only if the illegal votes can be proven to be outcome-determinative.[211] Because only a tiny handful of the illegal votes were affirmatively shown to have favored a particular candidate, and the rest could not be shown to have disproportionately favored either candidate, the election outcome withstood challenge despite the illegal votes.[212]

As discussed further below,[213] some statistical adjustments of vote totals may yet hold promise in correcting for unlawful votes. But without both substantial refinements of the demographic analyses required for such methods, and public acceptance of the application of these methods to election errors, such adjustments will continue to lack sufficient transparency to be perceived as fair and to insulate courts from accusations of political interference. Accordingly, if adjusting the vote totals is an unattractive option, the primary remedial choice, when improper votes are known to be included in the tallies but cannot be ascribed to a particular candidate's total, is whether to order a new election or to permit the flawed election to stand undisturbed.

Even if the election is close, as in Washington 2004, this choice should turn primarily on whether the amount of illegal voting is de minimis,

---

[204] *See* Borders Transcript, *supra* note 103, at 5.
[205] *See id.* at 19.
[206] *See id.*
[207] *See id.* at 13.
[208] *See id.* at 14.
[209] *See id.* at 16–17; *see also supra* note 85 and accompanying text.
[210] *See* Borders Transcript, *supra* note 103, at 24.
[211] *See* WASH. REV. CODE ANN. §29A.68.110 (West 2007).
[212] *See* Borders Transcript, *supra* note 103, at 9, 21–24.
[213] *See infra* Part IV.B.4.

or instead is aberrational. When the errors are of a de minimis sort, a new election offers no advantage—as similar errors are likely to recur—while bringing the drawbacks of delay, inefficiency, and cost. In contrast, when the errors are more extensive than the "background noise" that may be inevitable in most elections,[214] concerns for fairness and accuracy will justify a new election.

Of course, determining whether the errors are typical or acceptable may be problematic without some widespread agreement about the degree of tolerable imperfections. Nonetheless, courts ought to be capable of comprehending the practical difficulties of conducting flawless elections, even as election administrators continue to strive to reduce voting errors. Accordingly, one way of thinking about whether voting irregularities merit a new election is in terms of identifying an acceptable margin of election error, recognizing that we cannot and should not expect perfection in our election processes.[215] In essence, this reflects the inherent trade-off between the values of promptness and cost, on the one hand, and the value of accuracy, on the other. This may in turn require some public discussion and awareness, which ideally should occur independently of a particular election contest.

In addition to the 2004 Washington governor's race, another example of an accidental inclusion of improper votes is the 2004 local Ohio school levy in which two voters innocently voted twice.[216] These voters, husband and wife, voted first by absentee ballot, and then in person when they were told, incorrectly, that their absentee ballots had not arrived.[217] Given that the levy failed by one vote, their improper second votes were potentially determinative.

In the minds of some, this circumstance called for breaching the secrecy of the ballot and requiring the two voters to disclose how they voted.[218] This is a common approach taken when particular voters are shown to have voted fraudulently, on the theory that by their fraud they have forfeited their rights to ballot secrecy.[219] However, it is problematic to sacrifice the ballot secrecy of voters who have not engaged in any wrongdoing.[220] At the same time, requiring a new election in this circumstance seems grossly inefficient when simply questioning the two voters who voted twice could obviate the enormous expense, as well as the delay, of rerunning the election.

Instead, perhaps it is worth thinking of an election this close as a functional tie, and therefore being satisfied with a coin toss or the drawing of

---

[214] *See* Foley, *supra* note 33, at 109–10.
[215] *See id.*
[216] *See* Zachariah, *supra* note 41, at D8.
[217] *See id.*
[218] *See id.*
[219] *See supra* note 151 and accompanying text.
[220] *See supra* Part II.B.

lots as a fair, prompt means of settling a contest over a few improper ballots, just as most election systems call for some such resolution in the event of a mathematical tie.[221] No election system presently makes provision for such a resolution merely in the event of a "close" election, but we might be willing to accept such a remedy if the outcome is so close as to already be within that narrow margin of error that we should realistically be prepared to tolerate in any election, and no votes (or only a number smaller than the margin of error) are disputed.[222]

Of course, one alternative to drawing lots in these circumstances of a functional tie is instead to let the precontest result stand, and use it as the equivalent of a tie-breaking mechanism. Arguably, this is one way of interpreting the Washington statute that governed the 2004 governor's contest.[223] Thus, when the election is a functional tie, just as when the impact of a "typical" de minimis number of illegal votes cannot be identified, it may not be worth the cost, delay, and other trouble of rerunning the election. On the other hand, when the number of disputed votes is larger than this acceptable margin of error, it may be necessary to invalidate the election and rerun it for the sake of legitimacy and accuracy.

E. Absentee Ballot Fraud

Election tallies may also include unlawful votes, not through error or inadvertence, but as a result of deliberate fraud. Indeed, the potential for election fraud remains high on the list of public concerns about our election system. In recent years, the most frequent locus of fraudulent voting has been the absentee ballot.

For instance, in the 1997 Miami mayoral race, widespread absentee ballot fraud rendered the outcome uncertain.[224] Moreover, the fraud was shown to have advantaged a particular candidate, even though the candidate was not linked to the fraud.[225] Although one remedial option in these circumstances would be to void the entire election, a preferable approach, where possible, is to void only those portions of the election shown to be tainted by the fraud. Following this approach, the candidate who did not benefit from the fraud may often end up the victor, assuming the race was

---

[221] *See, e.g.*, ME. REV. STAT. ANN. tit. 21, § 732 (2007); MONT. CODE ANN. § 13-16-501 (2007); N.H. REV. STAT. ANN. § 669:36 (2007).

[222] The practical difficulty is that the candidate with the greater number of absolute votes will claim to be the outright "winner," even if the margin is too close to justify bestowing that label on one candidate with any greater confidence than on the other. We are likely to be able to treat very close races as functional ties only if candidates and the public recognize that within some narrow margin, we in fact cannot say with any confidence which candidate was truly the voters' preferred choice.

[223] *See supra* notes 209–211 and accompanying text.

[224] *See In re* Protest of Election Returns and Absentee Ballots in the Nov. 4, 1997 Election for Miami, 707 So. 2d 1170, 1172 (Fla. Dist. Ct. App. 1998).

[225] *See id.*

close prior to the exclusion of the fraudulent category of votes. By avoiding a new election, this remedy favors promptness over accuracy, and more important, it also protects the fairness and integrity of the election process by depriving those who have engaged in fraud of a second chance to win the election outright.[226] Although this remedy does involve the judiciary in determining the winner, the judiciary's reasoning in such cases typically should be quite transparent and acceptable.

* * *

The preceding examples—from problems of ballot design, to outright balloting failures, to polling place operations that either discourage lawful voting or permit unlawful voting, to voter fraud—begin to suggest how the fundamental values of democratic elections should shape the choice of remedy for election problems. In short, these values should be applied in light of the fact that elections are imperfect. In addition, the judiciary's role in the contest phase of an election ought to be to remedy only those failures that could not have been reasonably identified ex ante, or those that necessarily precluded voters from being able to express their true preferences.

IV. Potential Reforms to Election Remedies

It is hard to imagine that values such as fairness, legitimacy, promptness, finality, voter anonymity, accuracy, transparency, efficiency, cost, or separation of powers will become irrelevant to election policy in the near future. Among these values, the most obvious candidate for some readjustment is cost. In particular, if the public became sufficiently dissatisfied with the existing system, we might substantially increase the amount of public money that we were willing to spend to conduct our elections. But no such popular uprising seems imminent.

Even without a public uproar about the trustworthiness of our democratic processes, the integrity of our election systems is worthy of increased financial support. Yet not even substantial additional expenditures will eliminate all the potential for election failures.[227] Accordingly, the relatively marginal funding increases that are much more likely to occur will certainly not themselves produce dramatic reductions in the incidence of election problems.

The task therefore is to improve our methods for remedying election failures while taking current values as the touchstones for success. Using the sometimes conflicting values of fairness, promptness, secrecy, accuracy, and separation of powers as primary constraints, several key catego-

---

[226] *See supra* notes 92–93 and accompanying text.

[227] *Cf.* Borders Transcript, *supra* note 103, at 3 (opining that fixing the deficiencies in the state's election processes will "require more than just constructing new buildings and hiring new staff").

ries of potential reform deserve exploration. These reforms include: greater ex ante consideration of when to call new elections, model standards for election contest provisions, increased use of nonjudicial alternatives for resolving failed elections, and refining public expectations about the degree of perfection achievable in the democratic election processes of a large and complex society.

A. Ex Ante Consideration of When (and When Not) To Call New Elections

As suggested in Part III, many states' election contest processes often require courts to choose between conflicting values. Moreover, these election contests often put courts in the position of "kingmaker" without giving them clear, objective standards that might insulate them from charges of political meddling. In such circumstances, courts understandably may be reluctant to alter or invalidate official election results. It therefore is fundamentally important, both for courts and for the strength of our elections, that our political communities establish clearer and more objective standards for when and how courts should void or adjust a flawed election.

Just as many state election codes have a provision calling for an automatic recount when the margin of victory is within a specified range, the contest provisions of state codes similarly could specify circumstances in which courts must require a new election.[228] For instance, a code could specify that new elections would be mandatory whenever: (1) the voting process is shown to have necessarily excluded legal votes sufficient in number to create reasonable uncertainty about the reliability of the outcome; or (2) the official results are shown to include illegal votes sufficient in number to create reasonable uncertainty about the reliability of outcome, provided that the beneficiary of these illegal votes cannot be specifically identified[229] (in which case the official results should be adjusted accordingly), and provided that the victorious candidate did not aid or participate in acts of voting fraud giving rise to the illegal votes (in which case the runner-up should be declared the winner).

More generally, election contest statutes could specify in advance the authorized remedy for each of the circumstances most likely to arise in an election contest. The beginnings of one such framework are spelled out in Table 1, which associates a remedy with each of several categories of election contests. As Table 1 suggests, election contest provisions also might identify ex ante the circumstances in which a new election is *not* authorized, and in which instead the required remedy is either to adjust

---

[228] Most election contest provisions do comparatively little to constrain a court's choice of remedy once a court is satisfied that irregularities render the outcome unreliable. *See supra* notes 65–66 and accompanying text.

[229] Protecting ballot secrecy means that in most cases the specific beneficiary will not be known. *See supra* notes 150–153 and accompanying text.

the vote totals, to award the election to the runner-up, or to affirm the official results. Contest provisions further might specify only a narrow range of circumstances when a new election would be permitted (but not required) at the discretion of a court or other tribunal, perhaps in light of considerations of fundamental fairness and the egregiousness of the illegal voting.

TABLE 1

Contest Circumstance[230]	*Remedy*[231]
Legal votes necessarily excluded— # of excluded votes creates reasonable uncertainty:	Call new election
Illegal votes included, beneficiary of specific votes known— Non-fraudulent illegal voting: Fraudulent voting with victor participation:	Adjust totals accordingly Discretion to adjust totals accordingly or reverse outcome
Illegal votes included, beneficiary of specific votes not known— Non-fraudulent illegal voting— # of illegal votes creates reasonable uncertainty: # of illegal votes does not create reasonable uncertainty:	Call new election Uphold election

[230] "Reasonable uncertainty" is shorthand for a circumstance in which the number of errant votes is sufficient to create reasonable uncertainty about the validity of the election outcome. As applied to illegal votes resulting from mistake, rather than fraud, it is flexible enough to allow some degree of proportionality in assessing the likely impact (though not necessarily in implementing a remedy). In contrast, when fraudulent voting favoring one candidate is proven, the presumption may be that all the illegal votes accrued to that candidate, in which case the standard is expressed in terms of whether the number of illegal votes favoring the victorious candidate is "proven to exceed" the margin of victory, or if not so proven, whether the number of fraudulent votes "could reasonably exceed" this margin.

[231] "Reverse outcome" is shorthand for deducting sufficient votes from the victor to award the election to the runner-up. In some hypothetically extreme cases, it is possible that the runner-up may have received so few votes as to render becoming the victor problematic.

Illegal votes included, beneficiary of specific votes not known— Fraudulent voting with victor participation— # of illegal votes proven to exceed margin: # of illegal votes could reasonably exceed margin: # of illegal votes not likely to exceed margin:	 Reverse outcome Reverse outcome Discretion to uphold, reverse, or call new election
Illegal votes included, beneficiary of specific votes not known— Fraudulent voting without victor participation— # of illegal votes proven to exceed margin: # of illegal votes could reasonably exceed margin: # of illegal votes not likely to exceed margin:	 Reverse outcome Discretion to reverse or call new election Discretion to uphold or call new election
Act of God:	Discretion to uphold or call new election

The set of remedies proposed in Table 1 may not necessarily be the ideal method for cabining judicial discretion, and some might object to the table's "cookbook" approach. The table is intended merely as an example, to serve as a starting point for a discussion about how some systematic ex ante attention to typical election contest scenarios could liberate courts from making hard policy choices that resolve tight political contests. Furthermore, the circumstances listed in the table are not intended to be comprehensive, as several other scenarios might also merit inclusion in such a framework.[232]

Three other observations about this sample framework merit brief comment. First, this framework could be modified to permit a limited new election for only some subset of the original voters, if that would be sufficient to remedy a voting defect (as when an election failure has tainted only the votes of one precinct or county). Indeed, in response to a contested election that highlighted the benefits of such an approach,[233] the

[232] For instance, an election might have a combination of problems, such as voting fraud compounded by mistake, or fraud by both of the top two finishers in the election.

[233] *See* Gunaji v. Macias, 31 P.3d 1008 (N.M. 2001) (creating an equitable remedy of

New Mexico legislature recently revised its contest provisions to permit just such a partial new election in some circumstances.[234]

Second, states should consider what the desired remedy should be whenever the number of nonfraudulent illegal votes (or the number of excluded legal votes), although mathematically sufficient to have potentially affected the outcome, is not reasonably likely in fact to have had such an impact. We may be expecting an unreasonable degree of perfection from our elections if we insist on rerunning such an election. One alternative would be to determine the statistical probability that the errant votes would in fact alter the outcome.[235] If the probability were below some statistical threshold, a contest statute could require the reviewing court to dismiss the contest and uphold the election.[236]

Third, awarding an election to the runner-up because of fraudulent conduct by or on behalf of the victor can be problematic if the runner-up does not have broad electoral support. It therefore is hypothetically possible that in some instances the more democratically sound remedy for voting fraud is to hold a new election. Yet because it is impossible to guarantee that the new election will itself be free from fraud, and because merely voiding the election and rerunning it may be an insufficient political penalty for engaging in voting fraud, states might decide as a policy matter that the better approach is to award the race to the runner-up, even at the risk of installing a candidate without even strong plurality support.[237]

In addition to protecting the appropriate separation of powers by sparing courts from taking sides on a hard policy choice that will directly determine the outcome of a partisan battle, an ex ante framework also would

---

partial revote, in contrast to the then-existing code requirement of disregarding the entire precinct).

[234] *See* N.M. STAT. § 1-12-37-1 (2003) (allowing court to order county clerk to send new ballots to the voters identified as having voted with a defective ballot). Courts in other states have rejected such a remedy absent statutory authority. *See, e.g.*, Howell v. Fears, 571 S.E.2d 392, 393 (Ga. 2002) (concluding that the Georgia contest statute does not authorize limiting a new election to an isolated precinct with defective ballots).

[235] More than a generation ago, two mathematicians proposed a model that would estimate the probability that a specified number of irregular votes would have altered the outcome of the election had they not been irregular. *See* Michael O. Finkstein & Herbert E. Robbins, *Mathematical Probability in Election Challenges*, 73 COLUM. L. REV. 241 (1973). Unfortunately, there has been little consideration of matters relating to the implementation of this model, such as considering which statistical tools to use and who would be charged with administering them.

[236] For instance, a contest statute could require convincing the court that at least a ten percent probability exists that the irregular votes would alter the outcome before allowing the court to require a new election. Regardless of the exact magnitude of the threshold, it would provide the benefit of an objective standard.

[237] Of course, it is hard to worry too much about this result when turnout rates for many local elections hover around twenty percent. *See* Donald P. Green, Alan S. Gerber, & David W. Nickerson, *Getting Out the Vote in Local Elections: Results from Six Door-to-Door Canvassing Experiments*, 65 J. POL. 1083, 1083 (2003). The result is that even a landslide victor often cannot claim much of a true mandate.

further several other values. Most significantly, by insulating courts from deciding an election outcome and instead specifying the appropriate remedy in advance (when by definition the candidate who will benefit from the rule is not known), an ex ante framework also promotes the ideal of fairness in election administration. Furthermore, to the extent that it eases the burden on the reviewing tribunal, such a framework also permits a quicker disposition of election contests and promotes the interest of promptness. Also, the policy judgments that go into deciding when a new election is in order will in turn reflect a balance between the interests in an accurate tally and the need for a prompt, final result at an affordable cost. Accordingly, as a first step in improving our remedies for failed elections, each state's election code should clearly specify the circumstances in which particular remedies are appropriate.

B. Toward a Model Code for Election Contests

In most states, "contesting an election is purely statutory, and a strict observance of statutory requirements is essential to the exercise of jurisdiction by the court, as it is desirable that election results have a degree of stability and finality."[238] Keeping in mind the fact that election contests are statutory proceedings, is it practical to develop a comprehensive model for a statutory framework governing an election contest? The answer, given the rich variety in state government structures, electoral mechanisms, and judicial institutions, is probably not.[239] Reasonable people may differ about how a particular state should strike the balance between finality and certainty, for instance, in structuring an election contest procedure. Nevertheless, each state legislature should make explicit where it has set this balance, so that courts are freed from the need to resolve these policy questions after the fact.

Accordingly, with the continuing aim of developing a more systematic way of reducing the discretionary judgments that courts need to make about election remedies, this section identifies a few uniformly de-

---

[238] Republican Party of Garland County v. Johnson, 193 S.W.3d 248, 252 (Ark. 2004); *see also Ex parte* Vines, 456 So. 2d 26, 28 (Ala. 1984) (describing election contests as "strictly statutory"); Collin v. Knoblock, 25 La. Ann. 263 (La. 1873) (stating that courts have no jurisdiction over elections absent delegated authority).

[239] For similar reasons, prospects for successfully federalizing our election processes are slim. Even where nationalization of some aspects of our election processes (perhaps through Congress's authority to make or alter the regulation of the times, places, and manner of conducting congressional elections, *see* U.S. CONST. art. I, § 4, cl. 1, or through making federal funding contingent on states' satisfying specific requirements) might ease the problems of post-election litigation, American elections likely will continue to be administered primarily at the local level, with a wide range of personnel and processes. Nonetheless, the issue of nationalization has not gone undiscussed. *See, e.g.*, Paul Hernnson, *Improving Election Technology and Administration: Toward a Larger Federal Role in Elections?*, 13 STAN. L. & POL'Y REV. 147 (2002); Richard L. Hasen, *Beyond the Margin of Litigation*, *supra* note 14, at 964–73.

sirable features of election contest statutes. These include clearly defining procedural matters such as: (1) who can be a contestant; (2) what standard of evidence to require; and (3) how to expedite contests. This section also identifies several fundamental issues that any such statute should address, in addition to specifying the acceptable reasons or grounds for a contest. These include: (1) whether and in what circumstances to permit proportional or statistical adjustment of election results; (2) how readily to permit new elections to occur; and (3) whether races for different kinds of offices deserve different approaches to these and other issues. Another crucial matter is the extent to which primary elections ought (or need) to receive different treatment from general elections.

Although each jurisdiction may reach its own conclusion about each of these matters in light of its own unique election ecosystem and political culture, as well as its decision about how to balance fundamental values that may sometimes conflict, the following suggestions may serve as a starting point for election remedy reform.

1. Contestants

The issue of who can bring a contest action is a threshold question that all election contest statutes should address definitively. At present, the law on this issue varies from state to state. Some states allow only defeated candidates to commence an election contest,[240] while others permit any qualified voter, or even any taxpayer, to initiate a challenge.[241] In between these extremes, some states permit (or at one time permitted) government officers or groups of at least a specific number of electors to contest an election outcome.[242]

Understandably, states would not want just any voter to be able to commence an election contest if doing so were cost-free to the voter. Otherwise, dissatisfied voters might frequently contest elections simply to vent their pique at the outcome, rather than out of a genuine grievance about the process. The effect, in addition to burdening the judiciary, might be increased cynicism about both the reliability and finality of our elections.

On the other hand, because the injury occasioned by an election failure is a public injury, some might argue that any citizen should be able to

---

[240] *See, e.g.*, ARK. CODE ANN. § 7-5-801 (2000); IOWA CODE ANN. § 57.1(1)(a) (West 1999); TENN. CODE ANN. § 2-17-101(b) (2003).

[241] *See, e.g.*, CONN. GEN. STAT. ANN. § 9-328 (West 2002); FLA. STAT. ANN. § 102-168 (West 2006); N.J. STAT. ANN. 19:29-1 (1999); OHIO REV. CODE ANN. § 3515.08 (West 1994).

[242] *See, e.g.*, ALASKA STAT. § 15.20.540 (2006) (allowing ten voters to commence certain types of election contests); N.D. CENT. CODE § 16.1-16-02 (2004) (permitting ten qualified electors to contest election); OR. REV. STAT. ANN. § 258.016 (1991) (authorizing county clerk to bring election contest); Phillips v. Ericson, 80 N.W.2d 513, 523 (Minn. 1957) (describing former statute permitting twenty-five voters to commence election contest); State *ex rel.* Farnsworth v. McCabe, 35 N.E.2d 474, 476 (Ohio 1940) (describing former statute permitting any five voters to demand recount).

seek relief for it. Meanwhile, permitting citizens to bring election contests has the further advantage of letting candidates avoid appearing to be sore losers. Unless a candidate can foresee a strong chance of succeeding in a contest, political calculations about the potential reputational harm of appearing to be a sore loser may on occasion dissuade the candidate from commencing an election contest, even where genuine grounds for a contest exist.[243] The presumptions in favor of sustaining election outcomes presumably pressure candidates to forgo contests as a matter of "political accommodation," in order to preserve their viability in a subsequent election.[244]

An election contest statute therefore should not limit contestants to just defeated candidates. Even where the chances for success may not appear strong at the outset, a valid ground for a contest may still exist. Moreover, even when such contests are not ultimately successful in altering election outcomes, they may nonetheless help to purify and identify problems in our election systems, and also contribute to a deeper jurisprudence of election contests.

Accordingly, both candidates and groups of voters, perhaps groups with a minimum of 50 or 100 voters collectively, should be able to contest the election.[245] This type of numerical requirement would help guard against meritless actions by a rogue voter[246] and other abuses, and also would serve the interests of fairness and transparency by allowing the public to vindicate its right to a sound election. Meanwhile, to further protect against contests being filed in anger or disappointment without sufficient cause, states should assess some of the costs of the contest action on the contestant (unless the outcome of the contest voids the election or alters its result), and should require contestants to post a bond as a prerequisite to the contest.[247]

---

[243] *But see* Kyle Whitmore, *War on Dumb: No More Gracious Losers*, Birmingham Weekly Online, Aug. 3, 2006, *available at* http://www.bhamweekly.com/archived/pages/20060803_war%20on%20dumb.php (describing ugly and unwarranted primary contest as an example that "the era of the gracious loser died with the 2000 presidential election"); *supra* note 163 (describing possibility that high campaign costs may make some candidates unwilling to concede defeat easily).

[244] Voters, too, could in theory be discouraged from contesting an election by a concern for political accommodation, recognizing both that contests are quite stressful to democratic processes and that the prospect of a future election mitigates the inevitable imperfections of any particular election. *See generally* Foley, *supra* note 33. Indeed, such political realism may partially explain the strong presumption that election results are reliable. *See infra* note 254 and accompanying text.

[245] Alternatively, a numerical requirement could be structured in terms of a petition process, in which the complaint in an election contest must be supported by a petition signed by some number of voters. Illinois employs this approach, requiring a petition signed by the same number of voters as must sign a petition to nominate a candidate for the office in question. *See* 10 ILL. COMP. STAT. § 5/23-1.2a (2004).

[246] An individual voter aggrieved by an election failure might still seek to bring a civil rights action for damages or prospective injunctive relief rather than a contest action to invalidate the election outcome. *See supra* notes 121–131 and accompanying text.

[247] Most contest statutes that permit voters to contest elections have these provisions. *See* 26 AM. JUR. 2D *Elections* § 460 (1996).

2. Standard of Evidence

One recurrent ambiguity in the law of election contests involves the standard of evidence.[248] Many election contest statutes say nothing about the standard of evidence,[249] although they may recite that the ordinary civil rules should apply.[250] As a result, many states appear to rely on the simple "preponderance of the evidence" test of ordinary civil litigation.[251] Meanwhile, a few courts appear to have applied a "beyond a reasonable doubt" (or similar) standard to at least some elements of election contests.[252]

The most suitable test, however, and one presently employed in a number of states, is to require clear and convincing evidence of an election failure.[253] A clear and convincing standard is appropriate because our election processes, though imperfect, have earned a strong presumption of correctness.[254] To rebut this presumption, and thereby void or alter an official result, should require not just a fifty-one percent probability, but some higher confidence or likelihood that the official certification is not trustworthy.

If a mere preponderance of evidence could invalidate an election result, principles of both finality and efficiency would be sacrificed, as elections would have to be rerun more often. More important, principles of fundamental fairness could be compromised because of the destabilizing impact that a preponderance standard would produce. That is, provided that we continue to conduct elections in a manner in which they are ordinarily entitled to a presumption of correctness (as we must strive to do, for purposes of democratic legitimacy), then we should not create a circum-

---

[248] A set of related questions asks whether election contests should be heard by judges or by juries, whether appellate review should be available, and whether factual questions should be tried *de novo* on appeal. At least one court has answered in the affirmative to this last question. *See* Big Spring v. Jore, 109 P.3d 219, 222 (Mont. 2005).

[249] *See, e.g.*, NEV. REV. STAT. § 293.417 (2005) (providing only that a court shall alter an election outcome if it "finds from the evidence" that the outcome is incorrect); UTAH CODE ANN. § 20A-4-404 (2003) (providing only that a court shall alter an election outcome if it "determines" that the outcome is incorrect).

[250] *See, e.g.*, KAN. STAT. ANN. § 25-1446 (2000) (providing that election contests be tried as civil actions under applicable provisions of code of civil procedure); N.J. STAT. ANN. § 19:29-5 (West 1999) ("[Contest] proceedings shall be similar to those in a civil action."); N.D. CENT. CODE § 16.1-16-06 (2003) (providing that election contests "be tried as civil actions").

[251] *See, e.g.*, Maynard v. Hammond, 79 S.E.2d 295, 299 (W.Va. 1953); Pierce v. Harrold, 138 Cal.App.3d 415, 428 (Cal. Ct. App. 1982); *In re* Gen. Election of Nov. 5, 1991 for Office of Twp. Comm. of Maplewood, 605 A.2d 1164, 1170 (N.J. Super. Ct. Law Div. 1992).

[252] *See, e.g.*, Rogers v. Holder, 636 So. 2d 645, 650 (Miss. 1994) (using "beyond a reasonable certainty" standard).

[253] *See, e.g.*, MD. CODE ANN. [ELEC. LAW] § 12-204(d) (LexisNexis 2003); Wilks v. Mouton, 722 P.2d 187, 190 (Cal. 1986); Bazydlo v. Volant, 647 N.E.2d 273, 276 (Ill. 1995); *In re* Election of Nov. 6, 1990 for the Office of Att'y Gen. of Ohio, 569 N.E.2d 447, 450 (Ohio 1991); Quinn v. Tulsa, 777 P.2d 1331, 1341 (Okla. 1989); Concerned Citizens for Better Educ., Inc. v. Woodley, 623 S.W.2d 488, 491 (Tex. 1981).

[254] *See* 26 AM. JUR. 2D *Elections* § 439 (1996).

stance in which every close election turns into a question of which side has that scintilla of additional evidence that will determine whether the results stand. Instead we should treat the question as whether we can clearly see that the outcome is not reliable.

The clear and convincing test should be the standard for two discrete components of the contest: proof that some irregularity occurred in the election, and proof that this irregularity altered the outcome or at least rendered it uncertain.[255] In contrast, some states use what is termed a "direct evidence" requirement that in effect requires proof beyond a reasonable doubt that the irregularity altered the outcome,[256] in which case the remedy is to reverse the result. Requiring proof beyond a reasonable doubt that the outcome would have been different has the virtue of simplifying the question of the appropriate remedy (because mere uncertainty about a result's validity is insufficient to disturb the result), but has the disadvantage of leaving many election failures unremedied. In contrast, accepting clear and convincing evidence that the result is not reliable is likely to correct more election defects without destabilizing the system, but in turn calls for greater guidance about what remedy to impose.[257]

A related issue requiring attention is what showing to require in order to determine that a provisional ballot should be counted. Since at stake is a particular individual's right to vote, a simple preponderance test might suffice, especially since this decision ought to be made prior to an election contest as part of the ordinary canvass process. If questions about the eligibility of specific provisional ballots then become part of a contest proceeding, the same standard used in the initial determination of a provisional ballot's validity should continue to apply. In any event, this is an issue that has already received a fair amount of academic commentary.[258] Most important, states must establish the standard in advance, so that courts have clear guidance when the inevitable post-election litigation occurs.

3. Contest Timetables and Expedition

A third crucial element of a contest action is its timing. Election contests need to be both commenced and concluded as expeditiously as possi-

---

[255] *See, e.g.*, Taft v. Cuyahoga Cty. Bd. of Elections, 854 N.E.2d 472, 476 (Ohio 2006) (stating that in Ohio election contests, the clear and convincing standard applies to both proof of irregularity and proof that irregularity altered outcome).

[256] *See* State *ex rel.* Wahl v. Richards, 64 A.2d 400, 407 (Del. 1949); Jaycox v. Varnum, 226 P. 285, 289 (Idaho 1924); Brown v. Grzeskowiak, 101 N.E.2d 639, 656 (Ind. 1951); Wilkinson v. McGill, 64 A.2d 266, 274 (Md. 1949), State *ex rel.* Brogan v. Boehner, 119 N.W. 2d 147, 151–53 (Neb. 1963).

[257] Greater guidance is required given that the tribunal now must choose between a wider range of possible remedies—from ordering a full new election, to ordering a partial new election, to making some corrections or adjustments to the vote totals. *See supra* Part I.B.2&3.

[258] *See supra* note 143.

ble. However, it is difficult to specify a precise timetable suitable for all contest actions, because the ideal schedule will depend both on how soon after the election the victor is to assume office, and on whether the decision of the tribunal hearing the contest is conclusive or is subject to appeal.

One way to increase the speed with which election contests are concluded would be to try the contest action before a panel of judges whose decision is final and unappealable. Especially if some election contests receive *de novo* review in appellate courts anyway,[259] it might be both more efficient and most expeditious to combine the trial and the appeal into a single event.[260]

In any event, a contest should commence within days of certification of the election, in order to preserve the best evidence and expedite the final determination of the election,[261] and the contest statute should make clear what this period is. In addition, a strong contest statute should include other provisions designed to expedite resolution of the contest. These could include provisions permitting summary or informal pleadings,[262] as well as provisions requiring the court to give the contest action priority over other actions.[263]

Another procedural matter affecting the timing of election contests concerns the expedition of the discovery process. A prolonged period of discovery, while obviously conducive to a more accurate determination of election outcomes, is antithetical to the need for a prompt and final determination. Accordingly, courts must limit discovery and require parties to expedite the creation of the evidentiary record underlying the contest action.

Special timing considerations come into play with respect to federal elections, for which Congress is the ultimate judge of the outcome.[264] With respect to congressional elections, it is well settled that state courts may resolve issues arising in election contests, even though Congress will have the final word concerning the outcome.[265] But for state proceedings to be

---

[259] *See supra* note 248.

[260] A few states already grant jurisdiction over election contests to special panels of judges drawn from throughout the judiciary. *See* MISS. CODE ANN. § 23-15-935 (2001); VA. CODE ANN. § 24.2-801 (2006).

[261] The contest statute also should be structured to permit the contest to begin before administrative recounts are completed, in order to expedite the final determination of the outcome. The recount still may provide the official result that is the subject of the contest action, but in many cases the preliminary steps of a contest could begin before the recount has generated the final tally.

[262] *See, e.g.*, NEB. REV. STAT. § 32-1110 (2004) (directing court to hear contest in "summary manner without any formal pleading").

[263] *See, e.g.*, CAL. ELEC. CODE § 16003 (2003); TEX. ELEC. CODE ANN. § 231.009 (2003); VT. STAT. ANN. tit. 17 § 2603 (2006).

[264] Some similar issues may arise with respect to those state elections whose outcomes are adjudicated in the state legislature, although more typically it may be that the matter is heard exclusively in the legislature without any court assistance.

[265] *See* Roudebush v. Hartke, 405 U.S. 15, 25 (1972) ("A [state's] recount proceeding does not prevent the Senate from independently evaluating the election any more than the

of assistance to Congress's review of the election, the state proceedings must be concluded before Congress takes up the matter. Similarly, state efforts to clear up uncertainties in the election of presidential electors need to be concluded no later than before the electors cast their ballots, and preferably by the "safe harbor" deadline Congress has established, which is five weeks after election day.[266] Unfortunately, very few states appear to have structured their recount or contest provisions with this federal safe harbor deadline in mind, and in most states post-election remedial processes may not be up to meeting this deadline if the controversy is difficult.[267]

Finally, courts hearing election contests should be prohibited from entertaining claims that could have been litigated prior to the election.[268] In large part this is because post-election litigation is so inherently destabilizing to the democratic process that it should be avoided whenever possible.[269] In addition, candidates and their supporters are otherwise invited to game the system, waiting to see whether they win at the polls before deciding whether to try to win in court.

4. Alternatives to Proportional Adjustments

Given the range of potential election miscues that can preclude an accurate determination of the true will of the voters, and given the high social costs of resolving these failures through a new election, some sustained attention also should be given to the potential contributions of statistical and demographic analyses. Such analyses could identify the probability that an election outcome is not a reliable reflection of the voters' good-faith efforts to express their will. Although it is too early to predict whether professional consensus could be achieved in this area, and whether the public would accept it, in any event it would be preferable for states to undertake their own deliberate consideration of the potential value of sta-

---

[state's] initial count does. The Senate is free to accept or reject the apparent winner in either count, and, if it chooses, to conduct its own recount.").

[266] One aspect of the presidential election process that the 2000 election highlighted is that to prevent Congress from rejecting a state's slate of presidential electors, states must have their results finalized by the second week of December in a presidential election year. *See* 3 U.S.C.A. § 5 (2000). If a state misses this safe harbor deadline, Congress is no longer statutorily obligated to accept that state's certification of its electors. *See* 3 U.S.C.A. § 15 (2000). Florida's effort to meet this deadline persuaded the U.S. Supreme Court to conclude that Florida would not be willing to continue any recount after the safe harbor date. *See* Bush v. Gore, 531 U.S. 98, 113 (2000).

[267] *See* Foley, *supra* note 33, at 103–04; Steven F. Huefner, *Reforming the Timetable for the Electoral College Process*, ELECTION LAW @ MORITZ WEEKLY COMMENT, Nov. 30, 2004, http://moritzlaw.osu.edu/electionlaw/comments/2004/041130.php.

[268] For further discussion of this proposal, see Edward B. Foley, *The Promise and Problems of Provisional Voting*, 73 GEO. WASH. L. REV. 1193, 1203–04 & n.69 (2005).

[269] *See* Ross v. State Bd. of Elections, 876 A.2d 692, 705–06 (Md. 2005) (concluding that laches barred an election lawsuit brought after the election that could have been brought before the election).

tistical analysis, rather than to leave such consideration to a court in the course of deciding a specific contest.

For instance, imagine a two-candidate race with a margin of victory of 3000 votes out of 50,000 cast. Further imagine that some 4000 of the 50,000 votes cast have been irretrievably lost, without ever being counted, so that Candidate *A* has received 24,500 votes, in apparent victory, and Candidate *B* has received 21,500. Some would argue that because the number of lost ballots is substantially (33%) greater than *A*'s margin of victory, the election outcome is unreliable and a new election must be held. But suppose that statisticians could determine, after taking into account the circumstances of the lost votes, that the chances are ten thousand to one that more than 3500 of the lost 4000 votes would go to Candidate *B*, as would be necessary to alter the outcome. Given the small probability that the lost votes could alter the election, it might be more likely that some new error would occur in a new election. In that circumstance, it would make no sense to hold a new election.[270]

Although some courts have already recognized this,[271] others seem to be awaiting a statutory change that clarifies the propriety of this mode of analysis. Such a statute would need to specify the threshold probability that is sufficient to require a new election: Is it only five percent? Might it be as high as twenty percent? This is a policy question that courts are not well-suited to decide, but one which the legislature could easily (even if arbitrarily) set in advance.

5. Presumption Against New Elections

As previously noted, ordering new elections whenever an election is in doubt could minimize the judiciary's entanglement in the political process.[272] New elections may be desirable for other reasons as well. Sometimes they may provide the only opportunity to rectify an election prob-

---

[270] In certain circumstances, a similar approach could be used to handle a group of invalid (rather than missing or lost) votes. Rather than deducting the votes from candidates in proportion to the vote distribution of some representative class from which the invalid votes are drawn, as in traditional proportional deduction, the analysis would seek to ascertain the probability that these votes would be distributed in such a way as to change the outcome. This analysis also could include more sophisticated predictors of the likely beneficiaries of the invalid votes, to avoid the "ecological fallacy" that the trial court in the Washington gubernatorial race found fatal to proportional deduction, *see supra* note 103 and accompanying text, but would not purport to identify the specific number of invalid votes cast for each candidate because of the impossibility of knowing the precise number when voting is by secret ballot.

[271] *See, e.g.*, Boyes v. Allen, 32 A.D.2d 990, 991 (N.Y. App. Div. 1969) (finding no reasonable probability that twelve-vote margin of victory would be altered by eliminating thirteen disputed votes); Badillo v. Santangelo, 15 A.D.2d 341 (N.Y. App. Div. 1962) (refusing to overturn election because it was "highly unlikely" that eighty-three of ninety invalid votes were cast for victor, as would be necessary to overcome seventy-five-vote margin of victory).

[272] *See supra* Part II.F.

lem, given the anonymity of ballots. For instance, in some elections tainted by fraud, both the source and the impact of the fraud may be impossible to determine, and a clean election will be the only opportunity to remove the taint.

Florida's butterfly ballot was a classic case in which only a new election could have satisfactorily remedied the problem.[273] Almost certainly, thousands of voters had marked their ballots in a way that would register a vote for a candidate other than the candidate whom the voter meant to choose.[274] But which specific ballots were marked in error could not be identified, let alone which voters had cast those ballots. Only bringing all voters who had used the butterfly ballot back to the polls with a new ballot could undo the distorting effect of the flawed ballot design.

However, as previously discussed, holding a new election imposes huge costs.[275] New elections therefore should be the last resort, even though this places a greater burden on courts to determine the winner in some cases. Although courts justifiably should be reluctant to assume the responsibility of determining the victorious candidate, in some instances courts can legitimately declare a new winner, rather than order a new election, if they can be satisfied (under a clear and convincing evidence standard) that a plurality of voters freely and independently casting valid votes in fact intended to vote for this candidate.

In other instances, remedying an election failure may simply be too costly. For instance, the flawed design of the butterfly ballot did not necessarily prevent any particular voter from casting a valid vote that accurately reflected that voter's will.[276] Thus, although the balloting was undeniably less than ideal, whatever errant balloting in fact occurred might be attributed at least partly to voter error. Furthermore, the specific impact of this election failure on the vote totals was uncertain. And even though in hindsight it seems almost certain that the ballot design resulted in miscredited voter preferences, it would be asking too much of our election system to insist on a new election on account of every small distortion in the process that speculatively may have meant the difference in a tight race.[277] A new election should be conducted only when voters have been completely prevented from accurately registering their intended preference in numbers sufficient to affect the outcome.

---

[273] *See supra* notes 184–187 and accompanying text.

[274] *See* Wand et al., *supra* note 47, at 799–801.

[275] *See supra* Part II.F.

[276] *See supra* note 185 and accompanying text.

[277] For instance, many states now require that the order in which candidates are listed on the ballot be rotated, to prevent giving the candidate listed first a systematic advantage. *See supra* note 184. But it would be problematic for courts to require a new election for every close race conducted without a requirement of ballot order rotation. *See* Bradley v. Perrodin, 106 Cal. App. 4th 1153 (Cal. Ct. App. 2003) (overturning lower court vote reallocation based on testimony about the "primacy effect" of first ballot position).

In sum, new elections should occur only when three conditions are satisfied: (1) fraud, mistake, or an "act of God" has necessarily precluded the certified vote total from correctly aggregating all voters' independent, uncoerced, and unprocured preferences; (2) as a result of this irregularity the official outcome cannot be trusted to accurately reflect the will of the voters; and (3) a reliable outcome cannot be determined in a manner other than holding a new election.

6. Different Offices

One reason that it is difficult to develop a model election contest statute is that each state's election code must cover a complex mix of federal, state, and local elected offices in both the executive and legislative branches, and often in the judicial branch as well. The precise mix of elected offices covered in any particular state's code may affect the way in which that code could best address the issue of how to remedy a failed election. For instance, for elections to some of these offices, the state may already have decided that its legislature is to be the final judge.[278] In that event, the courts may have no role at all, or may play only an initial or advisory role. The choice to let courts play an initial, rather than advisory, role—for instance, in creating the evidentiary record, but leaving the legislature to decide whether an election outcome is reliable—may shape the authority they are given.

7. Primary Elections

A further difficulty in developing a model election contest statute is the question of whether to treat primary elections in the same manner as general elections. The answer to this question depends on such factors as the type of primary election used in the state,[279] the interval between the primary and the general election,[280] and the state's political traditions.[281]

---

[278] *See infra* notes 289–291 and accompanying text.

[279] Several nominating methods are in use in the United States today, from party conventions (in which delegates choose the party's nominees), to closed primaries (in which only registered party members may vote), to open primaries (in which any registered voter may vote, selecting in which party's primary to vote in the privacy of the voting booth), and variations thereon. *See* JOHN F. BIBBY, POLITICS, PARTIES, AND ELECTIONS IN AMERICA 155–64 (2003).

[280] This interval can range from as long as eight months, *see* OHIO REV. CODE ANN. § 3513.01 (West 2007) (mandating that primaries during years in which a presidential election is held occur on the first Tuesday after the first Monday in March), to as short as one to two months, *see, e.g.*, HAW. REV. STAT. ANN. § 12-2 (LexisNexis 2006) ("The primary shall be held at the polling place for each precinct on the second to the last Saturday of September in every even numbered year; provided that in no case shall any primary election precede a general election by less than forty-five days."); 10 ILL. COMP. STAT. ANN. 5/2A-1.1 (West 2006) ("In odd-numbered years, an election to be known as the consolidated election shall be held on the first Tuesday in April except as provided in Section 2A-1.1a of this Act; and an election to be known as the consolidated primary election shall

As a result, some states treat all elections the same, some have separate sets of rules, and still others leave the resolution of contested primaries to the political parties.[282] Perhaps the most critical factor in resolving primary elections is speed, as the victorious candidate needs to be identified in time to conduct a meaningful general election campaign. Some states' intervals between primary and general election are even shorter than the interval between the general election and the date that the winner takes office,[283] dramatically heightening the importance of a prompt resolution of a primary election contest.

C. Non-Judicial Alternatives

Any systematic effort to reform how election failures are remedied also ought to consider alternatives to the judiciary. Traditional courts are not the only or even the ideal forum for resolving all election contests. Other forums that may play some role include legislative bodies, administrative tribunals, and special election courts.

For instance, where elections to federal offices are concerned, Congress has the final say over who is the winner. In particular, the office of President of the United States, the only nationwide federal elected office, is filled when presidential electors cast their votes pursuant to federal constitutional and statutory authority[284] (including the "safe harbor" deadline for each state's selection of its slate of electors[285]). These provisions also confer on Congress the ultimate power to resolve a contested presidential election.[286]

---

be held on the last Tuesday in February.").

[281] A state's political traditions can influence such factors as the amount of competition at the primary stage, or the ability of party leaders to unify defeated candidates behind the party's nominee.

[282] *See, e.g.*, ALA. CODE §§ 17-13-70 to 17-13-89 and 17-16-40 to 17-16-76 (2006) (delineating separate rules governing primary and general elections); ALASKA STAT. § 15.25. 090 (2006) (treating primary and general elections the same); ARIZ. REV. STAT. ANN. § 16-671 (2006) (treating primary and general elections the same); MICH. COMP. LAWS ANN. §§ 168.574–168.588 and 168.641-168.799a (2007) (delineating separate rules governing primary and general elections); MISS. CODE. ANN. §§ 23-15-921, 23-15-923 (West 2006) (providing that party executive committee resolves primary election contests); WASH. REV. CODE. ANN. § 29A.52.121 (2007) (treating primary and general elections the same).

[283] *See supra* note 280. Several states have advanced or are contemplating advancing their primary date to provide more time to resolve primary election contests. *See* IMPROVING ARIZONA'S RECOUNT AND ELECTION CONTEST LAWS 5–8 (2005), *available at* http://www.azsos.gov/election/Brewer_Voting_Action_Plan/Election_Law_Advisory_Committee/Committee_Report_12-30-2005.pdf.

[284] *See* U.S. CONST. art. II, §1; U.S. CONST. amend. XII; 3 U.S.C. §§ 1–15 (2000) (prescribing process for selecting presidential electors, for voting by the electors, and for counting the elector's votes).

[285] *See supra* note 266 and accompanying text.

[286] *See* 3 U.S.C.A. § 15 (2000) (prescribing processes by which Congress counts electoral votes, including process for resolving objections to particular votes). However, as the 2000 election made clear, state and federal courts can play significant roles at earlier stages—roles that ultimately may obviate the need for Congress to settle an election controversy.

Similarly, with respect to elections for members of Congress, the U.S. Constitution provides that "[e]ach House shall be the Judge of the Elections, Returns and Qualifications of its own Members . . . ."[287] In addition, the Constitution provides that while each state legislature may prescribe the "Times, Places, and Manner of holding elections for Senators and Representatives," Congress may "make or alter such regulations."[288]

In some states the state legislature has a power analogous to Congress's Article I Section 5 power to judge the elections and qualifications of its own members.[289] In fact, a few state legislatures even have the final authority to review the elections of state officers elected on a statewide basis. For instance, in Colorado, a joint session of the state General Assembly resolves all contests concerning elections for state officers.[290] North Carolina has a similar process, under which the North Carolina General Assembly recently resolved a contest over the 2004 election of the Superintendent of Public Instruction.[291]

Legislatures obviously will resolve contested elections without even a patina of political neutrality.[292] But letting majoritarian institutions resolve questions about the majority's will in an election contest may be appropriate.[293] Even these partisan institutions ought to protect a minority party's candidate in the face of compelling evidence that an election result is unreliable,[294] rather than face the political consequences of being perceived to have thrown an election.[295] And when the election contest pre-

---

[287] U.S. CONST. art. I, § 5, cl. 1.

[288] *Id.* at art. I, § 4, cl. 1. However, when congressional elections are contested, Congress typically awaits the disposition of any state proceedings before making a final determination whether to accept the state's choice of candidate. *See, e.g.*, Roudebush v. Hartke, 405 U.S. 15, 19 (1972).

[289] *See, e.g.*, ALA. CONST. art. II, § 12; KY. CONST. § 38; LA. CONST. art. III, § 7; MICH. CONST. art. IV, § 16; *see also* Steven F. Huefner, *Echoes of* Bush v. Gore*: Courts Are Not Always the Right Forum for Election Contests*, ELECTION LAW @ MORITZ WEEKLY COMMENT, Jan. 10, 2006, http://moritzlaw.osu.edu/electionlaw/comments/2006/060110.php (describing contested Kentucky state senate race that state supreme court resolved contrary to state senate's resolution).

[290] *See* COLO. REV. STAT. ANN. §§ 1-11-205, 207 (West 2000).

[291] *See* N.C. GEN. STAT. ANN. § 120-10.10 (West 2006); *see also Last Unresolved State Election is Settled,* N.Y. TIMES, Aug. 24, 2005, at A14.

[292] *Cf.* Richard Pildes, *Democracy and Disorder*, 68 U. CHI. L. REV. 695, 715–16 (2001) (describing *Bush v. Gore* majority as distrustful of legislatures resolving highly political elections).

[293] *See* Samuel Issacharoff, *Political Judgments*, 68 U. CHI. L. REV. 637, 655 (2001) (discussing a "presumption . . . that vindication [of majority preference] lies in the political arena"). Because the political composition of a state legislature presumably approximates the political composition of the electorate of the state, it may be most appropriate to let the legislature resolve statewide election contests, as well as contests over its own membership, but not to task it with resolving local races.

[294] For instance, a Republican U.S. Senate ultimately rejected a challenge to the very close 1996 election of Democratic Senator Mary Landrieu of Louisiana, although not until after conducting an investigation over the objection of Senate Democrats. *See* Lizette Alvarez, *Senate Election Inquiry Clears Democrat from Louisiana*, N.Y. TIMES, Oct. 2, 1997, at A17.

[295] Because of their authority to judge the elections of their own members, legislatures

sents a close question, even a nakedly partisan outcome determined by a politically accountable branch may be as defensible as a judicial remedy under ambiguous statutory standards. Ambiguous statutes may leave courts relatively unmoored to resolve election contests, and with much less accountability than the legislative branch. Accordingly, this approach has the benefit of insulating courts from the political process.

Another alternative to judicial resolution of contested elections is an impartial administrative tribunal. The members of such a panel could be appointed in advance of each election on a neutral or bipartisan basis, ready to serve in the event of a contest. We might be comfortable giving these bodies more discretion than we are comfortable giving the legislature, or even judges (who may be partisan appointees, or who may themselves be subject to election), to craft a remedy that best recreates the will of the voters, as expressed in the election under review.[296]

A third option is a special election contest panel composed of members of the judicial branch. For instance, Kansas provides that a panel of three state district court judges will hear contests involving statewide elections,[297] while Iowa provides that a panel of four state district court judges and one supreme court justice will review contests involving its presidential electors or its congressional delegation.[298] A special panel also could be created akin to an arbitration panel, with the contestant picking one member, the contestee picking another member, and those two members jointly picking a third member.[299] In these and other configurations, special panels can be structured to reduce the potential that the outcome reflects the partisan bias of a particular judge or court.

Not only should states give some care to identifying the institution best suited to resolve a particular type of election contest, they also should think about the specific powers that this tribunal should have. For instance, unless their contest authority expressly includes a power to direct the boards of elections to conduct a partial new election when they deem it the best remedy for a particular election failure, neither administrative nor legislative tribunals will have this remedial tool, and even courts may be reluctant to use their equitable powers to fashion such a remedy absent statutory authority.[300] These decisions in turn may depend on other choices

---

of course are already fully capable of manipulating the outcomes of these elections, but this authority does not appear to have been frequently abused.

[296] A similar option is to charge the state board of elections with hearing the full merits of an election contest action, and thereafter subject their decision to judicial review. *See* 10 ILL. COMP. STAT. ANN. 5/23-1.8b (West 2006). As long as the board of elections is a bipartisan panel, this has the advantage of insulating the court from allegations of partisan favoritism if the court is ultimately able to uphold the administrative decision.

[297] *See* KAN. STAT. ANN. § 25-1443 (2000).

[298] *See* IOWA CODE § 60.1 (2003).

[299] *See* IOWA CODE § 57.7 (2003) (prescribing this method for contest of public measure).

[300] In many instances, when a legislature voids an election, the technical result is that the disputed office is temporarily vacant. The vacancy is then filled according to statute,

that a state makes about its contest processes, including what remedies to favor (or allow) and in what circumstances.

E. Concluding Thoughts About Perfect Elections

Often implicit and occasionally explicit in the foregoing discussion is that when it comes to election perfection, our reach may exceed our grasp. Any system that preserves local control, guarantees anonymous voting, and seeks to determine a fair but final election outcome promptly will sacrifice some amount of accuracy or certainty.[301] In part, this is because such a system will inevitably allow greater possibilities for voting fraud. But even absent fraud, it is harder to remedy mistakes in a system that favors speed over certainty, anonymity over auditability, and local control over uniformity.[302]

Accordingly, one helpful reform in the area of election remedies would be to educate the general public about the realistic limits of our election system. As long as we have done as much as is reasonable to minimize the potential for error, while ensuring that the residual errors that do occur do not systematically favor one candidate or party,[303] we should be content with the remaining imperfections in our elections.

It is probably a fiction to believe that in a very close election we can know with certainty which candidate was supported by the greatest number of eligible voters who attempted to vote.[304] The expected error rates in the processing and tabulating of tens of thousands of ballots can often dwarf the declared margin of victory in close races. Instead, we should recognize that in extremely close races, what we have is a functional tie, and that unless we have a true mathematical tie (in which case most codes provide for a coin toss or drawing of lots to decide the winner), the contest process effectively functions as our tiebreaking mechanism.

---

which may call for a new election. In that sense, the legislature itself is not technically calling for a new election. If the legislature does not want to void the entire election, however, its remedial options may be limited because it lacks a mechanism for triggering only a partial new election.

[301] *Cf.* GOV'T ACCOUNTABILITY OFFICE, THE NATION'S EVOLVING ELECTION SYSTEM AS REFLECTED IN THE NOVEMBER 2004 GENERAL ELECTION 31 (2006) (concluding that "administration of election systems will never be error free or perfect").

[302] For a fuller discussion of the impossibility of holding perfect elections, see generally Foley, *The Legitimacy of Imperfect Elections*, *supra* note 33.

[303] *Cf.* Black v. McGuffage, 209 F. Supp. 889, 891 (N.D. Ill. 2002) (describing equal treatment of voters, not perfect accuracy, as the goal of judicial supervision of elections); Foley, *supra* note 33, at 110 (stressing that voting system errors must be randomly distributed).

[304] *Cf.* Huggins v. Superior Court in and for County of Navajo, 788 P.2d 81, 86 (Ariz. 1990) (observing that in conducting elections, "we lack the luxury of perfection"). Shortly after the 2000 election, mathematics professor John Allen Paulos noted that our effort to determine who was the true victor of the presidential race in Florida (and hence in the nation) was like trying to measure bacteria with a yardstick. *See* John Allen Paulos, Op-Ed., *We're Measuring Bacteria with a Yardstick*, N.Y. TIMES, Nov. 22, 2000 at A27.

Of course, we also have to live with the fact that the contest process itself cannot be perfect, and may not truly identify the candidate who has the majority's support. Our main solace then may be that a new election cycle will be upon us shortly, when we can see if the voters have developed a clearer preference.

CONCLUSION

Although society may not be eager to embrace the notion that elections are necessarily imperfect, acknowledging this fact might allow us to be more satisfied with our reform efforts, and also might enhance the overall legitimacy of our election system. For instance, for some categories of election failures we might then be more comfortable with carefully designed (and predetermined) methods of statistical adjustment (just as we may already acquiesce in using a coin toss for a literal tie). This would increase both the speed and efficiency of remedying these failures. We might also be more open to partial new elections in some circumstances, even though it would mean that only some voters (those participating in the partial revote) would have an additional opportunity to learn more about, and be further wooed by, the candidates.

In part, such an understanding will depend on explicitly acknowledging the fact that our election processes serve multiple and sometimes conflicting values. For instance, any perfectly accurate election system will be excessively time consuming or cost-prohibitive, while any perfectly expeditious or cost-efficient election system will not be perfectly accurate. Furthermore, in those instances when an election is demonstrably inaccurate, sometimes there may be no perfect solution, given the secrecy of the ballot and the system's inability to recreate a purified version of a tainted election. That is, even if the cost, inconvenience, and other burdens of holding a new election could be eliminated, rerunning an election simply would not generate an accurate or corrected accounting of the will of the voters from the original election, but would instead generate an accounting of a new, or substitute, will of the voters.

Nevertheless, it is important to develop satisfactory, rather than perfect, methods of dealing with inevitable election disputes. These methods can be substantially improved through a clearer legislative articulation of election contest standards and remedies. In particular, election codes should specify in detail which remedial options—such as new elections, statistical (or other) adjustments to vote totals, or reversing the outcome—are appropriate for which types of election problems. In other respects as well, contest provisions should minimize the discretionary judgments that courts otherwise would make in the heat of a pitched partisan election controversy, such as what standard of evidence to use, and should generally reflect careful judgments about when and how courts and other tribunals should resolve various types of post-election controversies.

For at least the short term, these efforts to clarify election contest processes must occur at the state level, taking as a given the variety of local conditions that exist in each state. This approach capitalizes on our federal system's flexibility, whether thought of in terms of the innovation spawned by Justice Brandeis's laboratories of democracy,[305] or in terms of James Madison's solution to the tyranny of faction.[306] However, notwithstanding important state differences and the benefits of varied state election contest processes, most states share the problem of excessive judicial discretion in their procedures for resolving election contests. Accordingly, every state ought to seriously consider how it might adopt—and adapt—certain model election provisions, to protect both the judicial and the representative branches.

To that end, this Article has offered a framework for inventorying typical types of election failures, as well as possible remedies. It also has offered a framework for linking election failures and remedies together in a way that will protect the legitimacy of our democratic processes, in light of several competing values and priorities. These are difficult issues, and much additional work remains to be done, both in identifying the strengths of various state approaches and in analyzing possible reforms. But there should be little doubt about the importance of this enterprise.

[305] *See* New State Ice Co. v. Liebman, 285 U.S. 262, 311 (1932) (Brandeis, J., dissenting).
[306] *See* THE FEDERALIST No. 10 (James Madison).

ARTICLE

MISGUIDED ENERGY: WHY RECENT LEGISLATIVE, REGULATORY, AND MARKET INITIATIVES ARE INSUFFICIENT TO IMPROVE THE U.S. ENERGY INFRASTRUCTURE

JOSHUA P. FERSHEE*

This Article argues that recent legislative and regulatory attempts to address inadequate energy infrastructure in the United States are too limited in scope and rely too heavily on market-based initiatives to stimulate the urgent improvements that are necessary. The Article analyzes the likely effects of the Energy Policy Act of 2005, challenging the assumption that the provisions intended to remove potential impediments to investment--including those repealing the Public Utilities Holding Company Act and modifying the merger review authority of the Federal Energy Regulatory Commission ("FERC")--are likely to result in significant new investment in energy infrastructure. In addition to identifying remaining impediments to FERC's siting authority, the Article explains how FERC's increased use of market-based rates is insufficient to attract the necessary capital for infrastructure construction, considering the long lead times involved and the potentially tragic effects that can befall consumers in the interim. As a model of a successful regulatory approach, the Article examines the limited success of small-scale emergency orders in improving the energy infrastructure in targeted areas during the aftermaths of the California Energy Crisis and Hurricanes Katrina and Rita. The author concludes by calling for similar targeted intervention, albeit on a larger scale, and specifically recommends the expansion of federal siting authority for new construction and the provision of financial incentives tied to realistic deadlines to attract new investment.

I. INTRODUCTION

Soaring energy prices, natural gas supply shortages, and blackouts in major areas of the United States have led to a flurry of legislative and regulatory activity. Through this activity, lawmakers and regulators purport to resolve problems regarding natural gas and electricity supplies and service reliability.[1] A major goal of these actions has been to address

* Assistant Professor, University of North Dakota School of Law; J.D., Tulane Law School, 2003; B.A., Michigan State University, 1995. This Article was drafted primarily while the author was an attorney at Hogan & Hartson LLP, Washington, D.C., and was completed while the author was a visiting assistant professor at Penn State Dickinson School of Law. The opinions and conclusions in this Article are exclusively those of the author, as are any errors or omissions. The author wishes to thank his colleagues and mentors at Hogan & Hartson and Professor Jamie Colburn. He dedicates this Article to his wife, Kendra, and his son, Holden, who make every day amazing.

[1] Gasoline supply and prices are the most publicized of the U.S. energy issues; how-

the overall energy crisis by increasing investment in the U.S. energy infrastructure.[2] However, as is often the case with political remedies for difficult problems, what is being done and what legislators and policymakers claim is being done are two entirely different things. Recent legislative and regulatory policies are simply ill-equipped to have any substantial impact on the nation's energy infrastructure in the foreseeable future. Although some of the policies provide long-term hope for increasing the amount and sources of capital available for investment, they are not adequate solutions to a current, and progressing, energy crisis.

The goal of increasing investment in U.S. energy infrastructure is well-founded. The most notable recent infrastructure failure was the blackout of August 2003, which left more than fifty million people in Canada and the Great Lakes, New England, and mid-Atlantic regions without power and reportedly caused $10 billion in damage in the United States alone.[3] More recently, an unexpected 101-degree April day in Texas led to rolling blackouts affecting approximately 210,000 homes.[4] Such failures cause a "ripple effect of disruption and damage far beyond the energy industry's own domain"[5] because of the substantial economic investments that are based on electricity being available at predicted levels and costs.[6]

And the crisis likely will only get worse. The current energy infrastructure is insufficient in light of current demand, and the demand for energy in the United States is projected to increase at an average annual rate of 1.5% per year.[7] In 2005, the volume of electricity generation rose 2.1% and electricity sales increased by 3.2% from 2004 levels.[8]

Unfortunately, such failures are not shocking. A dearth of investment in electricity transmission infrastructure is a significant part of the prob-

---

ever, beyond a few brief comparisons, gasoline-related issues are beyond the scope of this Article.

[2] Energy infrastructure includes all infrastructure related to the production, generation, transmission, or distribution of energy. *See* Critical Energy Infrastructure Information, 71 Fed. Reg. 58,273, 58,274 (Oct. 3, 2006) (providing the definition of Critical Energy Infrastructure Information). However, the most glaring infrastructure need, and thus the primary focus of this Article, relates to the interstate electric transmission lines.

[3] *See* Lianne Elliott, *Brink of a Blackout?; Controversy over May Power Surge Spills over the Canada-U.S. Border*, KITCHENER-WATERLOO RECORD (Ontario), June 22, 2005, at A1.

[4] *See* Tom Fowler, *Rolling Blackouts as Texas Heats Up*, HOUSTON CHRON., Apr. 18, 2006, at A1; Barry Shlachter et al., *Shlachter, Perotin, Fuquay & Co.*, FORT WORTH STAR-TELEGRAM, Apr. 24, 2006, at C1 ("Even in Texas we might be forgiven for feeling a little nervous about energy supplies—and prices.").

[5] Robert C. Fellmeth, *Plunging Into Darkness: Energy Deregulation Collides with Scarcity*, 33 LOY. U. CHI. L.J. 823, 825, 830 (2002).

[6] *See id.* at 830 (stating that the prices paid for ineffective deregulation are "momentous" for small businesses, consumers, and energy infrastructure and that there are numerous other indirect victims).

[7] ENERGY INFO. ADMIN., ANNUAL ENERGY OUTLOOK 2007 WITH PROJECTIONS TO 2030, at 6 (2007), *available at* http://www.eia.doe.gov/oiaf/aeo/pdf/0383(2007).pdf.

[8] ENERGY INFO. ADMIN., ELECTRIC POWER ANNUAL 2005, at 1 (2006), *available at* http://www.eia.doe.gov/cneaf/electricity/epa/epa.pdf.

lem. Electric transmission investment, in real dollars, declined for the twenty-three years between 1975 and 1998.[9] Investment increased after 1998 but remains below 1975 levels.[10] In the same time period, demand for electricity has more than doubled.[11] Perhaps more illustrative of the lack of infrastructure investment is that "the interstate transmission system expanded [merely] by a total of 0.6 percent in circuit miles" in 2004.[12] There are indications that transmission investment has been growing considerably since 1999, but additional transmission investment remains necessary because current investment is not necessarily increasing efficiency.[13]

Rather than establishing an effective transportation grid, developers currently tend to invest in electricity transmission only where it is clearly necessary for reliability or where it lowers local costs.[14] Both of these are good reasons for investment, but new investment is also needed to create an effective nationwide transportation network that will facilitate long-distance electricity transportation. Such a network would provide economic benefit and improve reliability, and would facilitate a better energy market through increased energy source options.[15]

The amount of capital needed for infrastructure investment is staggering: "Energy industry spokespeople have called for grid investments of $56 billion, $100 billion, and even as much as $450 billion in total electricity infrastructure investments."[16] And these infrastructure investments are needed now.[17] Congress, regulators, and other leaders should address the problem directly because a failing infrastructure truly is a crisis. Yet, despite political rhetoric to the contrary, these actors are apparently reluctant to wage a more full-scale, direct attack on the problem, perhaps in

---

[9] Press Release, Fed. Energy Regulatory Comm'n ("FERC" or "Commission"), Commission Proposes Transmission Pricing Reforms to Increase Power Grid Investment (Nov. 17, 2005) [hereinafter Transmission Pricing Release], *available at* http://elibrary.ferc.gov/idmws/common/OpenNat.asp?fileID=10882511.

[10] *Id.* at 1.

[11] *Id.*

[12] Joseph T. Kelliher, Chairman Kelliher on Transmission Pricing Proposed Rules (Docket No. RM06-4-000) (Nov. 17, 2005), *available at* http://www. ferc.gov/press-room/statements-speeches/kelliher/ 2005/11-17-05-kelliher-pricing.pdf.

[13] *See* Elliot Roseman, *The Energy Policy Act of 2005: Striking the Right Federal-State Balance*, WORLD GENERATION, Sept.-Oct. 2005, at 18, 19, *available at* http://www.icfi.com/markets/energy/doc_files/energy-act-balance.pdf (arguing that to achieve greater efficiency the EPAct 2005 needed to include provisions designed to require building of certain types of transmission infrastructure because "[w]hile the reversal of the downward trend in transmission investment is welcome, quantity [alone] does not always maximize efficiency").

[14] *See id.* at 18.

[15] *See id.*

[16] UNION OF CONCERNED SCIENTISTS, LESSONS FROM THE AUGUST 2003 BLACKOUT (Aug. 10, 2005), http://www.ucsusa.org/clean_energy/clean_energy_policies/lessons-from-the-august-2003-blackout.html.

[17] *See* Claire Poole, *High Voltage Capital*, DAILY DEAL, Sept. 12, 2005 (stating that the energy "sector still needs capital, perhaps $100 billion over the next 10 years for upgrades and new generation and transmission so blackouts like the one in the Northeast in 2003 don't happen again").

part because it is not clear that appropriate short-term measures are readily available. Assertive measures, more expansive than the "emergency measures" taken by the Federal Energy Regulatory Commission ("FERC" or "Commission")[18] in response to the California energy crisis of 2000 to 2001 ("2000-2001 California Energy Crisis") and Hurricanes Katrina and Rita,[19] are necessary on a sustained, national scale.[20]

There are strong indications that infrastructure investment is key to alleviating many of the energy issues affecting the United States.[21] The 2003 blackout affecting much of the midwestern United States and the New York metropolitan area demonstrated that the current energy infrastructure cannot always satisfy peak demand[22] and lacks important redundancies that would improve reliability.[23] The chaos that accompanied these mass power outages[24]—the result of both infrastructure and operational failures[25]—indicates that the United States could be especially vulnerable to targeted and deliberate attacks on its power supplies.[26] This risk is yet another reason to support improvements in the U.S. energy

---

[18] Please note that citation of FERC orders and cases herein is consistent with FERC's guidance for the citation of FERC orders. *See* 2 FERC, FERC PRACTICE & PROCEDURE MANUAL ¶ 2003.05 (2007) (quoting Notice Regarding Paragraph Numbering in Commission Orders (Dec. 19, 2001) (unreported)). The manual provides one citation format for orders issued before June 26, 2002 and another for orders after that date. *See id.*

[19] *See infra* Part III.B.

[20] FERC regulates wholesale sales of electricity and natural gas, which is then sold via retailers to consumers under individual state regulatory schemes. *See* Schneidewind v. ANR Pipeline Co., 485 U.S. 293, 200–01 (1998). Such review of the activities at a federal level provides valuable insight regarding likely investment around the United States. Jim Rossi, *Transmission Siting in Deregulated Wholesale Power Markets: Re-Imagining the Role of Courts in Resolving Federal-State Siting Impasses*, 15 DUKE ENVTL. L. & POL'Y F. 315, 315–16 (2005) (arguing that "courts can play a positive role to facilitate the resolution of state-federal siting conflicts"). Of course, "[d]uring most of the twentieth century, state and local regulatory bodies coordinated the siting of power plants and transmission lines." *Id.* However, because of the many variations among the states and the relatively low-impact of other federal agencies on energy infrastructure investment, this Article focuses on federal legislation and FERC regulation.

[21] *See, e.g.*, Neil J. Numark & Robert D. MacDougall, *Nuclear Power in Deregulated Markets: Performance to Date and Prospects for the Future*, 14 TUL. ENVTL. L.J. 463, 464 (2001) ("Electricity deregulation in the United States is still in the first inning of play, and some worry the future will be darker if deregulation does not offer sufficient incentives for new power plant construction.").

[22] *See id.* (stating that "the final report of the U.S.-Canada Power System Outage Task Force . . . concluded that poor free maintenance along transmission lines in the First Energy (Ohio) service area, combined with inoperative computer software and operator errors, were the proximate cause of the blackout").

[23] *See* MATTHEW H. BROWN & RICHARD P. SEDANO, ELECTRICITY TRANSMISSION: A PRIMER 9 (2004), *available at* http://www.raponline.org/Pubs/ELECTRICITYTRANSMISSION.pdf ("If the transmission system is robust, with a certain amount of redundancy built in, it can withstand the failure of its most critical lines or other components.").

[24] *See, e.g.*, Phillip Coorey & Larry McShane, *Chaos Rocks Crisis Cities*, SUNDAY MAIL (Austl.), Aug. 17, 2003, at 38 (noting that millions were without power as a result of "America's worst power failure").

[25] *See* Numark & MacDougall, *supra* note 21, at 464.

[26] *See* Steven E. Roberts, *Protect the Power Grid*, 52 MIAMI DAILY BUS. REV. 6 (2005) ("Security experts have long warned that the electric grid is vulnerable to terrorism . . . .").

infrastructure. Improved energy infrastructure would not, of course, prevent terrorist attacks. But an improved infrastructure would help mitigate the damages and difficulties stemming from power outages, regardless of the cause.

The most high-profile portion of the Energy Policy Act of 2005 ("EPAct 2005")[27] represents a response to the complaints of business and corporate leaders that restrictions on corporate structures and other regulatory hurdles have limited the number of available sources of capital for investment in utilities. Prior to the enactment of EPAct 2005, business leaders continually argued that regulatory changes (that is, relaxed regulation) were essential to increase investment.[28] In response to the current and continuing "energy crisis,"[29] Congress took action[30] to remove several long-standing restrictions on the corporate structure and governance of U.S. utilities, and thereby improve, at least in theory, the U.S. energy infrastructure.[31] Politicians, regulators, and corporate leaders have all lauded these recent activities as crucial steps that will increase capital investment and "help modernize our aging energy infrastructure."[32]

Other less prominent portions of EPAct 2005 could increase investment more directly if the new or modified grants of power are actively and aggressively used. In one of the provisions providing the most promise, FERC was granted "backstop authority"[33] for siting interstate electric

---

[27] Pub. L. No. 109-58, 119 Stat. 594 (2005) (codified at 42 U.S.C.S. §§ 15801–16524 and scattered sections of 15 U.S.C.S and 16 U.S.C.S (LexisNexis 2006)).

[28] *See* Markian M. W. Melnyk & William S. Lamb, *PUHCA's Gone: What is Next for Holding Companies?*, 27 ENERGY L.J. 1, 1 (2006) ("[The Public Utilities Holding Company Act of 1935 ("PUHCA")] was widely believed to have discouraged investment in electric and gas utility infrastructure by companies that could not restructure to satisfy PUHCA's prohibition on the ownership of diversified business."); *see also infra* note 67 and accompanying text.

[29] *See, e.g.*, California *ex rel.* Lockyer v. FERC, 329 F.3d 700, 703, 711 (9th Cir. 2003) (denying petition seeking review of FERC's decision to expedite approval of a corporate reorganization and finding that FERC had acted properly "to mitigate the growing California energy crisis").

[30] Pamela A. MacLean, *9th Circuit Tackles Natural Gas Price-Fixing*, NAT'L L.J., Feb. 12, 2007, at 15 (stating that "Congress passed the [EPAct 2005] in response to the energy crisis").

[31] *See infra* Part II.A.

[32] *See* Press Release, White House, President Bush Signs Into Law a National Energy Plan (Aug. 8, 2005) [hereinafter White House Press Release I], *available at* http://www.whitehouse.gov/news/releases/2005/08/20050808-4.html.

[33] *See* EPAct 2005 § 1221, 16 U.S.C.S. § 824p (LexisNexis 2006); *see also* Roseman, *supra* note 13, at 18. This limited authority is called "backstop" because it is available only where states lack the authority or otherwise have failed to act. The term "backstop" has often been used to describe this type of federal authority in the energy industry. *See, e.g.*, 149 CONG. REC. H11415 (2003) (statement of Rep. Joe Barton (R-Tex.)) (referring to the Energy Policy Act of 2003, S. 2095, 108th Cong. (2003), which was never passed, as "federal backstop authority for siting of new transmission lines"); 147 CONG. REC. S9158 (2001) (statement of Sen. Jeff Bingaman (D-N.M.)) (arguing that "[w]hat is needed is to use federal eminent domain as a backstop to a more cooperative, regionally based approach to transmission and siting issues" in discussing "legislative solution[s]" to energy problems in 2001).

transmission facilities.[34] This limited authority is available only in areas the Department of Energy ("DOE") identifies as a "national interest electric transmission corridor" ("NIETC").[35] DOE will issue a report in which it will "designate any geographic area experiencing electric energy transmission capacity constraints or congestion that adversely affects consumers as a [NIETC]."[36] In addition, EPAct 2005 grants FERC exclusive jurisdiction over siting, construction, expansion, and operation of liquefied natural gas ("LNG")[37] terminals.[38] However, despite the promise of these provisions, similar past initiatives have failed to produce significant results.[39]

EPAct 2005 was passed, in part, because the United States faces both short- and long-term energy issues, and while the recent high-profile activities may provide some long-term benefit, they do not offer much promise for remedying the very real short-term problems. As a sponsor of EPAct 2005, Senator Pete Domenici (R-N.M.) admitted: "It's not a bill for today or necessarily tomorrow—it's for the future."[40] The suggestion that EPAct 2005 may not be especially effective is not a unique proposition.[41] It is, after all, the same bill that Senator John McCain (R-Ariz.), among others, dubbed "the No Lobbyist Left Behind Act of 2005."[42]

Given the political rhetoric regarding both public safety issues and energy prices,[43] it would seem that the nation's leaders would be eager to incentivize significant and immediate infrastructure investment. However, recent legislative and regulatory actions provide, at best, long-term prom-

---

[34] EPAct 2005 § 1221, 16 U.S.C.S. § 824p (LexisNexis 2006).

[35] *Id.*

[36] *Id.*

[37] LNG is natural gas that is condensed into a liquid after having being cooled to minus 260 degrees Farenheit or below. *See* Monica Berry, *Liquefied Natural Gas Import Terminals: Jurisdiction over Siting, Construction, and Operation in the Context of Commerce Clause Jurisprudence*, 26 ENERGY L.J. 135, 137 (2005).

[38] EPAct 2005 § 311(c)(2), 15 U.S.C.S. § 717b(e)(1) (LexisNexis 2006) ("[FERC] shall have the exclusive authority to approve or deny an application for the siting, construction, expansion, or operation of an LNG terminal."). This power is limited by certain rights retained by the states pursuant to: (1) the Coastal Zone Management Act, 16 U.S.C. §§ 1451-1466 (2000); (2) the Clean Air Act, 42 U.S.C. §§ 7401-7431 (2000); and (3) the Federal Water Pollution Control Act, 33 U.S.C. §§ 1251–1274 (2000).

[39] *See infra* Part III (discussing the limited success of FERC's emergency orders designed to remedy infrastructure problems related to Hurricanes Rita and Katrina).

[40] Jim VandeHei & Justin Blum, *Energy Bill Unlikely to Lower Prices*, CHARLOTTE OBSERVER, Aug. 9, 2005, at 1A, 7A.

[41] *See, e.g.*, Robert Westervelt, *Dow CEO Urges Action on Natural Gas "Crisis,"* CHEM. WEEK, Nov. 16, 2005, at 11 ("The Energy Policy Act of 2005 was a good start toward addressing the supply-demand problem that drives the U.S. natural gas crisis, and I commend [Congress] for it. But that is not enough. Congress must act now." (internal quotation marks omitted) (quoting Andrew Liveris, Chairman and CEO, Dow Chemical Co.)).

[42] Peter Van Doren & Jerry Taylor, *A Low-Voltage Energy Bill*, PUB. UTIL. FORTNIGHTLY, Oct. 2005, at 52.

[43] *See* 151 CONG. REC. H6949 (daily ed. July 28, 2005) (statement of Rep. Richard Pombo (R-Wash.) ("The [EPAct 2005] is a good first step in the effort to lower energy prices and reduce our dependence on foreign energy.").

ise rather than reasonably quick fixes to what is an imminent concern.[44] The political claims that EPAct 2005 comprehensively addresses the looming "energy crisis" [45] simply do not accurately describe the actions taken. Certainly, it partly addresses the energy crisis, but for all the discussion about the major problems, the proposed fixes are either small in scope or do not address the infrastructure problems.

There are three primary ways in which legislators and regulators have attempted to increase investment in energy infrastructure. First, with the hope of increasing utilities' access to capital, they have removed or relaxed several barriers to capital investment.[46] Second, they have permitted incentive pricing policies, including market-based rates (as opposed to traditional, cost-based rates)[47] and favorable tax treatment for new investments[48] in certain circumstances for both natural gas and electricity. Third, the regulatory approval processes for mergers and acquisitions and for construction of new facilities have been streamlined, at least in theory.[49]

The present programs designed to address major infrastructure are too vague and ill-defined to initiate major construction projects. Conversely, recent short-term efforts are so limited that vast infrastructure needs remain even in the targeted areas. To address a large-scale energy crisis, a coherent and comprehensive federal energy program is needed. A major program must be designed to identify energy infrastructure needs quickly and accurately, provide attractive financial incentives, and provide aggressive yet feasible deadlines to motivate investors. The authority of the fed-

---

[44] *See, e.g.*, OilOnline, *CEO Alliance Calls on Congress to Take Immediate Action to Increase Natural Gas Supplies*, OILONLINE INTERNET INQUIRER, Mar. 28, 2006, http://www.oilonline.com/news/headlines/internet/20060328.CEO_Alli.20760.asp ("I believe we are facing an energy crisis of epic proportions and in order to solve it, policy makers must take immediate action." (internal quotation marks omitted) (quoting Larry Downes, CEO, New Jersey Resources)).

[45] *See, e.g.*, 151 CONG. REC. S9344 (daily ed. July 29, 2005) (statement of Sen. Maria Cantwell (D-Wash.)) ("[EPAct 2005] also takes steps to respond to the disastrous western energy crisis, which extracted billions of dollars and hundreds of thousands of jobs from our regional economy."); 151 CONG. REC. S6885 (daily ed. June 21, 2005) (statement of Sen. George Voinovich (R-Ohio)) (stating that "[EPAct 2005] tries to address" the high cost of energy).

[46] *See infra* Part II.A.

[47] *See, e.g.*, *In re* Cal. Wholesale Elec. Antitrust Litig., 244 F.Supp. 2d 1072, 1076 (S.D. Cal. 2003) ("[T]he key feature of California's recently deregulated wholesale energy markets is the markets' reliance on 'market-based rates.' These rates are still subject to FERC oversight, but to a much lesser extent than traditional 'cost-based rates.'"); Alternatives to Traditional Cost-of-Service Ratemaking for Natural Gas Pipelines and Regulation of Negotiated Transportation Services of Natural Gas Pipelines, 61 Fed. Reg. 4633, 4633 (Feb. 7, 1996) (stating that FERC had been "exploring the criteria it should use when evaluating rates established through methods other than the traditional cost-of-service ratemaking method" and was "now providing the industry with guidance by stating the criteria it will consider when evaluating proposals for market-based rates"). *See generally infra* Part II.B.

[48] *See* 18 C.F.R. § 35.35(d)(1)(v), (vii) (2007) (providing for accelerated depreciation and deferred cost recovery).

[49] *See infra* Part II.C.

eral government must be expanded and exercised to ensure that regulatory delays do not impede the process.

This Article reviews recent federal legislative and regulatory activity seeking to improve energy infrastructure and analyzes the ability of each action to achieve this goal efficiently and effectively. In Part II, the Article reviews the current federal statutory scheme and related developments in market rates and pricing. Part II first briefly reviews EPAct 2005, which includes the historic repeal of the Public Utilities Holding Company Act of 1935 ("PUHCA").[50] This discussion considers the events leading to the passage of PUHCA and its subsequent repeal and assesses the likely impact (both positive and negative) that the repeal of PUHCA will have on timely investment in energy infrastructure. Part II next describes the use of pricing incentives and the advent of market-based rates in wholesale energy markets, and elucidates the apparent economic and political rationales behind such incentives and rates. This Part then summarizes recent and expected developments related to these incentive pricing programs. Part II concludes by looking at recent legislation and the subsequent regulatory actions related to approval of mergers and acquisitions under the Federal Power Act ("FPA").[51]

Part III explains that while significant legislative and regulatory activities abound, the actions to date provide little reason to expect significant changes in the near future. This Part considers recent FERC action in response to certain "emergencies" related to the 2000–2001 California Energy Crisis and Hurricanes Rita and Katrina and argues that these emergency actions provide yet another example of plans that fail to effectively improve the nation's energy infrastructure in a timely way. This Part argues that more aggressive action is necessary to implement the programs effectively.

Finally, Part IV concludes that it is time for an aggressive and innovative plan that will lead to immediate and sustained energy infrastructure enhancements. Even with financial incentives tied to specific deadlines for putting new facilities in service, such as those used in FERC's emergency orders, improvements in energy infrastructure under current policies are insufficient to effectuate real change. A targeted and comprehensive program is needed to ensure new facilities are built in a timely and effective manner.

---

[50] 15 U.S.C. § 79(a)–(z-6) (2000) (repealed 2005).

[51] Federal Power Act, 16 U.S.C. §§ 791a–797, 798–824a, 824b–825r (2000).

II. INDIRECT MARKET-BASED ENERGY POLICIES: HIGH HOPES, UNREALISTIC EXPECTATIONS

A. The Impact of PUHCA Repeal: New Investment or Just New Investors?

> Of all the laws regulating utilities today, PUHCA may well be the most antiquated. Its detailed provisions continue to inhibit the market discovery process and to ward off hobgoblins that have long ceased to exist.
>
> —R. Richard Geddes[52]

> Closing the barn door after the horses have fled is a futile act. Public-utility holding companies and their subsidiary companies are affected with a national public interest, and consumers and investors are harmed by the lack of effective public regulation to prevent abuses similar to those that gave rise to the enactment of PUHCA.
>
> —Rep. John D. Dingell (D-Mich.)[53]

When President Bush signed EPAct 2005 on August 8, 2005,[54] it signaled the removal of one of the longest standing and most significant regulatory hurdles facing investors in the U.S. energy industry: PUHCA.[55] Enacted in 1935, PUHCA was passed as part of President Franklin Delano Roosevelt's New Deal.[56] The goal of PUHCA was to regulate "for the equal benefit of the consumer and the investor."[57] PUHCA was created to protect investors and consumers—who had lost billions of dollars in the crash of Samuel Insull's utility holding company empire—"by au-

---

[52] R. Richard Geddes, *Time to Repeal the Public Utility Holding Company Act*, 16 CATO J. 63, 75 (1996).

[53] Press Release, Rep. John D. Dingell (D-Mich.), Ranking Member, H. Comm. on Energy and Commerce, GAO Commits to Report to Dingell, Markey by June 17, 2005, on Impact of Lax SEC Enforcement of PUHCA (Jan. 27, 2005), *available at* http://energycommerce.house.gov/Press_109/109nr1.shtml.

[54] Pub. L. No. 109-58, 119 Stat. 594 (2005).

[55] 15 U.S.C. § 79(a)–(z-6) (2000), *repealed by* EPAct 2005 § 1263, 119 Stat. 594, 974. *See* Foster Assocs., *Public Citizen Says FERC is Unprepared to Confront a Coming Tsunami of Oil-Gas-Power Generation Mergers*, FOSTER NATURAL GAS REP., Nov. 10, 2005, at 7 (quoting a citizen advocacy group that stated that PUHCA stood as a kind of "merger barrier reef" for more than seventy years and that FERC "must prepare for a likely tsunami of proposed non-utility acquisitions and other new types of public utility mergers made possible" by its repeal).

[56] *See* Pub. L. No. 74-333, 49 Stat. 803 (1935); *see also* Richard D. Cudahy, *70th Anniversary Celebration of the Federal Power Act*, 26 ENERGY L.J. 389, 390 (2005) ("PUHCA has, for better or for worse, become the black sheep of New Deal regulatory legislation.").

[57] Alan Richardson & John Kelly, *The Relevance and Importance of Public Power in the United States*, 19 NAT. RESOURCES & ENV'T 54, 56 (2005).

thorizing the SEC to regulate the financial and corporate transactions of registered holding companies that owned utility subsidiaries in more than one state."[58]

As of 1930, Insull controlled nearly ten percent of the electricity in the United States,[59] and was a monopoly service provider in the Chicago area.[60] Insull built his conglomerate, Commonwealth Edison, through a series of holding companies in a sort of a pyramid scheme,[61] which created enormous profits until just after the market crashed following the Great Depression.[62] The collapse of the "Insull monstrosity"[63] led to the passage of PUHCA and the resulting restrictions on the corporate structures of public utility holding companies would last for more than seventy years.[64]

PUHCA allowed utility holding companies to own electricity distribution systems in only a single state or region and prevented them from owning businesses that were not functionally or otherwise related to their energy business.[65] Each utility was to operate as a solitary, integrated system,[66] and thus, PUHCA significantly discouraged ownership of U.S. electric and natural gas utilities by domestic industrial and financial institutions and by foreign institutions.[67]

---

[58] Jacqueline Lang Weaver, *Can Energy Markets Be Trusted? The Effect of the Rise and Fall of Enron on Energy Markets*, 4 HOUSTON BUS. & TAX L.J. 1, 116 (2004).

[59] Richardson & Kelly, *supra* note 57, at 55.

[60] *See* David B. Spence, *The Politics of Electricity Restructuring: Theory v. Practice*, 40 WAKE FOREST L. REV. 417, 419 n.7 (2005) ("As founder of Commonwealth Edison in Chicago, Insull was the first to secure a charter to provide monopoly service in Commonwealth Edison's service area.").

[61] *See* Richard D. Cudahy & William D. Henderson, *From Insull to Enron: Corporate (Re)Regulation After the Rise and Fall of Two Energy Icons*, 26 ENERGY L.J. 35, 58 (2005) ("[I]nvestors near the top of the pyramid . . . bear the bulk of the risk that is typically associated with leverage—in good times, rapidly burgeoning profits; in bad times, the danger of losses . . . . Insull's highly leveraged empire was obviously highly dependent on the continued growth of the electric industry.").

[62] *See* Sidney A. Shapiro & Joseph P. Tomain, *Rethinking Reform of Electricity Markets*, 40 WAKE FOREST L. REV. 497, 505 (2005). Although he was eventually acquitted, Insull was accused of fraud and there were claims of stock manipulation. *See* Cudahy & Henderson, *supra* note 61, at 39.

[63] Richardson & Kelly, *supra* note 57, at 55.

[64] Note that some commentators believe that the Securities and Exchange Commission ("SEC") was not enforcing PUHCA anyway. *See, e.g.*, Weaver, *supra* note 58, at 116–17 ("[I]t appears that the SEC has not been actively enforcing PUHCA."). However, the courts served as a check on this apparent lack of oversight. *See, e.g.*, Nat'l Rural Elec. Coop. Ass'n v. SEC, 276 F.3d 609, 617 (D.C. Cir. 2002) (finding that SEC approval of a proposed acquisition by American Electric Power Company did not comply with PUHCA and stating that the SEC's determination could not "withstand even the most deferential review because the SEC had failed to make any evidentiary findings").

[65] *See* 15 U.S.C. § 79k(b)(1) (2000) (repealed 2005); *see also* N. Am. Co. v. SEC, 327 U.S. 686, 704 (1946) ("[H]olding companies are compelled to integrate and coordinate their systems and to divest themselves of security holdings of geographically and economically unrelated properties.").

[66] *See* 15 U.S.C. § 79k(a) (repealed 2005).

[67] *See, e.g.*, David L. Sokol, *Discarding PUHCA*, 27 ELECTRIC PERSP. 10, 11–12 (2002) (stating that PUHCA kept "new investment out of the energy industry").

PUHCA also required that any holding company register with the Securities and Exchange Commission ("SEC") if it owned subsidiaries that operated utilities in more than one U.S. state.[68] The business structures and operations of these registered holding companies were severely restricted. Such companies were required to maintain a specified capital structure, their relationships with affiliates were limited, and potential diversification activities were constrained.[69] Furthermore, registered holding companies faced additional potential liabilities that could be imposed by federal and state rate regulators.[70]

Without PUHCA's restrictions, widely dispersed utility companies can now be owned without regard to where each utility is located or whether the resulting entity can be operated as a single system following a merger or acquisition. This change will likely accelerate consolidation in the energy sector and presents increasing opportunities for foreign investors interested in acquiring U.S. utilities.[71] Prior to passage of EPAct 2005, some acquisitions were apparently negotiated and proposed with the full expectation that PUHCA repeal was imminent. For example, MidAmerican-Pacifi-Corp, approved by FERC in December 2005,[72] would not have been permitted under PUHCA because of restrictions on "cross-country" transactions—transactions that would merge geographically remote electric utilities.[73]

The repeal of PUHCA provides diverse and ample investment opportunities for non-U.S. investors because it allows new kinds of nonutility inves-

---

[68] *See* 15 U.S.C. § 79e (2000) (repealed 2005). The repeal of PUHCA removed all SEC utility-specific oversight. The only regulatory oversight that remains under SEC purview is that which exists for any company, regardless of industry. *See, e.g.*, Securities Act of 1933, 15 U.S.C. §§ 77a–77bbbb (2000); Securities Exchange Act of 1934, 15 U.S.C. §§ 78a–78mm (2000). Congress shifted some responsibilities, such as oversight of certain record-retention and accounting policies, to FERC. *See* EPAct 2005 § 1273, 42 U.S.C.S. § 16461 (LexisNexis 2006) ("All books and records that relate primarily to the functions transferred to [FERC] under this subtitle shall be transferred from the [SEC] to [FERC].").

[69] *See* Seth A. Kaplan & Gregory N. Racz, *It Seemed Like a Good Idea at the Time: Recent Trends in Mergers and Acquisitions in the Electric and Gas Utility Industries*, *in* TELECOMMUNICATIONS MERGERS & ACQUISITIONS: FINANCING, REGULATORY AND BUSINESS ISSUES 491, 497 (PLI Corp. Law & Practice, Course Handbook Series No. B0-0079, 1998) ("PUHCA imposes limits on mergers and acquisitions and regulates capital structures, equity and debt financings, internal restructurings, the formation of subsidiaries, transactions between affiliates, the composition of boards of directors and other matters.").

[70] *See* Alabama Elec. Coop. v. SEC, 353 F.2d 905, 907 (D.C. Cir. 1965) ("The purpose of the [PUHCA], as shown by its legislative history, was to supplement state regulation—not to supplant it.").

[71] *See* Foster Assocs., *supra* note 55, at 7; Robert Robinson & Branko Terzic, *New Energy Law to Influence Mergers*, ENERGYBIZ MAG., Nov.-Dec. 2005, at 14, 14 ("One of the most important aspects of the [EPAct 2005] is the repeal of this Depression-era law, which unleashes a new set of M&A possibilities and facilitates the attraction of much-needed investment capital into the industry.").

[72] *See* MidAmerican Energy Holdings Co., 113 FERC ¶ 61,298, at P 1 (2005).

[73] *Cf.* Nat'l Rural Elec. Coop. Ass'n v. SEC, 276 F.3d 609, 610 (2002) (vacating an SEC order approving American Electric Power Company's acquisition of four wholly owned operating subsidiaries of Central and South West Corporation because the SEC did not justify its conclusions that the proposed acquisition would "satisfy the single-area-or-region requirement" or "the interconnection requirement" mandated at that time by PUHCA).

tors to enter the market without these restrictions. Additionally, PUHCA's repeal could significantly impact the overall business structure of utilities in the United States, potentially leading to additional consolidation among U.S. utilities and diversification by utility companies and their affiliates, especially in light of the removal of the structural and geographic restrictions PUHCA imposed.[74]

However, PUHCA's repeal[75] does not eliminate all regulatory obstacles to utility-related mergers and acquisitions. FERC and state regulatory commissions will continue to review mergers and acquisitions in the energy industry. EPAct 2005 requires that public utility holding companies and their affiliates and subsidiaries maintain "books and records" and make them available to FERC to ensure that consumers are protected with respect to jurisdictional rates (that is, the rates over which FERC already has jurisdiction under the FPA and the Natural Gas Act).[76] EPAct 2005 also extended FERC's merger approval authority over electric utilities to mergers and acquisitions, including stock acquisitions, by holding companies.[77]

In particular, now that PUHCA no longer serves as an initial defense in preserving local control over utility operations, state regulators are expected to scrutinize acquisitions of electric and gas distribution utilities by geographically distant companies, whether such companies are utility or nonutility in nature.[78] Under EPAct 2005, to the extent necessary to discharge their duties, state commissions have access to books and records comparable to those possessed by FERC.[79] It is not clear whether these changes will lead to additional (or reduced) investment by current utility owners; additional mergers or acquisitions being attempted and completed; or a rise in the number of new investors.

In addition to this change in the regulation of mergers and acquisitions, the repeal of PUHCA might increase the amount of capital available for investment by increasing the number of potential investors. Upon signing EPAct 2005, President Bush stated that "[t]he bill removes outdated obstacles to investment in electricity transmission lines in generating facilities"[80] and that it would "modernize the electricity grid."[81] How-

---

[74] *See, e.g.*, Poole, *supra* note 17 ("Repeal of the [PUHCA] opens the switches for $100 billion in investments the utility industry urgently needs . . . .").

[75] Note that, as required by EPAct 2005, FERC issued a final rule related to the repeal of PUHCA that implemented the Public Utility Holding Company Act of 2005 ("PUHCA 2005"). Public Utility Holding Company Act of 2005, 18 C.F.R. pt. 366 (2006) [hereinafter PUHCA 2005 Final Rule]. Although EPAct 2005 refers to "PUHCA 2005," see EPAct 2005 §§ 1261–1277, the repeal of the 1935 law effectively eviscerated PUHCA as it was understood for more than seventy years. *See* Foster Associates, *supra* note 55, at 7.

[76] EPAct 2005 § 1264, 42 U.S.C.S. § 16452 (LexisNexis 2006).

[77] *See infra* Part II.C.

[78] *See* Robinson & Terzic, *supra* note 71, at 14.

[79] EPAct 2005 § 1265, 42 U.S.C.S. § 16453 (LexisNexis 2006).

[80] Press Release, White House, President Signs Energy Policy Act (Aug. 8, 2005), *available at* http://www. whitehouse.gov/news/releases/2005/08/20050808-6.html.

[81] White House Press Release I, *supra* note 32.

ever, it is not clear that removing barriers to investment will actually increase investment in new facilities. Given the relaxed regulation of consolidation, PUHCA's repeal could simply trigger a wave of consolidation through mergers and acquisitions, particularly in the near term. This might result in mere change in the ownership of current facilities rather than the planning and construction of new facilities by new investors.

Ultimately, to the extent industry consolidation is furthered by the repeal of PUHCA, the resulting mergers could conceivably lead to the exercise of market power and, thus, to increased prices. The repeal of PUHCA may lead to this power shift because, at least to some degree, PUHCA protected consumers from anticompetitive behavior by large utilities.[82] Without PUHCA, merged utilities can inappropriately maintain and use increased market power largely because competitors are unable to enter the market following a merger-induced price spike.[83] If competitors could easily enter the market they would (relatively) quickly drive prices back down,[84] but the long timeline for most energy projects makes this an impossible outcome. Furthermore, a merged utility would be less inclined to invest in new infrastructure because, without PUHCA's restrictions on nonutility investment, it may pursue higher-risk, higher reward investments first,[85] because utility mergers themselves can be risky investments, and "in turn both stockholders and customers feel the pinch as the utility seeks to compensate for its overvalued investment."[86]

---

[82] *See, e.g.*, 152 CONG. REC. H2439 (daily ed. Apr. 21, 2005) (statement of Rep. Marty Meehan (D-Mass.)) ("[The EPAct 2005] would repeal [PUHCA] which prevents big energy firms, like Enron, from driving smaller utilities out of business and monopolizing the energy market."); AM. PUB. POWER ASS'N, THE ELECTRIC UTILITY INDUSTRY AFTER PUHCA REPEAL: WHAT HAPPENS NEXT? 1 (2005), *available at* http://www.appanet.org/files/PDFs/APPAreportAfterPUHCARepeal.pdf ("The effect [of EPAct 2005] will likely be greater consolidation of the electric industry, greater concentration of ownership, more complex company structures, and more opportunities for the exercise of market power.").

[83] *See* HERBERT HOVENKAMP, FEDERAL ANTITRUST POLICY § 1.6 (3d ed. 2005) ("[A] barrier to entry is some factor in a market that permits firms already in the market to earn monopoly profits, while deterring outsiders from coming in.").

[84] U.S. DEP'T OF JUSTICE & FED. TRADE COMM'N, HORIZONTAL MERGER GUIDELINES § 3.0 (rev. ed. 1997), *available at* http://www.usdoj.gov/atr/public/guidelines/horiz_book/30.html ("A merger is not likely to create or enhance market power or to facilitate its exercise, if entry into the market is so easy that market participants, after the merger, either collectively or unilaterally could not profitably maintain a price increase above premerger levels."); *see also* Geoffrey A. Manne & E. Marcellus Williamson, *Hot Docs vs. Cold Economics: The Use and Misuse of Business Documents in Antitrust Enforcement and Adjudication*, 47 ARIZ. L. REV. 609, 632 n.91 (2005) ("Entry is important in antitrust analysis because the threat of post-merger entry, exacerbated by the attractiveness of putative monopoly pricing, may ameliorate the negative price effect of a merger . . . .").

[85] *See* AM. PUB. POWER ASS'N, *supra* note 82, at 16 (stating that, following the repeal of PUHCA, restrictions that prevent utilities from being used as a "cash cow" to support unregulated investments should be maintained).

[86] *See* AM. PUB. POWER ASS'N, THE POST-MERGER EXPERIENCE 1 (2005), *available at* http://appanet.org/files/PDFs/APPAreportPost-MergerExperience.pdf (stating that although it is often assumed that larger companies "will have more capital to invest in infrastructure development . . . [p]ost-merger analysis . . . does not confirm these optimistic expectations").

Several legislators believe that consolidated market power is at least one reason current oil and gas prices are so high. In an interesting move, several members of the Senate Judiciary Committee recently introduced the Oil and Gas Industry Antitrust Act of 2006 ("Oil and Gas Act"), which is designed "to improve competition in the oil and gas industry" and strengthen antitrust enforcement of energy industry mergers.[87] The Oil and Gas Act is targeted primarily at "the escalating price of gasoline," [88] which have risen more than seventy percent between January 2001 and March 2007.[89] Other targets include the increased prices for other petroleum products, such as heating oil and natural gas.[90] Senator Specter, for one, has claimed that energy industry price increases are linked primarily to "rapid consolidation in the oil and gas industry," which created a "collusive environment" and gave market power to the remaining entities.[91] This, according to Senator Specter, created the opportunity to increase prices for oil and gas supplies beyond the proper market price.[92]

Beyond the effects of consolidation, recent history provides another indicator that deregulation, at least in the energy sector, does not lead directly to new infrastructure investment. The 2000-01 California Energy Crisis followed massive deregulation in the form of Assembly Bill 1890 ("AB 1890") in 1996,[93] and the investment in new power plants that was predicted to follow never came to fruition.[94] Some commentators have argued that deregulation is a major hurdle to the success of the market, including increased infrastructure investment, when, as in California, it creates redundant regu-

---

[87] S. 2557, 109th Cong. § 1 (2006). Interestingly, several of the senators, including Arlen Specter (R-Pa.), Mike DeWine (R-Ohio), Patrick Leahy (D-Vt.), and Dianne Feinstein (D-Cal.), who introduced this legislation to scrutinize energy industry mergers more closely, also voted for EPAct 2005 and the repeal of PUHCA. *See* Govtrack.us, H.R. 6: Energy Policy Act of 2005 (Vote On Passage), http://www.govtrack.us/congress/vote.xpd?vote=s2005-158 (providing the voting record for EPAct 2005). Certainly, these senators may have supported EPAct 2005 for other reasons, but this voting pattern may also indicate that the senators had some second thoughts about the wisdom of repealing PUHCA.

[88] 152 CONG. REC. S3213 (daily ed. Apr. 6, 2006) (statement of Sen. Specter).

[89] 153 CONG. REC. S2911 (daily ed. Mar. 8, 2007) (statement of Sen. Bernard Sanders (I-Vt.)).

[90] *See id.*

[91] *Id.* at S3213–14.

[92] *See id.* at S3214.

[93] Assem. B. 1890, 1995-96 Gen. Assem., Reg. Sess. (Cal. 1996) (codified at CAL. PUB. UTIL. CODE §§ 330–398.5 (1996)).

[94] *See* Brian Orion, *Transmission in Transition: Analyzing California's Proposed Electricity Transmission Regulatory Reforms*, 56 HASTINGS L.J. 343, 344 (2004) ("The promise [of AB 1890] was compelling: competition among generators of electricity would bring lower costs and greater investments in the state's energy infrastructure."); *see also* Jim Chappell, *Deregulation, Re-Regulation, and California's Energy Future*, S.F. PLANNING AND URBAN RESEARCH ASS'N NEWSL. (S.F. Planning and Urban Research Ass'n, S.F., Cal.) Aug. 2003, at 2, *available at* http://www.spur.org/documents/pdf/030801_article_05.pdf ("[The] predicted investments in new power plants following deregulation failed to take place while cutbacks in funding for utility energy efficiency programs reduced the surplus of power supplies that existed in the mid-1990s.").

lation by both the federal and state government.[95] This potential hurdle may be particularly worrisome against the backdrop of the newly created state access to books and records that will coincide with FERC's review.[96]

Indeed, redundant regulation may have recently affected the success of two proposed mega-mergers. The merger of PSEG with Exelon and of Constellation Energy with FPL Group, failed despite, and perhaps in part due to, PUHCA's repeal.[97] The PSEG-Exelon merger was a deal valued at about $17.7 billion that would have resulted in the United States' largest utility.[98] Exelon and PSEG obtained approvals from FERC, the U.S. Department of Justice, and Pennsylvania state regulators before getting bogged down by "insurmountable obstacles, chiefly rate concessions and power-plant divestitures sought by public officials and consumer advocates in New Jersey, where PSEG is based."[99] The companies' offer of $600 million in rate concessions proved insufficient to satisfy New Jersey regulators and overcome concerns regarding the resulting market power of the merged entities.[100] The failure was costly: Exelon spent more than $100 million in pursuit of the doomed merger.[101] Similarly, Constellation Energy and FPL Group called off their $12.5 billion merger[102] in the face of "continued uncertainty over regulatory and judicial matters in Maryland and the potential for a protracted and open-ended merger review process."[103] These failures have dampened the enthusiasm of many industry executives and analysts, who once believed that PUHCA's repeal would lead to a friendlier merger environment.[104] It now appears that PUHCA's repeal

---

[95] *See, e.g.*, Orion, *supra* note 94, at 377 (stating that California's regulatory redundancies created following deregulation "jeopardize[ ] the significant economic, environmental, and reliability benefits" deregulation was intended to reap).

[96] *See* 18 C.F.R. § 366.2 (2006) (giving states access to utility and natural gas company books and records when "necessary or appropriate for the protection of utility customers with respect to jurisdictional rates"); *cf. infra* notes 205–214 and accompanying text (discussing the state and federal jurisdiction issues remaining for LNG siting).

[97] *See* Housley Carr & Ray Pospisil, *Constellation-FPL Merger Collapse to Hurt BGE Ratepayers, Chills Future Industry Mergers*, ELECTRIC UTIL. WKLY., Oct. 30, 2006, at 1 ("The Public Utility Holding Company Act of 1935—repealed last year—presented less of an obstacle to energy giants considering mergers than the uncertainties that companies face when they take their merger plans to federal and state regulators." (internal quotation marks omitted) (quoting Daniele Seitz, a utility analyst at Dahlman Rose & Company)).

[98] Rebecca Smith, *Exelon Abandons PSEG Acquisition, Faults New Jersey*, WALL ST. J., Sept. 15, 2006, at A3. The deal started at a price of $12 billion, but Exelon's escalating stock price raised the deal value by more than $5 billion. *See id.*

[99] *Id.* (internal quotation marks omitted).

[100] *See id.*

[101] *See id.*

[102] *See* Reuters, *Utilities Call Off A Merger Deal*, N.Y. TIMES, Oct. 26, 2006, at C15.

[103] Press Release, Constellation Energy Group, Inc. & FPL Group, Inc., FPL Group and Constellation Energy Terminate Plans to Merge (Oct. 25, 2006), *quoted in* Carr & Pospisil, *supra* note 97, at 1.

[104] *See* Carr & Pospisil, *supra* note 97, at 1 ("[Large utilities] need to do a lot more legwork on the front end [and convince state regulators] why these deals make sense and how they will benefit customers in the long term." (internal quotation marks omitted) (quoting David Ratcliffe, Chairman, President, and CEO, Southern Company)); *id.* ("One would expect

may have initially created excitement and incentives for additional mergers and acquisitions, yet the resulting regulatory redundancies have limited the likelihood such transactions would be completed.

Finally, even if PUHCA's repeal is actually successful in bringing new investors into the arena, such a "success" could, in fact, slow the actual development of new infrastructure that already was planned by a pre-PUHCA entity. Investors not familiar with the energy arena and its political and regulatory landscape may not be as effective in moving projects forward as current industry participants.[105] In addition, state and federal legislators may be skeptical of new market participants, which could delay their review and approval of much needed new construction. Over time, the increased numbers of capital sources may very well translate into infrastructure enhancements. But any such benefits are years from leading to even applications for initial construction, let alone putting new facilities into service.[106]

B. Incentive Pricing and Market Forces: A Waiting Game

> Transmission congestion has been rising steadily since 1998. Transmission underinvestment is a national problem. We need a national solution. Transmission pricing reform can be an important part of the solution.[107]

FERC has long used incentive-based pricing policies, including the use of market-based rates, in wholesale energy markets (i.e., those regulated by FERC) in order to increase efficiencies in the energy industry. FERC started harnessing market forces in the 1980s as a means to reduce wholesale power prices, permitting certain public utilities to move from traditional cost-based rates to market-based rates for wholesale power sales.[108] The goal of this policy shift was "to create competitive pressures

---

that with the repeal of PUHCA that merger activity would increase," and that in his view "with the deals falling apart, larger companies will think twice before getting into large mergers" (internal quotation marks omitted) (quoting Johnson Kachidza, Managing Principal, Knox Lawrence International)).

[105] *See* Robinson & Terzic, *supra* note 71, at 15 (stating that closing mergers and acquisitions under the new regulatory scheme "will depend on multiple factors, among which will be the ability to understand and successfully navigate the new regulatory environment created by PUHCA's successors at the state and federal levels").

[106] *See* Fellmeth, *supra* note 5, at 829 (noting that ideally supply increases as demand requires, but that the market problems are triggered "by a substantial time or cost component required where supply must increase to meet demand"); *see also* Alejandro Bodipo-Memba, *Report Expected to Push for More Power Plants*, DETROIT FREE PRESS, Jan. 3, 2006, at 3B ("It would take about $2 billion and nearly seven years to build a 1,000-megawatt coal-fired power plant, according to most experts."); George, *State Utilities Locked into Building Spree*, GAINESVILLE SUN, Apr. 2, 1983, at B4 (stating that "electrical power plants require approximately 5-6 years to build").

[107] Kelliher, *supra* note 12.

[108] *See* Market-Based Rates For Public Utilities, 107 FERC ¶ 61,019, at p. 61,088

that would improve efficiency and lower wholesale power prices."[109] The Commission's policy was based on the concept that a competitive market leads to a reasonable exchange and that prices will lead a seller to obtain a "normal return on its investment."[110] This policy assumes that no participant in a Commission-approved transaction will have excessive market power and that there is sufficient competition in the market.[111]

The Commission moved to market-based rates because it believed that traditional cost-of-service "rate regulation does not encourage the regulated utility to be efficient and provide service at a low cost."[112] Current FERC chairman Joseph T. Kelliher has stated that the Commission's market-based rate "policy was never intended to deregulate wholesale power markets."[113] FERC merely shifted its focus, according to Chairman Kelliher, and FERC now regulates energy markets instead of only regulating energy prices.[114] In limited circumstances, FERC's market rules have been used in establishing price caps to prevent FERC-jurisdictional power sellers (wholesale energy sellers) from exacting monopoly rents,[115] but such rules are clearly not the norm.[116]

Courts have upheld market-based rates as adequate to assure just and reasonable rates for both electricity[117] and natural gas.[118] A recent case in

---

(2004) ("Much has changed in the industry since the Commission began using the four-prong test in the 1980s, and we believe it is important not only to ensure that our test is sufficient to support market-based rates in today's energy markets, but also to provide clarity, by way of codified regulations, as to what applicants must demonstrate in order to obtain (and retain) authority to sell at market-based rates."); AEP Power Mktg., Inc., 97 FERC ¶ 61,219, at p. 61,969 (2001) ("Since beginning to grant market-based rates to public utilities in the 1980s, the Commission primarily focused on the applicant and employed the "hub-and-spoke" analysis to determine whether an individual entity and its affiliates have the ability to exercise generation market power.").

[109] Joseph T. Kelliher, *Market Manipulation, Market Power, and the Authority of the Federal Energy Regulatory Commission*, 26 ENERGY L.J. 1, 9 (2005).

[110] Tejas Power Corp. v. FERC, 908 F.2d 998, 1004 (D.C. Cir. 1990) (finding that FERC improperly approved a settlement agreement that was designed to "make a market" for natural gas sales via a gas inventory charge).

[111] *See id.*

[112] Kelliher, *supra* note 109, at 10; *see also* 58 FERC ¶ 61,234, at p. 61,753 (1992) ("Traditional cost-of-service rate regulation is not always adequate to meet these needs and, at times, competitive markets can provide more efficient, lower-cost capacity for the long term as well as lower-cost energy in the short term.").

[113] Kelliher, *supra* note 109, at 11.

[114] *See id.*

[115] *See* San Diego Gas & Elec. Co., 95 FERC ¶ 61,418, at p. 62,545 (2001) (establishing price mitigation—price caps—for sales in energy "spot markets throughout the West"), *reh'g granted in part* 97 FERC ¶ 61,275 (2001), *reh'g denied* 99 FERC ¶ 61,160 (2002).

[116] *See* New England Power Pool, 90 FERC ¶ 61,168, at p. 61,541 (2000) (Herbert, C.J., dissenting) (stating that the Commission has indicated "its extreme distaste for temporary price cap band-aids").

[117] *See, e.g.*, Louisiana Energy & Power Auth. v. FERC, 141 F.3d 364, 365 (D.C. Cir. 1998) (stating that FERC, in a competitive market, "may rely on market-based rates in lieu of cost-of-service regulation to ensure that rates" are just and reasonable).

[118] *See* Elizabethtown Gas Co. v. FERC, 10 F.3d 866, 870–71 (D.C. Cir. 1993).

the Ninth Circuit similarly upheld this market-based rate authority;[119] however, in upholding the market-based rate policy as just and reasonable, the court also stated that FERC cannot use market-based rates alone in carrying out its duties under the FPA.[120] That is, a mere finding that a seller lacks market power is not sufficient oversight to ensure just and reasonable rates. Additional oversight, such as FERC's reporting requirement,[121] is necessary.[122]

FERC continues to view pricing policies as a key way in which to improve the U.S. energy infrastructure. FERC's first goal in its Fiscal Year 2007 Congressional Performance Budget Request is to "Promote Development of a Robust Energy Infrastructure."[123] Admittedly, part of the reason FERC has focused on pricing programs to reach the goal of developing energy infrastructure is because, beyond hydroelectric power, FERC does not have direct jurisdiction over the development of electric generation capacity and natural gas reserves.[124] As discussed above, FERC's jurisdiction extends to the wholesale markets in which such products operate.

Beyond granting market-based rate authority, FERC's ability to impact these areas was increased in EPAct 2005, which expanded the Commission's role in electric transmission siting[125] and mandated a rulemaking proceeding to establish "incentive-based rates for the transmission of electric energy in interstate commerce."[126] EPAct 2005 included a section entitled Transmission Infrastructure Investment.[127] FERC accordingly amended its

---

[119] *See* California v. FERC, 383 F.3d 1006 (9th Cir. 2004).

[120] *See id.* at 1013.

[121] *See id.* ("FERC's system consists of a finding that the applicant lacks market power (or has taken sufficient steps to mitigate market power), coupled with strict reporting requirements to ensure that the rate is 'just and reasonable' and that markets are not subject to manipulation.").

[122] In fact, the Ninth Circuit recently took FERC to task for failing to exercise adequate oversight of several market-based rate contracts necessitated by the 2000–2001 California Energy Crisis when it applied the doctrine of United Gas Pipe Line Co. v. Mobile Gas Serv. Corp., 350 U.S. 332 (1956) and Fed. Power Comm'n v. Sierra Pac. Power Co., 350 U.S. 348 (1956), which permit FERC to order modifications to certain power contracts when they "affect the public interest." Pub. Util. Dist. No. 1 v. FERC, 471 F.3d 1053, 1057, 1080 (9th Cir. 2006) ("We hold that although market-based rate authority *can* qualify as sufficient prior review to justify limited *Mobile-Sierra* review, it can only do so when accompanied by effective oversight permitting timely reconsideration of market-based authorization if market conditions change." (emphasis in original)); *see also* Cal. Pub. Util. Comm'n v. FERC, 474 F.3d 587, 595 (9th Cir. 2006) ("FERC 'cannot use [its] choice [of regulatory regime] to excuse its duty to maintain effective oversight [of rates] and then invoke *Mobile-Sierra* as a ground for precluding ordinary rate review, including review of the propriety of market-based rate authority at the time the contracts became effective.'") (modifications in original) (quoting Pub. Util. Dist. No. 1, 471 F.3d at 1085).

[123] FERC, FISCAL YEAR 2007 CONGRESSIONAL PERFORMANCE BUDGET REQUEST 11 (2006) [hereinafter FERC 2007 BUDGET], *available at* http://www.ferc.gov/about/strat-docs/FY07-budg.pdf.

[124] *See id.* at 12.

[125] *See* EPAct 2005 § 1221, 16 U.S.C.S § 824p (LexisNexis 2006).

[126] FERC 2007 BUDGET, *supra* note 123, at 13.

[127] EPAct 2005 § 1241, 16 U.S.C.S. § 824s (LexisNexis 2006) (adding a new section 219 to the Federal Power Act).

regulations to establish "incentive-based (including performance-based) rate treatments" for electric energy transmission "for the purpose of benefiting consumers by ensuring reliability and reducing the cost of delivered power by reducing transmission congestion."[128] FERC's proposed rules provide for a variety of incentives for transmission investment. These include:

> (i) a rate of return on equity sufficient to attract new investment in transmission facilities;
> (ii) 100 percent of prudently incurred Construction Work in Progress (CWIP) in rate base;
> (iii) recovery of prudently incurred pre-commercial operations costs;
> (iv) hypothetical capital structure;
> (v) accelerated regulatory book depreciation;
> (vi) recovery of 100 percent of prudently incurred costs of transmission facilities that are cancelled or abandoned due to factors beyond the control of the public utility;
> (vii) deferred cost recovery; and
> (viii) any other incentives approved by the Commission, pursuant to the requirements of this paragraph, that are determined to be just and reasonable and not unduly discriminatory or preferential.[129]

However, FERC's rules are another example of initiatives with long-term potential but little short-term value. As discussed in more detail in Part IV, without clear authority to ensure a place to build the necessary infrastructure, these incentives are, at best, only half-measures.

These rules do demonstrate that FERC has recognized the infrastructure problem.[130] There remains, however, what seems to be a misplaced faith in the abilities of the current market—one that implies that the market is providing correct infrastructure investment cues right now. For example, FERC Commissioner Nora Brownell recently stated:

> I think it is important to recognize that scarcity pricing is the market response to a supply/demand imbalance that appropriately signals the need for infrastructure. For example, the high prices of 2000-2001 that reflected supply/demand fundamentals resulted

---

[128] Promoting Transmission Investment Through Pricing Reform, 70 Fed. Reg. 71,409, 71,410 (Nov. 29, 2005) (codified at 18 C.F.R. pt. 35 (2007)).

[129] *Id.* at 71,419–20 (codified at 18 C.F.R. § 35.35(d)(1) (2007)).

[130] As FERC Chairman Kelliher has plainly stated: "[u]nderinvestment in transmission is a national problem." Transmission Pricing Release, *supra* note 9, at 1. In FERC's fiscal year 2007 budget for "Energy Infrastructure Resources," FERC proposed to add eighteen full-time employees and increase its budget by more than $13 million over their fiscal year 2005 actual numbers. *See* FERC 2007 BUDGET, *supra* note 123, at 11.

> in the first new power plants being constructed in California in ten years; price risk being hedged through the use of long-term contracting; and renewed efforts to correct a flawed market design.[131]

Such an assessment is accurate if massive blackouts and power shortages are merely considered "appropriate signals" for initiating long-term construction projects,[132] but it is becoming more apparent that this type of reactive policy comes with significant costs.[133]

C. Mergers and Acquisitions: New Rules, New Process, New Investment?

The recent enactment of EPAct 2005 and the subsequent regulatory actions related to approval of mergers and acquisitions under the FPA have the potential to trigger improvements to the U.S. energy infrastructure, but not in the near term. Section 1289 of EPAct 2005, "Merger Review Reform," amended section 203 of the FPA to restrict certain elements of FERC authority while expanding others.[134] The amendment restricted FERC authority by raising the monetary threshold for FERC review of several types of transactions from $50,000 to $10 million.[135] Similarly, it limited FERC's

---

[131] Amendments to Blanket Sales Certificates, 68 Fed. Reg. 66,323, 66,338 (Nov. 26, 2003) (Brownell, Comm'r, concurring). *But see* Fellmeth, *supra* note 5, at 829 ("The most obvious flaw in deregulation arises from scarcity defects in the underlying market newly relied upon.").

[132] Advocates of truly free markets, of course, find such cues perfectly appropriate. Unfortunately, externalities, including social costs, are often improperly accounted for by free market actors. *See, e.g.*, Amy Lynne Bomse, Note, *The Dependence of Cyberspace*, 50 DUKE L.J. 1717, 1736 n.107 (2001) ("Market theory defines an externality as anything that causes a market to fail to reach pareto optimalism."); *see also infra* note 154 (discussing Pareto optimality). The presence of such externalities often provides the justification for government regulation. *See, e.g.*, Allan Kanner, *Toxic Tort Litigation in a Regulatory World*, 41 WASHBURN L.J. 535, 545 (2002) ("One explanation for public law and the need for regulatory agencies is that they address the failure of market exchange mechanisms. The classic example is the pollution externality, a social cost that common law and the free market arguably fail to force a firm to internalize."); *see also* Cheryl D. Block, *Overt and Covert Bailouts: Developing a Public Bailout Policy*, 67 IND. L.J. 951, 991–92 (1992) ("Markets may fail for numerous reasons, including inadequate flow of information, nonexistence of a market for certain goods, concentration of power in the form of monopolies, high transaction costs for certain exchanges, and spillover or externality effects of individual behavior that the market does not take into account.").

[133] There are legitimate, if not wholly satisfying, arguments that a free market could remedy the current inefficiencies in the energy industry. However, a truly free market, in which market participants would operate as they desired, runs contrary to the long-held "central charge of the Commission." *See* FERC 2007 BUDGET, *supra* note 123, at 3 ("Of the Commission's primary task there is no doubt, however, and that is to guard the consumer from exploitation by non-competitive electric power companies." (quoting NAACP v. FPC, 520 F.2d 432, 438 (D.C. Cir. 1975))). This central charge presumably encompasses acting to prevent electric power companies from profiting off avoidable market failures.

[134] *Compare* EPAct 2005 § 1289(a), 16 U.S.C.S. § 824b(a) (LexisNexis 2006), *with* 16 U.S.C. § 824b (2000).

[135] *Compare* EPAct 2005 § 1289(a)(1)–(2), 16 U.S.C.S. § 824b(a)(1)–(2) (LexisNexis 2006), *with* 16 U.S.C. § 824b(a) (2000).

review of public utility acquisitions of the securities of other public utilities to transactions valued at more than $10 million, whereas there was no such monetary threshhold before.[136] On the other hand, the amendment expanded FERC authority to include some previously exempt transactions involving the transfer of generation facilities and certain other holding company transactions that have a value in excess of $10 million.[137] Amended section 203 further requires FERC to examine cross-subsidization and pledges and encumbrances of utility assets when considering a transaction subject to section 203 review.[138] Finally, FERC was ordered to adopt procedures to expedite review of applications for section 203 approval of dispositions, consolidations, and acquisitions.[139] Even if these changes will make merger review more efficient in the long term, by simultaneously expanding and contracting FERC authority in these ways Congress has also created a new set of rules for energy companies to decipher before entering potential mergers.

Following a notice of proposed rulemaking,[140] FERC issued an order, Order No. 669, adopting a final rule on the new mergers and acquisitions authority granted by EPAct 2005.[141] In addition to implementing the rules related to FERC's authority under section 203 described above, the final rule granted "blanket authorizations for certain types of transactions, including foreign utility acquisitions by holding companies, intra-holding company system financing and cash management arrangements, certain internal corporate reorganizations, and certain investments in transmitting utilities and electric utility companies."[142] The rule also provides for the "expeditious consideration of completed applications for the approval of transactions that are not contested, do not involve mergers, and are consistent with Commission precedent."[143]

FERC stated that the goal of Order No. 669 was to ensure that all transactions subject to FPA section 203 were "consistent with the public interest and at the same time ensure that our rules do not impede day-to-day

---

[136] *Compare* EPAct 2005 § 1289(a)(1)(C), (a)(2), 16 U.S.C.S. § 824b(a)(1)(C), (a)(2) (LexisNexis 2006), *with* 16 U.S.C. § 824b(a) (2000).

[137] *Compare* EPAct 2005 § 1289(a)(1)(D), 16 U.S.C.S. § 824b(a)(1)(D) (LexisNexis 2006), *with* 16 U.S.C. § 824b(a) (2000).

[138] *Compare* EPAct 2005 § 1289(a)(4), 16 U.S.C.S. § 824b(a)(4) (LexisNexis 2006), *with* 16 U.S.C. § 824b(a) (2000).

[139] *Compare* EPAct 2005 § 1289(a)(5), 16 U.S.C.S. § 824b(a)(5) (LexisNexis 2006), *with* 16 U.S.C. § 824b(a) (2000).

[140] Transactions Subject to FPA Section 203, 70 Fed. Reg. 58,636 (Oct. 7, 2005).

[141] Transactions Subject to FPA Section 203, Order No. 669, 71 Fed. Reg. 1348 (Jan. 6, 2006) (codified at 18 C.F.R. pts. 2, 33 (2007)), *order on reh'g*, Order No. 669-A, 71 Fed. Reg. 28,422 (May 16, 2006), *order on reh'g*, Order No. 669-B, 71 Fed. Reg. 42,579 (July 27, 2006).

[142] Transactions Subject to FPA Section 203, Order No. 669, 71 Fed. Reg. at 1349; *see* 18 C.F.R. § 33.1 (2006). This blanket authority permits a company to enter such transactions without seeking separate approvals for each transaction. *See id.*

[143] Transactions Subject to FPA Section 203, Order No. 669, 71 Fed. Reg. at 1349; *see* 18 C.F.R. § 33.11 (2006).

business transactions or *stifle timely investment in transmission and generation infrastructure*."[144] FERC noted that it believed it had accomplished that result but that it would be addressing additional issues, such as the "appropriateness of blanket authorizations" and whether "additional steps are needed to protect against cross-subsidization and pledges or encumbrance of utility assets," in a technical conference announced in the PUHCA 2005 Final Rule.[145]

The U.S. energy industry is likely to face a new wave of mergers and acquisitions activity following the recent legislative and regulatory changes.[146] Of course, even before the repeal of PUHCA, the energy industry was experiencing significant merger activity, which started in the mid-1990s.[147] This trend toward consolidation has not led to adequate, if any, infrastructure improvements. Given that this trend has been in place for nearly ten years, there is little reason to believe that making the mergers and acquisitions process easier or open to more potential investors would have any direct impact on the state of the U.S. energy infrastructure, particularly in the near term.[148]

III. MARKET-BASED HOPE AND LIMITED SCOPE: LONG-TERM MARKET EVOLUTION AND EMERGENCY ORDERS ARE NOT ADEQUATE SOLUTIONS

Critics of U.S. energy policies often focus on decisions to implement short-term gap-filling policies instead of developing comprehensive long-term programs.[149] However, federal energy policies have actually "evolved"

---

[144] Transactions Subject to FPA Section 203, Order No. 669, 71 Fed. Reg. at 1349 (emphasis added); *see also* 18 C.F.R. § 33.2(g) (2006).

[145] Transactions Subject to FPA Section 203, Order No. 669, 71 Fed. Reg. at 1349; *see also* 18 C.F.R. 33.1(c)(5) (2006) (granting blanket authorizations of foreign utility company acquisitions subject to certain conditions to protect U.S. captive customers). The technical conference was announced in PUHCA 2005 Final Rule, *supra* note 75, at P 6.

[146] *See* Robinson & Terzic, *supra* note 71, at 14 (stating that, following the repeal of PUHCA, "the prognosis for utility M&A [is] strong over the next 12 to 24 months, driven by the increasing need to address the capital market's earnings growth and investment performance expectations for the sector").

[147] *See* Edison Elec. Inst., *One Plus One Doesn't Always Equal Two*, ELECTRIC PERSP., Jan.-Feb. 2002, *available at* http://www.eei.org/magazine/editorial_content/nonav_stories/2002-01-01-NT.htm; *see also* Robinson & Terzic, *supra* note 71, at 14 ("PUHCA repeal will not be the sole trigger for returning to M&A, with utility executives, investment bankers and regulators falling into many camps when describing the value creation potential of further consolidation.").

[148] As discussed above in Part II, it is not at all clear that EPAct 2005 actually made mergers easier to complete or more likely to occur. *See supra* notes 75–79 and accompanying text.

[149] *See, e.g.*, Amory B. Lovins, *Energy Strategy: The Road Not Taken?*, 55 FOREIGN AFF. 65, 65–66 (1976) (arguing that traditional U.S. energy policy has implemented "incremental past practices" instead of pursuing "long-term goals"); *see also* Jeffrey Rudd, *Restructuring America's Government to Create Sustainable Development*, 30 WM. & MARY ENVTL. L. & POL'Y REV. 371, 377 (2006) (stating that, in the 1970s, "Amory Lovins' prophetic emphasis on long-term energy policies fell on political ears deafened by industries' control over public policy").

to the point of including equally inadequate long-term and short-term gap-filling policies. In fact, the majority of recent legislative and regulatory actions are intended to provide long-term solutions by increasing energy infrastructure through various types of market evolution. Unfortunately, such indirect, long-term "solutions" have not worked over the past thirty years, and there is little to indicate that the new proposals will, on their own, fare any better.

Market-based rates, for one, without additional support, will not provide the necessary incentives to trigger infrastructure improvements in an efficient and acceptable manner. Barring almost complete deregulation, there is little indication that market-based rates can be successful at all. Social, political, and environmental costs, as well as safety and reliability concerns, make full deregulation imprudent and impractical.[150] Additionally, although an open market might lead to better rates for those in high-demand areas, rural and other isolated locations could suffer without mandatory service obligations.[151] Given the limited impact of recent market-based efforts, and the risks associated with any broader implementation, there is little reason to expect significant impact from such policies in the near future.

What is obvious is that the current energy infrastructure is insufficient for the current and ever-growing U.S. demand.[152] Recent FERC emergency

---

[150] While "free-market rhetoric" is strong among business leaders and corporate law scholars, very few executives, especially in the energy industry, would seek full deregulation, except perhaps when it comes to corporate governance. *See* Kent Greenfield, *September 11th and the End of History for Corporate Law*, 76 TUL. L. REV. 1409, 1420–21 (2002) ("[T]he very infrastructure of the market . . . is in large part a creation of government and government regulation.").

[151] This issue has long been recognized in the telecommunications industry. *See* Clinton Howard Brannon, *Reach Out and Tax Someone: What Does the Future Hold for the Taxation and Regulation of Voice Over Internet Protocol Telephone Services?*, 57 ALA. L. REV. 173, 180 (2005) (describing the purpose and effect of 47 U.S.C. § 254 (2000)). Regulations in the telecommunications industry include a Universal Service Fund ("USF"), which collects a fee from all telecommunications providers in the United States and then pools the fees into a "fund maintained by the federal government as a way to subsidize the high costs of providing telecommunications services to rural areas." *Id.* (footnote omitted). "The fund is later disbursed to telecommunications companies that provide service in rural areas to compensate them for the higher costs of providing access lines to their customers." *Id.*

[152] *See supra* notes 3–15 and accompanying text. It would seem sensible to consider reducing demand as another method to alleviate strains on the energy infrastructure. However, conservation is generally not a top priority of politicians. *See, e.g.*, Oliver Houck, *Can We Save New Orleans?*, 19 TUL. ENVTL. L.J. 1, 29–30 (2006) ("In more than 30 years, I do not believe I have heard a Louisiana politician say the words 'energy conservation.' By some gap in the neurons, the fact that reversing climate change will save coastal communities and the oil and gas infrastructure in Louisiana doesn't seem to reach the head."); Gary C. Bryner, *The National Energy Policy: Assessing Energy Policy Choices*, 73 U. COLO. L. REV. 341, 346 (2002) (stating that the Bush Administration's National Energy Plan "clearly emphasizes and gives priority to expanding the supply of traditional energy sources by opening new lands for exploration, streamlining the permitting process, easing regulatory requirements, and enlarging the nation's energy infrastructure") (citing NAT'L ENERGY POL'Y DEV. GROUP, NATIONAL ENERGY POLICY: RELIABLE, AFFORDABLE, AND ENVIRONMENTALLY SOUND ENERGY FOR AMERICA'S FUTURE, at viii (2001)).

orders tried incentives tied to specific deadlines to put new facilities in service.[153] These FERC actions, in response to energy emergencies caused by the 2000–2001 California energy crisis and Hurricanes Rita and Katrina, would seem to warrant consideration as a possible larger-scale solution. The EPAct 2005 legislation enacted some provisions similar to these emergency actions, and should provide some promise as well. But more is needed to effectively and immediately improve the nation's energy infrastructure, because even investment incentives tied to deadlines have not proven especially effective.

A. Market-Based Rates Come Up Short

Market-based rates are wholly inadequate as a short-term solution to the infrastructure problem. Because infrastructure construction itself is so time consuming, consumers necessarily suffer during the lag between market signals of infrastructure problems and the completion of infrastructure improvements prompted by such signals.

More fundamentally, it is unclear what the market is expected to provide in the first place. That is, most politicians and many court cases seem to imply that Pareto-optimal improvements[154] will result from allowing market forces to work. Certainly, this kind of economic outcome is what most consumers would expect from a "market system."[155] However, it seems more likely that any benefits would, at best, represent a Kaldor-Hicks improvement,[156] providing lower costs to industrial and other large users while raising individual rates for many consumers. To the extent consumers understand or believe that this is what is occurring, the result would be politically untenable.

---

[153] *See infra* notes 168 and 181 and accompanying text.

[154] "Pareto optimality" refers the point at which resources are distributed in a manner in which no change can be made making someone better off without making someone else worse off. *See* RICHARD A. POSNER, ECONOMIC ANALYSIS OF LAW 13 (4th ed. 1992). A Pareto improvement is possible in an inefficient market; that is, a market that would allow a change in which at least one person can be made better off without affecting anyone else. *See* Thad Kousser & Mathew D. McCubbins, *Social Choice, Crypto-Initiatives, and Policy Making by Direct Democracy*, 78 S. CAL. L. REV. 949, 962–66, 982 (2005); Russell Hardin, *Magic on the Frontier: The Norm of Efficiency*, 144 U. PA. L. REV. 1987, 1996–97 (1996).

[155] The concept of Pareto optimality, in this sense, "has enormous intuitive appeal." *See* Guido Calabresi, *The Pointlessness of Pareto: Carrying Coase Further*, 100 YALE L.J. 1211, 1216 (1991) (discussing the problems of seeking Pareto optimality and noting that "if Pareto optimality means a place where no improvement can be made without ex ante creating the possibility that there will be some losers, then we *are always* there") (emphasis added).

[156] "Kaldor-Hicks efficiency" is described by economists as the point at which social net benefits are maximized without regard to the particular distribution of benefits. *See* POSNER, *supra* note 154, at 13–14. A "Kaldor-Hicks improvement" has been described as "a change that increases social net benefits but does not necessarily make everyone better off." Spence, *supra* note 60, at 418 n.3.

Such problems are inherent in a market system. The market may eventually provide an infrastructure that will provide sufficient transmission capacity. Most likely, the market would actually provide excess capacity in certain parts of the United States at some point because investors will be clamoring to get in on the action once energy prices rise high enough. Once the capacity exceeds demand, prices will drop. It is this principle that leads many to argue for deregulation as a trigger for investment.[157]

Commentators have discussed the "market effect" deregulation had in the telecommunications industry in great detail.[158] In the wake of the recent failures of many telecommunications companies, some commentators have argued that there was not really any deregulation of the industry at all and that the present regulatory scheme negatively impacts the telecommunications market.[159] Regardless, it seems clear that the telecommunications market participants had some degree of freedom to build their systems, which they did at their own peril. Whether the trigger was deregulation or simply the changes in the subject markets (e.g., the Internet boom),[160] when demand for telecommunications took off and prices soared, companies like Tyco, WorldCom, and Global Crossing put billions of dollars into communications infrastructure around the world.[161] As the prices of the telecommunications services plummeted, along with corresponding stock valuations, many companies in the sector went bankrupt,[162] the infrastructure

---

[157] Note that with regard to energy infrastructure, Chairman Kelliher insists that deregulation is not what is occurring; rather, FERC is shifting toward having regulations that rely on the market. *See* Kelliher, *supra* note 109, at 9.

[158] *See, e.g.*, Larry E. Ribstein, *Bubble Laws*, 40 HOUS. L. REV. 77, 83 (2003) ("The recent boom and bubble probably began in the mid-1990s, as people started seeing limitless potential in the Internet and in the deregulation of telecommunications and other markets.").

[159] *See, e.g.*, Jerry Ellig, *Costs and Consequences of Federal Telecommunications Regulations*, 58 FED. COMM. L.J. 37, 44 (2006) ("[A] regulatory system that imposes through administrative mandate a set of prices that tries to mimic those that competition would have set does not thereby become any the less a regulatory process, nor any the more a competitive one."); Deborah Ellenberg et al., *Antitrust: Will It Change the Lives of Telecommunications Executives?*, 4 RICH. J.L. & TECH. 3 (1997), http:/www.richmond.edu/jolt/v4si/speech3.html ("Sad to say, the 1996 Telecommunications Act did not significantly reduce regulation despite Congress' professed ambition to do so. In the wake of the 1996 Act, there was an outpouring of new regulations from the FCC unequaled in the history of telecommunications regulation.") (quoting Glen Robinson, Associate Dean and Professor of Law, University of Virginia).

[160] *See* Ribstein, *supra* note 158, at 83. High prices, and the possibility of cashing in on such prices, may have been sufficient to move the market to some degree; however, deregulation helped, at least somewhat, to move the process along. *See* Numark & MacDougall, *supra* note 21, at 464 (stating that the deregulation of the airline, railroad, trucking, natural gas, and telecommunications industries brought "lower prices, expanded markets, and a smaller number of bigger, more competitive and more efficient producers and suppliers").

[161] BURTON G. MALKIEL, A RANDOM WALK DOWN WALL STREET 95 (8th ed. 2003) (stating that telecommunications companies laid enough fiber-optic cable "to circle the earth 1,500 times").

[162] *See* J. Gregory Sidak, *The Failure of Good Intentions: The WorldCom Fraud and the Collapse of American Telecommunications After Deregulation*, 20 YALE J. ON REG. 207, 216 (2003) ("Global Crossing's worldwide fiber optic network, which consumed $15

was sold off, and consumers finally reaped the benefits with lower prices.[163] Of course, these benefits were of little solace to stockholders and creditors of their dramatically devalued holdings.[164]

This demonstrates how, over the long term, markets might be able to provide adequate infrastructure and proper pricing (i.e., producers being able to set prices that recover their total cost over a cycle) but can impose undesirable, and perhaps unacceptable, societal costs over short-term cycles.[165] This is perhaps particularly true of energy markets. For instance, people often require a minimum amount of energy for survival, and extremely high costs over significant periods of time can have disastrous effects for consumers.[166] A low average cost over time does not assist paycheck-to-paycheck consumers needing heat or air conditioning to survive during times of crisis. Concerning overinvestment, there are potential environmental concerns as well: building unnecessary infrastructure can cause significant harms to wetlands and increased emissions without related net price or efficiency gains.[167]

B. Limited Small-Scale Success of Emergency "Remedies"

In March 2001, FERC issued an order announcing actions

> within its regulatory authorities under the Federal Power Act, the Natural Gas Act, the Natural Gas Policy Act, the Public Utility Regulatory Policies Act, and the Interstate Commerce Act to

---

billion in financing to construct in the late 1990s, was implicitly valued in March 2003 at only $406.5 million.").

[163] *See* Robert E. Litan, *The Telecommunications Crash: What To Do Now?*, at 1–3 (Brookings Inst. Policy Brief #112, 2002), *available at* http://www.brookings.edu/comm/policybriefs/pb112.pdf.

[164] The telecommunications industry is not directly analogous to the energy industry, in the sense that there is significantly more regulation in the energy sector. *See id.* at 1 (discussing the effects of deregulation under the Telecommunications Act of 1996). However, the economic result in the telecommunications industry underscores the potential problems of allowing the market alone to dictate infrastructure developments.

[165] *See* Michael E. Levine, *Price Discrimination Without Market Power*, 19 YALE J. ON REG. 1, 6–7, 12 (2002) (explaining that the recovery of total costs is normally achieved through prices that never fall below variable cost and that recover fixed costs through additional charges that vary with supply and demand).

[166] *See* Press Release, Sen. Susan Collins (R-Me.), Senator Collins' Statement on Release of Emergency LIHEAP Funds (Mar. 23, 2006), *available at* 2006 WLNR 4904591 ("Tragically, one Maine family was already lost earlier this year when after running out of heating oil, they sought to heat their home with a wood stove that led to the house catching fire. For low-income families, [Low Income Home Energy Assistance Program ("LIHEAP")] funds can literally be a matter of life and death.").

[167] *See* Certification of New Interstate Natural Gas Pipeline Facilities: Statement of Policy, 88 FERC ¶ 61,227, at p. 61,737 (1999) ("In considering the impact of new construction projects on existing pipelines, the Commission's goal is to appropriately consider the enhancement of competitive transportation alternatives, the possibility of overbuilding, the avoidance of unnecessary disruption of the environment, and the unneeded exercise of eminent domain.").

> help increase electric generation supply and delivery in the Western United States, in order to protect consumers from supply disruptions.[168]

Despite the lofty goals, FERC recognized its own limitations: "The Commission recognizes that the actions announced here, by themselves, will not solve the electricity crisis facing California and other areas of the West and will not prevent electricity blackouts in the summer of 2001."[169] Nonetheless, FERC initiated a plan to help alleviate energy supply concerns in the short term while attempting to provide "medium and longer term solutions, including new infrastructure that [could] help avert future recurrences of the current electric supply shortage in the West."[170]

To boost electric supply, FERC planned to provide premium returns on equity and a favorable depreciable life for facilities that could be placed in service quickly.[171] This effectively raised the available return on equity from 11.5% to as high as 14.5%.[172] FERC also provided similar incentives for electric transmission system upgrades that required new rights of way (providing a return on equity of 12.5% and a 15-year depreciable life if in service by November 1, 2001) and for new "facilities needed to interconnect new supply to the grid" (providing a return on equity of 13.5% if in service by November 1, 2001 and 12.5% if in service by November 1, 2002).[173]

FERC subsequently approved a similar 200-basis point return-on-equity adder[174] for a Pacific Gas & Electric ("PG&E") expansion project ("Path 15 Project") that would not be completed until late 2004.[175] The Path 15 Project was designed to reduce congestion on an eighty-four-mile segment of high-voltage transmission lines that connects southern and

---

[168] Removing Obstacles to Increased Electric Generation and Natural Gas Supply in the Western United States, 94 FERC ¶ 61,272, at p. 61,967 (2001) [hereinafter Removing Obstacles Order] (footnote omitted).

[169] *Id.*

[170] *Id.*

[171] *See id.* at p. 61,969 (providing a 300 basis-point premium and a 10-year depreciable life for projects in service by July 1, 2001 and a 200 basis-point premium and a 10-year depreciable life for projects in service by Nov. 1, 2001).

[172] *See id.* Return on equity applies only to cost-based rates (and not market-based rates), where a fixed-percentage return (in addition to actual costs) is calculated as part of the rate. *See, e.g.*, PJM Interconnection, LLC, 112 FERC ¶ 61,031, at P 57 (2005) ("The issues of return on equity and depreciation are concerns only with setting cost-based rates.").

[173] Removing Obstacles Order, *supra* note 168, at p. 61,970.

[174] A 200-basis-point adder translates to a 2% increase in the return on equity. One basis point equals 0.01%. *See* Stingray Pipeline Company, 98 FERC ¶ 63,004, at P 31 n.58 (2002) (explaining that "100 basis points equal 1%"). Thus, for example, if there is a uniform baseline return on equity of 11.5%, a 200-basis-point adder would provide a 2% premium, increasing the return on equity to 13.5%. *See* Removing Obstacles Order, *supra* note 168, at p. 61,970.

[175] *See D.C. Circuit Signs Off on FERC's California Path 15 Financial Incentives*, FOSTER ELECTRIC REP., May 19, 2004, at 8.

northern California.[176] Despite protests by the California Public Utilities Commission ("CPUC") that FERC had "unlawfully extended the deadline" in the Removing Obstacles Order, the D.C. Circuit found that FERC had appropriately approved the incentive "on a case-by-case basis."[177]

FERC's approval of the Path 15 Project is an isolated example of aggressive and appropriate action that helped ensure needed infrastructure was built where it was most needed. Despite fervent (and expected) challenges,[178] FERC recognized that construction was needed to alleviate congestion and also knew "that unless it approved the PG&E incentives, the project would likely not be built in the near future."[179] This represented an uncharacteristically bold step through which FERC exercised its limited authority "to foster 'the installation of critical transmission investment,' by offering incentives to increase the supply of energy."[180] In doing so, FERC demonstrated the aggressive action that is appropriate in responding to a "crisis," albeit on far too small of a scale.

More recently, in response to Hurricanes Katrina and Rita, FERC issued an order on November 18, 2005,[181] temporarily waiving its regulations and raising the limitations on the costs for projects that natural gas pipelines may construct without prior specific authorization under their Part 157, Subpart F blanket certificates.[182] The Commission stated that it was acting to help more natural gas reach the market to mitigate the cost impact on consumers.[183] To expedite the construction of infrastructure that might provide access to additional natural gas supplies, the Commission increased the costs of projects that can be constructed under the automatic provisions of blanket certificates from $8 million to $16 million and under the "prior-notice" provisions from $22 million to $50 million, thus eliminating several regulatory hurdles to infrastructure construction for larger scale projects.[184]

Importantly, FERC provided that these temporary waivers would apply only to those projects constructed and placed in service by October 31, 2006.[185] Recognizing "that projects which cannot be completed in time to

---

[176] *See id.*

[177] Cal. Pub. Util. Comm'n v. FERC, 367 F.3d 925, 930 (D.C. Cir. 2004).

[178] *See* Western Area Power Administration, 99 FERC ¶ 61,306, at PP 2–4 (2002) (recounting the several protests and motions to intervene filed in response to notice of filing for the Path 15 Project).

[179] *Cal. Pub. Util. Comm'n*, 367 F.3d at 929.

[180] *Id.* at 930 (citation omitted).

[181] Expediting Infrastructure Construction To Speed Hurricane Recovery, 113 FERC ¶ 61,179, at P 1 (2005) [hereinafter Katrina Relief Order I].

[182] 18 C.F.R. §§ 157.201–157.218 (2005). A blanket certificate permits the recipient to pursue certain construction, acquisition, operation, replacement, and miscellaneous rearrangement of facilities projects, as provided in the regulations, without seeking and obtaining separate authorizations for each project. *See id.* § 157.203(a).

[183] Katrina Relief Order I, *supra* note 181, at P 1 ("The more natural gas that reaches the market, the less the price impact will be for users of that gas.").

[184] *See id.*

[185] *Id.*

provide service at the start of the heating season might still deliver the benefits associated with additional gas supply if they are placed into service before the end of the heating season," the Commission extended the waivers to include projects built and placed into service by February 28, 2007.[186]

The Commission also expanded the definition of "eligible facilities" to include mainline facilities for this purpose.[187] Specifically, the Commission temporarily waived several of its regulatory provisions to include the following as "eligible facilit[ies]:" main lines; extensions of main lines; facilities, including compression and looping, that alter the capacity of main lines; and temporary compression that raises the capacity of main lines.[188] The cost-limit waivers described above were also permitted to apply to newly eligible facilities, and these temporary waivers were also put in effect through October 31, 2006.[189] The waivers are still, however, capped at $50 million.[190] As such, any eligible facilities built under these provisions would still be significantly limited in size and scope.

As FERC perhaps recognized, the emergency orders have not been especially effective. At the time the Commission issued the deadline extension, there had not been a single filing of a prior-notice application. That is not to say that the program was not helpful at all. However, while some smaller-scale projects ($16 million or less) may have been undertaken, not a single large-scale project ($16 million to $50 million) related to the emergency order had been announced five months after the order became effective.[191] Of course, this is not especially shocking: one year is a short time frame to plan and complete construction projects with costs approaching $50 million.

Incentives of the type used in FERC's emergency orders should be effective in increasing energy infrastructure, but often even those with strict in-service dates can fail to incentivize the appropriate investment in a timely manner.[192] As the D.C. Circuit noted in *California Public Utilities Commission*: "Although it was well-known that Path 15 was constrained and although this suggested a ready market if new transmission lines were built, no party stepped forward to construct upgrades."[193] Participants for the expansion project were found only after specific requests for proposals

---

[186] Expediting Infrastructure Construction To Speed Hurricane Recovery, 114 FERC ¶ 61,186, at P 2 (2006) [hereinafter Katrina Relief Order II].

[187] Katrina Relief Order I, *supra* note 181, at P 1.

[188] *Id.* at P 7 (waiving 18 C.F.R. § 157.202(b)(2)(ii) (A)–(C), (F) (2005)).

[189] *See id.*

[190] *See id.* at P 1.

[191] Companies cannot split their projects into smaller scale projects to avoid the prior-notice provisions because FERC's regulations prohibit the segmenting of "projects in order to meet the cost limitations." 18 C.F.R. § 157.208(a) (2005).

[192] *See, e.g.*, Katrina Relief Order II, *supra* note 186, at P 2 (extending the initial in-service deadline to place new facilities constructed pursuant to emergency waivers just three months after issuing the initial deadline).

[193] Cal. Pub. Util. Comm'n v. FERC, 367 F.3d 925, 929 (2004).

were issued, "and then only [after] incentives were offered" and the deadline was extended.[194]

The limited success of the Path 15 Project and the Katrina relief orders indicate that more is needed to trigger additional infrastructure investment, even for smaller scale projects. In addition to aggressive pricing incentives, realistic in-service deadlines (i.e., deadlines tied to the size and scope of a proposed project) and targeted outreach to current and potential industry participants are needed to promote much-needed infrastructure investment.

C. Small Solutions Are Inadequate for a Large Energy Crisis

Energy industry professionals are aware and agree that infrastructure needs exist all over the United States similar to those in California and the areas affected by Hurricanes Katrina and Rita,[195] yet persuading utility companies to make the needed infrastructure investments remains difficult. Motivating utilities to make the necessary infrastructure investments is imperative if reliable energy is to remain available in the fastest-growing regions of the country.[196]

EPAct 2005 in fact includes several provisions that are similar to FERC's recent emergency orders. The provisions requiring DOE to identify and report NIETCs[197] are very similar to those used to combat the California issues in 2000 to 2001.[198] The NIETC provision identifies key areas in need of investment and provides a clear deadline for both federal and state action.[199] This provision provides for federal intervention (via FERC) to ensure that transmission lines are built if a state cannot or will not act to move forward a project in the areas targeted by DOE.[200] The purpose of this provision is, in part, to address the not-in-my-back-yard ("NIMBY") problem, which often stops much-needed infrastructure development.[201] Even with the changes, though, it has been recognized that NIMBY issues could mean that necessary energy projects will not be

---

[194] *Id.*

[195] *See* William McCall, *More Power Lines Needed—Soon*, DESERET MORNING NEWS, Dec. 4, 2005, at M9 ("There is one thing that everybody agrees has to be done about the thousands of miles of electricity transmission lines that crisscross the West—build more of them.").

[196] *See id.* (stating that unless new transmission lines are built in the West, "the risk of a blackout like the one that left the East Coast in the dark in 2003 keeps rising").

[197] EPAct 2005 § 1221, 16 U.S.C.S. § 824p (LexisNexis 2006); *see also supra* text accompanying notes 34–36 (describing FERC's authority to identify NIETCs under these statutory provisions).

[198] *See supra* notes 168–180 and accompanying text.

[199] *See* EPAct 2005 § 1221(a), 16 U.S.C.S. § 824p(a).

[200] *See* EPAct 2005 § 1221(a), 16 U.S.C.S. § 824p(b).

[201] *See Brownell: With New EPAct Authority, FERC Will Push Infrastructure Development, Watch Markets*, INSIDE FERC, Mar. 6, 2006, at 16 (reporting comments of FERC Commissioner Nora Brownell at a Ziff Energy Gas Storage Conference).

built.[202] As FERC Commissioner Nora Brownell recently stated: "Nobody wants anything in their backyard. I don't want anything in my backyard either, but I want to turn the lights on when I flip a switch."[203]

EPAct 2005 also purportedly provided FERC with "exclusive" jurisdiction over onshore LNG siting.[204] Construction of new LNG terminals unquestionably faces a NIMBY problem, and proposed new construction has been vociferously opposed.[205] Congress acted to provide this "exclusive" jurisdiction in response to a dispute in California over a proposed LNG terminal in Long Beach.[206] The CPUC had asserted that California, not FERC, had the power to approve or deny the siting of an LNG terminal.[207] The case was dropped following the passage of EPAct 2005.[208]

However, while it is technically accurate that LNG siting is now solely a federal issue, the legislative history is replete with contradictory statements concerning the scope of FERC's exclusivity. The House and Senate committee reports assert that LNG siting is now completely under FERC's authority,[209] while acknowledging that a significant state role remains.[210]

---

[202] *See id.*

[203] *Id.*

[204] *See* EPAct 2005 § 311(c)(2), 15 U.S.C.S. § 717b(e)(1) (LexisNexis 2006).

[205] *See, e.g., Procedural Complaint, Environmental Review Highlight Latest Flap over SES LNG Project*, INSIDE FERC, Dec. 12, 2005, at 9 ("Although the Energy Policy Act settled the jurisdictional fight over the proposed liquefied natural gas terminal in Long Beach, Calif., tensions between California and federal officials appear anything but resolved.").

[206] *See id.*

[207] John A. Sullivan, *Greens Vowing to Continue Fight Against Long Beach LNG Project,* NAT. GAS WEEK, Dec. 5, 2005, at 1 (stating that both FERC and the CPUC claimed that "they had the final say in siting LNG terminals").

[208] *See* Order Instituting Investigation into the Proposal of Sound Energy Solutions to Construct and Operate a Liquefied Natural Gas Terminal at the Port of Long Beach, D.05-11-010, 2005 Cal. PUC LEXIS 477, at *1 & n.1 (Nov. 18, 2005) (order closing proceeding) (stating that FERC and the CPUC filed consent motions to dismiss the petition for review as moot in light of EPAct 2005); *FERC and Long Beach Request Comments on Draft Environmental Studies for LNG Terminal; California PUC Withdraws Challenge, Seeks Either Dismissal of Application or Hearing at FERC*, FOSTER NAT. GAS REP., Oct. 13, 2005, at 13 ("[T]he California Public Utilities Commission (CPUC) recently dropped its appeal in the U.S. Court of Appeals for the Ninth Circuit of FERC's order asserting exclusive jurisdiction over siting and operation of the terminal and the outlet pipe.").

[209] *See* H.R. REP. NO. 109-215, at 235 (2005) (stating that section 320, "Liquefaction or Gasification Natural Gas Terminals," of the original House version of the Energy Policy Act of 2005, which became section 311 of the final bill as modified and passed by both the House and Senate, "makes clear that FERC has preemptive authority to site liquefaction or gasification natural gas terminals to the extent the terminal involves foreign or interstate commerce"); S. REP. NO. 109-78, at 29 (2005) (stating that section 381 of the Senate version of the Energy Policy Act of 2005, S.10, 109th Cong., which was the parallel provision to the final section 311 of H.R. 6, both titled "Exportation or Importation of Natural Gas," "clarifies FERC's exclusive jurisdiction under the Natural Gas Act for siting, construction, expansion and operation of import/export facilities located onshore or in State waters").

[210] *See* H.R. REP. NO. 109-215, at 235 (2005) ("Section 320 also requires FERC to consult with the State commission of the state in which the liquefaction or gasification natural gas terminal is located regarding local safety considerations during the authoriza-

Statements by members of Congress indicate an even more nuanced sharing of authority with the states. For example, several representatives from California warned Governor Arnold Schwarzenegger: "The bill will hand over exclusive jurisdiction for the siting of [LNG] facilities to [FERC], preventing the states from having a role in approving the location of LNG terminals and the conditions under which these terminals must operate."[211] Yet the states still have a role in the process of bringing an LNG facility online, which could ultimately dilute the effectiveness of FERC's exclusivity:

> States retain their authority to issue or deny permits under federal statutes such as the Coastal Zone Management Act and the Clean Water Act. This bill takes away no state authority, as long as state permitting agencies issue timely decisions. Let me repeat: State permitting authority remains in place under [EPAct 2005]. States can still deny LNG facilities on their coasts. But they need a reason—Clean Air Act, Clean Water Act, or the Coastal Zone Management Act.[212]

As such, states cannot choose or deny a particular site,[213] but they still retain significant authority over an LNG terminal at a given site.[214] The scope of potential dilatory tactics has simply been reduced.

Additionally, EPAct 2005 created a new section of the FPA, which requires FERC to "establish, by rule, incentive-based (including performance-based) rate treatments for the transmission of electric energy in interstate commerce by public utilities for the purpose of benefitting [sic] con-

---

tion process."); S. REP. NO. 109-78, at 29 (2005) ("[Section 381] does not provide FERC eminent domain authority over siting LNG facilities. The Committee believes that State and local government involvement should be a critical part of the FERC siting process.").

[211] Letter from U.S. Reps. Eshoo, Waxman, Capps, Napolitano, Miller, and Solis to California Governor Arnold Schwarzenegger, *quoted in* 151 CONG. REC. H2186 (daily ed. Apr. 20, 2005) (statement of Rep. Anna Eshoo (D-Cal.)); *see also* 151 CONG. REC. H2399 (daily ed. Apr. 21, 2005) (statement of Rep. Edward Markey (D-Mass.)) ("The Republican bill eliminates the State and local participation in determining where a facility like this would be placed."); 151 CONG. REC. H2434 (daily ed. Apr. 21, 2005) (statement of Rep. Timothy Bishop (D-N.Y.)) ("The language included in H.R. 6 silences the voices of state governments, local municipalities, and environmental advocacy organization during the LNG terminal site selection process.").

[212] 151 CONG. REC. H2432 (daily ed. Apr. 21, 2005) (statement of Rep. Gene Green (D-Tex)).

[213] 151 CONG. REC. H2189 (daily ed. Apr. 20, 2005) (statement of Rep. Barney Frank (D-Mass.)) ("[T]his bill takes a limited State role in the siting of these [LNG terminals] and makes it a nonexistent State role.").

[214] *See* EPAct 2005 § 311, 15 U.S.C.S. § 717(b) (LexisNexis 2006); Sound Energy Solutions, 108 FERC ¶ 61,155, at P 9 (stating that FERC "anticipate[s] relying on . . . state agencies' efforts to confirm compliance with federal statutory requirements" in authorizing new LNG facilities); 151 CONG. REC. S6449 (daily ed. June 14, 2005) (statement of Sen. Rodney Alexander (R-La.)) (stating that EPAct 2005 "preserves States' authorities under the Coastal Zone Management Act and other acts").

sumers by ensuring reliability and reducing the cost of delivered power by reducing transmission congestion."[215] The rules must include provisions to: (1) promote capital investment in efficient and reliable generation and transmission; (2) provide an attractive return on equity to attract new investment; (3) encourage increases in the capacity of current facilities; and (4) permit the recovery of all prudent costs related to reliability and infrastructure investment.[216]

As discussed above, the incentives found in FERC's emergency orders have worked in some smaller-scale circumstances, but it is clear from these instances that the market required additional prodding and concessions to achieve the investment goals. As the Path 15 Project indicated, specific requests for action may be necessary to motivate investors to act.[217] In contrast, the repeal of PUHCA simply removes impediments to certain types of investors.[218] But there is no indication that these investors will actually initiate new construction. In fact, history has shown that market participants have failed to act until expressly asked (and then motivated through additional incentives), even when the market seemed to be sending appropriate investment signals.[219]

The introduction of market forces, the repeal of PUHCA, and other recent policy changes might assist in the process of enhancing infrastructure, but a more comprehensive and focused approach is needed. The availability of financial resources (i.e., investment capital) is clearly a prerequisite to infrastructure enhancement. Thus, Congress has made construction of new infrastructure more feasible by making new funding sources available through such measures as the PUHCA repeal. But this is only a small first step. Providing availability of new funding sources without providing direct incentives and specific locations for new infrastructure construction is like oiling a hamster wheel: the wheel will spin faster, but it still won't go anywhere. Bringing in new funding sources is unlikely to be effective in a market where the current funding sources are not willing to invest in an industry that already represents solid and stable investment.[220]

---

[215] EPAct 2005 § 1241, 16 U.S.C.S. § 824s (LexisNexis 2006) (creating FPA section 219).

[216] *See id.*

[217] *See supra* notes 175–180 and accompanying text.

[218] *See supra* Part II.A.

[219] *See, e.g.*, Cal. Pub. Util. Comm'n v. FERC, 367 F.3d 925, 928–29 (2004) (stating that despite a clear need for infrastructure construction on "a uniquely critical path," incentives were needed or "the project would likely not be built in the near future").

[220] *See* Rob Carrick, *Infrastructure: A Safe Road to Riches*, GLOBE & MAIL (Toronto), Mar. 31, 2007, at B8 ("Investing in infrastructure is increasingly popular with pension funds because of the stable returns."); *cf. supra* notes 93–95 and accompanying text.

IV. CONCLUSION

To improve the U.S. energy infrastructure effectively, additional measures are necessary. These measures must include the aggressive implementation of processes for identifying necessary infrastructure enhancements, such as those related to NIETCs,[221] and the use of incentives combined with realistic in-service deadlines so that investors will invest and initiate construction quickly.

Despite the touted "comprehensive" nature of EPAct 2005, there remains a need for action. A few recent proposals provide examples of the types of aggressive, innovative approaches that should be applied to improving energy infrastructure.

In the natural gas sector, for instance, a "unique alliance" of five CEOs representing natural gas consumers and producers in the United States has outlined an "immediate" proposal to increase U.S. natural gas supply.[222] The proposal calls for Congress to "[r]educe the permitting backlog and accelerate the processes for applications to work on onshore federal non-park, non-wilderness lands," open up certain lands in the Gulf of Mexico, and "push to lift the exploration moratoria on the East Coast, West Coast and offshore Alaska."[223] Although the wisdom of these proposals might be debatable, they are at least the kind of proposals that *could* have a direct and immediate impact on the U.S. energy supply.

In the electricity area, Congress could have taken a bold move toward improving the U.S. energy infrastructure but instead stopped short of implementing a proven and much-needed measure: granting FERC exclusive siting authority for transmission lines. Although Congress granted limited backstop authority to approve federal electric transmission line siting in a few specific circumstances,[224] the process is protracted and inefficient.[225] Congress should have granted FERC exclusive jurisdiction

---

[221] *See, e.g.*, *supra* note 197 and accompanying text.

[222] *See* OilOnline, *supra* note 44. The alliance is made up of CEOs from Anadarko Petroleum Corporation, CF Industries Holdings, Inc., Nucor Corporation, New Jersey Resources, and Devon Energy. *Id.* Characterizing this group as an alliance between consumers and producers gives the CEOs the benefit of the doubt in assessing whom they represent. While Nucor Corporation and CF Industries Holdings, Inc., are direct consumers, it can be persuasively argued that a consumer representative for everyday people would really be a state rate-payer advocate, e.g., the State Of New Jersey Division of Ratepayer Advocate, and not a local distributing company that provides natural gas directly to consumers. Nonetheless, it is accurate that the alignment of CEOs is somewhat unique given the potentially competing interests of the various parties.

[223] *Id.*

[224] *See* EPAct 2005 § 1221, 16 U.S.C.S. § 824p (LexisNexis 2006).

[225] *See* notes 34–36 and accompanying text; *cf.* Steven J. Eagle, *Securing a Reliable Electricity Grid: A New Era in Transmission Siting Regulation?*, 73 TENN. L. REV. 1, 46 (2005) (noting the limitations and uncertainty of the federal siting authority granted under EPAct 2005 but stating that "if the Act is well implemented it has the potential to encourage investment in new transmission capacity and to stave off a catastrophic electric transmission shortage").

over transmission siting, making the FPA mirror the Natural Gas Act ("NGA").[226]

Historically, electricity was believed to be a local commodity: one better generated and monitored locally.[227] When it comes to electricity transmission, transactions (buying and selling capacity on transmission lines) are inherently "interstate" in nature, and are exclusively federally regulated under the FPA.[228] However, when it comes time to site and build the transmission lines upon which that federally regulated capacity will be bought and sold, the states have authority to restrict the construction. Thus, the competitive wholesale market concept is being advanced by federal regulators who lack siting jurisdiction, and the states with the siting authority may lack the statutory authority (if they were to have the inclination) to promote that market concept.[229] Exclusive federal transmission siting is the surest way to change course and initiate new interstate transmission infrastructure where it is desperately needed.

EPAct 2005, current market-based rate programs, and FERC's limited emergency orders all lack the scope and focus needed to trigger significant infrastructure investment. Even where such initiatives show promise, recent programs have been too fragmented and isolated to lead to significant change. The current large-scale programs are too long-term and speculative to be an adequate response to an energy crisis; recent short-term emergency solutions are so limited in time and scope that vast infrastructure

---

[226] Natural Gas Act, 15 U.S.C. §§ 717–717z (2000). Exclusive federal control for siting interstate natural gas pipelines was codified as part of the NGA in 1938. *See* Schneidewind v. ANR Pipeline Co., 485 U.S. 293, 300–01 (1988). Under this exclusive control, natural gas companies have been far more efficient in building necessary infrastructure than electric utilities. The primary issue in the natural gas industry has been supply, not the ability to move the commodity throughout the country once it is obtained. *See* Berry, *supra* note 37, at 137–38 (discussing FERC's efforts to increase natural gas supplies in light of continually increasing demand). Although there is also need for continued infrastructure enhancements in the natural gas industry, when compared to the electric industry, the NGA has provided a much better structure through which needed construction is authorized. As noted above, there remains a recognized lack of infrastructure (LNG terminals) for actually putting more natural gas supply on the grid. Congress at least tried to address this by confirming that siting of LNG terminals is also part of FERC's exclusive jurisdiction under the NGA. *See supra* notes 205–214 and accompanying text.

[227] *See* Eagle, *supra* note 225, at 1–2 ("The United States now is undergoing a transition from local command-and-control electric production and distribution to regional market-controlled production and distribution. This profound transformation requires changes in federal and state regulatory regimes to ensure the availability of an adequate and reliable supply of electricity throughout the nation.")

[228] *See* New England Power Co. v. New Hampshire, 455 U.S. 331, 340 (1982) ("In 1935, Congress enacted Part II of the Federal Power Act, 16 U.S.C. §§ 824-824k (1976 ed. and Supp.IV), which delegated to the Federal Power Commission (now FERC) exclusive authority to regulate the transmission and sale at wholesale of electric energy in interstate commerce, without regard to the source of production."); *see also* Eagle, *supra* note 225, at 34–35 (providing a good description of the issues raised by state versus federal jurisdiction over transmission).

[229] *See* Ashley C. Brown & Damon Daniels, *Vision Without Site: Site Without Vision*, ELECTRICITY J., Oct. 2003, at 23, 24.

needs remain even in the targeted areas. A large-scale, coherent, and comprehensive federal energy program is needed. This program must quickly and clearly identify energy infrastructure needs, provide significant financial incentives and realistic deadlines to entice and enable investors, and expand and exercise all available federal authorities to ensure that regulatory delays do not impede the process.

Despite the political battles that might lie ahead, the nation needs programs and plans that directly address the nation's energy crisis by improving the U.S. energy infrastructure. Given the unquestioned need for additional generation facilities and transmission lines and increased access to natural gas supplies to avert potentially drastic energy outages, it is time for FERC, Congress, and the Administration to put forth an innovative plan, building upon EPAct 2005, which will lead to immediate and sustained energy infrastructure enhancements. The need for energy is too significant, and the time line for construction too long, to tolerate additional misguided policies.

ARTICLE

REGULATING FOLLOW-ON BIOLOGICS

BRYAN A. LIANG*

More than twenty years ago, Congress recognized that the high costs of brand name chemical medicines limited access to their benefits. It responded with the Hatch-Waxman Act, which sped generic copies to market using an abbreviated approval process. Today, biotechnology drugs, known as biologics, provide revolutionary life-saving treatment. However, like their chemical predecessors, biologics are expensive, and access to their benefits is limited. Rep. Henry Waxman recently introduced the Access to Life-Saving Medicine Act (ALSMA) to establish a mechanism for abbreviated approval of brand name biologic copies, known as follow-on biologics. This Article compares chemical and biological drugs and examines the current regulatory regimes for follow-on biologics in the U.S. and in the E.U. It concludes that ALSMA fails to take into account the unique aspects of biologics and focuses upon limited economic benefits at the cost of safety. It offers an alternative bill that proposes a regime for abbreviated approval of follow-on biologics while more fully accounting for patient safety and access concerns.

I. INTRODUCTION

Biotechnology drugs, also known as biologics,[1] make up one of the fastest-growing sectors of the pharmaceutical industry. In 2005, worldwide spending on drug therapy grew by seven percent and topped $600 billion.[2] However, sales of biologics have grown even more rapidly, with an

* Professor of Law and Executive Director, Institute of Health Law Studies, California Western School of Law; Co-Director and Adjunct Associate Professor of Anesthesiology, San Diego Center for Patient Safety, University of California, San Diego School of Medicine; Adjunct Associate Professor of Public Health, College of Health and Human Services, San Diego State University; Adjunct Professor of Aviation, College of Aviation, Western Michigan University. J.D., Harvard Law School, 1995; M.D., Columbia University College of Physicians and Surgeons, 1991; Ph.D., University of Chicago, 1989; B.S., Massachusetts Institute of Technology, 1983. Professor Liang is also vice president of the Partnership for Safe Medicines, a consortium of academic, industrial, professional, and patient organizations dedicated to ensuring the safety of the drug supply. Thanks to Shannon M. Biggs and Dr. James Class for multiple readings of the manuscript, and to Professor Marv Shepherd for information on biologics at risk for adulteration, diversion, and counterfeiting.

[1] Biologic drugs are those medicines that are produced through protein synthesis and are injectable, such as insulin and growth hormone. *See infra* Part II.A. These are distinguished from more traditional chemical drugs that one would typically purchase at a pharmacy, like aspirin. *See id.*

[2] The United States represented the greatest share of this amount at $252 billion. *See* Matthew Herper & Peter Kang, *The World's Ten Best-Selling Drugs*, FORBES.COM, Mar. 22, 2006, http://www.forbes.com (search "Search Forbes.com" for "ten best-selling drugs" and click on "The World's Ten Best-Selling Drugs" hyperlink).

increase of seventeen percent in 2005 and annual expenditures worldwide of greater than $50 billion.[3] By 2010, spending on biologics is estimated to grow to $105 billion,[4] with biologics making up nearly half of all newly approved medicines.[5] Reflecting their widespread use, the top five drugs, in terms of Medicare expenditures, administered in physician offices in fiscal year 2005 were all biologics.[6] Society at large, including the more than 325 million patients around the world who have been treated with biologics,[7] has benefited greatly from the ongoing development of new and more effective drugs.[8] However, as is becoming increasingly apparent, this benefit carries with it high financial and social costs.[9]

Due to the high costs involved in the production and consumption of many medicines, regulatory regimes have been created to balance the intellectual property interests and investments made by originator companies with the need for wider patient access through generic forms of the drugs.[10] In the traditional chemical drug market, such a regime was created by the

---

[3] *See id.* However, note that even with this growth and revenue, the biotechnology industry as a whole lost greater than $4 billion in 2005. *See* John K. Iglehart, *Biotech Drugs Come of Age; Policymakers Take Notice*, HEALTH AFF. Sept.–Oct. 1202, 1202 (2006).

[4] *See* Mark J. Belsey et al., *Biosimilars: Initial Excitement Gives Way to Reality*, 5 NATURE REV. DRUG DISCOVERY 535, 536 (2006).

[5] *See* Huub Schellekens, *Follow-on Biologics: Challenges of the "Next Generation,"* 20 NEPHROLOGY DIALYSIS TRANSPLANTATION, iv31, iv31 (2005).

[6] *See* Press Release, Henry Waxman, Backgrounder on Biologics, at 2 (Feb. 20, 2007), http://www.henrywaxman.house.gov/pdfs/biologicsbackground_2.14.07.pdf. These drugs, covered under Medicare Part B, are Epogen, Aranesp, Procrit, Remicade, and Neulasta. *Id.; see also* Rebecca Adams, *Drugmakers' Battle for Medicare Market Share*, 64 CQ WKLY. 2606, 2606 (2006).

[7] *See* Carl B. Feldbaum, President, Biotechnology Industry Organization, Address at BioIreland Conference: "It Was 20 Years Ago Today": U.S. Biotechnology Trends, Fall 2002 (Nov. 14, 2002) Health Care Overview, *available at* http://www.bio.org/speeches/speeches/20021114.asp.

[8] For an economic perspective see, for example, DAVID M. CUTLER, YOUR MONEY OR YOUR LIFE 63 (2004) (arguing that the benefits of medicine and medical care justify the costs); Samuel A. Bozette et al., *Expenditures for the Care of HIV-Infected Patients in the Era of Highly Active Antiretroviral Therapy*, 344 NEW ENG. J. MED. 817, 817 (2001) (concluding that antiviral therapy for HIV-infected patients is cost-effective); Frank R. Lichtenberg, *Are the Benefits of Newer Drugs Worth Their Cost? Evidence from the 1996 MEPS*, HEALTH AFF. Sept.–Oct. 2001, at 241, 241 (finding, based on data on prescribed medicines from the 1996 Medical Expenditure Panel Survey, that people consuming newer drugs were significantly less likely to die by the end of the survey, were significantly less likely to experience work-loss days, and tended to spend less on all types of nondrug medical needs, resulting in a substantial net reduction in the total cost of treatment); *cf.* Iglehart, *supra* note 3. Of course, medicines also are critical in improving quality of life, or even in saving lives.

[9] This high financial cost reflects the substantial costs of development. It is estimated that, on average, a new biologic drug costs $1.2 billion to develop and requires 97.7 months for clinical development and regulatory review. *See* Press Release, Tufts Ctr. for the Study of Drug Dev., Average Cost to Develop a New Biotechnology Product Is $1.2 Billion (Nov. 9, 2006), *available at* htpp://csdd.tufts.edu/NewsEvents/NewsArticle.asp?newsid=69.

[10] Originator companies are those that developed the originally approved drugs. Sometimes these are known as "innovator" companies; however, recognizing that both originator companies and companies that follow the originator must innovate to develop products, the term originator is used herein.

Drug Price Competition and Patent Term Restoration Act of 1984,[11] more commonly known as the Hatch-Waxman Act ("HWA"). The HWA added section 505(j) to the Federal Food, Drug, and Cosmetic Act ("FDCA").[12] This section and its accompanying regulations created the Abbreviated New Drug Application ("ANDA") process, which was designed to provide independent generic firms with a strong incentive to develop and introduce lower-cost generic drugs to consumers.[13] By virtually all accounts, the HWA has been highly successful in bringing cheaper generic products to the market while maintaining incentives for the development and discovery of new drugs.[14]

Because of the effectiveness of the HWA approach in the context of traditional chemical drugs, some have called for applying a similar regime for biologics, given the similarly high costs of this class of drugs.[15] In particular, Rep. Henry Waxman (D-Cal.), one of the original authors of the HWA, has sponsored the Access to Life-Saving Medicine Act ("ALSMA").[16] To determine whether such a legal infrastructure is appropriate for regulating these "follow-on" biologics,[17] a thorough policy assessment is necessary.

This Article attempts to perform such an assessment by identifying potential information gaps in the science of biologics production that may have an impact upon regulatory considerations, examining the proposed ALSMA regime to determine the relative vulnerability of the polity that bears the risk of policy failure, and considering the harms that

---

[11] Pub. L. No. 98-417, 98 Stat. 1585 (codified as amended at 21 U.S.C. § 355 (2000); 35 U.S.C. §§ 156, 271, 282 (2000 & Supp. IV 2004)).

[12] Food, Drug, and Cosmetic Act ("FDCA") § 505(j), 21 U.S.C. § 355(j) (2000 & Supp. IV 2004).

[13] *See, e.g.*, *In re* Barr Labs, Inc., 930 F. 2d 72, 76 (D.C. Cir. 1991) (stating that the goal of the Hatch-Waxman Act was to "get generic drugs into the hands of patients at reasonable prices—fast").

[14] *See, e.g.*, Jon Leibowitz, Comm'r, Fed. Trade Comm'n, Address at the Second Annual In-House Counsel's Forum on Pharmaceutical Antitrust: Exclusion Payments to Settle Pharmaceutical Patent Cases 3 (Apr. 24, 2006), *available at* http://www.ftc.gov/speeches/leibowitz/060424PharmaSpeechACI.pdf.

[15] *See, e.g.*, Citizen Petition Filed on Behalf of the States of Kansas, Minnesota, Vermont and Wisconsin, Docket No. 2006P-0306, *available at* http://www.fda.gov/ohrms/dockets/dockets/06p0306/06p-0306-cp00001-vol1.pdf (calling for the FDA to "issue guidance that will facilitate the availability of more affordable, therapeutically equivalent versions of [human growth hormone]"); Daniel Costello, *Generic Biotech Drugs Backed,* L.A. TIMES, Nov. 11, 2006, at C1 (discussing congressional interest in legislation setting up an approval process for generic versions of biologics).

[16] H.R. 1038, 110th Cong. (2007).

[17] The terms "follow-on biologic," "follow-on protein product" (used by the FDA), "biosimilar" (used in the E.U.), or "biogeneric," are used for subsequent legal versions of a biologic drug developed independently from the originator. *See* Simon D. Roger, *Biosimilars: How Similar or Dissimilar Are They?*, 11 NEPHROLOGY 341, 341 (2006). Note, however, that the Spanish term *biosimilares* denotes illegally produced drugs—biologics and nonbiologics—in countries with limited intellectual property regimes. Interview with Lew Kontnik, Director, Brand Protection and Business Continuity, Amgen Inc., in San Diego, Cal. (June 9, 2006).

might be associated with that potential failure. A foundational principle for this analysis is the principle of "greater and higher"—the greater the information gaps, the higher the vulnerability of the polity shouldering the risk of failure, and the greater the potential harm, the more the policy should emphasize consumer safety over its other potential benefits.

Part II begins this analysis by reviewing the science of biologics and their production, setting forth the substantive differences in research, form, complexity, and manufacture between biologics and traditional chemical drugs. Part III considers the patient safety issues associated with follow-on biologics, including heightened risk of counterfeit, diverted, and adulterated biological products in the U.S. supply chain. As a foundation for the analysis of ALSMA, Part IV explores current regulatory mechanisms for biologics, including existing U.S. regulation and the European Union ("E.U.") scheme, including the latter's regulation of follow-on biologics, or "biosimilar biologics," as they are known in the E.U. This Part notes that there is currently no general abbreviated pathway in the United States for follow-on biologics approval, and it discusses some themes associated with biosimilar approvals and rejections in the E.U. Next, Part V reviews and assesses ALSMA's proposed approach for regulating follow-on biologics in the U.S. This Part raises significant concerns over the proposed legislation's failure to take into account the complexities of biologics, the vulnerability of the population that would be subject to the policy risks from the proposed legislation, the tremendous magnitude of harm that could result from failure, and the legislation's potential adverse impact on innovation. Part VI offers an alternative legislative proposal for a regulatory structure that addresses these policy concerns. Finally, Part VII offers some concluding remarks.

II. The Science and Development of Biologics

Biotechnology companies generally spend significantly more on research and development than do their traditional chemical drug counterparts.[18] And for good reason: the development of biologics is highly complex and involves appreciably greater and more diverse resources for drug creation and production.[19] Chemical medicines are well known to us—they are the pills and capsules that we buy at a pharmacy. Yet most of us are also familiar with some biologic drugs as well, like injectable drugs such as insulin, growth hormone, and vaccines. Differences in the composition and production of chemical and biologic drugs provide important insights on the regulation of biologic drugs.

---

[18] *See* Rodney J. Y. Ho & Milo Gibaldi, Biotechnology and Biopharmaceuticals 4 (2003) (stating that biologics producers spend more than 20% of revenues on research and development, compared to the 6% to 18% spent by large pharmaceutical companies).

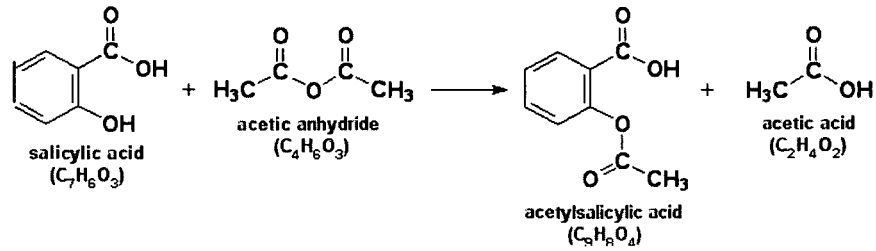
[19] *See id.*

A. Chemistry Versus Biology

Chemistry, an offshoot of now-archaic alchemy, is an ancient science that deals with the properties, composition, and structures of inanimate matter.[20] Biology, from the Greek word *bìo*, meaning "life," is, indeed, the study of life.[21] This difference is the foundation of all the issues that arise between chemical medicines and biologics.

Chemistry looks at the composition and interactions of matter. In general, scientists create chemical molecules by relying on static formulations of compounds and specific ingredients, along with knowledge of the way in which discrete chemical reactions take place. This allows for a formulaic stepwise approach to creation of chemical medicines.

Making generic aspirin, which is chemically known as acetylsalicylic acid, illustrates this concept. The chemical process is as follows:[22]



Stepwise, a chemist would perform this synthesis by simply placing the salicylic acid and acetic anhydride ingredients into a vessel, which, under standard laboratory conditions, would cause a chemical reaction to consistently create the final acetylsalicylic acid product (aspirin), plus the side product of acetic acid (the prime ingredient of household vinegar).[23] The molecule is easily characterized by the chemical formula indicated above.

Biological processes and the making of biological drugs, however, are significantly more complex. Biological medicines are not made from synthetic chemical ingredients discretely added and mixed together to form an easily characterized product such as aspirin. Instead, biologics are made using the very machinery of life: these proteins are synthesized from living organisms housed in life-sustaining, protected environments.

---

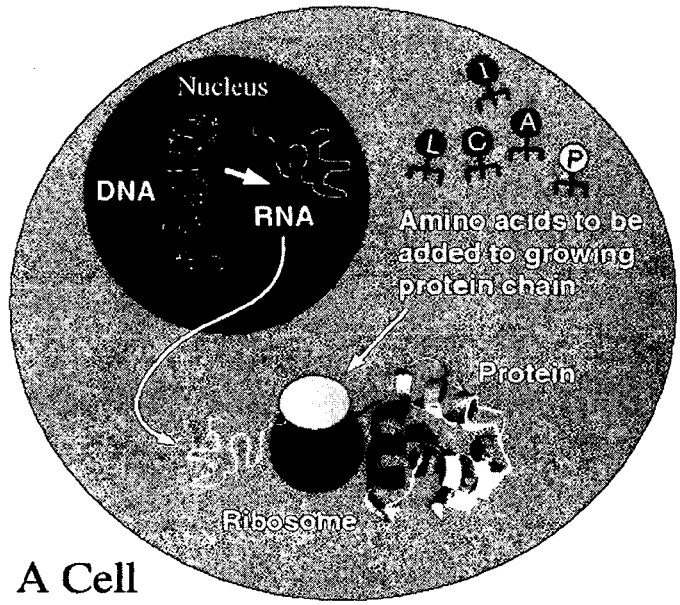
[20] *See, e.g.*, *Chemistry*, ENCYCLOPEDIA BRITTANICA ONLINE, http://www.britannica.com/eb/article-9108655/ chemistry (last visited Apr. 20, 2007).

[21] *See, e.g.*, *Biology*, ENCYCLOPEDIA BRITTANICA ONLINE, http://www.britannica.com/ebc/article-9357300 (last visited Apr. 20, 2007).

[22] *See* Chem. Heritage Found., Making Aspirin, http://www.chemheritage.org/educational services/pharm/asp/asp31.htm (last visited Feb. 23, 2007). The diagram is also from this website.

[23] *See* Anne Marie Helmenstine, How to Make Aspirin–Acetylsalicylic Acid–Procedure, http://chemistry.about.com/od/demonstrationsexperiments/ss/aspirin_3.htm (last visited Apr. 19, 2007).

Protein synthesis is an exceedingly complex process,[24] and so what follows is merely a brief description of that process. After a cell receives a signal to produce a protein product, the deoxyribonucleic acid ("DNA") in the cell's nucleus then unwinds, and through a process known as transcription, the cell produces an intermediary compound known as messenger ribonucleic acid ("mRNA").[25] This mRNA acts as a protein template; each three-base triplet on the mRNA corresponds to a unique amino acid.[26] Upon completion of the mRNA synthesis in the nucleus, the mRNA is then sent outside the nucleus to specialized proteins known as ribosomes.[27] In a subsequent process, known as translation, the mRNA interacts with both the ribosomes and a molecule called transfer ribonucleic acid (tRNA).[28] The tRNA carries specific amino acids and attaches to the mRNA triplets on the ribosome to form a growing chain of amino acids.[29] The protein synthesis process is schematically depicted below:[30]



[24] *See, e.g.*, Jeremy M. Berg et al., BIOCHEMISTRY (5th ed. 2002) at 813–33.
[25] *See id.* at 783–89.
[26] *See id.* at 826–27.
[27] *See id.* at 829.
[28] *See id.* at 830–35.
[29] *See id.*
[30] Laboratory of Molecular Physics, Univ. of Oxford, How are proteins made?, http://biop.ox.ac.uk/www/mol_of_life/How_Are_Proteins_Made.html (last visited Apr. 20, 2007).

In biologics production, scientists create uniform and unique "cell lines" of a specific bacterium, mammalian organ, or another cellular source.[31] This process begins by introducing the DNA that codes for the protein product of interest into the cell line; this DNA travels to the cellular nucleus, and the cellular machinery from the cell line reads the DNA and creates the protein.[32] This process must take place within a highly controlled environment to ensure the appropriate materials and conditions for maintenance of cell line life and the successful production of the biologic.[33]

Concomitant with their different manufacturing processes and functions, biologics and chemical medicines are tremendously different in size. A biologic, with thousands to millions of atoms forming a highly interconnected group of hundreds to thousands of amino acids aggregated into chains and subgroups, is much larger than a chemical drug, which typically consists of just dozens of atoms forming a single molecule.[34] Thus, unlike in the context of chemical drugs, one cannot characterize a biologic simply on the basis of a chemical formula. Indeed, efforts to characterize a biologic must include primary, secondary, tertiary, and in some cases quaternary descriptions, and even those may not completely characterize the product.[35] Further complicating these descriptive efforts are various additional molecules that are attached to particular sites of the protein, which also affect its characterization and irrevocably alter its structure and bioactivity.[36]

As an illustration, aspirin's molecular weight is approximately 180 Daltons.[37] In comparison, a single amino acid, the fundamental building block for a protein, has a similar molecular weight (between 75 Daltons and 204 Daltons).[38] Since a large number of amino acids connect together to form a protein, however, common biologic drugs often have weights that are orders of magnitude greater than aspirin. For example, insulin, a small

---

[31] *See* Deborah M. Shelton, *Moving Toward Biogenerics in the U.S.*, PHARMA AND BIOINGREDIENT, Apr. 2005, *available at* http://www.pharmabioingredients.com/articles/2005/04/feature2.

[32] *See* Roger, *supra* note 17, at 342.

[33] *See* Shelton, *supra* note 31.

[34] *See id.*

[35] *See* Berg, *supra* note 24, at 51–63.

[36] *See* Sheldon, *supra* note 31; *see also* Roger, *supra* note 17, at 342. A protein can undergo a large array of post-translational chemical modifications. *See* Len Packman, *A Database of Protein Post-translation Modifications*, ASSOC. OF BIOMOLECULAR RES. FACILITIES, *available at* http://www.abrf.org/index.cfm/dm.home?AvgMass=all; *see generally* Daan J. A. Crommelin et al., *Shifting Paradigms: Biopharmaceuticals Versus Low Molecular Weight Drugs*, 266 INT'L J. PHARMACEUTICALS 3 (2003) (describing the limited means for characterizing biologics).

[37] *See* MARGARET A. WHEATLEY, DRUG DELIVERY: ENGINEERING BIOTECHNOLOGY GATEWAY PROJECT 10, http://www.gatewaycoalition.org/files/Engineering_Biotechnology%5CHtmls/drugdelivery.pdf (last visited Apr. 17, 2007).

[38] *See* ProtScale Tool: Molecular Weight of Each Amino Acid, http://www.expasy.org/tools/pscale/Molecular weight.html (last visited Feb. 23, 2007).

biologic, has an approximate molecular weight of 5800 Daltons;[39] growth hormone has a molecular weight of approximately 22,000 Daltons;[40] and daclizumab, which suppresses the immune system for transplant patients, has a molecular weight of approximately 142,600 Daltons.[41]

The large size of proteins reflects the biological need to create specific three-dimensional structures that function in exquisitely sensitive human physiological roles. These roles include acting as enzymes, sustaining structural components of the body, serving as molecular channels for diverse functions, acting as signals for physiologic processes, providing for adaptive immune function, and engaging in many other essential life processes.[42] Due to the nature of the large, complex molecules associated with these life processes, very minor chemical changes can completely eliminate or severely alter a biologic's function.[43]

B. Replicating the Drug Product

Because of the differences in production and size between biologics and chemical drugs, as well as the unique cellular source of biologics, it is nearly impossible to make truly identical copies of a protein using two different production cell lines.[44] Just as humans differ from each other in their metabolism and other physiological characteristics, cells from different cell lines exhibit diversity.[45] This diversity is present to an even greater degree between cell lines from different living organisms, such as bacteria,

---

[39] *See, e.g.*, *Chemistry*, ENCYCLOPEDIA BRITTANICA ONLINE, *supra* note 20.

[40] Felix Hepner et al., *Mass Spectrometrical Analysis of Recombinant Human Growth Hormone (Genotropin®) Reveals Amino Acid Substitutions in 2% of the Expressed Protein*, 3 PROTEOME SCI. 1, 2 (2003), *available at* http://www.proteomesci.com/content/pdf/1477-5956-3-1.pdf.

[41] *See* DrugBank: Daclizumab, http://redpoll.pharmacy.ualberta.ca/drugbank (search "Search Drugbank for:" for "daclizumab" and click on "BIOD00007" hyperlink) (last visited Apr. 17, 2007).

[42] *See* Bergen County Academies, Functions of Proteins, http://www.bergen.org/AAST/Projects/Gel/profunc1.htm (last visited Apr. 20, 2007).

[43] *See Biotech Drugs: Hearing Before the S. Comm. on Health, Education, Labor and Pensions*, 110th Cong. 3 (Mar. 8, 2007) (statement of Jay P. Siegel, M.D., Group President of Biotechnology, Immunology and Oncology, Research and Development, Johnson & Johnson) *available at* http://help.senate.gov/Hearings/2007_03_08/Siegel.pdf [hereinafter *Biotech Drugs Hearing*]. For example, the sickled cells of sickle cell anemia are a result of a single incorrect amino acid in the protein chains of the hemoglobin molecule. *See id.* at 11; *see also Anemia, Sickle Cell*, *in* NAT'L CTR. FOR BIOLOGIC INFO., GENES AND DISEASE, http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=gnd.section.98&ref=sidebar (last visited Apr. 17, 2007).

[44] Roger, *supra* note 17, at 341 ("It should be stressed that it is not possible for another manufacturer to duplicate the original production process of the innovator, thus the terms 'generic biosimilars' or 'generic biopharmaceuticals' are misleading. . . . [B]y definition, similar biological medicinal products are not generic medicinal products . . . .") (citation omitted).

[45] *See, e.g.*, *id.* at 342 ("The various cell lines that are used to produce the proteins may have an impact on the overall structure of the protein, and may affect post-translational modifications . . . .").

mammalian organs, yeast, and other sources.[46] Cell lines do not always create the full, functional protein efficiently; the biologic "product" from a cell line is actually a heterogeneous and complex mix of materials that includes the desired product as well as side products and cell materials.[47] Hence, even though the DNA sequence that codes for a protein is identical, it may make a different product in a different cell line. Accordingly, complete and specific characterization of the products from a specific cell line creating a particular biologic formulation is often impossible.[48] The complexity of the biologic molecule, its sensitivity to production, and the challenges associated with characterization result in its being defined primarily in terms of its manufacturing method.[49]

Contrast this situation with that of chemical medicines. Generic manufacturers can make identical copies of chemical medicines either by using the same known chemical processes or through independently developed processes.[50] It is possible to make atom-by-atom comparisons between the original manufacturer's product and a generic product through a wide array of chemical and functional tests, to verify that the products have identical chemical compositions.[51] Purifications can be performed on discrete, individual molecules, resulting in a highly uniform composition of the product.[52]

In comparison, a follow-on biologic by its very nature will be a different product from the originator's biologic.[53] This makes both originator and follow-on biologic drug production much more challenging than generic chemical drug production and results in unique expenditures not faced by chemical drug manufacturers.[54]

---

[46] *See id.*; *see also Biotech Drugs Hearing*, *supra* note 43, at 3 ("[B]iologics frequently can bind themselves to form pairs or aggregates, can change their shape over time or with minor changes in conditions, and can interact with materials in their containers and packaging."); Lisa J. Raines, *Bad Medicine: Why the Generic Drug Regulatory Paradigm Is Inapplicable to Biotechnology Products*, 5 J. BIOLAW & BUS. 6 (2002) (explaining that variations in the size of the proteins within a cell may affect the product).

[47] *See Biotech Drugs Hearing*, *supra* note 43, at 3.

[48] *See* FDA, Frequently Asked Questions About Therapeutic Biological Products (July 21, 2006), http://www.fda.gov/cder/biologics/qa.htm ("In contrast to chemically synthesized small molecular weight drugs, which have a well-defined structure and can be thoroughly characterized, biological products are generally derived from living materials—human, animal, or microorganism—are complex in structure and thus are usually not fully characterized.").

[49] Melissa R. Leuenberger-Fisher, *The Road to Follow On Biologics: Are We There Yet?*, 23 BIOTECHNOLOGY L. REP. 389, 393 (2004).

[50] *See Biotech Drugs Hearing*, *supra* note 43, at 3.

[51] *See id.*

[52] *See generally, e.g.*, HANDBOOK OF ANALYTICAL SEPARATIONS: SEPARATION METHODS IN DRUG SYNTHESIS AND PURIFICATION (Klára Valcó ed., 2000) (providing an overview of the multitude of ways chemical drugs can be purified).

[53] *See supra* notes 44–49 and accompanying text.

[54] *See, e.g.*, Biotech. Indus. Org., A Brief Primer on Manufacturing Therapeutic Proteins, http://www.bio. org/healthcare/pmp/factsheet1.asp (last visited Apr. 17, 2007) ("Cell culture facilities, which take, on average, three to five years to construct, cost $250 million to $450 million to construct and must be individually approved and certified by the FDA

C. Manufacturing Issues

Because of the sensitivity and complexity of using living organisms to produce biologics, subtle changes in conditions associated with production have substantive effects on the products. In general, changes in biologic manufacturing processes can occur through alteration of cell lines, environment, or extraction or purification conditions, even in production by the same company. Because of these changes, the properties of the resultant product may also change,[55] justifying FDA requests for manufacturer submission of additional comparability analysis.[56]

For example, when the manufacturing process changes, a litany of changes can occur in the structure of the protein alone: the amino acids that make up the protein may have their order shuffled, the product may be altered when amino acids inadvertently link together, hydrocarbon rings may form ("molecular cyclization") when new terminal portions of the protein are created, chemical bonds are cleaved, and additional molecules may be attached, all of which have the potential to alter the clinical effects of the drug.[57] These differences may also distort the conditions required to preserve the product, including the temperature control requirements for its distribution.[58]

The FDA has noted that biologic drugs from different manufacturers might not be readily interchangeable. It has stated that:

> [A]s of today, the FDA has not determined how interchangeability can be established for complex proteins. Different large pro-

---

prior to full-scale operation."). In addition, compared with chemical medicines, which require scientists to conduct 40 to 50 tests during manufacturing, scientists creating biologics must undertake 250 or more tests. *See* European Assoc. of Bioindus., Biological and Biosimilar Medicines, Healthcare Biotech Fact Sheet (Jan. 2005), http://www.europabio.org/documents/FS-Biosimilar.pdf.; *see also* Leuenberger-Fisher, *supra* note 49, at 393 (describing the greater challenges facing the production of biologics in contrast with chemical drugs).

[55] *See infra* Part III (discussing patient safety issues associated with changes in biologics products). For example, even for relatively simple biologics such as growth hormone, six products from the same company with the same 191 amino acids were metabolized by patients at different rates (from 1.75 hours in the patient with the quickest metabolism to seven to ten hours in the patient with the slowest metabolism). *See* Raines, *supra* note 46, at 10. For example, differing versions of erythropoietin made in different parts of the world show differences with respect to the specific molecular forms each contains. *See generally* Huub Schellekens, *Biosimilar Epoetins: How Similar Are They?*, 3 EUR. J. HOSP. PHARMACY 43 (2004).

[56] *See* FDA, ICH Q5E: COMPARABILITY OF BIOTECHNOLOGICAL/BIOLOGICAL PRODUCTS SUBJECT TO CHANGES IN THEIR MANUFACTURING PROCESS (2003), *available at* http://www.fda. gov/cder/guidance/6003dft.pdf.

[57] *See* Arthur J. Chirino & Anthony Mire-Sluis, *Characterizing Biological Products and Assessing Comparability Following Manufacturing Changes*, 22 NATURE BIOTECH. 1383, 1384–86 (2004).

[58] *Id.* at 1385 ("Protein stability can be altered and can affect aggregation depending on formulation . . . and the physical environment (hence the need for strict maintenance of a cold chain during product distribution).").

> tein products, with similar molecular composition may behave differently in people and substitution of one for another may result in serious health outcomes. . . .[I]t is important that patients and physicians be aware that protein products with similar molecular composition may indeed not be interchangeable.[59]

These substantive scientific issues associated with the complexity of biologics and the related manufacturing processes create very important concerns for patient safety. These concerns are justified by reports from the medical community regarding clinical effects.[60]

III. PATIENT SAFETY ISSUES

A. Clinical Concerns

1. Product Excipients

Patients have experienced adverse consequences related to the delicacy of biological products and variation in production processes. Such products include excipients, the nontherapeutic materials accompanying the drug.[61] Although chemical drug excipients are also associated with clinical concerns,[62] biologic-associated events are more serious, more difficult to predict, and warrant policy vigilance to ensure patient safety.

Consider the following prominent example. Immune globulin ("IgG") is a biologic used to treat a wide array of disease states, including primary immunodeficiencies, immune-mediated low platelets, recent bone marrow transplantation in young patients, specific leukemias, and pediatric human immunodeficiency virus type 1 ("HIV-1") infection.[63] This biologic has been made by different manufacturers since the FDA first approved it in 1981.[64]

In a review covering the years from 1985 to 1998, the Centers for Disease Control and Prevention ("CDC") reported that excipients added during the manufacturing process by some companies to stabilize and help preserve the life of the biologic resulted in severe renal adverse allergic reac-

---

[59] Press Release, FDA, U.S. FDA Considerations: Discussion by National Regulatory Authorities with World Health Organization (WHO) on Possible International Non-proprietary Name (INN) Policies for Biosimilars (Sept. 1, 2006), *available at* http://www.fda.gov/cder/news/biosimilars.htm [hereinafter U.S. FDA Considerations].

[60] *See infra* Part III (discussing patient safety issues relating to biologics).

[61] *See* Thomas A. Wheatley, *What Are Excipients?, in* EXCIPIENT TOXICITY AND SAFETY 1 (Myra L. Weiner & Lois A. Kotkoskie eds., 2000).

[62] *See generally* Robert E. Osterberg & Norman A. See, *Toxicity of Excipients—A Food and Drug Administration Perspective*, 22 INT'L J. TOXICOLOGY 377 (2003).

[63] *See* Ctrs. for Disease Control & Prevention, *Renal Insufficiency and Failure Associated with Immune Globulin Intravenous Therapy—United States, 1985-1998*, 48 MORBIDITY & MORTALITY WKLY REP. 518, 518 (1999).

[64] *See id.*

tions among patients.[65] The CDC postulated that the interaction between IgG and manufacturing material excipients may have caused the reactions, although the mechanism still remains unclear.[66]

Yet even with this CDC warning and the intense study that has accompanied it, IgG manufacturing continues to be associated with clinical concerns. Most recently, in late 2005, one manufacturer sent out letters to healthcare professionals reporting that its immunoglobulin biologic was associated with falsely elevated glucose readings and intravascular hemolysis.[67] Intravascular hemolysis's relation to the biologic product "indicates that the etiology is complex and . . . is not clearly understood."[68] The falsely elevated glucose readings appear to be related to the excipient used in the product, and became the subject of a formal FDA warning.[69] The FDA took this step because "there have been reports of the inappropriate administration of insulin [because of falsely elevated glucose readings] and consequent life-threatening/fatal hypoglycemia in response to erroneous test results . . . ."[70] Such distortions of clinical test results from excipient materials have occurred in the United States and internationally.[71]

There have been additional reports associating biologics' excipients with negative clinical consequences. For example, a hepatitis B[72] detection kit with a biologic component was recalled because the FDA found that "[a]n unknown component in the diluting solution used to test blood and serum samples may produce 'Not Confirmed' results for samples found to be positive with the initial test, which can cause some results to be classified as false negatives."[73] The FDA therefore enacted a Class I recall, "the most

---

[65] *See id.*

[66] *See id.* at 520.

[67] *See* Letter from Maurice Genereux, Med. Dir., Cangene Corp. and Richard Schiff, Global Med. Dir., Baxter Healthcare Corp. to Healthcare Professionals, (Dec. 5, 2005), *available at* http://www.fda.gov/Medwatch/safety/2006/WinRho_deardoc_FINAL_07-Dec-2005.pdf [hereinafter Maurice Generaux Letter]. Intravascular hemolysis is the destruction or toxic damage of red blood cells within the circulation. *See* Dep't of Pathology, Univ. of Va. Health Sys., Hemolytic Anemia, http://www.med-ed.virginia.edu/courses/path/innes/rcd/hemo.cfm (last visited Apr. 17, 2007).

[68] Maurice Genereux Letter, *supra* note 67, at 1.

[69] FDA, Important Safety Information on Interference With Blood Glucose Measurement Following Use of Parenteral Maltose/Parenteral Galactose/Oral Xylose-Containing Products (Nov. 9, 2005), *available at* http://www.fda.gov/cber/safety/maltose110405.htm.

[70] *Id.*

[71] *See id.* (stating that the falsely elevated glucose reading results from "a known drug-device interaction," and citing sources from Australia and the United Kingdom, as well as the United States).

[72] Hepatitis B is a serious viral infection of the liver that can be the precursor to liver cirrhosis and hepatocellular carcinoma (liver cancer). *See* Kenneth W. Lin & Jeffrey T. Kirchner, *Hepatitis B*, AM. FAM. PHYSICIAN (Am. Acad. of Family Physicians, Leawood, Kan.), Jan. 2004, at 75, *available at* http://www. aafp.org/afp/20040101/75.pdf.

[73] FDA, Class 1 Recall: Ortho-Clinical Diagnostics VITROS® Immunodiagnostic HBsAg Confirmatory Kit (Dec. 15, 2005), *available at* http://www.fda.gov/cdrh/recalls/recall-121505.html.

serious type of recall and involv[ing] situations in which there is a reasonable probability that use of the product will cause serious injury or death."[74] In this case, the FDA noted the importance of the clinical issue:

> False negative results may prevent some patients infected with or carrying the hepatitis B virus from receiving necessary treatment. This is especially true for pregnant women whose tests show false negative results. When their fetuses are born, they will be presumed negative, and not treated with HBIG (hepatitis B immunoglobulin) and hepatitis B vaccine. Such infants have a 90% chance of progressing to chronic hepatitis B virus infection resulting in possible liver transplantation and early death.[75]

This case further illustrates that biologics face difficult manufacturing issues associated with excipients, which may have a tremendous impact upon patients' present and future well-being.

2. *Product Immunogenicity and Related Issues*

The inherent complexity of biologic drugs and their manufacture can create potential differences across versions of the same product, presenting additional problems aside from those involving excipients. These differences have resulted in severe clinical reactions with lasting and devastating impact, raising important patient safety concerns.

Due to the relatively large size of biologics, there is one central concern for all drugs in this category that is not present for chemical medicines: the potential for the product to induce an adverse immunologic reaction in a patient whose body sees the drug as a foreign invader, such as a virus or a bacterium.[76] In this situation, the drug is acting as an "immunogen."[77] The relative propensity, or "profile," of a molecule to induce immune reactions is known as immunogenicity.[78] Understanding immunogenicity of biologics is essential for safety and clinical monitoring pur-

---

[74] *Id.*

[75] *Id.*

[76] *See* Paul Chamberlain, *Immunogenicity of Therapeutic Proteins: Part I: Causes and Clinical Manifestations of Immunogenicity*, 5 REG. REV. 4, 4 (2002), *available at* http://www.egagenerics.com/doc/Chamberlain-Immuno_ pt1.pdf ("A major regulatory consideration in the development of therapeutic proteins is the assessment of *undesired* host immune responses to the drug product that may abrogate efficacy and give rise to potentially fatal adverse reactions.").

[77] *See id.* 4 ("An immunogen is any substance that can induce an immune response, either a humoral (B-lymphocyte secreting antibody), or a cell-mediated (T-lymphocyte), or both (most common situation)."). Antibodies are formed against the biologic as a result of this immune response. *See id.*

[78] *See id.*

poses. Reactions can be severe, and they can have serious negative outcomes for patients.[79]

For example, erythropoietin is a naturally occurring substance in the body as well as a biologic drug that promotes red blood cell growth.[80] Under a licensing agreement, both Amgen and Johnson & Johnson purportedly used the same methodology to produce Amgen's original erythropoietin drug. The former was produced in the United States and known as Epogen, while the latter was produced in the E.U. and known as Eprex.[81] Johnson & Johnson made several changes to the formulation of Eprex that were initially considered minor.[82]

Unfortunately, roughly two years after Johnson & Johnson made these changes, multiple patients taking Eprex in the E.U. developed a rejection reaction to the drug that resulted in pure red cell aplasia,[83] a severe and life-threatening condition in which the bone marrow ceases production of red blood cells.[84] There have been no reports of a similarly increased incidence of pure red blood aplasia among patients taking Epogen.

Clinical investigations revealed that the new formulation of Eprex had a different immunogenicity profile than Epogen and that patients developed antibodies to the new formulation.[85] Although this circumstance alone was cause for significant alarm, the new immunogenicity also created antibodies against the patients' own naturally occurring erythropoietin.[86] Moreover, it caused a cross-reactivity response to other pharmaceutical forms of erythropoietin, such as Epogen, resulting in an inability for the patients to rely on other forms of the drug.[87] The outcome was that patients could not use their own erythropoietin or commercial versions of

---

[79] *See id.* ("[E]ven if they may occur very rarely, such [immune] responses can have a very serious outcome—such as in the case for patients developing cross-reactive humoral responses to endogenous thrombopoietin[, an immune response against the patient's own naturally occurring thrombopoietin], resulting in irreversible thrombocytopenia [an inability of the body to manufacture red blood cells]").

[80] *See* Medterms Medical Dictionary, Erythropoietin, http://www.medterms.com/script/main/art.asp?articlekey=7032 (last visited Apr. 17, 2007).

[81] *See* Nicole Casadevall et al., *Epoetin-Induced Autoimmune Pure Red Cell Aplasia*, 16 J. AM. SOC'Y NEPHROLOGY S67, S67 (2005).

[82] *See* Charles L. Bennett et al., *Pure Red-Cell Aplasia and Epoetin Therapy*, 351 NEW ENG. J. MED. 1403, 1406 (2004).

[83] *See id.* at 1405–06; Casadevall, *supra* note 81, at S67; Huub Schellekens, *Erythropoiesis-Stimulating Agents—Present and Future*, BUS. BRIEFING: EUROPEAN ENDOCRINE REV. 2006 1, 3 (2006) *available at* http://www.touchbriefings.com/pdf/1711/ACF278B.pdf (stating that approximately 225 patients experienced pure red cell aplasia).

[84] *See* WebMD, *Pure Red Cell Aplasia, Acquired*, A–Z HEALTH GUIDE, http://www.webmd.com/hw/anemia/nord506.asp (last visited Apr. 17, 2007).

[85] *See* Charles L. Bennett et al., *Long-Term Outcome of Individuals with Pure Red Cell Aplasia and Antierythropoietin Antibodies in Patients with Recombinant Epoetin: A Follow-up Report from the Research on Adverse Drug Events and Reports (RADAR) Project*, 106 BLOOD 3343, 3343 (2005); Casadevall et al., *Epoetin-Induced Autoimmune Pure Red Cell Aplasia*, *supra* note 81, at S68.

[86] *See* Bennett, *supra* note 85, at 3343.

[87] *See id.*

exogenous erythropoietin to produce new red blood cells; several patients died, and others became permanently transfusion-dependent.[88] Clearly, seemingly minor changes in a biologic can have a tremendous clinical impact.

Unfortunately, the prediction, investigation, and characterization of adverse immunogenicity mechanisms are challenging. The immunogenicity of biologic drugs appears to be related to a broad array of factors, including the biologic's structure, the patient's genetic attributes, the type of biologic in question, impurities in the product, the route of administration, and the frequency of use.[89] The human immune response to a biologic product is difficult to predict generally, and this is even more difficult in the face of changes to manufacturing processes.[90]

Indeed, despite significant study of this important issue, there has been little in the way of discernable patterns that provide scientists with a clear conception of what changes to a molecule impact immunogenicity, or whether the magnitude of that effect will be great or small in a given case.[91] In the Eprex case, for example, intense research efforts still have not led to a clear causal explanation, despite a wide array of theories offered (including different carbohydrate structures,[92] route of administration,[93] changes in the stabilizer,[94] and a combination of these and other factors).[95] Most recently, it appears that at least part of the explanation for the change in Eprex's immunogenicity profile comes from a very unlikely source—a

---

[88] *See id.* at 3345.

[89] *See, e.g.*, Chamberlain, *supra* note 76, at 65; Arthur J. Chirino et al., *Minimizing the Immunogenicity of Protein Therapeutics*, 9 DRUG DISCOVERY TODAY 82, 82–84 (2004); E. Koren et al., *Immune Responses to Therapeutic Proteins in Humans—Clinical Significance, Assessment and Prediction*, 3 CURRENT PHARM. BIOTECH. 349, 351 (2002); Huub Schellekens, *Bioequivalence and the Immunogenicity of Biopharmaceuticals*, 1 NATURE REV. DRUG DISCOVERY 457, 458–59 (2002); Meenu Wadhwa et al., *Immunogenicity of Granulocyte-Macrophage Colony-Stimulating Factor (GM-CSF) Products in Patients Undergoing Combination Therapy with GM-CSF*, 5 CLINICAL CANCER RES. 1353, 1357, 1359–60 (1999). Even the same molecule tolerated at one time may create negative immunogenic reactions when readministered. *See* James E. Tcheng et al., *Abciximab Readministration: Results of the ReoPro Readministration Registry*, 104 CIRCULATION 870, 873–74 (2001) (reporting an increase in a certain type of antibody levels after biologic readministration, with profound thrombocytopenia as a result).

[90] *See, e.g.*, Chirino & Mire-Sluis, *supra* note 57, at 1386. For example, the manufacturer of an immune globulin product relocated its production to a different plant but continued to use the same manufacturing process. The manufacturer encountered problems producing the protein with uniform quality. *See* Amgen, Follow-on Biologics and Patient Safety 1 (Apr. 25, 2005) (unpublished article, on file with the author).

[91] *See* Amgen, *supra* note 90.

[92] *See* Steven S. Guest & Lee Levitt, Letter to the Editor, *Pure Red-Cell Aplasia Secondary to Antierythropoietin Antibodies*, 349 NEW ENG. J. MED. 2572, 2572 (2003).

[93] *See* Iain C. Macdougall, *Pure Red Cell Aplasia with Anti-Erythropoietin Antibodies Occurs More Commonly with One Formulation of Epoetin Alfa than Another*, 20 CURRENT MED. RES. & OPINION 83, 83 (2004).

[94] *See* Sabine Louet, *Lessons from Eprex for Biogeneric Firms*, 21 NATURE BIOTECH. 956, 956 (2003).

[95] *See* Janice M. Smiell, Letter to the Editor, *Equivocal Role of Micelles in Eprex Adverse Events*, 21 NATURE BIOTECH. 1265, 1265 (2003).

modification in the rubber stopper used in product packaging.[96] Yet even this explanation has been challenged, and the search continues for an explanation for the severe reactions associated with the biologic.[97] The Eprex case profoundly illustrates the obstacles inherent in predicting adverse clinical reactions from biologic manufacturing changes.[98] In this and other cases, severe clinical problems have occurred even though the research, testing, and technology for the affected biologics were well known to the regulatory authorities and to the industry, and full product reviews were performed.[99] Indeed, even the same manufacturer may experience difficulties in its products when conditions change.[100]

---

[96] *See* Katia Boven et al., *The Increased Incidence of Pure Red Cell Aplasia with an Eprex Formulation in Uncoated Rubber Stopper Syringes*, 67 KIDNEY INT'L 2346, 2350 (2005) (stating that the changed rubber stopper apparently interacted with a stabilizer to release immunogenic leachates into the product); *cf.* Letter from Frederick W. Telling, Vice President, Pfizer Inc., to the FDA 5–6 (Nov. 12, 2004), http://www.fda.gov/OHRMS/DOCKETS/DOCKETS/04n0355/04n-0355-c000006-01-vol1.pdf (describing changes in the immunogenicity profile of Somatonorm due to a bottle stopper change).

[97] *See* Huub Schellekens & Wim Jiskoot, Letter to the Editor, *Eprex-Associated Pure Red Cell Aplasia and Leachates*, 24 NATURE BIOTECH. 613, 614 (2006) (indicating that previous explanations of immunogenicity are flawed, and warning that "[w]ith the imminent advent of the era of biosimilar epoetins, understanding what caused the Eprex-associated [pure red cell aplasia]—where a minor change in formulation led to very serious adverse events—is imperative for patient safety"). *Compare* Basant Sharma et al., Letter to the Editor, *Reactions to Eprex's Adverse Reactions*, 24 NATURE BIOTECHNOLOGY 1199, 1200 (2006) (arguing that extant explanations for the cause of the pure red cell aplasias are suitable) *with* Schellekens & Jiskoot Response, 24 NATURE BIOTECHNOLOGY 1200, 1200 (2006) (criticizing the analysis of Sharma et al.).

[98] *See, e.g.*, Christopher Joneckis, Remarks at the Comparability Studies for Human Plasma-Derived Therapeutics Workshop 41 (May 30, 2002), *available at* http://www.fda.gov/cber/minutes/plasma053002.pdf ("[Eprex] is an example where making a change in a process that would not be predicted to cause this effect has resulted in the development of an antibody against the . . . molecule . . . .").

[99] There have been several other reports of immune reaction cases associated with biologics. *See, e.g.*, K. Oberg & G. Alm, *The Incidence and Clinical Significance of Antibodies to Interferon-α in Patients with Solid Tumors*, 10 BIOTHERAPY 1 (1997) (stating that the administration biologic interferon has resulted in the body creating antibodies to it, inhibiting its biological activity); Ibrahim C. Haznedaroglu et al., *Thrombopoietin as a Drug: Biologic Expectations, Clinical Realities, and Future Directions*, 8 CLINICAL APPLIED THROMBOSIS/HEMOSTASIS 193, 202 (2002) (stating that clinical trials of megakaryocyte growth factor were terminated when treatments were associated with thrombocytopenia, or deficiency of platelets). Genentech's experience with Raptiva, further illustrates the unpredictability involved in human reactivity to biologics. *See* DRUG INFO. ASS'N, FDA/DIA SCIENTIFIC WORKSHOP ON FOLLOW-ON PROTEIN PHARMACEUTICALS, BREAKOUT SESSION E: IMMUNOGENICITY STUDIES 61–65 (2005), *available at* http://www.fda.gov/cder/regulatory/follow_on/200502/200502_transcripts_0215bse2.pdf; (stating that, after transfer of the production from one location to another, and although two versions of the drug were apparently similar in analytic and animal testing, they resulted in different concentrations when used in humans); *see also* William Alpert, *Biotech's Next Challenge*, SMART MONEY, May 22, 2006, *available at* http://www.smartmoney.com/barrons/index.cfm?story=20060522 (describing the Raptiva experience).

[100] *See, e.g.*, Amgen, Follow-on Biologics and Patient Safety 1 (Apr. 25, 2005) (stating that a manufacturer of an immune globulin product who relocated its production to a different plant but continued to use the same manufacturing process nevertheless encountered problems producing the protein with uniform quality).

B. Diversion and Counterfeit Drugs

The diversion and counterfeiting of drugs are additional patient safety issues. Unfortunately, these issues are particularly significant for biologic drugs. Since biologics are exquisitely sensitive to environmental conditions during their manufacture, storage, and transport, any weakness in the supply chain of biologics is an important safety concern.[101] Further, because biologics are injectable, they are provided as clear fluids in standard vials, making them easily faked.[102] Patients with such conditions as cancer, AIDS, and other life-threatening illnesses,[103] who are often the least able to weather the negative consequences associated with deviations, risk obtaining adulterated, illegitimate, or improperly stored and transported drugs. Hence, it is important to understand the risks of counterfeits and diversion for patient welfare and public health purposes.

Most of the prescription drugs in the United States reach community pharmacies, hospitals, and nursing homes on a direct line from the manufacturer through one of three large wholesalers: AmerisourceBergen, Cardinal Health, or McKesson Corporation, which are sometimes known as the "Big Three."[104] However, a small but substantial portion of the U.S. drug supply reaches consumers after passing through the "gray market" of pharmaceutical trade.[105] Unfortunately, counterfeiters have exploited this secondary market to divert legitimate supplies and introduce their counterfeit versions.[106]

---

[101] *See, e.g.*, FDA, Impact of Severe Weather Conditions on Biological Products, http://www.fda. gov/cber/weatherimpact.htm (last visited Apr. 17, 2007) ("Most biological products require specific storage conditions, as indicated in the product labeling, to maintain their safety, purity, and potency."); Michael J. Akers et al., *Formulation Development of Protein Dosage Forms*, *in* 14 DEV. AND MANUFACTURE OF PROTEIN PHARMACEUTICALS 47, 49 (Steven L. Nail & Michael J. Akers eds., 2002) (stating that biologics degrade when in storage using different methods of degradation, further complicating any prediction of stability).

[102] *See* Bryan A. Liang, *Fade to Black: Importation and Counterfeit Drugs*, 32 AM. J.L. & MED. 279, 283–84 (2006) [hereinafter Liang, *Fade to Black*].

[103] The major biologic cancer drugs Rituxan, Herceptin, Avastin and Erbitux saw sales grow by fifty-seven percent in 2006. *See Biologic Cancer Drugs Fastest Growing Class of Biologics in 2006*, BIOLOGICS DRUG REP., Feb. 2007, http://www.biologicdrugreport.com/News/news-022607.htm. Meanwhile erythropoietin, which is used extensively to battle related problems with anemia, weighed in as the largest class of biologics. *See id.* Similarly, human growth hormone, a biologic, is an important tool in the battle against AIDS. *See, e.g.*, U.S. DEP'T OF HEALTH AND HUMAN SERVS., *Somatropin*, AIDSINFO, http://aidsinfo.nih.gov (search for "somatropin" and click on "Somatropin" drug factsheet hyperlink) (last visited Apr. 17, 2007).

[104] *See, e.g.*, KATHERINE EBAN, DANGEROUS DOSES: A TRUE STORY OF COPS, COUNTERFEITERS, AND THE CONTAMINATION OF AMERICA'S DRUG SUPPLY 89–90 (2006).

[105] *See* Liang, *Fade to Black*, *supra* note 102, at 287 (estimating that approximately ninety percent of the pharmaceutical trade involves direct sales from large wholesalers to pharmacies and other facilities, while the gray market makes up the other ten percent of sales).

[106] *See id.* All of the "Big Three" have been affected by counterfeit drug sales. *See* Mary Pat Flaherty & Gilbert M. Gaul, *Lax System Allows Criminals to Invade the Supply Chain*,

The gray market is characterized by multiple transactions between both large and small wholesalers and various providers.[107] If there is a short-term shortage in supply, the "Big Three" may buy back products from smaller secondary wholesalers.[108] In addition, many products are sold in bulk and then sent to repackagers via middlemen to receive their direct-to-consumer packaging.[109] These middlemen then move the products to places where they will sell quickly due to impending expiration dates.[110] Because of cash flow exigencies and a healthy arbitrage market, there is a constant flow of stock among pharmacies and between pharmacies and secondary wholesalers.[111] This system creates opportunities during which the drugs could be negligently damaged or deliberately faked: they may be diverted; they may be subjected to substandard conditions during storage and transport, ruining their active ingredients; or they may be replaced by or salted with counterfeits at some point during this process.[112]

The defects in the supply chain and the risks associated therewith are exacerbated by the limited regulation of gray-market sales. States play a large role in regulating the distribution, repackaging, dispensing, and return of pharmaceuticals by purchasers,[113] while the federal government predominantly regulates the approval and manufacturing of chemical drugs and biologics.[114] Because both systems suffer from inadequate staff and lack of coordination, the regulatory safety net of pharmaceutical distribution contains gaping holes.[115] Nefarious individuals take advantage of the holes and smokescreens provided by the gray market to capitalize on the highly lucrative pharmaceutical market by diverting drug supplies and introducing tainted medicine into the supply chain.[116]

Other risks increase the potential that patients will not receive the appropriate drug. If state governments' Internet purchasing policies allow purchases of Canadian and European biologics, as they currently do for chemical medicines, significant patient safety risks will likely result. These kinds of programs pose severe safety risks to patients because trade across Euro-

---

WASH. POST, Oct. 22, 2003, at A1, *available at* http://www.washingtonpost.com/ac2/wp-dyn/A61473-2003Oct21?.

[107] *See* EBAN, *supra* note 104, at 39.

[108] *See id.* at 89–91.

[109] *See* Liang, *Fade to Black*, *supra* note 102, at 288.

[110] *See id.*

[111] *See id.*

[112] *See id.* at 285 ("Salting occurs when legitimate drugs . . . are mixed or 'salted' with fake versions of the drug. In this way, even if patients, pharmacists, or government authorities are attempting to detect counterfeits, these fake drugs may elude detection due to a legitimate sample or fake with the active molecule being pulled for testing.").

[113] *See id.*

[114] *See id.*

[115] *See, e.g.*, John Theriault, *Counterfeit Pharmaceuticals: Understanding the Threat*, 9 J. BIOLAW & BUS. 14, 17 (2006).

[116] *See* Liang, *Fade to Black*, *supra* note 102, at 288; *see generally* EBAN, *supra* note 104 (providing an overview of the complexities of the gray market and its impact upon counterfeit drug distribution and sale).

pean borders ("parallel trade") in markets such as the E.U. is highly unregulated,[117] enables criminals to sell diverted and fake drugs virtually undetected and without accountability,[118] is not generally covered by health and safety regulatory oversight,[119] and enables purveyors of poor-quality or fake drugs to easily obtain parallel-trade licenses through fraud.[120] The result is that the weaknesses of the E.U.'s drug safety programs can result in damage to European biologics imported into the United States.

Unfortunately, these are not theoretical concerns. Cases of fake and diverted biologics, with resulting adverse consequences, have been reported in the United States. For example, Maxine Blount, a breast cancer patient who needed Procrit, a version of erythropoietin, was sold a diluted version of the drug.[121] Tim Fagan, a sixteen-year-old liver transplant patient, was also the victim of fake erythropoietin.[122] After buying the drug Epogen from his local CVS pharmacy, he was injected with the material he thought was the drug and experienced severe spasms.[123] Unfortunately, he continued to inject the counterfeit drug for several weeks before it was discovered to be a counterfeit.[124]

Another investigation discovered diluted erythropoietin, but only ten percent of the registered shipment was recovered, while ninety percent presumably remained on the market.[125]

Estimates are that as many as 25,000 cancer and HIV patients were treated with the ineffective fake biologic but never knew.[126] The problem is not limited to the United States; fake forms of the drug were also reportedly shipped into Canada.[127]

---

[117] *See* Bryan A. Liang, *Parallel Trade in Pharmaceuticals: Injecting the Counterfeit Element into the Public Health*, 31 N.C. J. INT'L L. & COM. REG. 847, 854–57 (2006) [hereinafter Liang, *Parallel Trade*] (describing the presence of counterfeit fake drugs in parallel trade in the E.U.).

[118] *See id.*

[119] *See* Liang, *Fade to Black*, *supra* note 102, at 297.

[120] *See* Liang, *Parallel Trade*, *supra* note 117, at 856 n.22.

[121] *See* EBAN, *supra* note 104, at 16, 334. Maxine Blount's family mourned the loss of quality of life due to the inefficacy of the diluted erythropoietin, and some members directly attributed her death to it. *See id.* at 113.

[122] *See* Melinda Murphy, *Targeting Phony Pharmaceuticals*, CBS NEWS, May 9, 2005, http://www.cbsnews.com/stories/2005/05/09/earlyshow/contributors/melindamurphy/printable693799.shtml; *see also* EBAN, *supra* note 104, at 333 (outlining the investigation of fake Epogen, including that taken by Tim Fagan).

[123] *See* Murphy, *supra* note 122.

[124] *See id.*

[125] *See* Flaherty & Gaul, *supra* note 106, at A1.

[126] *See id.* For another example of criminal drug dilution, see Bob LaMendola, *Two Florida Men Given Prison Time, House Arrest in Counterfeit Drug* Case, S. FL. SUN-SENTINEL, Aug. 30, 2003, at 5B; Chris Hansen, *Inside the World of Counterfeit Drugs*, NBC DATELINE, June 4, 2006, http://www.msnbc.msn.com/id/13137839 (reporting sentences against counterfeiters who had sold a form of a biological drug diluted with bacteria-contaminated water).

[127] *See* Matthew Herper, *Bad Medicine*, FORBES, May 23, 2005, at 202, 204, *available at* http://www.forbes. com/home_europe/free_forbes/2005/0523/202.html ("There are tens of thousands of Web sites that sell prescription drugs of unknown origin to Americans illegally

Other fake biologics have also been discovered. A seventeen-year-old boy who needed growth hormone was injected with what turned out to be insulin; a pharmacist discovered the counterfeit batches.[128] An HIV patient who purchased and injected a purported anti-wasting drug for debilitating weight loss associated with the infection reported that "it burned like hell and raised a knot the size of a quarter."[129] Upon analysis, the product turned out to be a steroid.[130]

A Florida grand jury initiated an investigation into the abuses associated with fake drugs and porous protections.[131] It found that:

> Many of the wholesalers in Florida's market are unqualified, inexperienced, irresponsible and incompetent to properly handle, store or deal in pharmaceuticals. Some even have criminal records, though how many is impossible to know since Florida only does minimal background checks before issuing wholesale permits. Any drugs that come into the possession of these wholesalers, whether acquired legally or illegally, are likely to become tainted due to improper handling and storage.[132]

Importantly, they also noted that the drugs most targeted by criminals were high-cost drugs used to treat patients suffering from cancer or AIDS or undergoing transplants—i.e., biologics.[133] Unfortunately, as the grand jury acknowledged, patients who need these biologic medicines are some of the most vulnerable, and are unable to withstand the burden of fake medicines.[134]

As startling as these examples of patients encountering counterfeit medicines are, what is of deeper concern is that these are the very few cases actually uncovered by patients, medical providers, and authorities.[135] Yet counterfeit medications are not manufactured one at a time. For every report of a counterfeit medication, there are countless other counterfeits in that same and other batches circulating throughout the supply chain.[136] Exacerbating the problem is that providers and patients often do not suspect or

---

. . . . A Web site based in Canada may get its products from India or China, or may traffic in counterfeits . . . . Argentinean export records seem to show tens of thousands of doses of drugs, including knockoffs of erythropoietin and the cancer drugs Eloxatin and methotrexate, making their way from Argentina into Canada.").

[128] *See id.*

[129] *See id.*

[130] *See id.*

[131] *See* SUPREME COURT OF THE STATE OF FLA., FIRST INTERIM REPORT OF THE SEVENTEENTH STATEWIDE GRAND JURY (2003), *available at* http://myfloridalegal.com/grandjury17.pdf.

[132] *Id.* at 2–3.

[133] *See id.* at 3.

[134] *See id.*

[135] *See* Thomas T. Kubic and Sebastian J. Mollo, *Pharmaceutical Counterfeiting Trends: Understanding the Extent of Criminal Activity*, 9 (4) J. BIOLAW & BUS. 19, 19 (2006).

[136] *See id.*

consider counterfeit, diverted, or adulterated medicines when therapeutic failure occurs. Physicians and nurses generally attribute negative outcomes to human variation or to the underlying ailment, since the affected patients in many cases are frail, elderly, or very ill.[137] Further, the packaging of the counterfeit product can be of deceptively high quality, with the counterfeit product appearing identical to the actual medicine.[138] Patients and their families also may not know that they have been victimized by counterfeit products due to lack of clinical knowledge and unsophisticated counterfeit detection systems. Moreover, providers rarely ask where drugs were purchased, and patients may be reluctant to disclose that medicines were bought from a suspect source.[139]

Detecting counterfeit, adulterated, and diverted biologics is an immense challenge. Medication packaging is thrown away; the patient's own body disposes of the material by metabolizing it; and lab tests are normally not available to expose counterfeit medicines.[140] This situation makes it exceedingly problematic to investigate forensically where, how, and what changes occurred.[141] Such challenges are illustrated by the recent frustrating investigations into counterfeit drug deaths in Canada.[142] In that case, seven elderly individuals who filled their prescriptions for the heart drug Norvasc at the same pharmacy later died; however, because there was little forensic evidence, the Canadian authorities struggled to build a case against the defendants.[143]

Recognizing the importance of the issue, the National Association of Boards of Pharmacy has warned providers and patients of the risks of counterfeit and diverted drugs by issuing a list of susceptible fake drug products.[144] Although biologics make up only a small fraction of the medicines prescribed every year,[145] ten of the thirty-two products listed as vulnerable to adulteration, counterfeiting, and diversion are biologics.[146] Currently, security measures aimed at addressing the problems of counterfeiting and diversion through track and trace or technologically based authentication

---

[137] *See* Liang, *Fade to Black*, *supra* note 102, at 289.

[138] *See id.* at 290.

[139] *See id.* (referring to purchases over the Internet or from a foreign country). This reluctance may be due to embarrassment or stigma associated with a particular disease state, or frustration with access to the care desired. *See* Jim Thompson, Chief Executive, Ctr. for Mental Health, Address at the Centre for Reform: Stigma? What Stigma? (Sept. 6, 2005), *available at* http://www.ehiprimarycare.com/comment_and_analysis/index.cfm?ID=100.

[140] *See* Liang *Fade to Black*, *supra* note 102, at 289 (explaining that there is currently no mechanism in place for a patient or provider to easily have a medication tested to determine whether it is counterfeit).

[141] *See id.*

[142] *See* Luma Muhtadie, *Fake-Drug Case: Huge Forensic Challenge*, HAMILTON SPECTATOR (Ontario), July 16, 2005, at A01.

[143] *See id.*

[144] *See* NAT'L ASS'N OF BDS. OF PHARMACY, NATIONAL SPECIFIED LIST OF SUSCEPTIBLE PRODUCTS (2004), *available at* http://www.nabp.net/ftpfiles/NABP01/List.pdf.

[145] *See* Herper & Kang, *supra* note 2.

[146] *See id.* at 3.

methods have significant weaknesses, emphasizing the need for a heightened focus on safety.[147]

IV. THE REGULATORY REGIME

A. The United States

1. Chemical Medicines

a. New Drug Applications

The evaluation and regulation of new chemical drugs in the United States is based on the New Drug Application ("NDA") process under the FDCA.[148] Every new chemical drug is subject to the NDA premarketing approval process before it may be sold in the United States.[149]

The NDA process is the means by which chemical drug sponsors, usually brand name pharmaceutical companies, formally request FDA approval of a new chemical entity for sale and marketing in the United States.[150] The NDA includes data gathering by sponsors through animal studies and human clinical trials of the chemical formula.[151]

The fundamental goal of the NDA process is to provide enough information to the FDA for it to make the following key assessments: (1)whether the drug is safe and effective for its proposed uses; (2) whether the drug's benefits outweigh its risks; (3) whether the drug's proposed labeling is appropriate; and (4) whether the methods used in manufacturing the drug are adequate to assure its identity, strength, quality, and purity.[152]

---

[147] *See* Bryan A. Liang, *Structurally Sophisticated or Lamentably Limited? Mechanisms to Ensure Safety of the Medicine Supply*, 16 ALB. L.J. SCI. & TECH. 483, 499 (2007) (reviewing authentication and track and trace technologies for pharmaceuticals and concluding that they are not ready for broad application due to the unreliability of the technology and the ease of counterfeiting the results). Recently, a wholesaler completed a pilot study using radio-frequency identification ("RFID") tags and found significant variation in readability. *See* Emile Reymond, *Cardinal's RFID Pilot Shows Promise and Perils*, IN-PHARMATECHNOLOGIST.COM, Nov. 16, 2006, http://www.in-pharmatechnologist.com/news/ng.asp?n=72120-cardinal-health-rfid-drug-counterfeit. The high failure rate led the wholesaler to conclude that, despite the promise of these technologies, further development was necessary. *See id.*

[148] *See* 21 U.S.C. § 355(b)(1) (2000 & Supp. IV 2004); 21 C.F.R. § 314 (2006); *see also* Leuenberger-Fisher, *supra* note 49, at 391–92 (providing an overview of the historical shifts of responsibility for reviewing chemical and biological drugs). Currently, approval of most biologics is the responsibility of the FDA's Center for Drug Evaluation and Research ("CDER"), with the notable exceptions of insulin and human growth hormone, which are within the purview of the FDA Center for Biologics Evaluation and Research ("CBER"). *See* 68 Fed. Reg. 38,067 (2003).

[149] *See* CDER, New Drug Application (NDA) Process, http://www.fda.gov/Cder/regulatory/applications/nda.htm (last visited Apr. 17, 2007).

[150] *See* 21 U.S.C. § 355.

[151] *See id.* § 355(i)(2)(B); *see also* 21 C.F.R. § 314.50 (stating the contents of an NDA).

[152] *See* 21 C.F.R. § 314.127. In general, NDA applicants examine their compounds through several steps before submitting a completed NDA. First, applicants perform preclinical animal

An NDA is geared toward giving the applicant an opportunity to provide the FDA with the chemical entity's "story," detailing the drug's ingredients; the results of the animal studies; the findings of the clinical tests and trials; the pharmacological characteristics of the drug; and how it is manufactured, processed, and packaged.[153] This story also includes the bioavailability of the drug, or the fraction of the administered drug that is absorbed into the systemic circulation.[154] Bioavailability data from the NDA sets out a series of profiles showing different levels of exposure to the chemical drug and can be used as a benchmark for subsequent formulation changes by the applicant. It also may be useful as a reference for future bioequivalence studies by generic firms.[155]

From the FDA's perspective, a chemical drug product is one that is comparable to a brand name chemical drug product in dosage form, strength, labeling, route of administration, quality, performance characteristics, and intended use.[156] The focus of the assessment is upon the absolute chemical identity of the drug.[157] All approved products, both NDA

---

studies focusing upon the molecular entity's pharmacology and toxicology to establish that it will not present unreasonable risks in early-stage human clinical studies. *See* C.F.R. 21 § 312.23; *see also* Leuenberger-Fisher, *supra* note 49, at 392. These studies are submitted to the FDA for its approval under an Investigational New Drug ("IND") application. *See* 21 C.F.R. § 312.23. If the FDA approves, the applicant must conduct three phases of human clinical trials. *See id.* Phase I trials are small studies generally using healthy volunteers to assess metabolic and pharmacologic activity of the drug when administered to humans, and to provide information to design Phase II trials. *See id.* § 312.21. Phase II trials are controlled clinical studies with a limited number of patients to assess in a preliminary sense whether the drug is effective for a particular indication of a specific disease. *See id.* If there appears to be evidence of efficacy from the Phase II studies, Phase III trials, which focus upon collecting additional data on the safety and efficacy of the drug, will be performed. *See id.* In combination, Phase II and Phase III trials provide information on the relative risks and benefits of the drug and data on appropriate dosages and drug labeling requirements. *See* CDER, FDA, Phase 3 Clinical Studies, http://www.fda.gov/cder/handbook/phase3.htm (last visited Apr. 17, 2007). After this process has been completed, the applicant may submit an NDA, which includes all IND and clinical trial information, as well as information on product manufacturing that comports with standardized FDA requirements (i.e., Current Good Manufacturing Practice). *See* 21 C.F.R. § 314.50.

[153] *See* 21 C.F.R. § 314.50.

[154] *See id.* Bioavailability is defined by the federal regulations as:

> the rate and extent to which the active ingredient or active moiety is absorbed from a drug product and becomes available at the site of action. For drug products that are not intended to be absorbed into the bloodstream, bioavailability may be assessed by measurements intended to reflect the rate and extent to which the active ingredient or active moiety becomes available at the site of action.

Id. § 320.1(a); *see also infra* notes 166–172 (discussing the use of bioavailability testing in the ANDA approval process).

[155] *See infra* note 164–172 and accompanying text (discussing bioequivalence and bioavailability testing in the ANDA context).

[156] *See* 21 C.F.R. § 314.92 (2006).

[157] By comparison, characterization of biologics is based on an array of manufacturing and processes-oriented measurements. *See* 21 C.F.R. 25.31 (2006); *see infra* notes 187–201 and accompanying text (reviewing biologic medicine approval).

brand name and ANDA generic, are listed in the FDA's *Approved Drug Products with Therapeutic Equivalence Evaluations*, also known as the *Orange Book*.[158]

If an NDA is approved, the brand name company may then market its product for the approved indication.[159] Further, under the HWA, an NDA applicant receives five years of "data exclusivity" in which no other manufacturer may use its safety and efficacy data to support another (usually generic) drug application.[160]

b. Abbreviated New Drug Application

In contrast to NDAs, which are used for new drug development by brand-name firms, the Abbreviated New Drug Application ("ANDA") focuses upon generic drug development. In 1984, the HWA added section 505(j) to the FDCA, creating the ANDA process.[161] The ANDA process eliminated the duplicative tests that had been required for a generic drug to obtain approval from the FDA, allowing the generic drug to gain approval based upon the earlier tests submitted to the FDA for approval of the brand name product.[162] ANDA applications are termed "abbreviated" because they are generally not required to include preclinical (animal) or clinical (human) data to establish safety and effectiveness.[163] Instead, generic applicants must show that their chemical product is "bioequivalent" to the referenced brand name chemical product, meaning that it acts in the same way as the brand name form of the drug.[164] Because chemical drugs are easily characterized, assessed, and manufactured to high levels of purity for comparison with the originator's brand name drug, the ANDA process under

---

[158] FDA, APPROVED DRUG PRODUCTS WITH THERAPEUTIC EQUIVALENCE EVALUATIONS, *available at* http://www.fda.gov/cder/ob/default.htm [hereinafter "ORANGE BOOK"].

[159] *See* 21 C.F.R. § 314.

[160] *See* 21 U.S.C. § 355(j) (2000 and Supp. IV 2004); 21 C.F.R. § 314.108. Note that the United States has one of the most limited data exclusivity periods in the world. *See* Jonathan de Ridder, *Data Exclusivity: Further Protection for Pharmaceuticals*, FINDLAW AUSTRALIA, June 2003, http://www.findlaw.com.au/articles/default.asp?task=read&id=9200&site=GN.

[161] *See* Pub. L. No. 98-417, 98 Stat. 1585 (codified as 21 U.S.C. § 355(j) (2000 and Supp. IV 2004)).

[162] *See* CONG. BUDGET OFFICE, HOW INCREASED COMPETITION FROM GENERIC DRUGS HAS AFFECTED PRICES AND RETURNS IN THE PHARMACEUTICAL INDUSTRY xii, at (1998).

[163] *See* 21 C.F.R. 314.94.

[164] *See* 21 U.S.C. § 355(j)(8)(B) (2000 & Supp. III 2003). Bioequivalence is defined by the federal regulations as:

> the absence of a significant difference in the rate and extent to which the active ingredient or active moiety in pharmaceutical equivalents or pharmaceutical alternatives becomes available at the site of drug action when administered at the same molar dose under similar conditions in an appropriately designed study.

21 C.F.R. § 320.1(e).

section 505(j) generally presumes approval unless the application does not include certain enumerated information.[165]

For the key bioequivalence showing, a common method by which generic firms demonstrate this requirement is through measuring bioavailability.[166] Bioavailability data provide an estimate of the relative fraction of the administered drug that is absorbed into the systemic circulation.[167] In addition, bioavailability studies yield other important pharmacokinetic ("PK")[168] information related to the distribution, elimination, and effects of nutrients on absorption of the drug.[169]

To measure ANDA drug bioavailability so as to establish bioequivalence, PK measurements that assess release of the drug substance from the drug product over time into the patient's circulatory system are often used.[170] This bioavailability methodology usually measures the time it takes the generic drug to reach the bloodstream in twenty-four to thirty-six healthy volunteers.[171] This provides the rate of absorption—i.e., the bioavailability—of the generic chemical drug, which can then be compared to the bioavailability of the brand name chemical drug on a reasonable statistical grounding.[172] For bottom line ANDA purposes, the generic version must deliver the same amount of active ingredients into a patient's bloodstream in the same amount of time as the brand name form.

In addition, beyond focusing on bioequivalence and bioavailability data, the FDA will also review an ANDA application for chemical and microbiological data as well as manufacturing plant inspection and drug labeling information when assessing the application for adequacy and ultimately for potential approval.[173] However, this process is relatively straightforward, and generally does not involve any new clinical trials.[174]

---

[165] *See* 21 U.S.C. § 355(j); 21 C.F.R. § 314.

[166] *See* 21 C.F.R. § 314.50.

[167] *See* FDA, GUIDANCE FOR INDUSTRY: BIOAVAILABILITY AND BIOEQUIVALENCE STUDIES FOR ORALLY ADMINISTERED DRUG PRODUCTS—GENERAL CONSIDERATIONS DRAFT GUIDANCE (2002), *available at* http://www.fda.gov/cder/guidance/4964dft.htm [hereinafter FDA, GUIDANCE FOR INDUSTRY].

[168] Pharmacokinetics ("PK") is the branch of pharmacology that studies the time course of drugs within the body, and particularly involves "the study of the bodily absorption, distribution, metabolism, and excretion of drugs." MERRIAM-WEBSTER'S MEDICAL DESK DICTIONARY 536 (1993).

[169] *See id.*

[170] *See id.*

[171] *See* FDA, ABBREVIATED NEW DRUG APPLICATION (ANDA) PROCESS FOR GENERIC DRUGS, http://www.fda.gov/cder/regulatory/applications/ANDA.htm (last visited Apr. 17, 2007).

[172] *See* FDA, GUIDANCE FOR INDUSTRY, *supra* note 167.

[173] *See id.*

[174] *See* Leuenberger-Fischer, *supra* note 49, at 396 ("[D]etermining what studies should be done in support of a chemical drug product subject to a change in manufacturing from a listed drug product is generally straightforward and rarely requires even limited clinical trials. If it looks the same and acts the same in pharmacokinetic studies, it is bioequivalent.").

The HWA also attempted to create a sensitive balance of incentives between intellectual property concerns of the originator manufacturers and the goal of increasing access to cheaper and safe generic chemical drug forms.[175] These concerns are evident in the different components of the HWA. For example, an ANDA applicant must acknowledge the existence of any patents listed by an NDA holder during the approval process.[176] Another exclusivity provision delays generic entry for three years when a brand name firm's application for new clinical use is approved but requires further testing.[177] On the other hand, generic companies are not required to generate their own preclinical and clinical trial data and may rely on the data of an NDA.[178]

An important example of this balance is through the HWA's provisions that encourage generic firms to challenge invalid patents to speed generic entry and provides incentives for being the first generic firm to do so.[179] This provision rewards the generic firm for bringing a cheaper product to the market earlier than if generic firms had waited until patent expiration. However, because early entry by generic manufacturers could lead to significant economic losses for the relevant brand name manufacturer,

---

[175] *See* Andrx Pharms. v. Biovail Corp., 276 F.3d 1368, 1370–71 (Fed. Cir. 2002) (noting that the HWA is a compromise between encouraging innovative research by brand-name companies and approving low-cost generic products created by generic firms).

[176] *See* 21 C.F.R. § 314.95. The ANDA applicant must make one of four contentions regarding the status of each existing patent relevant to the product. The application must state one of the following: (1) that the relevant patent information has not been filed ("paragraph (I)"); (2) that the patent has expired ("paragraph (II)"); (3) that the patent will expire on a particular date ("paragraph (III)"); or (4) that the patent is invalid or will not be infringed by the drug for which approval is being sought ("paragraph (IV)"). 21 U.S.C. § 355(j)(2)(A)(vii)(I)–(IV).

A certification under paragraphs (I) or (II) allows the FDA to approve the ANDA immediately, if it is otherwise eligible. *See id.* § 355(j)(5)(B)(i). A certification under paragraph (III) indicates that the ANDA may be approved on the patent expiration date. *See id.* § 355(j)(5)(B)(ii). Applicants may enter the market to sell their generic product on the dates specified by their approval. For discussion of paragraph (IV) certifications, see *infra* note 184 and accompanying text.

[177] 21 C.F.R. § 314.108.

[178] *See* 21 C.F.R. § 314.94.

[179] *See* 21 C.F.R. § 314.52; *see also* Aidan Hollis & Bryan A. Liang, *An Assessment of the Effect of Authorized Generics on Consumer Prices*, 9 J. BIOLAW & BUS. (forthcoming) (2006). When an NDA applicant applies based on a paragraph (IV) certification, challenging the legitimacy of a brand name firm patent or claiming to have developed a noninfringing means of producing a drug under patent, the HWA provides that the first generic ANDA filer obtains a 180-day period during which it is the only authorized generic seller on the market. *See* 21 C.F.R. § 355(j)(5)(B)(iii)–(iv). This marketing exclusivity period is critical, because during this time the generic firm typically is able to recoup much of its investment before the entry of other generic competitors drives the price of the drugs much lower and greatly reduces profits. *See, e.g.*, *The Generic Drug Maze: Speeding Access to Affordable, Life Saving Drugs: Hearing Before the S. Spec. Comm. on Aging*, 109th Cong. 142–43 (2006) (statement of Heather Bresch, Senior Vice President, Mylan Labs) [herinafter *Lifesaving Drugs Hearing*] (stating that the entry of additional generic competition "can all but eliminate the financial benefit for the first generic filer"); David Reiffen & Michael R. Ward, *Generic Industry Dynamics*, Feb. 2002, at 11, 35, *available at* http://www.ftc.gov/be/healthcare/wp/11_Reiffen_WP248_GENERICDRUGINDUSTRYDYNAMICS.pdf.

the HWA balanced this provision with robust protections for brand name firms.[180]

In this process, the generic firm may submit an ANDA application claiming an invalid brand name patent; in turn, the brand name firm may challenge the generic firm's ANDA by filing a claim that the generic product violates the brand name firm's patent rights.[181] If the brand name firm files a patent infringement action within forty-five days of notification, the FDA may not approve the ANDA for thirty months, or until the patent dispute has been resolved, whichever is sooner.[182] Yet this thirty-month stay, its associated patent litigation, and the generic product development process are expensive events for generic firms to weather.[183] However, this is the primary rationale as to why the generic firm is granted a marketing exclusivity period if they are the first to successfully challenge a listed patent.[184]

The endpoint of the ANDA application submitted by generic chemical drug companies is review and potential approval of a generic drug product. If approved as "bioequivalent" to the originator drug,[185] the product is deemed "AB-rated," meaning that the generic and brand name chemical products are bioequivalent and substitutable.[186] In this case, if intellectual

---

[180] An ANDA applicant who files a paragraph (IV) certification with the FDA must notify the patent owner (and the NDA holder for the listed drug, usually the brand name firm) that it has filed an ANDA containing a patent challenge. *See* CDER, FDA, Small Business Assistance: 180-Day Generic Exclusivity, http://www.fda.gov/cder/about/smallbiz/generic_exclusivity.htm (last visited Apr. 17, 2007). This notice must include the factual and legal basis for why the relevant patent is not valid or will not be infringed. *See id.*

[181] *See* 21 U.S.C. § 355(j)(5)(B)(iii) (2003).

[182] *See id.*

[183] The FTC has found that the thirty-month stay process has been abused by brand name firms. *See* FTC, GENERIC DRUG ENTRY PRIOR TO PATENT EXPIRATION: AN FTC STUDY ii (2002), *available at* http://www.ftc.gov/os/2002/07/genericdrugstudy.pdf. According to its recent study, one of the most common ways that patent-holding companies were able to further delay the market entry of generic drugs is through multiple patent listings in the Orange Book. *See id.* The study found that brand name firms had listed patents in the Orange Book after an ANDA had been filed, requiring the generic filer to recertify, and resulting in another patent litigation suit filed by the brand name firm, triggering another thirty-month stay. *See id.* Recognizing these concerns, in 2003 Congress limited brand name firms to only one thirty-month stay under the Medicare Modernization Act. *See* Pub. L. No. 108-173, 117 Stat. 2066 (2003) (codified at 26 U.S.C. §§ 139A, 223, 4980G (2000 & Supp. IV 2004) and in scattered sections of 42 U.S.C. (2000 and Supp. IV 2004)).

[184] Overall, according to the FDA, "an ANDA applicant whose ANDA contains a paragraph (IV) certification is protected from competition from subsequent generic versions of the same drug product for 180-days . . . ." CDER, FDA, GUIDANCE FOR INDUSTRY: 180-DAY GENERIC DRUG EXCLUSIVITY UNDER THE HATCH-WAXMAN AMENDMENTS TO THE FEDERAL FOOD, DRUG, AND COSMETIC ACT 2 (1998), *available at* http://www.fda.gov/CDER/GUIDANCE/2576fnl.pdf (interpreting the 180-day generic exclusivity provision, 21 U.S.C. § 355(j)(5)(B)(iv)).

[185] *See* 21 U.S.C. § 355(j)(8)(B); 21 C.F.R. §§ 320.1(e), 314.94(a)(7); *see also supra* notes 164–172 and accompanying text (describing bioequivalence).

[186] *See* 21 C.F.R. § 320.1(a); *see also supra* notes 164–167 and accompanying text (describing bioequivalence and bioavailability); "ORANGE BOOK," *supra* note 158.

property rights allow, the generic firm may manufacture and market the generic chemical drug product in the United States.

2. *Biologic Medicines*

The regulatory process for biologic medicines is more complicated.[187] Currently, biologics are regulated under both the FDCA[188] and the Public Health Service Act ("PHSA").[189] The reason for this dual regime is that the FDCA governs "drugs" while the PHSA governs "biological products."[190]

"Drugs," according to the FDCA, are "articles intended for use in the diagnosis, cure, mitigation, treatment, or prevention of disease . . . and . . . articles (other than food) intended to affect the structure or any function of the body of man or other animals."[191] Biologics are clearly drugs and therefore subject to the FDCA. A "biological product" is defined as "a virus, therapeutic serum, toxin, antitoxin, vaccine, blood, or blood component or derivative, allergenic product, or analogous product . . . applicable to the prevention, treatment or cure of a disease or condition of human beings."[192] Biologic medicines are also clearly biological products and therefore subject to the PHSA as well.

The FDA regulates and approves, as drugs, relatively simple biologic molecules representing natural source protein products such as insulin and growth hormone.[193] These biologic products are generally approved under a parallel NDA application process delineated by section 505(b)(2) of the FDCA.[194] This section allows the FDA to approve a drug under NDA au-

---

[187] Nevertheless, regulating biologics is not a new phenomenon. Biologics have been regulated for over 100 years. *See* Biologics Control Act, Pub. L. No. 57-244, 32 Stat. 328 (1902) (codified as amended at 42 U.S.C. § 262 (2000 & Supp. IV 2004)). For a review of the history of biologics regulation, see David M. Dudzinski, *Reflections on Historical, Scientific, and Legal Issues Relevant to Designing Approval Pathways for Generic Versions of Recombinant Protein-Based Therapeutics and Monoclonal Antibodies*, 60 FOOD & DRUG L.J. 143, 145–79 (2005).

[188] *See* 21 U.S.C. § 355 (b)(1)(2000 & Supp. III 2003).

[189] 42 U.S.C. § 262 (2000 & Supp. IV 2004); *see also* 21 C.F.R. § 601.2 (2006) (detailing regulatory requirements).

[190] *See* 42 U.S.C. § 262 (2000 & Supp. IV 2004); 21 C.F.R. § 601.2 (2006).

[191] 21 U.S.C. § 321(g)(1) (2000).

[192] PHSA § 351(i), 42 U.S.C. § 262(i)(2004); 21 C.F.R. § 600.3 (2004).

[193] *See The Law of Biologic Medicine: Hearing Before the S. Comm. on the Judiciary*, 108th Cong. 5 (2004) (statement of Lester M. Crawford, Acting Commissioner of Food and Drugs, Department of Health and Human Services) [hereinafter *Biologic Medicine Hearing*].

[194] *See* FDCA § 505(b)(2), 21 U.S.C. § 355(b)(2) (2000 & Supp. IV 2004). Any drug may be approved under section 505(b)(2) authority, and any biologic can be approved under traditional NDA processes. Originally, the FDA contemplated that drugs with changes only in dose, strength, formulation, route of administration, regimen, combination products, and differing chemical forms of already approved active ingredients would be suitable for the section 505(b)(2) process. *See* FDA, GUIDANCE FOR INDUSTRY: APPLICATIONS COVERED BY SECTION 505(B)(2), at 4–6 (Oct. 1999) (draft), *available at* http://www.fda.gov/cder/guidance/2853dft.pdf (interpreting 21 U.S.C. § 505(b)(2)). However, many biologics

thority if supported by scientific literature or by a previous FDA finding that a drug is safe and effective.[195] Note that under the section 505(b)(2) mechanism, which is an NDA process, the drug is literally different from the brand name or originator product and therefore is not a generic.[196] Indeed, reflecting this difference, these drugs may require additional human studies prior to FDA approval.[197] These additional actions are needed because bioequivalence studies like those used in chemical ANDA applications are not generally applicable to biologics.[198]

---

regulated as drugs are necessarily approved under this section, including follow-on products. *See infra* notes 195–198 and accompanying text (describing the 505(b)(2) process and the biologics approved under it).

[195] *See* FDCA § 505(b)(2), 21 U.S.C. § 355(b)(2) (2004); *see also* CDER, FDA, DRAFT GUIDANCE FOR APPLICATIONS COVERED BY SECTION 505(b)(2) 3 (1999) (providing that in application review assessments the FDA is permitted to "rely on [its] findings of safety and effectiveness" for an approved drug to the extent such reliance would be permitted under the generic drug approval provisions of section 505(j) of the NDA process); 21 U.S.C. § 355(b)(2) (permitting submission of an NDA for which the safety and effectiveness investigations "relied upon by the applicant for approval of the application were not conducted by or for the applicant and for which the applicant has not obtained a right of reference or use from the person by or for whom the investigations were conducted"). The FDA has noted that:

> [The FDA's] longstanding interpretation of section 505(b)(2) is intended to permit the pharmaceutical industry to rely to the greatest extent possible under the law on what is already known about a drug . . . . The 505(b)(2) pathway permits sponsors and FDA to determine what studies are necessary to support the approval of the new aspect of the drug. It then allows sponsors to target drug development resources to studies needed to support the proposed difference or innovation.

Letter from Janet Woodcock, Dir., CDER, to Katherine M. Sanzo et al. 3–4 (Oct. 14, 2003) *available at* http://www.bio.org/healthcare/followon/20060530.pdf [hereinafter Woodcock Letter].

[196] *See* Letter from Steven K. Galston, Dir., CDER, to Kathleen M. Sanzo et al. 6 (May 30, 2006), *available at* http://www.bio.org/healthcare/followon/20060530.pdf [hereinafter FDA Response] ("To the extent that the listed drug and the drug proposed in the 505(b)(2) application differ, the 505(b)(2) application must include sufficient data, including clinical or nonclinical data, as appropriate . . . to demonstrate that the proposed drug meets the statutory approval standard for safety and effectiveness.") (citations omitted). Note that the 505(b)(2) process contemplates that it is not to be used for mere duplicates, but instead for drugs that may be different or changed from a listed drug. *See id.* at 6 n.16 ("Our regulations, which expressly provide that we may refuse to review duplicates under section 505(b)(2) of the Act . . . , reflect an intended use of the 505(b)(2) approval pathway for products that include changes from a listed drug."); *see also id.* at 20 (noting that section 505(b)(2) "specifically contemplates that drug products (including proteins products regulated under the [FDCA]) that are submitted in 505(b)(2) applications will . . . represent a change in an already approved drug [that is] supported by a combination of literature or new clinical investigations and the agency's finding that a previously approved drug is safe and effective") (citations and quotations omitted).

[197] See *Biologic Medicine Hearing*, *supra* note 193, at 5.

[198] The inapplicability of bioequivalence studies to follow-on biologics stems from the fact that biologics are large, complex molecules; they have complex mechanisms of action; the relationship between such studies of the follow-on and the product's clinical effect is not clear; biologics often have multiple targets of action; assays for biologics are challenging to perform and often ambiguous; and efficacy markers for biologics are not clear. *See, e.g.*, J. C. Ryff & H. Schellekens, *Immunogenicity of rDNA-derived Pharmaceuticals*, 23

Other biologics, such as blood factors, recombinant DNA-created proteins other than insulin and growth hormone, and monoclonal antibodies are regulated under the PHSA.[199] The PHSA has a "licensure," as opposed to an FDCA "approval," process; however, both the Biologics License Application ("BLA") under the PHSA as well as the traditional NDA process under the FDCA require animal and clinical data to support their applications.[200] BLA applicants must show that the biologic drug is "safe, pure, and potent" and manufactured using processes and facilities that maintain such characteristics.[201]

With respect to follow-on biologics, there is neither an ANDA-equivalent nor section 505(b)(2)-like process for abbreviated biologics approval under the PHSA. For "relatively simple peptide or protein products" such as insulin and growth hormone regulated under the FDCA, current technology permits the FDA to assess follow-on products and hence approve them under the 505(b)(2) mechanism.[202] However, the FDA has concluded that it does "not believe such authority exists for follow-on biologics application under [the BLA process] of the PHS Act that relies on the prior approval of the biological product or on data submitted by another sponsor."[203]

TRENDS PHARMACOLOGICAL SCI. 254 (2002).

[199] *See Biologic Medicine Hearing*, *supra* note 193.

[200] *See id.*

[201] *See* 42 U.S.C. § 262(a)(2)(B)(i)(I) (2004); *see also Applications of Current Statutory Authorities to Human Somatic Cell Therapy Products and Gene Therapy Products*, 58 Fed. Reg. 53,248 (Oct. 14, 1993) (noting that section 351(a) of the PHSA "requires premarket approval for biological products. Licenses are to be issued upon a showing that the establishments and products meet standards, designed to insure the continued safety, purity, and potency of such products," a standard similar to the NDA process for chemical drugs) (quotations omitted).

Note, however, that the showing of these requisite characteristics may be more challenging for biologics. An important concern is the manufacturing process. Good Manufacturing Practices ("GMP"), standardized across the industry, exist for chemical drugs. *See* Leuenberger-Fisher, *supra* note 49, at 392–393 (citations omitted) ("GMPs for chemical drugs can be standardized in this way because chemical drug composition and purity generally can be determined quite easily by chemical analysis").

In contrast,

> not every component of a biological product can be easily identified or measured. Thus, the PHSA . . . provides for specific manufacturing controls designed to assure safety, purity, potency, and effectiveness of products that often cannot be completely characterized. In recognition of this, both biological products and the facilities that manufacture them have historically had to meet specific licensing requirements under the PHSA.

Id. at 393.

[202] *Biologic Medicine Hearing*, *supra* note 193.

[203] *Id.*

C. The Omnitrope Case

Indeed, some follow-on biologics have been approved using the section 505(b)(2) process.[204] A notable example is the case involving an NDA application for Omnitrope, a recombinant-DNA created human growth hormone.[205] The Omnitrope case has garnered significant industry and policymaker attention because of the use of the 505(b)(2) process to approve a follow-on, recombinant protein drug.[206]

Sandoz, the applicant, submitted its 505(b)(2) NDA application for Omnitrope on July 30, 2003.[207] The application used Pfizer's Genotropin as its reference drug.[208] Sandoz was applying to the FDA for approval to market and sell Omnitrope for two of the same clinical applications as that of the Pfizer product.[209]

Initially, in August 2004, the FDA indicated that it could not reach a final decision on the Omnitrope application because of uncertainty regarding the scientific and legal issues related to follow-on biologics.[210] Because of this delay, Sandoz filed suit against the FDA in September 2005, claiming that the FDA was statutorily mandated to rule on its application within 180 days of submission under the terms of the Prescription Drug User Fee Act.[211] The federal district court agreed, and on April 10, 2006, ordered the FDA to make an assessment and hold a hearing on the Omnitrope application.[212] The FDA approved the Omnitrope application under section 505 (b)(2) on May 30, 2006.[213]

In its NDA application, Sandoz provided nonclinical pharmacological and toxicological data, results from human PK and pharmacodynamic

---

[204] Examples include GlucaGen (glucagon recombinant for injection), Hylenex (hyaluronidase recombinant human), Hydase and Amphadase (hyaluronidase), and Fortical (calcitonin salmon recombinant nasal spray). *See* FDA, Omnitrope (Somatropin [rDNA Origin]) Questions and Answers, *available at* http://www.fda.gov/cder/drug/infopage/somatropin/qa.htm; *see also* FDA Response, *supra* note 196, at 45. Since 1987, the FDA has approved 178 applications through the section 505(b)(2) process. *See id.* at 7.

[205] *See* FDA, Omnitrope (Somatropin [rDNA Origin]) Questions and Answers, *supra* note 204.

[206] Because of its small size and simple form—only thirty-two amino acids—Omnitrope is actually an oligopeptide rather than a protein and is not considered a biologic by some commentators. *See, e.g.*, Christopher Webster et al., *Biologics: Can There Be Abbreviated Applications, Generics, or Follow-on Products?*, BIOPHARM INT'L, July 1, 2003, at Q.9, *available at* http://www.biopharm-mag.com/biopharm/article/articleDetail.jsp?id=73785.

[207] *See* Sandoz v. Leavitt, 427 F. Supp. 2d 29, 32 (D.D.C. 2006).

[208] FDA Response, *supra* note 196, at 8.

[209] *Id.* These applications were long-term treatment of pediatric patients for growth failure associated with growth hormone deficiency ("GHD") as well as long term replacement therapy in adults with growth failure resulting either from childhood or adult-onset GHD etiology. *See id.*

[210] *See Sandoz*, 427 F. Supp. 2d at 32.

[211] *See id.*

[212] *See id.* at 41.

[213] *See* FDA Response, *supra* note 196, at 1.

("PD")[214] studies (the latter of which describe the effects of the drug), as well as the results of three original Phase III clinical trials testing a variety of Omnitrope forms in pediatric patients with growth failure.[215] Throughout the studies, Sandoz also compared the active ingredient in Omnitrope with Pfizer's Genotropin.[216]

Several entities objected to the Omnitrope NDA application through the citizen petition process, including Pfizer, Genentech, and the Biotechnology Industry Organization ("BIO").[217] Pfizer objected to any FDA reliance upon Pfizer's product and the original Genotropin NDA for approval of Omnitrope on both legal as well as scientific grounds. It also claimed that the Omnitrope NDA did not meet safety, efficacy, or appropriate manufacturing criteria for the drug.[218] The Genentech and BIO petitions argued that the FDA process of assessing the Omnitrope NDA was legally unsupportable.[219] In particular, they objected that the FDCA did not permit the FDA to use data from an originator's NDA product and that it was scientifically unsound for assessment of follow-on biologics.[220]

---

[214] Pharmacodynamics is the branch of pharmacology that "deals with the reactions between drugs and living systems." MERRIAM-WEBSTER'S MEDICAL DESK DICTIONARY 536 (1993). It includes the study of biochemical and physiological drug effects, the mechanisms of drug action, and the relationship between drug concentration and effect. These can be divided into primary PD studies, which analyze the mode of action or effects of a substance in relation to its therapeutic target, and secondary PD studies, which analyze the mode of action or effects not related to the therapeutic target. CBER, FDA, GUIDANCE FOR INDUSTRY: S7A SAFETY PHARMACOLOGY STUDIES FOR HUMAN PHARMACEUTICALS 10 (2001), *available at* http://www.fda.gov/cber/gdlns/ichs7a071201.pdf.

[215] *See id.* at 10. Sandoz also provided the FDA with results from a fourth Phase III clinical trial. *See id.* at 10 n.26.

[216] *See id.* at 10 n.26. The studies submitted by Sandoz on which the FDA relied used three versions of Omnitrope. *See id.* at 10.

[217] Citizen Petition submitted on behalf of Pfizer, May 14, 2004 and supplement to the petition dated August 4, 2004 Docket Nos. 2004P-023 11CP1 and SUP 1 [hereinafter Pfizer Petition]. Citizen Petition submitted on behalf of the Biotechnology Industry Organization on April 23, 2003 2003P-0176lCP 1 and EMC 1 [hereinafter BIO Petition]. Citizen Petition submitted on behalf of Genentech on April 8, 2004, 2004P-0171lCP1 [hereinafter Genentech Petition]. The FDA's citizen petition regulations give external parties a formal means to contact FDA and seek its action or response on a specific matter or concern. *See* 21 C.F.R. § 10.30 (2001). For example, the process can be used by a drug company to request a change in the approval standards for a generic competitor, denial of approval of a product, establishment of exemptions from certain package labeling requirements, or a tightening of regulations for a particular product. *See id.* Citizen petitions are submitted to FDA Dockets Management Branch for processing and referral to the appropriate office, and FDA regulations require it to issue a tentative or final response within 180 days after receiving the citizen petition. *See id.*

[218] *See* Pfizer Petition, *supra* note 217; FDA Response, *supra* note 196, at 1–3.

[219] *See* Genetech Petition, *supra* note 217; BIO Petition, *supra* note 217.

[220] *See* Genetech Petition, *supra* note 217; BIO Petition, *supra* note 217; FDA Response, *supra* note 196, at 1–2. Specifically, these petitions argued that follow-on biologics may not be approved using originator drug data or information from NDA filings because there are constitutional takings issues associated with such a regime. The FDA, in approving Omnitrope, did not address this issue because it did not rely on the originator drug NDA in its approval. *See* FDA Response, *supra* note 196, at 38 n.70.

Despite these concerns, the FDA noted that it was following its long-standing interpretation of the FDCA and the section 505(b)(2) process in reviewing the Omnitrope NDA and approving Omnitrope.[221] This process included using the FDA's previous finding of safety and efficacy of an originator product.[222]

Although the FDA acknowledged that it cannot use or disclose trade secrets or confidential commercial information of an originator's NDA, it indicated that it did not need to do so to assess and review the Omnitrope NDA.[223] According to the FDA, the Omnitrope NDA application was approved under 505(b)(2) in part because human growth hormone is a relatively simple recombinant protein, with only a single chain.[224] Moreover, the product is not complicated by additional bound molecules;[225] there are available bioassays and biomarkers for human growth hormone; and the drug's clinical mechanism of action and human toxicity profile are well known and extensively described in the scientific literature.[226] Thus, the FDA indicated it could readily make an assessment of the Omnitrope product[227] without reference to any other manufacturer's proprietary data.[228]

The FDA also indicated that this product's simplicity obviated the need to address the broader questions raised by the Genentech and BIO citizen petitions regarding potential Takings Clause issues that might be implicated by using originator drug data.[229] Further, the molecule's simplicity avoided scientific and clinical issues associated with more complex biologics that have unknown or multiple active ingredients, that have unknown mechanisms of action, that are difficult to classify or characterize, and that have molecular systems that make analysis problematic.[230] It should also be

---

[221] *See* FDA Response, *supra* note 196, at 6.

[222] *See id.*; 64 Fed. Reg. 688, 697; CDER, 1999 DRAFT GUIDANCE: APPLICATIONS COVERED BY SECTION 505(B)(2) 2 (1999), *available at* http://www.fda.gov/cder/guidance/2853dft.pdf.

[223] *See* FDA Response, *supra* note 196, at 37.

[224] *See id.* at 8.

[225] *See id.*

[226] *See id.*

[227] *See* FDA Response, *supra* note 196, at 8.

[228] *See id.* at 15, 37.

[229] *See* Genetech Petition, *supra* note 217; BIO Petition, *supra* note 217; FDA Response, *supra* note 196. The FDA did note, however:

> Because, under the Hatch-Waxman Amendments, the Agency's finding of safety and effectiveness for a listed drug may be relied upon for approval of an ANDA or a 505(b)(2) application, the finding of safety and effectiveness is not confidential commercial information that must be protected.

FDA Response, *supra* note 196, at 38 (citations omitted). It nonetheless concedes that "the applicant is entitled to expect that information in the application will be protected . . . when that information falls within the definition of trade secret and confidential commercial information." *Id.* at 38 (citations omitted). The FDA noted that, with respect to Omnitrope, there was no support for the argument that the FDA was required to reference the originator's trade-secret data in assessing follow-on similarity. *Id.* at 38 n.71.

[230] *See* FDA Response, *supra* note 196, at 8.

noted that the Omnitrope approval was not simply a paper NDA without applicant clinical trials or clinical data. The FDA required a significant amount of nonclinical and clinical testing, including several new clinical trials, before approving Omnitrope.[231]

Further, the FDA assessed Omnitrope's NDA based upon whether it was "sufficiently similar" to other previously approved forms such as Genotropin.[232] According to the FDA, "sufficient similarity" constitutes a lower standard than, say, the standard of "sameness" required for ANDA applicants under 505(j) applications.[233] However, the FDA did not specifically define "sufficient similarity" or "scientifically justified."[234]

With regard to immunogenicity, the FDA pointed out that the Omnitrope NDA included substantial original data establishing its profile as similar to Genotropin, and also included product modifications made during successive clinical trials that showed reductions in previously reactive patients.[235] Moreover, the fourth Phase III clinical trial, submitted with Sandoz's safety update on the drug, indicated no patient immunogenicity reaction to Omnitrope, "demonstrat[ing] that Omnitrope has a low and acceptable level of immunogenicity that is consistent with other approved [recombinant-DNA human growth hormone] products, including Genotropin."[236]

Thus, the FDA could rely in part on its previous finding of Genotropin's safety and effectiveness to approve Omnitrope because (1) Omnitrope was a simple molecule; (2) Sandoz conducted supporting clinical trials; (3) the scientific information submitted by Sandoz provided the FDA with a sound basis to conclude that Omnitrope and Genotropin were of "high similarity"; and (4) the FDA had experience with similar molecules.[237] It then approved Omnitrope, but with a BX rating; that is, the FDA concluded that Omnitrope is not therapeutically equivalent to Genotropin and therefore not substitutable.[238]

Importantly, the FDA emphasized that this methodology and the Omnitrope approval process did not set a precedent for future applications regarding more complex follow-on biologic approvals. Specifically, the FDA noted that biologics licensed under section 351 of the PHSA as well as

---

[231] *See id.* at 10.

[232] *Id.* at 19.

[233] *Id.* at 6 n.16. The FDA also noted that demonstrations of the safety and efficacy of the applicant's drug can rely on the originator's product when "scientifically justified." *Id.* at 19.

[234] *See* FDA Response, *supra* note 196, at 12, 19.

[235] *Id.* at 34.

[236] *Id.* at 35.

[237] *See id.* at 14.

[238] *See id.* at 4. A BX rating is given to drug products for which data are insufficient to determine therapeutic equivalence. *See* 21 C.F.R. § 320.1(a) (2005); *see also* Orange Book, *supra* note 158 (describing the *Orange Book* rating system).

even some FDCA-regulated biologics could not utilize the 505(b)(2) process.[239] As noted by the FDA:

> The approval of Omnitrope does not signal that the Agency has concluded that—regardless of the nature and complexity of the active ingredient and the indications for use—every protein product approved under section 505 of the [FDCA] is an appropriate candidate for reference by an applicant seeking approval of a follow-on protein product through an abbreviated pathway. Further, this decision does not address the distinct legal and regulatory issues related to approving follow-on versions of products licensed under the PHSA or the scientific challenges that may be posed by more complex and less well-understood licensed biological products.[240]

Hence, although the section 505(b)(2) process has resulted in approval of Omnitrope and several other follow-on protein products, it is unlikely that it provides an abbreviated process akin to ANDA for the majority of follow-on biologics. The emphasis upon the simplicity of the molecule approved indicates that more extensive study and concomitant regulatory structure will be necessary for approval of more complex follow-on biological products.

D. European Union

1. Overview of Regulation

The primary agency responsible for approving drugs in the E.U. is the European Medicines Evaluation Agency ("EMEA").[241] The EMEA is headquartered in London and began operations in 1995.[242] Its mission is similar to that of the FDA in the United States: to ensure and promote public health and safety by utilizing scientific resources to evaluate medicinal products as well as to advise on research and development within that mission.[243]

The EMEA, as a single agency, coordinates the evaluation of medicinal products in the E.U. It performs this function by utilizing the sci-

---

[239] *See* FDA Response, *supra* note 196, at 45 n.89.

[240] *Id.* at 52. The FDA has also noted that the majority of biologics are licensed under the PHSA, which has no abbreviated application process. *See* FDA, *Omnitrope* (somatropin [rDNA origin]) Questions and Answers, *supra* note 204.

[241] *See* EMEA, About the EMEA—Structure, *available at* http://www.emea.eu.int/htms/aboutus/emeaoverview.htm [hereinafter About the EMEA].

[242] *See id.*

[243] *See id.*

entific resources of 3500 scientific experts, in 25 E.U. member states, that reside in a network of 42 national authorities.[244] It also cooperates with international partners in an effort to advance global harmonization.[245]

The EMEA provides for two procedures in drug assessment and approval: (1) a centralized authorization procedure; and (2) a decentralized procedure based on mutual recognition of individual national authorizations.[246] Certain drugs, including all biotechnology drugs, must go through the centralized process, while other drugs, including those with new active substances that have not been authorized by any E.U. member state, are merely eligible for the EMEA centralized process.[247] Like the FDA, the EMEA must issue marketing authorization to all drugs that go through the centralized process before manufacturers can commence sales.[248]

In general, companies submit one single marketing authorization application to the EMEA in the centralized process.[249] For human drugs, the EMEA Committee for Medicinal Products for Human Use ("CHMP") conducts a single evaluation.[250] The CHMP is responsible for conducting the initial assessment of medicinal products for the E.U.-wide marketing authorization application,[251] and also bears responsibility for some post-authorization and maintenance activities, including review of modifications or extensions of the existing marketing authorization.[252]

If the CHMP concludes that quality, safety and efficacy of the particular product are sufficiently proven, it adopts a positive opinion in favor of approval.[253] This opinion is sent to the European Commission ("EC"), which may then issue a market authorization that is valid for the entire E.U.[254] The process from the beginning of assessment to the rendering of an

---

[244] *See id.*

[245] *See id.*

[246] About the EMEA, *supra* note 241. In the decentralized process, drug marketing approval is based on mutual recognition of national approval authority for most traditional medicinal products. *See id.* This approval system involves the EMEA as arbiter. If one member state does not recognize another member state's authorization, the EMEA is tasked with the role of finding a compromise. *See id.*

[247] *See* E.U. Business, E.U. Guide—Pharmaceuticals, *available at* http://www.eubusiness.com/guides/Pharmaceuticals.htm. The centralized process covers roughly ten percent of drugs in the E.U. *See id.*

[248] *See id.*

[249] *See* Irene Eckart, *The European Medicines Agency (EMEA)—From Research to Therapies*, Bridges, Sept. 2006, *available at* http://www.ostina.org/content/view/1439/617.

[250] *See id.*

[251] *See id.*

[252] *See* Council Regulation (EC) No. 726/2004, 2004 O.J. (L 136) 1; About the EMEA, *supra* note 241.

[253] *See* Council Regulation (EC) No. 726/2004, 2004 O.J. (L 136) 1; About the EMEA, *supra* note 241.

[254] *See* Council Regulation (EC) No. 726/2004, 2004 O.J. (L 136) 1; About the EMEA, *supra* note 241. Note that the EC may reject the recommendation of EMEA, as it did initially for Omnitrope. *See* Novartis AG, Annual Report (Form 20-F), at 54 (Jan. 28, 2005).

opinion is theoretically 210 days.[255] The European Commission is then given 90 days to render a final decision based on the EMEA opinion.[256]

2. *Biosimilars*

In contrast to the United States, the E.U. has established a regulatory pathway for follow-on biologics, known in the E.U. as biosimilars.[257] So far under this regime, two follow-on products, both human growth hormone forms, have been approved.[258]

The process through which the E.U. has established biosimilar regulation took several years. The first express notation of biosimilars appears to be in June 2003, when the class of "similar biological medicinal products" was indicated in European Parliament legislation.[259] In the next year, the legal basis for approval of such products was issued.[260] These documents create the overall regulatory infrastructure that allows approval of biosimilars in the E.U.

Although there were some false starts with respect to the issuance of EMEA guidelines for biosimilars,[261] by 2005, the EMEA had issued general guidelines based on CHMP discussions in June 2004, which were put

---

[255] *See* Council Regulation (EC) No. 726/2004, 2004 O.J. (L 136) 1. Note, however, that this time period may be suspended if additional information is requested, and hence, the process may actually be much longer. *See* Gary Walsh, *Drug Approval in the European Union and the United States*, *in* PHARMACEUTICAL BIOTECHNOLOGY: DRUG DISCOVERY AND CLINICAL APPLICATIONS 201, 205 (Oliver Kayser & Rainier H. Müller eds., 2004).

[256] *See* Council Regulation (EC) No. 726/2004, 2004 O.J. (L 136) 1; *see also Drug Approval in the European Union and the United States*, *supra* note 255, at 205.

[257] *See* Roger, *supra* note 17, at 341.

[258] *See* EMEA, EMEA/H/C/607, EUROPEAN PUBLIC ASSESSMENT REPORT (EPAR): OMNITROPE, EPAR SUMMARY FOR THE PUBLIC, at 2 (2006), *available at* http://www.emea.eu.int/humandocs/PDFs/EPAR/Omnitrope/060706en1.pdf [hereinafter Omnitrope EPAR]; Global Insight, Second Biosimilar Drug Approved by European Commission (May 2006), http://www.globalinsight.com/SDA/SDADetail5845.htm [hereinafter Global Insight].

[259] *See* Commission Directive (EC) No. 2003/63, 2003 O.J. (L 159) 46, 78–79.

[260] *See* Parliament and Council Directive (EC) No. 2004/27, 2004 O.J. (L 136) 34.

[261] Initially, in December 2003, EMEA implemented guidelines for biosimilars, focusing on proving comparability between biologic drug forms. *See* Comm. for Proprietary Medicinal Products ("CPMP"), EMEA, CPMP/3097/02/Final (2003), GUIDELINE ON COMPARABILITY OF MEDICINAL PRODUCTS CONTAINING BIOTECHNOLOGY-DERIVED PROTEINS AS ACTIVE SUBSTANCE: NON-CLINICAL AND CLINICAL ISSUES, *available at* http://www.emea.europa.eu/pdfs/human/ewp/309702en.pdf; CPMP, EMEA, CPMP/BWP/3207/00 (2003), GUIDELINE ON COMPARABILITY OF MEDICINAL PRODUCTS CONTAINING BIOTECHNOLOGY-DERIVED PROTEINS AS ACTIVE SUBSTANCE: QUALITY ISSUES, *available at* http://www.emea.europa.eu/pdfs/human/bwp/320700en.pdf. However, during the development of these guidelines, the International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use ("ICH"), a cooperative international initiative of E.U., Japanese and U.S. public and industry experts in pharmaceuticals was engaged in assessing specific issues of scientific and technical considerations on product registration. *See* http://www.ich.org. This ultimately led to additional guidance on comparability that substantively changed a large proportion of the 2003 guidance. *See* CPMP, EMEA, CPMP/ICH/5721/03, NOTE FOR GUIDANCE ON BIOTECHNOLOGICAL/BIOLOGICAL PRODUCTS SUBJECT TO CHANGE IN THEIR MANUFACTURING PROCESS (2005), *available at* http://www.emea.europa.eu/pdfs/human/ich/572103en.pdf [hereinafter NOTE FOR GUIDANCE].

into effect on October 30, 2005.[262] More specific EMEA guidance on nonclinical issues, clinical issues,[263] and quality issues[264] with respect to biosimilars took effect in June 2006.

Most recently, the EMEA has decided to provide product-specific guidance for biosimilars. Hence, it has issued guidelines on insulin,[265] somatropin (human growth hormone),[266] erythropoietin,[267] and granulocyte-colony stimulating factor.[268]

3. Nonclinical and Clinical Issues

Several themes emerge from the EMEA guidelines. First, the clear focus of EMEA is to ensure comparability between the biosimilar and originator biologic products.[269] As such, for any abbreviated process, it is apparent that the acceptance of biosimilar products will depend on successfully characterizing the molecule in relation to the originator (or "reference") biologic product.[270] If EMEA believes that the characterization is insufficient for its purposes, it is likely that an application with complete nonclinical and clinical data will have to be submitted.

The EMEA guidelines indicate that although "official" data (such as pharmacopoeial monographs or other scientific data) can initially be used

---

[262] *See* CHMP, EMEA, CHMP/437/04, GUIDELINE ON SIMILAR BIOLOGICAL MEDICINAL PRODUCTS (2005), *available at* http://www.emea.europa.eu/pdfs/human/biosimilar/043704en.pdf.

[263] *See* CHMP, EMEA, CHMP/BMWP/42832/2005, GUIDELINE ON SIMILAR BIOLOGICAL MEDICINAL PRODUCTS CONTAINING BIOTECHNOLOGY-DERIVED PROTEINS AS ACTIVE SUBSTANCE: NON-CLINICAL AND CLINICAL ISSUES (2006), *available at* http://www.emea.europa.eu/pdfs/human/biosimilar/4283205en.pdf [hereinafter GUIDELINE: NON-CLINICAL AND CLINICAL ISSUES].

[264] *See* CHMP, EMEA, CHMP/BMWP/49348/2005, GUIDELINE ON SIMILAR BIOLOGICAL MEDICINAL PRODUCTS CONTAINING BIOTECHNOLOGY-DERIVED PROTEINS AS ACTIVE SUBSTANCE: QUALITY ISSUES (2006), *available at* http://www.emea.europa.eu/pdfs/human/biosimilar/4934805en.pdf [hereinafter GUIDELINE: QUALITY ISSUES].

[265] *See* CHMP, EMEA, CHMP/BMWP/32775/2005, GUIDANCE ON SIMILAR MEDICINAL PRODUCTS CONTAINING RECOMBINANT HUMAN SOLUBLE INSULIN (2006), *available at* http://www.emea.europa.eu/pdfs/human/biosimilar/3277505en.pdf [hereinafter GUIDELINE: INSULIN].

[266] *See* EMEA, CHMP/BMWP/94528/2005, GUIDANCE ON SIMILAR MEDICINAL PRODUCTS CONTAINING SOMATROPIN (2006), *available at* http://www.emea.europa.eu/pdfs/human/biosimilar/9452805en.pdf [hereinafter GUIDELINE: SOMATROPIN].

[267] *See* CHMP, EMEA, CHMP/BMWP/94526/2005, GUIDANCE ON SIMILAR MEDICINAL PRODUCTS CONTAINING RECOMBINANT ERYTHROPOIETINS (2006), *available at* http://www.emea.europa.eu/pdfs/human/biosimilar/452605en.pdf [hereinafter GUIDELINE: ERYTHROPOIETINS].

[268] *See* CHMP, EMEA, CHMP/BMWP/31329/2005, GUIDANCE ON SIMILAR MEDICINAL PRODUCTS CONTAINING RECOMBINANT GRANULOCYTE-COLONY STIMULATING FACTOR (2006), *available at* http://www.emea.europa.eu/pdfs/human/biosimilar/3132905en.pdf [hereinafter GUIDELINE: GRANULOCYTE-COLONY STIMULATING FACTOR].

[269] *See* GUIDELINE: QUALITY ISSUES, *supra* note 264.

[270] *See id.*

to support the application, this paper basis alone is not adequate for approval.[271]

Instead, comparability tests, using the biosimilar applicant drug and the reference product, must show that both are similar in quality, safety, and efficacy.[272] The EMEA guidelines do specifically allow for nonclinical testing to be abbreviated, although some such testing must be performed.[273] Further, nonclinical testing is to be aimed at finding differences between the biosimilar and originator product, not simply seeking to show a demonstrable response.[274]

The requirements for clinical efforts are a function of extant knowledge about the reference biologic and its therapeutic indication.[275] EMEA recognizes that manufacturing methods may change during development but strongly recommends using the final manufacturing process to generate clinical data for comparisons of the biosimilar with the reference product.[276] The guideline states that "[a]ny deviation from this recommendation should be justified and supported by adequate additional data,"[277] signaling that additional clinical testing requirements may be imposed by EMEA if the recommendation is not fulfilled. Importantly, the need for clinical trials will be assessed on a case-by-case basis by EMEA.[278]

Clinical comparability to the reference biologic is done through a stepwise process using PK and PD testing followed by clinical safety and efficacy trials, or in some cases, additional PK and PD studies.[279] As in nonclinical testing, applicants must also justify PK and PD testing approaches.[280]

After these assessments, applicants may perform clinical trials to show comparability and efficacy.[281] However, at this stage applicants may also use an abbreviated process. They may substitute comparative PK and PD testing between the biosimilar and the reference biologic, so long as they meet certain criteria—generally concerning the extent of knowledge of the reference biologic's characteristics.[282]

---

[271] *See id.* at 3.

[272] *See* GUIDELINE: NON-CLINICAL AND CLINICAL ISSUES, *supra* note 263, at 3.

[273] *See* GUIDELINE: QUALITY ISSUES, *supra* note 264.

[274] *See id.* at 4.

[275] *See id.*

[276] *See id.*

[277] *Id.*

[278] *See* GUIDELINE: QUALITY ISSUES, *supra* note 264.

[279] *See* GUIDELINE: NON-CLINICAL AND CLINICAL ISSUES, *supra* note 263, at 5.

[280] *See id.*

[281] *See id.*

[282] *See id.* at 5–6. These criteria are that (1) the reference biologic PDs are well characterized; (2) the PD properties of the reference biologic are also sufficiently known; (3) the relationship between dose and efficacy of the reference biologic is sufficiently characterized; and (4) one or more PD markers are accepted as a surrogate marker for drug efficacy, and the dose for the biosimilar product and the surrogate marker is known. *See id.*

If the efficacy of the biosimilar and reference drugs is shown to be comparable, the applicant must still meet safety and pharmacovigilance[283] requirements.[284] To that end, EMEA requires that prelicensing safety data be provided.[285] This data must show results covering

> a number of patients sufficient to address the adverse effect profiles of the test and the reference medicinal product . . . [and] to compare the type, severity and frequency of the adverse reactions between the similar biological and the reference biological medicinal products.[286]

The biosimilar applicant must provide a risk-specification and pharmacovigilance program approach with its application to the EMEA, including a description of potential safety issues associated with the medicinal product that may be a result of differences between the manufacturing process of the biosimilar and that of the originator biologic.[287] The EMEA also recognizes that premarket authorization data is usually not sufficient to identify every potential difference between the reference and the biosimilar product.[288] Hence, it requires that the biosimilar applicant monitor the clinical safety of its drug continuously during the post-approval period.[289]

The EMEA guidelines directly address the critical issue of immunogenicity. Because of the wide array of factors associated with immunogenicity, inter-individual variability, and potential life-threatening consequences, immunogenicity must be investigated by the biosimilar applicant.[290] Such investigation includes clinical trials.[291]

---

[283] "Pharmacovigilance encompasses surveillance of side effects after short-term and long-term use of medicines." *See* EUROPEAN PUB. HEALTH ALLIANCE, *Pharmacovigilance* (draft) 1 (2003) *available at* http://www.epha.org/IMG/doc/DRAFT_PAPER_ON_PHARMACOVIGILANCE_12_08_2003.doc. The FDA defines pharmacovigilance as "all scientific and data gathering activities relating to the detection, assessment, and understanding of adverse events." FDA, CDER, GUIDANCE FOR INDUSTRY GOOD PHARMACOVIGILANCE PRACTICES AND PHARMOCOEPIDEMIOLOGICAL ASSESSMENT 4 (2005), *available at* http://www.fda.gov/Cder/guidance/6359OCC.pdf.

[284] *See* GUIDELINE: NON-CLINICAL AND CLINICAL ISSUES, *supra* note 263, at 6.

[285] *See id.*

[286] *See id.*

[287] *See* EUROPEAN PUB. HEALTH ALLIANCE, *Pharmacovigilance* (draft) 1 (2003) *available at* http://www.epha.org/IMG/doc/DRAFT_PAPER_ON_PHARMACOVIGILANCE_ 12_08_2003.doc.

[288] *See* GUIDELINE: NON-CLINICAL AND CLINICAL ISSUES, *supra* note 263, at 6.

[289] *See id.* Also, in line with this pharmacovigilance requirement, the specific biosimilar must be traceable so that effective tracking can be accomplished in the event of a clinical concern. Because this is an area of great concern to the EMEA, "pharmacovigilance obligations will be closely monitored." *Id.*

[290] *See* GUIDELINE: NON-CLINICAL AND CLINICAL ISSUES, *supra* note 263, at 7; *see also supra* notes 76–100 and accompanying text (discussing the threats posed by immunogenicity and the difficulties in avoiding such risks).

[291] *See* GUIDELINE: NON-CLINICAL AND CLINICAL ISSUES, *supra* note 263, at 7.

Importantly, the general guidance indicates that because of the unpredictability of immunogenicity reactions, the biosimilar applicant must present long-term results of antibody monitoring at specific intervals.[292] Further, for drugs that are chronically administered, the EMEA requires one-year follow-up data assessing immunogenicity before any licensing will be approved.[293] Applicants must also justify testing strategies in this area.[294] If clinical testing unearths a different immune response to the biosimilar than to the reference biologic, the EMEA requires additional analysis for the biosimilar applicant.[295]

Finally, in another gesture toward abbreviated processes, there is the potential to extrapolate safety and efficacy from one clinical indication to that of another. In general, the safety and efficacy of the biosimilar drug must be shown for each clinical use: however, the biosimilar applicant may extrapolate from clinical experience, scientific literature, identical mechanism of action, and the same drug target to demonstrate safety and efficacy for additional clinical uses.[296]

4. Quality Issues

The EMEA also provided guidelines with respect to manufacturing processes that are critical to characterizing the biologic product.[297] The EMEA recognizes that biologics are significantly defined by both their manufacturing processes and their molecular composition. Thus, the biosimilar applicant must demonstrate the consistency and robustness in its manufacturing.[298]

The EMEA requires that studies be performed that assess an array of issues associated with manufacturing.[299] These issues include the biosimilar product's stability and compatibility with nonclinical materials such as excipients, diluents, and packaging materials, as well as identification of the active ingredient.[300] Any change in manufacturing processes must be accompanied by a comparability assessment during the developmental

---

[292] *See id.*

[293] *See id.*

[294] *See id.*

[295] *See id.* at 8. Clinical, safety, efficacy, and PD assessments will be required, particularly with respect to products where the immune response could adversely affect the body's own protein product and its function. *See id.* This is consistent with the Eprex immunogenicity issues that created reactions to the body's own erythropoietin. *See supra* notes 76–98 and accompanying text (discussing Eprex and the associated risk of pure red cell aplasia).

[296] *See* GUIDELINE: NON-CLINICAL AND CLINICAL ISSUES, *supra* note 263, at 3.

[297] *See* GUIDELINE: QUALITY ISSUES, *supra* note 264, at 4.

[298] *See id.*

[299] The EMEA guidelines on biosimilars are not the only guidance with respect to changes in manufacturing for a given product. Manufacturing guidelines are addressed by a wholly separate set of requirements. *See* NOTE FOR GUIDANCE, *supra* note 261.

[300] *See* GUIDELINE: QUALITY ISSUES, *supra* note 264, at 4.

stages of the biosimilar, beyond when comparing the biosimilar product to the reference biologic.[301]

The EMEA emphasizes that quality issues should be a fundamental component of comparison between the biosimilar and the reference biologic.[302] Although recognizing that the quality characteristics of the biosimilar and the reference biologic will be different, the EMEA may allow impurity profiles and minor structural differences in the active substance due to molecular variability.[303] However, applicants must provide justification for such differences,[304] and the EMEA indicates that they would be assessed on a case-by-case basis, allowing it to determine the amount of nonclinical and clinical data required.[305]

With respect to quality issues, as with clinical and nonclinical guidance, the EMEA focuses upon ensuring the comparability of the biosimilar and the originator biologic. An E.U. member state must authorize the reference biologic;[306] applicants must provide scientific justification when selecting the reference biologic; and applicants must use the same product for quality, safety, and efficacy comparability testing.[307] Note that the EMEA likely requires applicants to test active substances alone (that is, in the absence of nonclinical materials) to show that the active substance in the biosimilar is "representative" of the active substance in the reference biologic.[308] In addition, of course, applicants must perform further testing using the complete biosimilar formulation and the reference biologic to show that that they are comparable.[309]

Procedures used in testing comparability on the quality side include a variety of means and methods. The EMEA indicates that the analytic methods should be state of the art, be validated, provide for physicochemical comparisons, determine biological activity, and assess purity and impurity profiles qualitatively and quantitatively.[310] The biosimilar applicant will not generally have access to the originator biologic drug information; however, the EMEA indicates that even so, "the level of detail [of purity and impurity profiles of biosimilar and originator biologics] must be such that firm conclusions on the purity and impurity profiles can be made."[311]

---

[301] *See id.*
[302] *See id.* at 5.
[303] *See id.*
[304] *See id.*
[305] *See* GUIDELINE: QUALITY ISSUES, *supra* note 264, at 5.
[306] *See id.*
[307] *See id.*
[308] *See id.* at 6.
[309] *See id.*
[310] *See* GUIDELINE: QUALITY ISSUES, *supra* note 264, at 5.
[311] *Id.*

5. Specific Product Annex Guidance

Particular product-specific annexes to the EMEA general guidance add nonclinical and clinical requirements to the general rules.[312] Overall, the more complex the molecule, the greater the requirements to show comparability, particularly with respect to clinical trials.[313] Though all product-specific EMEA annexes require twelve months of clinical trials, the specific clinical requirements vary between products.[314]

The EMEA reviews proposed extensions of a biosimilar's indications for use based upon specific disease states chosen by biosimilar applicants.[315] Generally, all uses of a biosimilar product must be demonstrated by "appropriate justification."[316]

6. First Approval: Omnitrope

As noted previously, the EMEA has approved market authorization for two biosimilar molecules.[317] These approvals provide some guidance on the substantive biosimilars approval process and the efforts that are nec-

---

[312] *See, e.g.*, GUIDELINE: INSULIN, *supra* note 265, at 4; GUIDELINE: SOMATROPIN, *supra* note 266, at 5; GUIDELINE: GRANULOCYTE-COLONY STIMULATING FACTOR, *supra* note 268, at 5; GUIDELINE: ERYTHROPOIETINS, *supra* note 267, at 5.

[313] For example, under the EMEA guidelines, insulin—the simplest biologic issued annex guidance—does not by itself require any *in vivo* PD assessment. *See* GUIDELINE: INSULIN, *supra* note 265, at 4. Further, with respect to biosimilar efficacy testing, biosimilar insulin applicants are not required to perform such testing. *See id.* at 5. Yet somatropin applicants would be required to perform at least one study, see GUIDELINE: SOMATROPIN, *supra* note 266, at 5, granulocyte-colony stimulating factor applicants would need to perform a two-arm comparability study, see GUIDELINE: GRANULOCYTE-COLONY STIMULATING FACTOR, *supra* note 268, at 5, and erythropoietin applicants would need to perform at least two studies, see GUIDELINE: ERYTHROPOIETINS, *supra* note 267, at 5.

[314] Insulin applicants are required to perform a clinical trial of at least twelve months using subcutaneous administration, including a comparative phase of at least six months completed before approval. *See* GUIDELINE: INSULIN, *supra* note 265, at 5. Somatropin applicants are required to gather twelve-month immunogenicity data from patients who also participated in efficacy trials, with at least a three-month sampling. *See* GUIDELINE: SOMATROPIN, *supra* note 266, at 5. Granulocyte-colony stimulating factor applicants must collect immunogenicity data for at least twelve months with follow-up for chronic administration patients and any nonclinical and clinical findings. *See* GUIDELINE: GRANULOCYTE-COLONY STIMULATING FACTOR, *supra* note 268, at 5. Erythropoietin applicants are required to provide twelve-month immunogenicity data from patients who are also participating in efficacy trials. *See* GUIDELINE: ERYTHROPOIETINS, *supra* note 267, at 6.

[315] *See* GUIDELINE: QUALITY ISSUES, *supra* note 264, at 4.

[316] *See id.* There is no extension of clinical indications for insulin biosimilars. *See id.* For somatropin, efficacy and safety shown in growth hormone-deficient children may allow extrapolations to other indications. *See id.* For granulocyte-colony stimulating factor, demonstrable clinical comparability in chemotherapy-induced neutropenia (i.e., low white cell count) may allow extrapolation to other clinical indications if the referenced medicinal product of the mechanism of action is the same. *See* GUIDELINE: GRANULOCYTE-COLONY STIMULATING FACTOR, *supra* note 268, at 5. For erythropoietin, efficacy and safety in renal-based anemia may allow extrapolations to other clinical indications. *See* GUIDELINE: ERYTHROPOIETINS, *supra* note 267, at 6.

[317] *See* Omnitrope EPAR, *supra* note 258; Global Insight, *supra* note 258.

essary to obtain market authorization in the E.U., as well as giving a helpful perspective to U.S. regulators.

Sandoz obtained market authorization for its human growth hormone, Omnitrope, in April 2006 after receiving a positive opinion from CHMP in January of that same year.[318] The process took roughly seventeen months.[319] Recall that Omnitrope is the molecule that was approved under the section 505(b)(2) process in the U.S.[320]

Sandoz utilized Pfizer's Genotropin as the reference biologic in its EMEA filing, as it did in its U.S. effort.[321] Sandoz filed an extensive number of documents and studies with its application to EMEA.[322] Assessing this package, it appears that Sandoz provided much more information than the EMEA somatropin guideline required.[323]

Nonetheless, at least some of the additional information provided was based upon EMEA mandates. For example, an additional Phase III safety study was required, perhaps because the Omnitrope product in the clinical efficacy study contained a high concentration of impurities from the cells making the growth hormone, resulting in sixty percent of patients showing antibodies to the drug compared with two percent of patients taking the reference biologic.[324] Apparently, after a change in manufacturing to address this issue, the adverse antibody formation disappeared and the EMEA did not later require a repeat comparative clinical trial.[325]

The EMEA also appeared to have reservations about the pharmacovigilance plan submitted by Sandoz, which was apparently the standard pharmacovigilance effort the company uses for its other drugs.[326] Sandoz

---

[318] *See* Omnitrope EPAR, *supra* note 258, at 2.

[319] Sandoz filed its new biosimilar application in July 2004 on the basis of the new guidelines. *See* EMEA, EUROPEAN PUBLIC ASSESSMENT REPORT: OMNITROPE, BACKGROUND INFORMATION ON THE PROCEDURE (2006), *available at* http://www.emea.eu.int/humandocs/PDFs/EPAR/Omnitrope/060706en7.pdf. Assessment began in August 2004 and a positive opinion was rendered by CHMP in January 2006. *See id.*

[320] *See* Omnitrope Case, *supra* notes 204–240, and accompanying text.

[321] *See* EMEA, EUROPEAN PUBLIC ASSESSMENT REPORT: OMNITROPE, SCIENTIFIC DISCUSSION 1 (2006), *available at* http://www.emea.eu.int/humandocs/PDFs/EPAR/Omnitrope/060706en6.pdf [hereinafter SCIENTIFIC DISCUSSION].

[322] Of note, these included the results of two nonclinical *in vivo* PD comparisons using two different animal models, a nonclinical repeat-dose toxicity study, three PK/PD clinical studies, an ongoing Phase III comparative clinical efficacy study with continuous patient follow-up (and a promise that a follow-up report for the first twenty-four months would be submitted). Additionally, Sandoz provided an ongoing Phase III open label noncomparative clinical safety study, standard clinical safety data based on the two Phase III studies, twelve-month immunogenicity data from the two Phase III studies, a promise to submit final twenty-four-month data from the clinical safety study, a detailed description of a pharmacovigilence system, and a risk management plan. *See id.*

[323] *See* GUIDELINE: SOMATROPIN, *supra* note 266.

[324] *See* SCIENTIFIC DISCUSSION, *supra* note 321, at 9. Note, however, that the antibodies formed against Omnitrope were non-neutralizing; that is, they did not neutralize the effect of the drug. *See id.*

[325] *See id.*

[326] *See id.* at 22, 25.

agreed to address these issues before marketing Omnitrope.[327] The EMEA did allow extrapolation of the Sandoz application data to other clinical applications associated with Genotropin.[328]

Sandoz has since begun marketing Omnitrope in Germany and Austria,[329] and it believes its price discount will be roughly twenty percent.[330]

7. Second Approval: Valtropin

The BioPartners biosimilar Valtropin was given market authorization in April 2006 based on a positive opinion from CHMP in February 2006,[331] after a process that took roughly eighteen months.[332]

BioPartners used Eli Lilly's Humatrope as its reference biologic drug.[333] Although it appears that BioPartners submitted somewhat less information than Sandoz in its biosimilar application, the EMEA was careful to adhere to its guideline requirements and mandated additional actions on the part of BioPartners.[334] During the process of review, the EMEA found that two forms of Humatrope were used in the application studies: one E.U. approved, one U.S. approved.[335] Since the guidelines indicate that only E.U. approved drugs may be used for biosimilar applications, the EMEA required that BioPartners resubmit data using subpopulations from the key clinical trial; BioPartners also submitted a second clinical efficacy study.[336]

EMEA approved Valtropin and, as it did for Omnitrope, allowed extrapolation to cover clinical indications of the reference biologic.[337]

8. Rejection: Alpheon

Although EMEA has approved two growth hormone biosimilars, it did not rubber-stamp these biosimilar applications. Not only has the EMEA

---

[327] *See id.* at 22.

[328] *See id.* at 25.

[329] *See* Novartis AG, Current Report of Foreign Issuer (Form 6-K), 2–3 (Apr. 19, 2006).

[330] *See* Global Insight, *supra* note 258.

[331] *See* EMEA, EMEA/H/C/607, EUROPEAN PUBLIC ASSESSMENT REPORT (EPAR), VALTROPIN, EPAR SUMMARY FOR THE PUBLIC (Mar. 2006), *available at* http://www.emea.eu.int/humandocs/PDFs/EPAR/valtropin/EPARSummary-en.pdf; EMEA, EUROPEAN PUBLIC ASSESSMENT REPORT, VALTROPIN, BACKGROUND INFORMATION ON THE PROCEDURE (2006), *available at* http://www.emea.eu.int/humandocs/PDFs/EPAR/valtropin/EPARPS%20Before-en.pdf.

[332] BioPartners submitted its application in June 2004. *See* EMEA, EMEA/H/C/607, EUROPEAN PUBLIC ASSESSMENT REPORT (EPAR): VALTROPIN, EPAR SUMMARY FOR THE PUBLIC, *supra* note 331.

[333] *See* EMEA, EUROPEAN PUBLIC ASSESSMENT REPORT: VALTROPIN, SCIENTIFIC DISCUSSION 1 (2006), *available at* http://www.emea.eu.int/humandocs/PDFs/EPAR/valtropin/EPARScientificDiscussion-en.pdf.

[334] *See id.*

[335] *See id.* at 18.

[336] *See id.*

[337] *See id.* at 28–29.

taken longer periods to review biosimilar applications than those for standard products and sent both biosimilar applicants back for more studies, it has also rejected one biosimilar application.[338] The rejected product was a biosimilar of interferon alpha-2a, known as Alpheon by BioPartners, the company that was successful in obtaining approval for Valtropin.[339]

Since there is no product-specific guidance for alpha interferons,[340] the EMEA evaluated the biosimilar application for the drug under its general guidelines. BioPartners used Roche's Roferon-A as its reference biologic.[341] EMEA ultimately rendered a negative opinion on Alpheon because it failed in all three key areas: quality, safety, and efficacy.

With respect to quality, it appeared that differences in Alpheon and Roferon-A included impurities as well as insufficient data on the stability of both the active substance and the medication itself.[342] Furthermore, there was inadequate validation of the final manufacturing process as judged by EMEA.[343] With regard to safety, in the clinical study, side effects were more prevalent in Alpheon as compared to the reference biologic.[344] Further, the testing method for immunogenicity was not validated.[345] Finally, with regard to efficacy, the number of patients with hepatitis C responding to treatment with Alpheon and the reference biologic were similar in the pivotal clinical study. However, critically, a greater number of patients relapsed when treatment was terminated in the Alpheon group.[346]

Hence, on the basis of these concerns, the EMEA rejected the Alpheon application.[347] BioPartners has indicated it will resubmit the application.[348]

---

[338] *See* EMEA, EMEA/190896/2006 (2006), QUESTIONS AND ANSWERS ON RECOMMENDATION FOR REFUSAL ON MARKETING APPLICATION FOR ALPHEON 2, *available at* http://www.emea.europa.eu/pdfs/human/opinion/19089606en.pdf [hereinafter QUESTIONS AND ANSWERS].

[339] *See id.*

[340] Product guidance may be forthcoming in draft form; a concept paper has been issued. *See* CHMP, EMEA, Comm. for Medicinal Products for Human Use, *Concept Paper on Similar Biological Medicinal Products Containing Recombinant Alpha-Interferon*, CHMP/BMWP/7241/2006 (2006), *available at* http://www.emea.eu.int/pdfs/human/biosimilar/724106en.pdf.

[341] *See* CHMP, EMEA, JUNE 2006 PLENARY MEETING MONTHLY REPORT, EMEA/222629/2006 1 (2006), *available at* http://www.emea.europa.eu/pdfs/human/press/pr/22262906.pdf.

[342] *See* QUESTIONS AND ANSWERS, *supra* note 338, at 2.

[343] *See id.*

[344] *See id.*

[345] *See id.*

[346] *See id.*

[347] *See* QUESTIONS AND ANSWERS, *supra* note 338, at 2.

[348] *See* Press Release, BioPartners, BioPartners Remains Positive After Alpheon CHMP Ruling (Apr. 5, 2006), http://www.biopartners.ch/news/300606.htm.

V. THE ACCESS TO LIFE-SAVING MEDICINE ACT (ALSMA)

At present, no regulatory regime exists for approving follow-on biologics in the United States.[349] However, Rep. Henry Waxman (D-Cal.), co-author of the HWA,[350] has introduced ALSMA, a proposal modeled on the HWA intended to speed access to follow-on biologic medicines.[351] This proposal provides an opportunity to assess key policy issues in drafting follow-on biologic approval procedures.

A. Overview

ALSMA would amend section 351 of the PHSA and provide for an expedited process to review follow-on biologics through application to the Secretary of Health and Human Services ("HHS").[352] It is quite evident that this legislation takes as its model the legislation crafted for chemical drugs. Like the HWA, ALSMA creates a system that is heavily weighted in favor of approval of the follow-on applicant drug.[353]

For the purposes of classifying biologics, molecules with several "minor"[354] differences could be considered "comparable" under the statute.[355] These deviations from the reference biologic would be permitted to include differences in heterogeneity profile, degradation patterns, impurities, post-translational events, translation, transcription, and amino acids.[356]

The proposed legislation would require that a follow-on biologic have no "clinically meaningful differences" from the originator biologic, as well as "highly similar principal molecular structural features."[357] Furthermore, an applicant must demonstrate comparability by showing that the follow-on product and the reference biologic have the same mechanism of action

---

[349] To date, the FDA has not issued guidance on approving follow-on biologics. Although there have supposedly been FDA announcements on guidance for insulin and human growth hormone in 2001, and rumors of draft documents in 2002, the FDA has not officially released any such documents. *See* Susan J. Ainsworth, *Biopharmaceuticals*, CHEM. & ENGINEERING. NEWS, June 6, 2005, at 21, 24, *available at* http://pubs.acs.org/cen/coverstory/83/8323biopharmaceuticals.html; Gary C. Messplay & Colleen Heisey, *Follow-On Biologics*, CONTRACT PHARMA, June 2006, at 20, 22, *available at* http://www.hunton.com/files/tbl_s47Details/FileUpload265/1509/ContractPharma_June06.pdf.

[350] Pub. L. No. 98-417, 98 Stat. 1585 (codified as amended at 21 U.S.C. § 355 (2000); 35 U.S.C. §§ 156, 271, 282 (2000 & Supp. IV 2004)).

[351] *See* Access to Life-Saving Medicine Act, H.R. 1038, 110th Cong. (2007).

[352] *See id.* § 3(a)(2)(k)(11). The FDA will likely be the primary reviewer, subject to the Secretary's guidance. *See id.* However, because the bill continuously refers to the Secretary in its provisions, that convention is adopted here.

[353] *See supra* note 165, and accompanying text (describing the presumption in favor of approving generic chemical drugs under the HWA); *infra* notes 370–372 (describing the presumption in favor of approving generic biologic drugs under ALSMA).

[354] H.R. 1038 § 3(a)(2)(k)(4)(A)(ii) (including "minor differences in heterogeneity profile, impurities, or degradation patterns") (emphasis added).

[355] *See id.* § 2(4).

[356] *See id.* § 3(a)(2)(k)(1).

[357] *See id.* § 2(a)(2), § 3(a)(2)(k)(1)(B).

(if known).[358] If the mechanism of action of the reference drug is known, the follow-on applicant must demonstrate comparability for at least one proposed clinical use; if not, the follow-on applicant must demonstrate comparability for each proposed condition of use.[359]

In addition, the product label must indicate that the Secretary of HHS has approved the follow-on product for at least one of the clinical indications of the reference drug and that the method of administration, dosage form, and strength are the same as in the originator biologic.[360] The applicant must also demonstrate good manufacturing practices.[361] A follow-on applicant may use "publicly-available information" from the Secretary's previous determinations or other sources, in order to show that the reference biologic is safe, pure, and potent.[362]

Under the legislation, the applicant would meet with the Secretary of HHS or a designee to agree on what testing, clinical trials, and other requirements are necessary for review of the follow-on molecule.[363] The agreement is then binding on the Secretary.[364]

If the originator biologic firm was earlier required to conduct postmarketing safety studies, the follow-on biologic firm may, at its option, agree with the Secretary to conduct similar studies. However, the Secretary may not, as a condition of approval, compel any postmarketing studies.[365]

The bill allows firms to submit applications for follow-on products that constitute a different form of, or incorporate a change to, the originator biologic.[366] This is permissible if the application contains "sufficient information to establish the safety, purity, and potency" of the follow-on product.[367]

Once the application is submitted and accepted, the Secretary would have 180 days after submission or eight months after acceptance, whichever is earlier, to issue a decision.[368] The final action date may be extended by joint agreement between the Secretary of HHS and the applicant.[369]

Because the proposed legislation is modeled on the HWA, it establishes a similar "approval unless" background rule.[370] The Secretary would be required to approve a follow-on biologic application unless, in relation to the originator biologic: there is insufficient evidence of comparability;

---

[358] *See id.* § 3(a)(2)(k)(1)(C).
[359] *See* H.R. 1038 § 3(a)(2)(k)(4)(A).
[360] *See id.* § 3(a)(2)(k)(1)(C)–(E).
[361] *See id.* § 3(a)(2)(k)(1)(F).
[362] *See id.* § 3(a)(2)(k)(1)(G).
[363] *See id.* § 3(a)(2)(k)(3)(B).
[364] *See* H.R. 1038 § 3(a)(2)(k)(3)(C).
[365] *See id.* § 3(a)(2)(k)(5).
[366] *See id.* § 3(a)(2)(k)(2).
[367] *See id.* § 3(a)(2)(k)(10).
[368] *See id.* § 3(a)(2)(k)(12)(A).
[369] *See* H.R. 1038 § 3(a)(2)(k)(12)(B). Any extension must be signed by both parties no later than thirty days prior to the final action date. *See id.*
[370] *See supra* note 165; *cf.* 21 U.S.C. § 355(j)(4) (2000) (explaining that ANDA applications would also be approved unless certain conditions met, similar to HWA).

there is insufficient evidence of molecular similarity; there is insufficient evidence of mechanism similarity; the dosage, strength, or method of administration differs; or there is insufficient information regarding the comparability of conditions of use.[371] The Secretary could also deny the follow-on biologic application if: follow-on excipients are not safe; manufacturing controls were inadequate; the originator drug has been withdrawn from the market for safety reasons; or the follow-on application contains false statements.[372]

Furthermore, the bill states that the Secretary cannot delay final action on a follow-on biologic request beyond the final action date on the grounds that a third party requested a delay.[373] The bill also states that a court may not enjoin the Secretary from taking final action or stay an approval except by permanent injunction.[374] Such an injunction may not issue unless the entity seeking the injunction demonstrates that it will suffer an injury greater than irrecoverable economic loss.[375] The bill also indicates that any petition[376] or civil action[377] to delay a follow-on drug application requires exhaustion of administrative procedural review.[378] The Secretary may not delay approval of the follow-on application unless the Secretary determines within thirty days after receiving a petition that a delay is necessary to protect the public health. Also, consideration of any petition is to be "separate and apart" from the application review and approval process.[379] The Secretary must report to Congress and the President if he or she grants an extension on the follow-on biologic application or fails to take action by the final action date.[380]

If the Secretary approves the follow-on biologic application, he or she may designate an official name for the biologic product that is the same as the reference product if "[he or she] determines that designation of an official name for a comparable biologic product is necessary or desirable in the interests of usefulness or simplicity."[381] This provision is limited to those biologic products approved under the more stringent definition of "comparable" included in the proposal,[382] rather than the looser approval

---

[371] *See* H.R. 1038 § 3(a)(2)(k)(4)(A)(i)–(ix).

[372] *See id.* In the case of an application containing false statements, the Secretary must provide the follow-on applicant with a detailed explanation for its decision. *Id.* § 3(a)(2)(k)(4)(A)(ix).

[373] *See* H.R. 1038 § 3(a)(2)(k)(13).

[374] *See id.*

[375] *See id.* § 3(a)(2)(k)(13)(B).

[376] A petition includes "any request to the Secretary, without regard to whether the request is characterized as a petition." *Id.* § 3(a)(2)(k)(18)(F).

[377] *See id.* at § 3(a)(2)(k)(18)(B).

[378] *See* H.R. 1038 § 3(a)(2)(k)(18)(B).

[379] *See id.*

[380] *See id.* § 3(a)(2)(k)(14). All approvals made and approvals delayed by petitions (as well as the number of days of delay) must be reported annually to Congress. *See id.* § 3(a)(2)(k)(18)(D).

[381] *See id.* § 3(a)(2)(k)(6).

[382] *See id.*

standard which requires only that "the application and any other information available to the Secretary are sufficient to establish the safety, purity, and potency of the comparable biological product relative to the reference product."[383]

Under the proposed legislation, follow-on applicants may establish "interchangeability" for marketing approval.[384] An interchangeable product is defined as a product that (1) "is comparable to the reference product;" and (2) "can be expected to produce the same clinical result as the reference product in any given patient."[385] Presumably, either the Secretary would be required to make this determination for all applications, or follow-on biologic applicants could request such an assessment.[386] The nature of

---

[383] *See* H.R. 1038 § 3(a)(2)(k)(7). Note that the newer bill contains errors in internal references, likely associated with previous and current versions of the bill. H.R. 1038 references "subparagraphs (A) through (E) of paragraph (4)(A)" in H.R. 1038. *Id.* However, those subparagraphs do not exist. The previous Waxman Proposal, H.R. 6257, states:

> (6) OTHER APPROVAL PROVISIONS. The Secretary shall approve, under the provisions of paragraph (5), an application for a license submitted under paragraph (2), except that the Secretary shall approve such an application that would otherwise be disapproved by reason of one or more of subparagraphs (A) through (E) of paragraph (5), if the application and any other information available to the Secretary contains sufficient information to establish the safety, purity, and potency of the comparable biological product relative to the reference product for the proposed condition or conditions of use for such product.

H.R. 6257, 109th Cong. § 3(a)(2)(k)(6) (2006). The subparagraphs listed in H.R. 6257 correspond to subparagraphs (i), (ii), (iii), (iv), and (v) of paragraph 4(A) of H.R. 1038. *See* H.R. 1038 § 3(a)(2)(k)(4)(A).

[384] *See* H.R. 1038 § 3(a)(2)(k)(5)(B).

[385] *Id.*

[386] This vagueness likely stems from a change in the bill from its previous iteration, H.R. 6257. The current bill, H.R. 1038, added the following section:

> (B) DETERMINATIONS ON INTERCHANGEABILITY.—Subject to subparagraph (C) and paragraph (10), upon issuing a product license for a biological product under subparagraph (A), the Secretary shall make and publish on the following determinations:
> (i) Such product is interchangeable with the reference product for one or more specified conditions of use prescribed, recommended, or suggested in the labeling of the biological product.
> (ii) Interchangeability has not been established.

H.R. 1038 § 3(a)(k)(2)(4)(B). However, H.R. 1038 employs language similar to that of the previous proposed bill, H.R. 6257, in allowing a follow-on biologic manufacturer to request an interchangeability determination by the Secretary:

> (7) INTERCHANGEABILITY DETERMINATIONS FOR COMPARABLE BIOLOGICAL PRODUCTS.—An applicant may request in an original application or supplement to an application that the Secretary make a determination as to the interchangeability of a comparable biological product and the reference product. An applicant may withdraw a request for a determination at any time. A request for an interchangeability determination submitted after the filing of an application shall be considered a major amendment to the application.

this ambiguity may be related to issues associated with differences between statutory drafting in an earlier version and the current proposal.[387] In any event, if the Secretary concludes that interchangeability exists between the follow-on and the reference biologic,[388] the legislation indicates that the follow-on product's label may state that it is interchangeable with the originator biologic for the approved clinical use.[389] ALSMA would eventually require the Secretary to develop guidance for interchangeability, but in the interim he or she could use broad general authority to approve follow-on products as interchangeable.[390]

Like the HWA, ALSMA also contains a provision that grants a 180-day period of marketing exclusivity, here to the first follow-on biologic applicant to show interchangeability.[391] During this time, the Secretary cannot make any follow-on interchangeability biologic determination or approve any other follow-on biologic for the same reference drug.[392] However, unlike the HWA, this legislation would expressly prohibit reference biologic com-

---

H.R. 6257 § 3(a)(2)(k)(7). Compare the relevant provision in H.R. 1038:

> (8) ESTABLISHING INTERCHANGEABILITY FOR COMPARABLE BIOLOGICAL PRODUCTS.—
> (A) IN GENERAL.—In an original application or a supplement to an application under this subsection, an applicant may submit information to the Secretary to demonstrate the interchangeability of a comparable biological product and the reference product. An applicant may withdraw an interchangeability submission at any time. A request for an interchangeability determination submitted after the filing of an application shall be considered a major amendment to the application.

H.R. 1038 § 3(a)(2)(k)(8)(A).

[387] *See* H.R. 1038 § 3(a)(2)(k)(7). This is further evidenced by H.R. 1038's treatment of meetings with follow-on applicants to determine the scope of activities for an application. *Compare* H.R. 1038 § 3(a)(2)(k)(3)(B) ("The Secretary shall meet with a sponsor of an investigation or an applicant for approval of a comparable or interchangeable biological product."), *with* H.R. 6257 § 3(a)(2)(k)(4)(B) ("The Secretary shall meet with a sponsor of an investigation or an applicant for approval of a comparable biological product.").

[388] *See* H.R. 1038 § 3(a)(2)(k)(8)(B).

[389] *See id.* § 3(a)(2)(k)(9).

[390] *See id.* § 3(a)(2)(k)(8)(B) (indicating that "[t]he Secretary may make determinations of interchangeability under paragraph (4)(B) prior to issuing guidance [for interchangeability]"). Paragraph (4)(B) simply provides that the Secretary will assess the follow-on biologic application generally, and ensure that it is subject to exclusivity provisions for the first approved interchangeable biologic. *See id.* § (3)(a)(2)(k)(4)(B).

[391] *See id.* § 3(a)(2)(k)(10).

[392] If the Secretary has approved an interchangeable follow-on biologic application, the Secretary would be permitted to approve a subsequent follow-on biologic license using the same reference product but would be mandated to defer "any determination of interchangeability as to the subsequent biological product until the [first interchangeable follow-on biologic] exclusivity period . . . has expired." *Id.* § 3(a)(2)(k)(4)(C)(ii). In addition, the bill would prohibit approval of a subsequent follow-on biologic interchangeability determination until the earliest of: 180 days of commercial marketing by the first comparable biological product approved as interchangeable; one year after a final court decision or dismissal of all patent infringement cases associated with the drug; thirty-six months after approval if patent litigation is ongoing; or one year after approval if no patent litigation is involved. *See id.* § 3(a)(2)(k)(10).

panies from authorizing production of "generic" versions of their drugs through licensing agreements during the exclusivity period.[393]

Follow-on companies may request a list of relevant patents from the originator biologic firm, and the originator firm must respond within sixty days with such a list.[394] During the two years following the request, the originator firm must update that list within thirty days of the issuance of a new, related patent or license.[395]

These follow-on firms may also notify the patent holder of the reference biologic that a follow-on application is being submitted and that it intends to challenge one or more patents from the list provided by the reference biologic firm.[396] If notice is sent, it must provide a detailed statement specifying the legal basis of why the identified patents are invalid, are unenforceable, or have not been infringed. The follow-on form must also name a judicial district where it consents to being sued.[397] If the reference biologic firm believes its patents have been infringed, it must institute a patent infringement suit on the patents specified in the notice within forty-five days and in the specified judicial district.[398]

If the reference biologic firm does not disclose a relevant patent, it cannot later enforce it against the follow-on applicant.[399] Furthermore, if it discloses a patent but does not pursue an infringement action within forty-five days of being notified, the remedy for any enforcement action by the originator biologic firm is limited to royalties.[400] Prior to the commercial marketing of the follow-on biologic, the recipient of the follow-on firm's notice may not bring an action for a declaration of patent infringement, validity, or enforceability[401] with respect to other patents.[402] Unlike in the HWA, there do not appear to be any data exclusivity or thirty-month stay provisions in the bill, so follow-on firms could theoretically enter at any time after the originator approval.

If the Secretary does not approve a follow-on biologic application, the applicant must be given notice and an opportunity for a hearing within ninety days.[403] Funding for review of these applications appears to be through user fees.[404]

---

[393] *See* H.R. 1038 § 3(a)(2)(k)(10)(A).

[394] *See id.* § 3(a)(2)(k)(17)(A)(i). The patent holder may charge up to $1,000 for the listing. *See id.* § 3(a)(2)(k)(17)(A)(ii).

[395] *See id.* § (3)(a)(2)(k)(17)(A)(iii).

[396] *See id.* § (3)(a)(2)(k)(17)(B).

[397] *See id.*

[398] *See* H.R. 1038 § (3)(a)(2)(k)(17)(C).

[399] *See id.* § 3(b)(1)(C).

[400] *See id.* § (3)(b)(1)(B)(6)(B). The same result occurs if an action is brought within forty-five days but is not maintained through a final decision or if the claim is dismissed with prejudice. *See id.*

[401] *See* 28 U.S.C. § 2201 (2000).

[402] *See* H.R. 1038, § 3(a)(2)(k)(17)(C).

[403] *See id.* § 3(a)(2)(k)(11).

[404] *See id.* § 2(b) (citing FDCA § 735(1)(C), 21 U.S.C. 379(g)(1)(C)); *see also* Jill Wechs-

B. Policy Assessment

1. Overview

As noted above, ALSMA adopts an approach that follows Congress's efforts to regulate chemical drugs. Its overarching theme is a presumption in favor of approving biologic applications. Indeed, the Secretary must approve the application as a background rule, and can reject such applications only if certain conditions are met.[405] This approach not only fails to take into account the unique difficulties and dangers inherent in biologics, but also raises additional significant policy concerns.

In weighing the proposed legislation, policymakers should consider information gaps related to substantive policy issues, the vulnerability of the polity that must bear the risk of policy failure, and the degree of potential harm if the policy fails. Here, the "greater and higher" principle suggests policymakers should err on the side of safety rather than the side of potential economic benefit.[406]

First, there are significant information gaps in our understanding of biologics. These include challenges in the science of follow-on biologics, as well as difficulties in comparability, characterization, and immunogenicity.[407] The FDA itself has indicated that knowledge surrounding critical safety issues is incomplete, particularly for newer, larger protein products.[408] Clearly, the situation for follow-on biologics is extremely different from that of chemical drugs. Chemical drugs are much better understood, characterizable, homogeneous, and have no immunogenicity issues. A background rule favoring the approval for biologics, like that in the proposed legislation, builds an unacceptable risk of policy failure into the system.

Secondly, a policy failure—including adverse clinical reactions and the risk of counterfeit or diverted, ineffective drugs—would place significant burdens upon vulnerable patient populations. This would occur at two levels.

---

ler, *The Push for Generic Drugs Accelerates*, PHARM. TECH., Dec. 2006, at 32, *available at* http://www.pharmtech.com/pharmtech/article/articleDetail.jsp?id=390981&sk=&date=&&pageID=2 (outlining the rationale for ALSMA user fees in the previous bill, H.R. 6257). Each follow-on biologic drug manufacturer desiring FDA assessment would pay a specific fee per molecule. *Id.*

[405] *See* H.R. 1038 § 3(a)(2)(k)(4)(A)(i)–(ix) (enumerating conditions for denial).

[406] This approach is akin to the well-known "BPL Formula" applied to policymaking. The BPL conception assesses the presence or absence of negligence based upon expected costs and benefits; if B, the benefit associated with an intervention, is greater than the product of P, the probability of harm, multiplied by L, the magnitude of the potential loss, the intervention should be put into place. *See* United States v. Carroll Towing Co., 159 F.2d 169, 173 (2d Cir. 1947); *see also* William B. Schwartz & Neil K. Komesar, *Doctors, Damages, and Deterrence: An Economic View of Medical Malpractice*, 298 NEW ENG. J. MED. 1282 (1978).

[407] *See supra* Part II.B (describing the difficulties in creating follow-on biologics).

[408] See FDA, Omnitrope (somatropin [rDNA origin]) Questions and Answers, *available at* http://www.fda.gov/cder/drug/infopage/somatropin/qa.htm (explaining FDA opinion on larger protein product assessments).

First, follow-on forms are likely to be of great interest to the elderly and those at a high risk of being uninsured, such as minority patients.[409] Furthermore, as previously noted, biologics treat some of the most challenging and serious diseases, such as cancer and HIV/AIDS.[410] Hence, the risk of a follow-on policy that assumes approval of these drugs falls squarely on the most vulnerable populations. This result, too, stands in stark contrast to the case of chemical drugs. Since the scope of chemical drugs encompasses a broad array of diseases, those with the most sensitive conditions do not shoulder the entire risk of ineffective generic drugs. Similarly, the U.S. polity as a whole uses a wide range of generic drugs, so highly price-sensitive individuals and underserved vulnerable groups are not assuming the bulk of the risk associated with an abbreviated drug approval regime.

Third, the potential harm from a biologics policy failure is severe. From the damage associated with excipients to the specter of life-threatening immunogenicity, biologics can come with many unknowns and safety issues attached to their use. Indeed, these safety concerns arise even when the drugs are subjected to full, unabbreviated review and scrutiny, and are manufactured by cooperative entities that clearly have an incentive to avoid harm. Again, this circumstance is qualitatively different from that of chemical drugs. Of course, there are potential significant adverse reactions to chemical medications as well. But the issue of immunogenicity and its concomitant severe clinical adverse events are attendant upon all biologics and have proven highly difficult to predict, even with years of intensive research.

Thus, although a policy analysis of abbreviated approvals for chemical drugs conceivably justifies the risk of a regime that results in relatively more approvals in exchange for cheaper prices, an analogous analysis for biologics argues for a different result. The greater and higher principle demands such a conclusion. Important information gaps, the risk to vulnerable patients, and life-threatening harm dictate that policymakers should direct follow-on biologics policy towards more deliberate approval processes, more development of data, and more protections and monitoring. This will ensure vigilance in detecting problems and will avoid a process that causes undue harm in an effort to obtain potentially reduced prices.

---

[409] *See* Healthcare Equality and Accountability Act, H.R. 3459, 108th Cong. §2 (1st Sess. 2003); *see also* Robert Menendez, *We Must End Minority Health Disparity*, Oct. 23, 2003, *available at* http://www.dems.gov/index.asp?Type=B_PR&SEC=%7BF36E20AD-AFDC-4594-8028-5A703D1DB93B%7D&DE=%7BA78DFFC6-4DB1-4F60-9376-E1D61DD16CC2%7D (announcing the Healthcare Equality and Accountability Act and citing a Kaiser study in support of its contention that "[m]inorities are less likely to have health insurance and are less likely to receive appropriate health care services"); Rene F. Rodriguez, *Drug Importation and the Hispanic Physician*, 36 CAL. W. INT'L L.J. 117, 124 (2005) (arguing that alternative drug programs, such as drug importation, create a two-tier system that puts the brunt of policy risk upon the poor).

[410] *See Biologic Cancer Drugs Fastest Growing Class of Biologics in 2006*, *supra* note 103.

An HWA regime of abbreviated approval design, designed for chemical drugs, is simply inappropriate for biologics.

2. Process of Review Under ALSMA

The pervading theme of "approval unless" is illustrated in the process of review put forth by ALSMA. The proposed legislation attempts to solidify the terms of follow-on biologic review before any substantive assessment is begun or data submitted by the follow-on applicant. Because ALSMA compels the Secretary to follow the agreed-on testing and review parameters,[411] the review lacks the flexibility necessary to evaluate a biologic for safety in favor of bureaucratic norms, all while emphasizing less data consideration rather than more.[412]

The legislation takes this approach at the risk of jeopardizing safety. The written agreement between the Secretary and the follow-on applicant can be changed on the basis of certain scientific issues, but only if "a substantial scientific issue essential to determining the safety, purity, and potency of the biological product" is shown.[413] The agreement is presumed immutable, thereby preventing the introduction of additional data and information to evaluate the follow-on product, serving only to accelerate the approval process.[414]

Beyond the process of review, there is wide latitude for the Secretary to approve follow-on biologics. ALSMA permits potential "comparable" biologics to contain "minor differences in heterogeneity profile, impurities and degradation patterns," including differences from the originator biologics in amino acid sequence.[415] Further, follow-on biologics with differences in post-translational modifications are still considered "highly similar" to the originator product.[416]

As a safety and policy matter, this latitude is highly problematic. Differences in amino acid sequences and post-translational changes, as well as other changes, may not in fact be minor for biologics. These changes and differences can completely alter the scientific, medical, and immunogenicity profile of the drug.[417] This clearly creates significant safety issues for patients who are exposed to the product.

---

[411] *See* Access to Life-Saving Medicine Act, H.R. 1038, 110th Cong. § 3(a)(2)(k)(3)(C) (2007). "Any agreement regarding the parameters of design and size of the studies of a biological product . . . reached between the Secretary and a sponsor or applicant . . . shall not be changed after the testing begins." *Id.*

[412] *See id.* § 3(a)(2)(k)(3)(F).

[413] *See id.* § 3(a)(2)(k)(3)(C)(ii).

[414] *See id.* § 3(a)(2)(k)(3)(C).

[415] *Id.* § 3(a)(2)(k)(1)(B).

[416] *See* H.R. 1038, § 3(a)(2)(k)(1)(B)(i)–(iv).

[417] *See, e.g.*, *supra* note 43(describing the single amino acid change in the hemoglobin protein that causes sickle cell anemia).

Indeed, ALSMA in its current form departs even from previous versions of the bill with respect to the rigor of analysis required of a follow-on applicant when showing comparability. Although the previous version of ALSMA mandated that follow-on applicants "demonstrate[ ] [comparability] by thorough characterization of the [follow-on and reference] products"[418] as a part of the follow-on biologic application, this "thorough characterization" provision is absent from the current proposal.[419] The current version of ALSMA has removed this characterization requirement from two key provisions. First, in submitting follow-on applications, thorough characterization is no longer required for a showing that the follow-on biologic and the originator product have highly similar principal molecular features.[420] Second, the Secretary must now accept an application even if it does not justify a conclusion, demonstrated through thorough characterization, that the reference and follow-on biologic are comparable.[421] In fact, although "thorough categorization" is retained in the definitions section of the current proposal,[422] the term appears nowhere else in the bill. This excision of an important requirement creates tremendous challenges for appropriate review and assessment to ensure safety of follow-on biologics.

Additionally, the current proposal creates poor innovation incentives. Critically, there are no data exclusivity provisions for the originator firm, and follow-on firms are able to rely on "publicly available data" to assist the Secretary in assessing that the product is comparable.[423]

These provisions create perverse incentives for potential innovators. Rather than enter the market as an originator, a firm (even if it is the first to develop a biologic) may wait in the wings until a second biologic developer engages in the expensive preclinical, nonclinical, and clinical trials required for BLA approval. Once approved, the first firm might then displace the recognized originator by using a slightly different but comparable molecule with "minor" differences. This is arguably the case in all follow-on situations, because the two firms are using different manufacturing processes or cell lines. Because the process is not subject to any data exclusivity time periods, the follow-on producer is immediately eligible for the abbreviated process, including less testing, fewer clinical trials, and lower costs. The proposed regime tends to favor being second in the marketplace rather than first, turning the incentive for innovation on its head.

---

[418] *See* H.R. 6257, 109th Cong. § 3(a)(2)(k)(1)(B) (2006).

[419] *See generally* H.R. 1038.

[420] *Compare* H.R. 1038, § 3(a)(2)(k)(1)(B), *with* H.R. 6257 § 3(a)(2)(k)(1)(B).

[421] *Compare* H.R. 1038, § 3(a)(2)(k)(5)(A)(i), *with* H.R. 6257 § 3(a)(2)(k)(5)(A).

[422] *See* H.R. 1038 § 2(a)(2)(6).

[423] *See id.* § 3(a)(2)(k)(1)(H). This may raise the issue that was avoided in the Omnitrope case: what is "publicly available data?" *See supra* notes 219–229 and accompanying text (discussing how the FDA avoided a Takings Clause issue while using proprietary data in approving Omnitrope). Without statutory attention, this will be another issue under ALSMA that will necessarily need to be litigated.

In addition, ALSMA also has a gaping hole that allows "other applications" from follow-on manufacturers for molecules that need not show comparability or demonstrate the very characteristics that are supposedly required by the bill.[424] This additional application provision would allow follow-on biologic firms to "submit an application for a biologic product that differs from, or incorporates a change to, the reference product with respect to one or more characteristics including a difference in safety, purity, or potency."[425]

This provision is extremely troubling from a safety perspective. Once again, the focus is on approval rather than other concerns. A follow-on molecule with purportedly "minor" differences from the originator product—differences that could have potentially major clinical implications—may be approved even though it is not comparable to an originator product.[426]

Hence, the proposed legislation creates an express presumption of approval beyond the broad definition of "comparable" or "principal molecular features." Even molecules otherwise ineligible for approval could be approved without the requisite data or information normally mandated for follow-on applications. This is because the Secretary

> shall approve such an application that would otherwise be disapproved by reason of one or more of subparagraphs (A) through (E) of paragraph (4)(A) if the application and any other information available to the Secretary are sufficient to establish the safety, purity, and potency of the comparable biological product relative to the reference product for the proposed condition or conditions of use for such product (emphasis added).[427]

Hence, even though such drugs need not meet all of the criteria for approval under the bill, are not comparable to the reference biologic, and would not meet the standards required for an original BLA application, the Secretary is mandated to approve them. Interestingly, the Secretary must approve the biologic on the basis of the application "and any other information available"[428]—a vague standard that constitutes an about-face from

---

[424] *See* H.R. 1038 § 3(a)(2)(k)(2). These characteristics include: "data on comparability, comparability of principal molecular structure, posttranslational events, infidelity of translation or transcription, amino acid sequence, polysaccharide repeating units, glycosylated protein product structure, polynucleotide purine and pyrimidine bases, partly definable biological products, mechanisms of action, conditions of usage, route of administration, dosage, and strength." *Id.* § 3(a)(2)(k)(1)(A)–(H).

[425] H.R. 1038 § 3(a)(2)(k)(2). This provision does not apply to the reference biologic firm. *See id.* § 3(a)(2)(k)(2).

[426] *See supra* Part III.A.2 (discussing the potential for immunogenicity reactions to therapeutic proteins and the broad spectrum of characteristics associated with them).

[427] *See* H.R. 1038 § 3(a)(k)(4)(C)(7).

[428] *See id.*

the desired bright line limits for challenges[429] and from drafters' approval perseveration.

These provisions create additional perverse policy incentives beyond those represented by comparable follow-on applications. They allow approval of very different molecules from an originator drug through an abbreviated process.

By permitting such an extensive array of differences between the follow-on molecule and the reference drug,[430] ALSMA would incentivize firms to wait for others to weather the extensive BLA process and then ride their application approval coattails by claiming comparability based upon the loose statutory definitions of acceptability. The approval process hence extends the disincentive to be first from comparable molecular forms to a wide array of substantively noncomparable forms.

The subordination of safety to approval is also apparent in other portions of ALSMA. The bill's virtual exemption of follow-on biologics from postmarketing safety studies seriously compromises patient safety. Unlike the E.U. system, which has recognized the need for and designed its sytem to incorporate postmarketing studies, the Secretary is limited to requesting such studies and cannot condition approval on cooperation from the follow-on firm.[431] In addition, this permissive assent by the applicant is only applicable if the originator firm was required to perform safety studies. With the high cost of bringing a biologic to market, it is unlikely that any follow-on firms would agree to willingly perform such extensive and expensive studies and activities. Yet pharmacovigilance is essential in the follow-on market because the scientific information gap makes it difficult to predict adverse clinical events resulting from seemingly innocuous changes.[432] Again, it should be emphasized that the E.U. mandates such pharmacovigilance studies as well as risk management studies and plans as a standard component of a biosimilar application.[433] To create a system in which the Secretary is prohibited from doing likewise is to ignore the very real dangers inherent in follow-on biologics.

---

[429] *See infra* note 440 and accompanying text (describing the limited information that may be considered by the Secretary in follow-on civil challenges).

[430] *See supra* note 424 and accompanying text (describing the allowable differences in follow-on molecules eligible for approval through the "other applications" abbreviated process).

[431] *See* H.R. 1038 § 3(a)(k)(4)(C)(5). "If the Secretary has agreed with the sponsor of the *reference* product, at the time of approval or any time thereafter, that the sponsor shall conduct one or more postmarketing safety studies, a person submitting an application for a biological product under paragraph (1) may agree . . . to conduct a similar post-marketing safety study . . . . The Secretary shall not, as a condition of approval, propose any additional postmarketing studies for such biological product." *Id.*

[432] *See supra* Parts III.A and III.B (discussing patient safety issues associated with biologic excipients and substantive proteins).

[433] *See European Public Health Alliance Draft Pharmacovigilance, supra* note 287 (stating that pharmacovigilance and risk management efforts must be included in E.U. biosimilars applications).

Policy and safety concerns also attend the citizen petition provisions of the ALSME. Although Congress has recognized that there have been abuses associated with the citizen petition processes,[434] the burdensome standard in ALSMA for citizen petitions may be permanently damaging to smaller firms and safety.[435] For example, courts may not enjoin the Secretary from taking final action on a follow-on biologic except by a court's permanent injunction, "based on an express finding of clear and convincing evidence."[436] Furthermore, those

> seeking to have the Secretary refuse to take or otherwise to deter final action by the final action date—
> (A) [must have] prevailed on the merits of the person's complaint against the Secretary;
> (B) [must show they] will suffer imminent and actual irreparable injury, constituting more than irrecoverable economic loss, and that will also threaten imminent destruction of such person's business; and
> (C) [must show they] ha[ve] an interest that outweighs the overwhelming interest that the public has in obtaining prompt access to a comparable biologic product.[437]

This provision is hasty and potentially irresponsible. Biotechnology firms subject to potential follow-on competition may have only a single product and limited resources for a legal fight, which could cause them to be bankrupted by litigation. This provision could destroy firms because it is unlikely that they could show all three factors the bill requires in time for effective court entry and impact. It provides incentives for larger generic and brand name firms to prey on smaller biotechnology firms that

---

[434] Since there are no penalties or sanctions for filing meritless citizen petitions, and generic drug approval cannot occur until these petitions are resolved, the incentives for filing frivolous petitions are high. This is particularly true when one considers that twenty of twenty-one brand-name firm FDA citizen petitions filed since 2003 have been rejected without ramifications to the brand-name firms. Yet long delays for generic firms in resolving these petitions are common. *See* Lower Prices Reduced with Increased Competition and Efficient Development of Drugs Act, S. 2300, 109th Cong. (2006); Marc Kaufman, *Petitions to FDA Delay Generic Drugs, Critics Say*, WASH. POST, July 3, 2006, at A1. An FTC study also identified FDA citizen petitions as a potential area of competitive concern. *See* FTC, GENERIC DRUG ENTRY PRIOR TO PATENT EXPIRATION: AN FTC STUDY (2002), *available at* http://www.ftc.gov/os/2002/07/genericdrugstudy.pdf. Congressmen on both sides of the aisle have expressed concern about the use of FDA citizen petitions and have introduced legislation to ban this practice, noting that the practice costs consumers millions of dollars a month. *See* S. 2300 § 5; Kaufman, *supra*, at A1 (discussing legislation presented by Sen. Deborah Stabenow (D-Mich.) and Sen. Trent Lott (R-Miss.), and estimating that the delay in approving a generic version of the antidepressant Wellbutrin XL alone is costing consumers $37 million a month).

[435] *See* H.R. 1038 § (3)(a)(2)(k)(13).

[436] *Id.* § (3)(a)(2)(k)(13).

[437] *Id.*

have only single products and limited resources for legal challenges. Such a regime may also stifle innovation and stop firms from entering into the market.[438]

Further, the bill indicates that any citizen petition or other civil action shall be separate from approval of the follow-on biologic, and any delays shall only be for the protection of public health.[439] Beyond solidifying the incentive for follow-on product firms to target smaller biotechnology companies, this provision creates an incentive to claim a safety or public health concern in each petition. Real and substantive safety concerns will be discounted as standard parts of a petition from an originator company. Unfortunately, this issue is exacerbated because

> The Secretary shall take final agency action on the petition not later than 180 days after the date on which the petition is submitted and that the Secretary shall not extend such period, even with the consent of the petitioner, *for any reason, including based on submission of comments relating to the petition or supplemental information supplied by the petitioner* (emphasis added).[440]

In addition, recall that the Secretary "shall approve [a comparable biologic product] . . . if the application and *any other information* available to the Secretary are sufficient to establish the safety, purity, and potency of the comparable biological product."[441] These provisions encourage the Secretary to make a decision irrespective of whether he or she has obtained the information necessary to make an informed decision.

Moreover, the bill would virtually guarantee patent infringement suits. The follow-on firm must notify the originator firm when a follow-on application is filed. The originator firm must act within forty-five days of receiving such notice by filing a patent infringement suit,[442] or lose the opportunity to obtain treble damages in court. The bill, however, does not provide for a thirty-month stay like the HWA.[443] Hence, the originator firm is in effect mandated to sue the follow-on applicant to avoid being limited to royalties.[444] Since there are no other mechanisms for addressing the issue of conflict or potential conflict over patent rights, suit is the only alternative. This is a highly wasteful and inefficient means to address the issue

---

[438] *See, e.g.*, Bryan A. Liang, *The Anticompetitive Nature of Brand Name Firm Introduction of Generics Before Patent Expiration*, 41 ANTITRUST BULL. 599, 615 (1996) (discussing the use of litigation to "discipline" competitors that might be better off in some other market).

[439] *See* H.R. 1038 § (3)(a)(2)(k)(18)(A)(i)(I).

[440] *Id.* § (3)(a)(2)(k)(18)(A)(ii).

[441] *Id.* § (3)(a)(2)(k)(7) (emphasis added).

[442] *See id.* §§ (3)(a)(2)(k)(17)(B)–(C), (3)(b)(1)(6)(B).

[443] *See* 21 U.S.C. § 355(j)(5)(B)(iii) (2000).

[444] *See* H.R. 1038 § (3)(b)(1)(6)(B).

because it does not correct poor incentives that already exist under the current HWA regime.[445]

ALSMA proposes to fund its regulatory regime through user fees.[446] This, too, is highly problematic. User fees have been intensely criticized by many commentators.[447] To have follow-on applicants fund their own review process, at current prices of around a million dollars a molecule,[448] would make the entity performing that review exceedingly dependent upon the industry whose products they are responsible for rigorously assessing. Such a situation might subject the agency to congressional budget cuts and conflict of interest charges, which would impede the agency from filling its mission effectively. Hence, the user fees approach, which has already resulted in potential safety concerns in other pharmaceutical contexts, has no place in the challenging and high stakes review of follow-on biologics.

Finally, ALSMA focuses only on the potential entry of follow-on firms and products. This legislation, and the HWA chemical predecessor, do not address issues of access to essential medications. Even assuming a large percentage discount for biologics in the follow-on market—a dubious assumption at best[449]—the cost of these drugs would still be too high for many vulnerable patients. Hence, ALSMA perpetuates disparities of care that continue to plague the U.S. health delivery system.

C. Naming

Because of the significant safety concerns associated with biologics and the difficulty of predicting critical clinical reactions, aggressive risk management and pharmacovigilance are imperative. One area of concern that has been properly addressed by neither the EMEA nor ALSMA is the naming of follow-on biologics.

---

[445] Under the current system, when brand-name firms do not sue, the generic firm cannot seek a declaratory judgment to address the potential uncertainty with respect to patent rights. *See infra* note 519 and accompanying text (discussing declaratory judgment issues and the resulting incentives).

[446] *See* H.R. 1038 § (2)(b); *see also* Jill Wechsler, *The Push for Generic Drugs Accelerates*, PHARMACEUTICAL TECHN., Dec. 2, 2006, at 3, *available at* http://www.pharmtech.com/pharmtech/article/articleDetail.jsp?id=390981&sk=&date=&&pageID=2 (outlining rationale for user fees in previous version of ALSMA, H.R. 6257, 109th Cong. (2006)).

[447] *See, e.g.*, Phil B. Fontanarosa et al., Editorial, *Postmarketing Surveillance—Lack of Vigilance, Lack of Trust*, 292 J. AM. MED. ASS'N 2647, 2647 (2004); Gary W. Lawson, Letter to the Editor, *FDA Dependence on Drug Industry*, 97 J. NAT'L MED. ASS'N 1039, 1039 (2005); Marcia Angell, Op-Ed, *What Ails the FDA? Payola*, BOSTON GLOBE, Mar. 10, 2005, at A15; Alexandra Marks, *How Drug-Approval Woes Crept up on FDA: Critics Charge Conflict of Interest in a System Where Pharmaceutical Giants Fund the Regulatory Process*, CHRISTIAN SCI. MONITOR, Nov. 26, 2004, at 2.

[448] *See* Prescription Drug User Fee Rates for Fiscal Year 2007, 71 Fed. Reg. 43,780 (Aug. 2, 2006) (announcing user fees for NDAs of roughly $900,000 per review).

[449] *See infra* note 499 (noting that discounts from originator prices may be only roughly ten percent).

Traditionally, the standard means for identifying a particular pharmaceutical is through its International Nonproprietary Name ("INN").[450] Under the auspice of the World Health Organization, INNs are internationally recognized nonproprietary "generic" names of drugs that allow providers to clearly identify particular pharmaceutical substances.[451]

Because a drug's trade name may vary between manufacturers and countries, the INN system is essential to patient safety. By providing a standard name for the active ingredient of drugs, it ensures that physicians will prescribe the right drugs to patients. The system also facilitates communication about a particular drug around the world.[452]

Although such an approach may be appropriate for easily characterized chemical drugs,[453] it is often not appropriate for biologics. The FDA has noted that: "INNs should not be used to imply pharmacologic interchangeability of products with the same active ingredient(s) when no credible scientific data exist that [so] demonstrate . . . ."[454] Unfortunately, at the present time, the INN system applies the same name across different forms of biologics.[455] For example, the brand-name forms of human growth hormone, Genotrope and Humatrope, have the same INN: somatropin.[456] Further, the biosimilars that used these branded forms as their reference molecules in the E.U., Omnitrope and Valtropin, also share that same INN.[457] Similarly, under ALSMA, the Secretary may allow follow-on biologics to have the same name as that of the originator drug.[458] Indeed, as with the E.U. system, there would be no need to demonstrate interchangeability before identical names are applied to the follow-on product.

But because identical INNs may cause some countries to mandate or encourage the use of cheaper follow-ons,[459] biologics and follow-ons with the same INN may present a significant patient safety risk. For example, from a practical patient and provider point of view, a patient would receive a prescription from her physician for somatropin. The patient, depending upon the regulatory regime to which he or she is subject, would be given one of a variety of brand name or follow-on forms. Critically, the provider would not know which form of the drug was given to the patient. The problem would be exacerbated if the patient took different forms

---

[450] *See* U.S. FDA Considerations, *supra* note 59.

[451] *See id.*

[452] *See id.*

[453] *See id.* ("With small molecular products, there is a long history to support the use of various scientific approaches to establishing 'bioequivalence' between products with the same active ingredient(s) produced by different manufacturers.").

[454] *Id.*

[455] U.S. FDA Considerations, *supra* note 59.

[456] *Id.*

[457] *Id.*

[458] *See* Access to Life-Saving Medicine Act, H.R. 1038, 110th Cong. § (3)(a)(2)(k)(6) (2007).

[459] *See* U.S. FDA Considerations, *supra* note 59. ("The FDA is concerned that some countries may be using the INN as an indicator of interchangeability.").

of the drugs because of varying supplies in the pharmacy, used multiple pharmacies, or was on the drug for an extended period of time. If the patient experienced an adverse reaction, the physician would only be able to report the INN, not the specific form of the drug. Because of the potential delay before significant clinical reactions arise, it might require significant resources to trace the specific drug forms taken by the patient, assuming the information is acquirable at all.

Varying state laws regarding drug substitution by pharmacists make the situation even more complex.[460] Although a common INN in the United States does not necessarily drive drug choice, therapeutic substitutions may occur, and pharmacists may dispense drugs that are chemically different (i.e., not AB-rated forms) but theoretically result in the same therapeutic outcome.[461] Managed care and other plans use this technique to decrease costs.[462] Yet as a patient safety concern, even for chemical medicines,

> [w]ithin a particular class of medications, there are often many drugs available to physicians for their patients. In one patient, only one of those medications may be tolerated and be of benefit, while another patient may only tolerate and benefit from another of the drugs available. With the well-recognized individual variability in response to medications, there is no way of knowing, other than through a thorough approach to each person's particular circumstances, which drug or drugs will be of benefit to an individual patient, or which will not have deleterious side effects.[463]

If drugs are using the same INN, therapeutic substitution could result in significant patient safety concerns, since if there is an adverse clinical event, follow-on forms using the same INN could not be easily tracked or identified.[464] Such a system would hamper any efforts to rapidly identify and address important drug-associated reactions across large groups of patient populations. Hence, different names for originator and follow-on biologics are vital for effective tracing and monitoring, particularly since follow-ons will have limited safety trials and early detecting of adverse

---

[460] *See id.*

[461] *See* Norman V. Caroll, *How Effectively Do Managed Care Organizations Influence Prescribing and Dispensing Decisions?*, 8 AM. J. MANAGED CARE 1041, 1042 (2002).

[462] Jesse C. Vivian, *Legal Aspects of Therapeutic Interchange Programs*, 28 U.S. PHARMACIST 58, 58 (2003).

[463] Am. Coll. of Rheumatology, Position Statement: Therapeutic Substitution (Mar. 13, 2004), http://www.rheumatology.org/publications/position/therasubs.asp. Therapeutic substitution is a longstanding practice and includes a large fraction of drugs. *See* Paul L. Doering et al., *Therapeutic Substitution in a Health Maintenance Organization Outpatient Environment*, 22 ANNALS OF PHARMACOTHERAPY 125, 126 (1988) (reporting that about thirty percent of HMO pharmacy plans allow therapeutic substitution).

[464] *See* U.S. FDA Considerations, *supra* note 59.

patient reactions tied to specific follow-on forms are necessary to ensure patient safety.[465]

VI. A Proposed Regulatory Regime

Scientific, safety, population, and policy concerns must be considered to effectively address the issues that arise from follow-on biologics. In particular, a proposal should expressly take into account information gaps, vulnerable populations, and the high potential for harm. In addition, Congress should consider the lessons learned by the E.U. when promulgating its follow-on biologics policy. This section presents an annotated[466] federal legislative proposal describing a regulatory regime that attempts to attend to these critical areas.

A Bill

H. R. _____

To amend the Public Health Service Act to provide for the licensing of comparable biological products, and for other purposes.

---

[465] The FDA has noted: "The issue of interchangeability is not an issue of nomenclature, but a scientific question that needs to be decided on its own merit. The question of nomenclature is more relevant to concerns about pharmacovigilance and the prevention of inappropriate substitution." Unfortunately, the FDA is U.S.-centered in its policy concerns. The FDA has indicated it sees no reason to change INN methodology for naming because "[i]t would be the U.S. FDA's preference that INNs continue to be granted based only on the molecular characteristics and pharmacologic class of the active ingredient(s). Regarding similar protein products, this view is predicated on the situation in the U.S. . . . These mechanisms may not exist in other countries." *Id.* This is a shortsighted and potentially dangerous view given the issues associated with importation and adulterated, diverted, or counterfeit drugs. *See supra* notes 101–147 and accompanying text (describing instances of fake, diverted, and adulterated biologic drugs in the United States). Indeed, this is particularly problematic in light of the FDA's acknowledgment that "[a]s of today, FDA has not determined how interchangeability can be established for complex proteins." U.S. FDA Considerations, *supra* note 59.

Other commentators have indicated the particular importance of identifying follow-on products in the post marketing period:

> The onus for detecting [patient safety] problems relies very much on postmarketing surveillance. It will not be possible to do accurate postmarketing surveillance unless biosimilars are clearly differentiated from the innovator product, to allow problems to be traced back to the correct source.

Roger, *supra* note 17, at 343.

[466] An unannotated version of the bill appears in the appendix.

A BILL

To amend the Public Health Service Act to provide for the licensing of comparable biological products and for other purposes.

Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled,

Section 1. Short Title.

This Act may be cited as the "Biologic Drug Safety and Access to Medicines Act."

Section 2. Findings.

Congress makes the following findings:

(1) Biological drugs, also known as biologics, created through the use of recombinant DNA technologies and other biotechnology means have provided significant and important health benefits to patients, including those with the most severe and debilitating diseases such as cancer, AIDS, and hepatitis.

(2) Biologics are highly complex, and their research, development, and manufacture are uniquely characterized and distinguished from the research, development, and manufacture of chemical drugs.

(3) Because of their complexity, biologic characterization and their mechanisms of action may be much more difficult, if not impossible, to describe compared with smaller, less complex chemical drugs.

(4) Because of their complexity and size, biologics may induce severe, adverse, and unwanted immunologic reactions in patients.

(5) Biologic drugs have intellectual property patent protection, which has promoted their development and rewarded these originator companies through market exclusivity rights.

(6) Important biologic drugs have patent terms that have ended or are to end within the next several years.

(7) For the greatest benefits to inure to patients, biologics with patent protections that have ended should be subject to competition from appropriately safe, comparable biologic products, also known as follow-on biologic drugs.

(8) Many patients, particularly minorities, the uninsured, and the underinsured, cannot access biologic products at current prices.

(9) Since follow-on biologic products are more complex than traditional chemical medicines, follow-on biologic products require a review of applications that is different from chemical medicines as well as tailored specifically for biologic molecules.

(10) Other areas of the world, including the European Union and Australia, have created regulatory systems to assess, approve, and deny follow-on biologic applications and products.

(11) Biologic drugs are generally supplied as clear fluids and are sensitive to temperature and other environmental conditions.

(12) Biologic drugs have been counterfeited and harmed patients.

(13) Biologic drugs require secure and traceable supply chains to ensure appropriate transport and authenticity, and should not be accessed using nontraditional or nonstandard supply means.

This section of findings is the basis for the rest of the Act. It illustrates the context of biologics, including their regulation under the Public Health Service Act and the issues associated with their unique safety and complexity concerns. In addition, it expressly recognizes the cost of biologics and the effect that has on vulnerable patients—both in terms of diseases and access—who are in need of these drugs. The findings also indicate that regulating such drugs is different than regulating traditional chemical drugs, which necessitates legislative attention.

Definition of terms are provided next.

Section 3. Definitions.

Section 351(i) of the Public Health Service Act (42 U.S.C. 262(i)) is amended—

(1) by striking "In this section, the term biological product means" and inserting the following:

"In this section:

(1) The term 'biological product' means"; and (2) by adding at the end the following:

> **"(2) The term 'comparable biological product application' means an application for a license of a biological product containing the same, or similar, active ingredient as a biological product for which a license has been approved under subsection (a). A comparable biologic application is a human drug application under section 735(1)(C) of the Federal Food, Drug, and Cosmetic Act.**
>
> **"(3) The term 'reference product' under this Act means the single licensed biological product, approved under subsection (a) or subsection (k), against which a comparable biologic product is evaluated for demonstration of safety, potency, or purity."**

Up until this point, this bill's definitions are similar to that within ALS-MA, with the exception that the proposed section (2) above does not include the term "abbreviated," since this follow-on review process is substantively different from ANDA applications and will likely involve some human clinical testing. Continuing with definitions,

> **"(4) The term 'comparable' in reference to a comparable biologic product application means the absence of clinically meaning-**

ful differences between the comparable biological product and the reference product in terms of the safety, purity, and potency associated with the clinical indication for which the comparable biologic product application is applying for approval. Such absence of clinically meaningful differences between the comparable biological product and the reference product shall be assessed based upon—

"(A) data derived from chemical, physical, and biological assays, other non-clinical laboratory studies; and

"(B) data from clinical study or studies sufficient to confirm safety, purity, and potency for appropriate conditions of use that the comparable biologic is applying for, which the reference product is licensed and intended to be used for.

"Any studies under subparagraph (B) shall be designed to avoid duplicative and unethical clinical testing.

"(5) The term 'thorough characterization' means an analysis of structural and functional features based upon appropriate analytical and functional testing sufficient to identify, and should be focused upon determining, differences between the comparable biologic product and the reference biologic product relevant to safety, purity, potency, and use."

In this section, in contrast to the approaches of ALSMA and the E.U. regulation,[467] follow-on biologic molecules are only considered as comparable for specific applied-for uses. Hence, nonclinical and clinical testing must expressly show comparability for each clinical use desired for marketing approval. This ensures that any extension of an approved use that is not clinically tested will not be permitted and sets the policy focus on caution rather than indiscriminate approval.

This section, however, mirrors ALSMA, which expressly notes that all clinical studies should be designed to avoid duplicative and unethical testing.[468] Further, with respect to "thorough characterization," this proposal goes beyond either the HWA or ALSMA and includes both structural and functional features. It mandates that the focus of testing should be identifying and determining differences relevant to safety, purity, and potency, as well as use of the molecule. Consistent with the E.U. approach, follow-on biologics must therefore be tested for differences in function to ensure the safety of the molecule through identification of potential differences in unwanted reactivity and clinical efficacy.[469]

Interchangeability is also an important concept to define, and this proposal adopts a definition similar to ALSMA:

---

[467] *See* Access to Life-Saving Medicine Act, H.R. 1038, 110th Cong. § (3)(a)(2)(k)(4)(A) (2007).

[468] *See Id.* § (2)(4)(B).

[469] *See* GUIDELINE: NON-CLINICAL AND CLINICAL ISSUES, *supra* note 263, at 4.

"(6) The term 'interchangeable' means that a biological product contains an active ingredient or ingredients with principle molecular structural features comparable to the reference product, and that the comparable biological product can be expected to produce the same clinical result as the reference product in any given patient in the condition or conditions of use for which both products are labeled."

With respect to administrative definitions:

"(7) The term 'process for the review of a comparable biological application' means, with respect to a comparable biological product application, the procedural activities of the Secretary, with respect to the review of human drug applications and supplements as defined in section 735(6) of the Federal Food, Drug, and Cosmetic Act, except as otherwise defined herein.

"(8) The term 'final action' means, with respect to a comparable biological product application, the Secretary's issuance on the final action date of a final action letter to the sponsor of a comparable biological product application under this Act, which may—

"(A) approve the application; or

"(B) disapprove the application and set forth in detail an enumeration of the specific deficiencies in the particular application and of the specific, enumerated actions the sponsor would be required to take in order for the sponsor to receive a final action letter that addresses the deficiencies of the application, which then may lead to approval of such application.

"Under subsection (B), addressing specific deficiencies of the particular applications and engaging in the specific enumerated action does not require the Secretary to approve the application, if other deficiencies are identified associated with the original application, or on the basis of further, submitted information.

"(9) The term 'final action date' means, with respect to a comparable biological product application, the date that is eighteen calendar months following the Secretary's acceptance of a sponsor's submission of such application, except that the final action date hereunder may be extended for such period of time as is agreed to by the Secretary and the sponsor of such application in a jointly executed written agreement that is countersigned by the Secretary and the sponsor of such application no later than thirty days prior to the final action date provided for by this subsection.

"(10) The term 'reviewing division' means the division responsible for the review of an application for approval of a biological product (including all scientific and medical matters, chemistry, manufacturing, and controls)."

The proposal herein differs from ALSMA in several important ways. First, the definition of "final action" is not merely approval or disap-

proval along with steps necessary to obtain approval.[470] Instead, to protect against a claim by applicants that because they have addressed all issues in a disapproval, their application must therefore be approved, this proposal states specifically that any additional information discovered by the Secretary in the original materials, or that arises from the resubmitted application, can be grounds for denying approval of the follow-on biological application. This allows flexibility in the follow-on application assessment that is absent in ALSMA. In addition, the final action date is changed to reflect E.U. experience in reviewing biosimilars. Because of the virtually identical time of roughly eighteen months taken by the E.U. regulators to assess, review, and approve both biosimilar versions of human growth hormone,[471] this experience is used here as a reasonable standard for U.S. regulatory approval time. The "reviewing division" section is identical to ALSMA.[472]

To provide appropriate information to those firms that wish to enter into the follow-on biologic market and to assist them in doing so, it is imperative that the law provides clear guidance for the application process. Contrary to the individualized approach adopted in the E.U., which provides guidance only by specific molecule and does not categorize requirements by drug characteristics,[473] facilitation of follow-on drug applications should focus on the relevant extant knowledge of the biologic, which then can drive scientific requirements for assessment of follow-on forms. Hence, the first step for such guidance is for biologics to be appropriately categorized.

Section 4. Categorization of Certain Biological Products.

(a) IN GENERAL.—Section 351 of the Public Health Service Act (42 U.S.C. §262) is amended—

(1) in subsection (a)(1)(A), by inserting after "biologics license" the following: ," or comparable biologics license,"; and

(2) by adding at the end the following subsection:

"(k) REGULATION OF COMPARABLE BIOLOGICAL PRODUCTS.—

"(1) BIOLOGICAL PRODUCT CATEGORIZATION.—The Secretary shall categorize all currently approved reference biological products. The Secretary shall:

"(A) utilize the following categories that shall be amended, changed, or revised from time to time, subject to subsection (C) herein:

"(I) Presumptively Well-Known Mechanism(s) of Action

---

[470] Access to Life-Saving Medicine Act, H.R. 1038, 110th Cong. § (2)(a)(7) (2007).

[471] *See supra* note 332 and accompanying text.

[472] H.R. 1038 § (2)(a)(9).

[473] *See, e.g.*, GUIDELINE: INSULIN, *supra* note 265; GUIDELINE: SOMATROPIN, *supra* note 266; GUIDELINE: GRANULOCYTE-COLONY STIMULATING FACTOR, *supra* note 268; GUIDELINE: ERYTHROPOIETINS, *supra* note 267.

"(A) Without Known Adverse Immunogenicity;
"(B) With Known Adverse Immunogenicity;
"(II) Partially Known Mechanism(s) of Action
"(A) Without Known Adverse Immunogenicity;
"(B) With Known Adverse Immunogenicity;
"(III) Unknown Mechanism(s) of Action
"(A) Without Known Adverse Immunogenicity;
"(B) With Known Adverse Immunogenicity.

"(B) allow any person to petition the Secretary to have a currently approved reference biological product moved from one category to another.

"(C) review the categories for approved reference biological products biannually to ensure substantive scientific relevance, effectiveness, and efficiency for review of comparable biological product applications."

Currently there are more than 300 biologic drugs approved in the United States.[474] However, as previously noted, the mechanisms of action may not be known for all these drugs, and issues of immunogenicity represent significant safety issues for patients who take these drugs and manufacturers who make them.[475] It is clear that well-known, well-understood drugs with no or low immunogenicity are less of a risk than newer drugs without determined mechanisms of action that have been associated with immunogenicity. Hence, these critical characteristics must be identified for the purpose of determining what regulatory basis should be applied to each.

For example, under category (I)(A), a biologic that is well known and has little or no adverse immunogenicity, such as virtually all of the FDCA-approved biologic drugs, would need less nonclinical and clinical information in a follow-on biologic application than a category (III)(B) drug, whose mechanism is not well known and has been reported to be associated with adverse immunogenicity.

Further, given the fact that the science of testing and medical assessment advances over time, the categories will need to be altered and biologics moved from category to category. Section 4 allows such actions to be taken either through petition by external parties or by periodic internal review. The identification of specific categories and categorical requirements provides flexibility and would allow regulators and industry to understand the specific needs associated with a particular molecule, and therefore to be more responsive to changes in scientific and medical knowledge.

---

[474] *See* Biotech. Industry Org., Biotechnology Industry Facts, http://www.bio.org/speeches/pubs/er/ statistics.asp (last visited Apr. 13, 2007).

[475] *See supra* Part III.A.2 discussing safety issues associated with immunogenicity.

On the basis of this categorization system, follow-on biologic drug applicants must receive guidance as to what kinds of studies and information are relevant for each category.

"(2) GUIDANCE BY CATEGORY.—The Secretary shall issue guidance by categories denoted in subparagraph (1) herein on the general requirements that all comparable biological product applications must fulfill for application review. This guidance shall include information and requirements associated with comparable biological product applications, including:

"(A) pharmacokinetic studies;

"(B) pharmacodynamic studies;

"(C) pharmacokinetic/pharmacodynamic studies;

"(D) clinical comparability studies;

"(E) immunogenicity studies;

"(F) adverse reaction studies;

"(G) pharmacovigilence monitoring;

"(H) traceability methodology;

"(I) manufacturing methodology;

"(J) quality assessment methodology;

"(K) post-marketing studies; and

"(L) any other assessments necessary to evaluate the comparability, safety, purity, potency, and function of the comparable biological product.

"(3) SPECIFIED INFORMATION.—Notwithstanding the previous section requirements, the Food and Drug Administration shall require the comparable biological product applicant to provide evidence of investigation of immunogenicity issues including data from investigation prior to the comparable biological product application; to monitor the clinical safety of its drug during the post approval period; and within its application, to provide for a risk specification and pharmacovigilance program plan that includes a description of potential safety issues that may be a result of differences in the manufacturing process from the originator biologic."

This section parses out and formalizes mandated sections employed by the E.U. biosimilar application regime.[476] Importantly, for U.S. follow-on applicants, it specifies the individual characteristics for any application and gives the Secretary, the FDA, and industry a cogent list of factors to focus upon in initiating, evaluating, and submitting such an application. Importantly, in (2)(L), it incorporates the requirement that the function of the molecule also be considered, in addition to the safety, purity, and potency of the product, as in ALSMA. Further, because it is roughly modeled upon the E.U. system, the Secretary or the FDA may wish to consult formally or informally with their European counterparts to discuss and deter-

[476] *See* GUIDELINE: NON-CLINICAL AND CLINICAL ISSUES, *supra* note 263, at 5–6.

mine best strategies in disseminating guidance for each factor. This section replaces ALSMA's statutorily defined conception of comparable or highly similar molecular structures[477] and instead relies on the scientific expertise available to the Secretary and FDA to define these critical and scientific determinations.[478]

Importantly, the proposal would not leave the requirements for these studies to the discretion of the applicant. Instead it mandates that the FDA determine the requirements based on scientific need. In addition, it eliminates the provision of ALSMA that "[t]he Secretary shall not, as a condition of approval, propose any additional post-marketing studies."[479] Because that provision potentially allows approval under adverse safety conditions, it undermines the Secretary's authority to ensure patient welfare and inappropriately takes safety assessments outside the scientific arena. Further, this section does not permit wholesale differences between the reference biologic product and the follow-on product,[480] nor does it allow approval of noncomparable products that are substantively different, as does ALSMA.[481] Finally, the section adopts the E.U. approach of focusing on safety considerations and mandates a plan to determine and monitor immunogenicity, pharmacovigilance, and risk management as part of the follow-on application and assessment.[482]

The administrative process is a concern for both the agency reviewing the applications and applicants applying for follow-on marketing authority. Because of the expertise of the FDA in evaluating biological products, it should have the authority to evaluate follow-on applications.

"(4) Authority of the Food and Drug Administration.—

"(A) The Secretary shall designate the Food and Drug Administration with the authority to review comparable biological drug applications.

"(B) The Food and Drug Administration may utilize its own agency expertise in evaluating comparable biological drug applications, and may utilize other experts for the purpose of evaluating these applications."

---

[477] *See* Access to Life-Saving Medicine Act ("ALSMA"), H.R. 1038, 110th Cong. § (a)(4) (2007).

[478] Indeed, it would seem apparent that had the terms "bioequivalence," "bioavailability," "manufacturing practice," and others been closely defined in the statutory provisions of the HWA, rather than through discussion and comments led by scientific expertise within the FDA and other agencies, the result would have been a freezing of scientific assessment and the means of determining appropriate issues such as comparability.

[479] H.R. 1038 § (3)(a)(2)(k)(2).

[480] *See id.* § (3)(a)(2)(k)(5) (allowing differences in profile, impurities, degradation patterns, post-translational events, infidelity of translation or transcription, amino acid sequence, and number of polysaccharides associated with a follow-on biologic).

[481] *See id.* § (3)(a)(2)(k)(7).

[482] *See* GUIDELINE: NON-CLINICAL AND CLINICAL ISSUES, *supra* note 263, at 6–8.

This section replaces ALSMA's requirement that the Secretary issue guidance for individuals who review follow-on applications.[483] Clearly the FDA has the expertise internally, or has access to the necessary expertise, and hence the section of ALSMA that designates authority to the Secretary is unnecessary for the appropriate review of these applications.

Guidance is important if follow-on biologic firms are to understand what a scientifically appropriate strategy for a robust application should be. Consistent with current practice, the FDA should meet with follow-on applicants to provide such guidance in good faith, both on a formal basis as well as on informal bases.[484]

"(5) Meetings with Applicants.—

"(A) GUIDANCE FOR SCIENTIFIC EVIDENCE.—The Food and Drug Administration, upon the written request of any applicant intending to submit a comparable biological product application, shall meet with such applicant to determine the type of valid scientific evidence that will be necessary to demonstrate for purposes of approval of an application the safety, purity, potency, and function of a comparable biological product and the conditions of use proposed by such applicant. The written request shall include a detailed description of the comparable biological product, a detailed description of the proposed conditions of use of the comparable biological product, and a proposed plan for determining whether there is a reasonable assurance of safety, purity, potency, and function of the comparable biological product. Within thirty days after such meeting, the Food and Drug Administration shall specify in writing the type of valid scientific evidence that will provide a reasonable assurance that the comparable biological product is safe, pure, potent, and functional under the conditions of use proposed by such person.

"(B) LEAST BURDENSOME MEANS.—The Food and Drug Administration shall consider, in consultation with the applicant, the least burdensome appropriate means of evaluating the comparable biological product's safety, purity, potency, and function that have a reasonable likelihood of resulting in application approval."

---

[483] *See* H.R. 1038 § (3)(a)(2)(k)(3)(A).

[484] *See* FDA, CDER Data Standards Manual, *available at* http://www.fda.gov/cder/dsm/drg/Drg00917.htm (outlining many meeting types that can be requested by applicants for FDA guidance at various stages of drug review); *see also* Felix Frueh & Larry Lesko, *Guidance, New Review Group to Spur "Personalized Medicine,"* NEWS ALONG THE PIKE, Apr. 2005, http://www.fda.gov/cder/pike/April2005.htm (encouraging voluntary, nonbinding data submissions that will "create[ ] an opportunity for early informal meetings with FDA pharmacogenomics experts," offer "flexibility in review and meeting process," and provide a forum where "[s]ponsors receive informal peer-review feedback on pharmacogenomic issues and questions").

It is essential that the FDA and the follow-on applicant agree on the testing, information, and data necessary to assess the application. Communication between the follow-on firm and the FDA should be initiated early so that discussions of these issues can occur. By providing for the same type of meeting as the Food and Drug Administration Modernization Act of 1997,[485] this proposal allows applicants to obtain a better understanding of what areas of focus will be necessary to address FDA needs, while also providing the applicant with insights into the thinking of the FDA with respect to requirements. Further, by mandating that the FDA consider the most effective, efficient, and expedient means for the application, it will reduce the burden on the applicant.

For there to be effective reliance upon the FDA assessment, provisions for recording and establishing the binding nature of such an assessment are important.

> **"(C) BINDING NATURE.—Any written response by the Food and Drug Administration under the provisions of paragraph (A), or in consultations with the applicant under the provisions of paragraph (B), regarding the parameters of valid scientific evidence of a comparable biological product, shall be reduced to writing and made part of the administrative record. Such agreement shall not be changed after collection of such scientific evidence begins, except—**
>
> **"(i) with the written agreement of the sponsor or applicant; or**
>
> **"(ii) pursuant to a decision, made in accordance with subparagraph (D) by the director of the reviewing division, that a substantial scientific issue essential to determining the safety, purity, potency, or function of the comparable biological product has been identified after the testing has begun. Any challenge by a comparable biological product applicant to a determination that there is a substantial scientific issue must be shown by clear and convincing evidence.**
>
> **"(D) SUBSTANTIAL SCIENTIFIC ISSUE MEETING.—A decision under subparagraph (C)(ii) by the director shall be in writing and the Secretary shall provide to the sponsor or applicant an opportunity for a meeting at which the director and the applicant will be present and at which the director will document the scientific issue(s) involved.**
>
> **"(E) WRITTEN DECISIONS AND FIELD AND COMPLIANCE PERSONNEL.—The written decisions of the reviewing division shall be binding upon, and may not directly or indirectly be changed by, the**

---

[485] Food and Drug Administration Modernization Act of 1997, Pub. L. No. 105-115, 111 Stat. 2296 (codified as amended in scattered sections of 21 U.S.C.). The provision for the device manufacturing review is used here. *See id.* at 2336–38.

field or compliance division personnel unless such field or compliance division personnel demonstrate to the reviewing division why such decision should be modified.

"(F) DELAY.—The FDA may delay the reviewing division action because of the unavailability of information, or for any reason to assure the marketing of a safe, pure, potent, and functional drug."

Here, the proposal outlines consultation with the follow-on, as well as the scope and binding nature of the FDA review. First, to promote commerce, this proposal requires a binding written accord and consultation, along with a formal record. These provisions are similar to those in ALSMA. However, in contrast to that proposal, the exception to the binding nature of any agreement goes beyond a simple determination that the safety, purity, and potency of the drug may be implicated. Here, as an important additional safety measure, the function of the follow-on product is added to ensure that changes associated with the functionality of the protein in clinical testing are enough to warrant a substantial scientific issue. Such an approach adopts lessons from the E.U. rejection of Alpheon, where functional differences between it and the reference biologic were grounds to reject the biosimilar application.[486]

In addition, to ensure that any decisions err on the side of safety, this proposal establishes a standard of clear and convincing evidence for follow-on applicant challenges of any such scientific grounds for altering the earlier written agreement between the FDA and the follow-on applicant. Further, contrary to ALSMA, which does not permit delay in reviewing the follow-on product application due to unavailability of information,[487] this bill would expressly allow the FDA to delay any decision if there was incomplete information or any other reason for the agency to be concerned about whether the drug is safe, pure, potent, or functional.

"(6) APPROVAL OF COMPARABLE BIOLOGICAL PRODUCTS.—

"(A) The Food and Drug Administration shall review the information submitted in the comparable biological product application and any other information available, subject to subsections (B) and (C) of this paragraph.

"(B) The Food and Drug Administration, in its review under subsection (A), shall not disclose information deemed a trade secret of the reference biological drug.

"(C) The Food and Drug Administration, in its review under subsection (A), may use reference biological drug information, including conclusions regarding safety and efficacy as well as conditions of its approval of the reference biological drug, in its assessment of the comparable biological product application."

---

[486] *See* QUESTIONS AND ANSWERS, *supra* note 338.

[487] *See* H.R. 1038 § (3)(a)(2)(k)(3)(F).

The initial conditions of information assessment are outlined in these subsections. The FDA must review the application, but it cannot disclose any trade secret information submitted by a reference biologic drug firm in its original application.[488] However, subsection (C) expressly allows the FDA to use reference biological safety and efficacy information and conditions of approval. The use, but not disclosure, of information by the FDA is consistent with the ability of ANDA applications to rely on the FDA's findings of safety and effectiveness.[489]

Then the FDA must undertake a substantive assessment of the applications.

> **"(D) The Food and Drug Administration shall not approve a comparable biological product application for any conditions of use relating to the reference product if the comparable biological product is not shown as comparable under terms and provisions of this section, and shall provide the applicant with a written, detailed explanation for its decision.**
>
> **"(E) The Food and Drug Administration shall not approve a comparable biological product application for any conditions of use relating to the reference product unless the comparable biological product application:**
>
> **"(i) shows the comparable biological product and the reference product have comparable principal molecular structural features as demonstrated by thorough categorization of the two products;**
>
> **"(ii) shows that the comparable biological product is comparable to the reference product for the conditions of use prescribed, recommended, or suggested in the labeling proposed in the application;**
>
> **"(iii) shows that the comparable biological product and reference product use the same mechanism(s) of action for the conditions of use prescribed, recommended, or suggested in the labeling proposed for the comparable biological product. However, this section shall not apply if the mechanism(s) of action is not known for the reference product for such conditions;**

---

[488] Trade secret information is "of a type customarily held in strict confidence or regarded as privileged and not disclosed to any member of the public by the persons to whom it belongs . . . ." 21 C.F.R. § 20.61(b) (2006). FDA employees are subject to 21 U.S.C. § 331(j), which prohibits disclosure of information acquired under FDCA § 505 except to FDA employees and a reviewing court. *See id.*

[489] *See* Woodcock Letter, *supra* note 195, at 3, 14 (discussing how the FDA allows reliance "to the greatest extent possible . . . on what is already known about a drug," and how "to the same extent an ANDA applicant may rely [on this information]").

"(iv) shows that the route of administration, the dosage form, and the strength of the comparable biological product are the same as those of the reference product;

"(v) shows that the conditions of use prescribed, recommended, or suggested in the labeling proposed for the comparable biological product are limited to one or more of the same use(s) as have been previously approved for the reference product;

"(vi) shows that the inactive ingredients of the comparable biological product are not unsafe for use under the conditions prescribed, recommended, or suggested in the labeling proposed for the biological product, or the composition of the comparable biological product is not unsafe under such conditions because of the type or quantity of inactive ingredients included or the manner in which the inactive ingredients are included;

"(vii) shows that the facility in which the comparable biological product is manufactured, processed, packed, or held meets standards designed to assure that the comparable biological product continues to be safe, pure, potent, and functional; and

"(viii) the Secretary has not withdrawn or suspended the license of the reference product, for safety or effectiveness reasons, or has published a notice of opportunity for hearing to withdraw such license for safety or effectiveness reasons, or has determined that the reference product has been withdrawn from sale for safety or effectiveness reasons; and

"(ix) the application does not contain an untrue statement of material fact; and

"in any event, shall provide the applicant with a written, detailed explanation for the decision.

"(F) If the Food and Drug Administration does not approve a comparable biological product application, within its written, detailed explanation for the decision, it shall give the applicant notice of an opportunity for a hearing before the Secretary on the question of whether the application is approvable. If the applicant elects to accept the opportunity for a hearing by written request within thirty days after such notice, such hearing shall commence not more than ninety days after the expiration of such thirty days unless the Secretary and the applicant otherwise agree. Any such hearing shall thereafter be conducted on an expedited basis and the Secretary's order thereon shall be issued within ninety days after the date fixed by the Secretary for filing final briefs. The Secretary shall take a final action on a comparable biological product application by the final action date."

These provisions mandate denial of follow-on biologic applications unless there is a demonstrable showing of comparability under the terms of the statute. This presumption of denial is necessary to ensure safety in light of the complex nature of biologics, the frequency of information gaps, the vulnerability of affected populations, and high potential harm. However, the enumerated reasons for denying the application are duplicated from ALSMA.[490] In addition, as with ALSMA, the denial of any application requires the Secretary to provide the applicant with a detailed explanation of the decision, as well as a hearing process relating to the application.[491]

Naming is an important concern for traceability, pharmacovigilance, dealing effectively with adverse events, and notification purposes for any follow-on biologic drug.[492]

> **"(G) NAMING OF COMPARABLE BIOLOGIC PRODUCTS.—If a comparable biological product is approved, the Food and Drug Administration shall assign the comparable biologic product a unique name. The Food and Drug Administration shall, within 180 days of the passage of this Act, work with international authorities to harmonize naming protocol to allow unique names for comparable biological products approved under this section."**

This provision establishes a naming system that facilitates rapid determination of the uses and users of specific follow-on biologics. Such traceability will facilitate identification of specific follow-on biologic forms that may be associated with adverse reactions. It also mandates that the FDA work to harmonize the protocol for follow-on products for international traceability.

If an applicant demonstrates that a follow-on biologic is interchangeable with an originator biologic, that alleviates significant safety concerns. In those cases, the follow-on molecule should have the same INN as the originator for substitutability purposes. However, because of the unpredictability of at least some biologics, there should be an additional identifier to indicate the follow-on nature of the product.

> **"(H) INTERCHANGEABLE COMPARABLE BIOLOGICAL PRODUCTS.—**
>
> **"(i) A comparable biological product applicant may request in an original application or in a supplement that the Food and Drug Administration make a determination that the comparable biological product is interchangeable with the reference product.**
>
> **"(ii) If a comparable biological product applicant requests the Food and Drug Administration to make a determination that**

---

[490] *See* H.R. 1038 § (3)(a)(2)(k)(4)(a).

[491] *See Id.* § (3)(a)(2)(k)(3)(D).

[492] See *supra* Part V.C for discussion on the naming of follow-on biologics.

the comparable biological product is interchangeable, the Food and Drug Administration shall first determine if the comparable biological product is comparable under the provisions of this Act.

"(iii) If the Food and Drug Administration approves the comparable biological product, it will then make a determination of the interchangeability of the comparable biological product if requested under section (H)(i). The Food and Drug Administration may approve a comparable biological product for marketing before a decision is made on the interchangeability of the comparable biological product.

"(iv) If the Food and Drug Administration deems the comparable biological product interchangeable, it shall assign the comparable biological product an identifier and publish a therapeutic comparability code indicating that the comparable biological product is interchangeable with the reference product.

"(v) For interchangeable products, the comparable biological product found interchangeable with the reference product shall be named using the INN protocol with the reference product INN; however, a numerical appendage shall be made on the basis of the order in which the comparable biological product was shown interchangeable, such as INN-1 for the first comparable biological product that was found interchangeable, INN-2 for the second comparable biological product that was found interchangeable.

"(vi) An interchangeable comparable biological product may include a statement on its label indicating that it is interchangeable with the biological reference product to which the sponsor of the comparable biological product application has demonstrated comparability to the reference product for the conditions of use prescribed, recommended, or suggested in the labeling proposed for the comparable biological product."

This section of the proposal outlines the process of showing interchangeability. The comparability issue of course must be assessed first, since clearly a follow-on molecule would not be interchangeable if it is not comparable to the originator biologic. This section expressly notes that comparability approval allows for marketing and sale before any assessment of interchangeability. However, if the FDA deems a follow-on product interchangeable, then for substitutability purposes, that product is named using the originator INN. However, to ensure traceability, a numerical appendage is used so that the specific follow-on can be identified if it was used. Labeling of the follow-on biologic as interchangeable is

permitted, as it is under ALSMA,[493] but requires an interchangeability determination, in contrast to both ALSMA and E.U. practice.

A common issue for many minorities and vulnerable patient populations is access to affordable medicines. Any program providing financial benefits and exclusivity for drugs and their manufacturers should take into account the patients who are most in need. A program that ensures data exclusivity but links incentives for follow-on production with programs to address the needs of underserved patient groups addresses the critical issues of access rather than merely the question of price.

"(7) ACCESS TO DRUGS AND DATA EXCLUSIVITY.—

"(A) The Secretary shall direct the Department of Health and Human Services Office of Minority Health to—

"(i) identify public and private low and no-cost drug programs in the United States, including those with culturally competent language translation services, and identify all state-level Offices of Minority Health;

"(ii) develop a nationwide program for low and no-cost drugs for minority and vulnerable patient populations under 400% of the federal poverty level utilizing and expanding the programs identified in section (A)(i) above, with the assistance of the Department Advisory Committee on Minority Health, state-level Offices of Minority Health, and industry members and groups, as appropriate;

"(iii) work with state governments to utilize the program developed in (A)(ii) to enroll participants into eligible health programs, including Medicaid, state children's health insurance programs, Supplemental Security Income, and state high risk insurance programs; and

"(iv) develop appropriate education, terms, and conditions of participation to ensure that access to biological drugs is provided to minority and vulnerable patient populations and that identification of any adverse reactions or events associated with these drugs are noted, reported, and disseminated.

"(B) No comparable biological product application may rely upon investigations of a reference drug application under section 351 of the Public Health Service Act performed for a successful Biological License Application that were not conducted by or for the comparable biological product applicant for approval of the comparable biological product application and for which the comparable biological product applicant has not obtained a right of reference or use from the person by or for whom the investigations were conducted, before five years from the date of the approval of the reference drug Biological License Application; and the Food

---

[493] *See* H.R. 1038 § 3(a)(k)(4)(C).

and Drug Administration shall not use or consider such studies in any comparable biological product application under section (6), except a comparable biological application may be submitted and investigations used under section (6) after the expiration of four years from the date of approval of the reference drug application if it contains a certification of patent invalidity or noninfringement described in section 9(C), subject to subsection (C) of this section.

"(C) The Food and Drug Administration may consider reference drug applicant investigations without regard to the five-year exclusion noted in section (B) in a comparable biological product application under subsection (6) if the reference drug Biological License Application applicant in section (B) does not participate in the nationwide drug program described in section (A) through provision of the approved reference drug to the nationwide drug program.

"(D) For any applicant that has a comparable biological product application approved by the Food and Drug Administration under section (B) or (C), such applicant must participate in the nationwide drug program described in section (A) within eighteen months of the date of application approval.

"(E) Subject to the provisions and terms of subsection (A)(v), to obtain the data exclusivity period of five years noted in section (B), a reference drug applicant must agree to participate in the nationwide program for a minimum of fifteen years after the date of approval of its Biological License Application or until the reference drug is withdrawn from the market; and subject to the provisions and terms of section (A)(v) and section (D), the producer of the comparable biological product must agree to participate in the nationwide program for a minimum of ten years after the date of approval of its comparable biological products application or until the comparable biological product is withdrawn from the market."

This section links data exclusivity, a provision absent in ALSMA, with access through a nationwide low-cost/no-cost drug access program. The Department of Health and Human Services ("DHHS") Office of Minority Health ("OMH") has significant insight and contact with minority and vulnerable patient populations. State-based Offices of Minority Health and their equivalents have significant local information on populations in need. Further, both have expertise on issues involving cultural competency and health literacy.[494] In addition, the DHHS Advisory Committee on Minority Health

[494] *See* Press Release, DHHS, HHS Secretary Appoints Advisory Committee on Minority Health (Apr. 22, 2005), *available at* http://www.hhs.gov/news/press/2005pres/20050422.html.

is composed of experts in programmatic health care systems who can facilitate the creation of a nationwide drug access program.[495]

This effort to create a nationwide program is consistent with the existing programs of individual drug companies as well as the Pharmaceutical Research and Manufacturers Association ("PhRMA") of America low-cost/no-cost drug program.[496] Note that this existing industry effort covers both chemical and biological drugs. Hence, these programs can be developed for an extensive set of medicines and will ensure access to some of the most expensive drugs that treat some of the most severe diseases for those in greatest need. Further, those accessing the nationwide program for drugs who are also eligible for public health insurance programs—a significant fraction[497]—but for whom the barriers of language, literacy, or culture prevent enrollment, will be provided with an infrastructure to obtain the benefits to which they are entitled. This may be an important outreach mechanism for government efforts to enroll eligible benefi-ciaries into public programs.

This proposal requires originator biologic companies to participate in a nationwide drug access program to gain access to a five-year data exclusivity grant. Importantly, as noted above, most originator companies already run their own low-cost/no-cost drug programs; hence, this participation should not necessarily be onerous.[498] But because of the expanded and centralized nature of the program contemplated by this proposal, the originator companies' efforts under this proposal would be more extensive and impact a greater number of patients. Participation by originator companies is a reasonably equitable exchange for a period of data exclusivity. The period of fifteen years would allow access to these drugs during the entire patent applicability period, and hopefully will overlap with the approval of follow-on forms of the drug.

Since follow-on biologics do not garner the discounts of chemical based drugs,[499] firms producing these drugs would also be mandated to par-

---

[495] *See id.*

[496] The program is known as the Partnership for Prescription Assistance, and it also offers its information and enrollment services in over 100 languages. *See* Partnership for Prescription Assistance, https://www.pparx.org/Intro.php [hereinafter Partnership for Prescription Assistance](last visited Apr. 13, 2006).

[497] For example, it has been estimated that more than two-thirds of uninsured children in California are eligible for public health insurance coverage. *See* Gregory D. Stevens et al., *Enrolling Vulnerable, Uninsured but Eligible Children in Public Health Insurance: Association with Health Status and Primary Care Access*, 117 PEDIATRICS e751, e752 (2006), *available at* http://pediatrics.aappublications.org/cgi/content/full/117/4/e751.

[498] *See* Partnership for Prescription Assistance, *supra* note 496.

[499] *See* Global Insight, *supra* note 258 (noting biosimilar discounts of only roughly twenty percent from originator product); Partnership for Prescription Assistance, *supra* note 496 (reporting range of discounts from ten to twenty percent). *But see* ENGEL & NOVITT, LLP, POTENTIAL SAVINGS THAT MIGHT BE REALIZED BY THE MEDICARE PROGRAM FROM ENACTMENT OF LEGISLATION SUCH AS THE ACCESS TO LIFE-SAVING MEDICINE ACT (H.R.6257/S. 4016) THAT ESTABLISHES A NEW cBLA PATHWAY FOR FOLLOW-ON BIOLOGICS 2 (2007), http://www.pcmanet.org/newsroom (scroll down under "January 2007" head-

ticipate in the nationwide drug access program to ensure benefits associated with the follow-on approval process are obtained by those who need it most. Note, however, that follow-on applicants are given eighteen months without participation to allow infrastructural sales networks to be created and some initial costs to be covered by initial sales. Mandating a period of ten years in the nationwide program serves the purpose of ensuring coverage of this drug for these patients until new, improved therapeutic forms are produced.

The nationwide program tied to data exclusivity and the follow-on biologic approval process would be a step towards providing vulnerable and poor patients access to drugs and will give many patients access to health insurance for which they are eligible but not enrolled. Moreover, because of the risks associated with follow-on biologics, such as immunogenicity and other complications, the nationwide program will also provide an infrastructure for reports on adverse reactions and potential drug issues. Such information could be kept in a nationwide database, such as MEDWATCH, which would allow it to be disseminated quickly to relevant providers and patients.[500] Because these patients are the most vulnerable, they should be monitored closely for any adverse reactions to biological drugs. Further, this monitoring is also imperative because clinical trials are notorious for limited minority participation, and primary and side effects of these drugs are not well known for these groups.[501]

---

ing and click on "Engel & Novitt Follow-on Biologics Paper") (projecting that the estimated cost savings by Medicare from legislation such as ALSMA would be $14.1 billion over ten years); California Healthcare Foundation, *Governors, Lawmakers Seek Access to Generic Insulin*, CAL. HEALTHLINE, Jan. 11, 2007, http://www.californiahealthline.org/index.cfm?Action=dspItem&itemID=129276; Stephanie Saul, *States, Bridling at Insulin's Cost, Push for Generics*, N.Y. TIMES, Jan. 11, 2007, at A1. Note, however, that the Engel-Novitt analysis assumes limited clinical trials and does not take into account pharmacovigilance, risk management, and immunogenicity activities that are required in other follow-on biologics regulatory regimes, such as the E.U. regime. *See supra* notes 257–296 and accompanying text (describing E.U. biosimilars regulatory requirements). Further, economic analysis indicates that savings from follow-on biologics entry and competition may be highly limited because of high fixed costs associated with clinical trial, capital investment outlay requirements, and manufacturing costs. *See* Henry G. Grabowski et al., Entry and Competition in Generic Biologicals, 10–11 (undated) (unpublished manuscript, on file with author) *available at* http://faculty.fuqua.duke.edu/health_sector_management/hsac/files/oct%2020_26_2006/Biogenerics_Ridley%20et%20al.pdf (last visited Feb. 23, 2007). Grabowski et al. conclude that

> [G]eneric biologics will be relatively close in price to branded biologics for the foreseeable future. Policy makers should be cautious in projecting large financial benefits from generic biologics for consumers and payers based on the experiences of generic pharmaceuticals. They should consider how generic biologics will differ in terms of economics as well as scientific and regulatory factors.

Id. at 26.

[500] MEDWATCH is the FDA's safety information and adverse event reporting system. *See* MEDWATCH, http://www.fda.gov/medwatch/ (last visited Apr. 13, 2007).

[501] *See, e.g.*, Dorie Hightower, *Minority Participation in Clinical Trials*, BENCHMARKS, Sept. 6, 2006, http://www.cancer.gov/newscenter/benchmarks-vol6-issue4 (noting that minori-

Follow-on biologics which are truly interchangeable with the originator products would be highly beneficial for patients, as long as safety issues are addressed. To address the numerous safety concerns and to ensure that the products are indeed interchangeable, it is likely that the FDA would impose additional requirements and testing. If the follow-on firm performs these additional tests and studies, they should be rewarded for this extra effort. Thus, the following is proposed:

"(8) MARKETING EXCLUSIVITY.—

"(a) If the Food and Drug Administration approves a comparable biological product, and the Food and Drug Administration approves such comparable biological product as the first interchangeable version of the reference drug, the Food and Drug Administration shall not approve a second interchangeable or subsequent interchangeable comparable biological product application, and no holder of a biological product license under subsection (a) shall manufacture, market, sell, or distribute a rebranded interchangeable biological that is interchangeable with the reference product, until the earlier of—

"(i) Ninety days after the first commercial marketing of the first interchangeable comparable biological product to be approved as interchangeable for that same reference product; or

"(ii) one year after—

"(I) a final court decision on all patents in suit in an action instituted under paragraph (9) against the applicant for the first approved interchangeable comparable biological product; or

"(II) the dismissal with or without prejudice of an action instituted under paragraph (9) against the applicant that submitted the application for the first approved interchangeable comparable biological product; or

"(iii) either—

"(I) Thirty-six months after approval of the first interchangeable comparable biological product if the applicant has been sued under paragraph (9) and such litigation is still ongoing within such thirty-six month period; or

"(II) one year after approval in the event that the first approved interchangeable comparable applicant has not been sued under paragraph (9).

"(b) For the purposes of this section—

"(i) Notwithstanding the foregoing provision, the applicant for a subsequent comparable biological product application that has demonstrated interchangeability with the reference product may elect, at its option, to have the product approved as a noninterchangeable com-

---

ties are particularly underrepresented in cancer clinical trials).

parable biological product whose approval will not be delayed by operation of this paragraph.

"(ii) A 'final court decision' means a final decision of a court from which no appeal, other than a petition to the United States Supreme Court for a writ of certiorari, has been or can be taken.

"(iii) A 'rebranded interchangeable biologic' means any rebranded interchangeable version of a reference product that the holder of the biological product license approved under subsection (a) for that reference product seeks to commence marketing, selling, or distributing, directly or indirectly, but does not include any product to be marketed, sold, or distributed by an entity eligible for exclusivity with respect to such product under this paragraph or after expiration of any exclusivity with respect to such product under this paragraph."

This section differs in several important ways from ALSMA. First, it clarifies that only subsequent interchangeable follow-on biologic applications are excluded under the section, rather than excluding all similar follow-on biologics applications, including those that are not claiming interchangeability.[502] Second, the period of exclusivity proposed here is 90 days rather than 180 days. The 180-day exclusivity period mirrors the period under the original HWA.[503] However, because it appears that follow-on biologics will only garner roughly a twenty percent discount from the brand-name form,[504] compared to a fifty percent discount for generics of chemical medicines when full entry is allowed,[505] a reduced exclusivity period is proposed here. Although full discounting of the time period to reflect the reduction in equilibrium price would equate to shorter than ninety days,[506] the proposal provides a ninety-day exclusivity period[507] to take into account additional potential costs associated with interchangeability activities.

This proposal then adopts language from the Medicare Modernization Act for various periods associated with patent challenges and approval of interchangeable follow-on applications by the FDA.[508] Like ALSMA,[509]

---

[502] *See* Access to Life-Saving Medicine Act, H.R. 1038, 110th Cong. § 3(a)(2)(k)(2) (2007).

[503] *See* 21 U.S.C. § 355(j) (2000 and Supp. IV 2004); 21 C.F.R. § 314.108 (2006).

[504] *See supra* note 499 and accompanying text.

[505] *See* David Reiffen & Michael R. Ward, *Generic Industry Dynamics*, 87 REV.ECON. & STAT. 37, 44 (2005).

[506] Since the price discount is only 20% compared with 50% in the chemical drug market, the exclusivity period should be (20%/50%) x 180 days, or 72 days. Note that this assumption of a 20% discount may actually be generous, since other experiences with biosimilars (*e.g.*, Omnitrope in Australia), as well as business and economic analysis, have estimated that follow-on prices would be close to or only approximately a 10% discount off brand prices. *See supra* note 499; Susan J. Ainsworth, *Biopharmaceuticals*, CHEM. & ENG. NEWS, June 6, 2005, at 21, *available at* http://pubs.acs.org/cen/coverstory/83/8323biopharmaceuticals.html (dealing with business analysis); Grabowski et al., *supra* note 499, at 19 (looking at economic analysis).

[507] This eighteen-day extension represents a 10% greater period of exclusivity.

[508] *See* Medicare Modernization Act, Pub. L. No. 108-173, 117 Stat. 2066 (2003).

[509] *See* Access to Life-Saving Medicine Act, H.R. 1038, 110th Cong. § (3)(a)(2)(k)(10)

it also prohibits "rebranded interchangeable biologic[s]" until the exclusivity period has ended. This provision prohibits the originator firm practice of marketing "authorized generics"—a strategy that weakens incentives for generics to enter the chemical medicine market.[510]

The patent provisions associated with the approval process of interchangeable follow-on biologic applications are an important part of the proposal.

"(9) PATENTS.—

"(A) REQUEST FOR PATENT INFORMATION.—

"(i) IN GENERAL.—At any time, including at the initial stages of development, but no later than the date upon which the applicant files its comparable biological product application with the Secretary, an applicant or prospective comparable biological product applicant may send a written request for patent information to the holder of the approved application for the reference product. Within sixty days of receipt of such request, the holder of the approved application for the reference product shall provide to the applicant or prospective comparable biological product applicant a list of all patents owned by, or licensed to, the holder of the approved application that the application holder in good faith believes relate to the reference product, including patents that claim the approved biological product, any method of using such product, any component of such product, or any method or process of manufacturing such product or component.

"(ii) COSTS OF COMPLYING WITH REQUEST.—The application holder may demand payment not exceeding $1,000 to offset the cost of responding to the information request, subject to adjustment from time to time by the Secretary to reflect increased costs of fulfilling such requests.

"(iii) UPDATES.—For a period of two years from the date of the request for information, the holder of the approved application for the reference product shall update its response to the request for information by identifying newly issued or licensed relevant patients. The updates must be provided within thirty days of patent issuance for newly issued patents, and within thirty days of obtaining a license for newly licensed patents.

"(iv) ADDITIONAL REQUESTS.—The comparable biological product applicant may submit additional requests for patent information, subject to the requirements of this paragraph, at any time."

(A) (2007).

[510] *See* Hollis & Liang, *supra* note 179, at 1.

This section is similar to ALSMA,[511] with the additions that the Secretary may adjust the cost of complying with the requests, and the proposal requires that requests for patent information from the originator occur no later than the date upon which the follow-on biologic application is filed.

However, with respect to patent notifications, this proposal mandates notification to the reference biologic firm.

> **"(B) PATENT NOTIFICATIONS.—The comparable biological product applicant must provide a notice under this subparagraph with respect to any one or more patents provided by the holder of the reference product, if the provisions under subparagraph (B)(ii) apply. Each notice shall—**
>
> **"(i) be sent to the holder of the approved application for the reference product and to the owner of the patent identified pursuant to subparagraph (A)(i);**
>
> **"(ii) include a detailed statement of the factual and legal bases for the applicant's belief that the patents included in the notice are invalid, unenforceable, or will not be infringed by the commercial sale of the product for which approval is or has been sought; and**
>
> **"(iii) identify the judicial district or districts in which the applicant consents to suit being brought in response to the notice.**
>
> **"A comparable biological product applicant, within the comparable biological product application, shall certify that, if applicable, such notice has been sent to the parties noted in subparagraph (B)(i)."**

Here, in contrast to ALSMA,[512] the patent notification by the follow-on applicant to the originator firm is mandated and noted in the follow-on biologic application. This is similar to the requirements under the HWA for paragraph (iv) certifications.[513] Requiring such notice will provide clarity for both parties as well as the FDA as to the patents and legal question at issue. If such notice was merely optional, significant uncertainty would arise as to patents and patent claims.

With respect to actions for patent infringement, ensuring clarity and certainty should also be encouraged.

> **"(C) ACTION FOR INFRINGEMENT.—**
>
> **"(i) TIMEFRAME FOR BRINGING ACTION.—Within forty-five days of receipt of notice described in subparagraph (B), the holder of the approved application for the reference product, or the owner of the patent, may bring an infringement action solely with respect to the patent or patents included in such notice.**
>
> **"(ii) APPROPRIATE JUDICIAL DISTRICT.—Notwithstanding section 1391 of title 28, United States Code, an infringement action**

---

[511] *See* H.R. 1038 § (3)(a)(2)(k)(17)(A).
[512] *See id.* § (3)(a)(2)(k)(17)(B).
[513] *See* 21 U.S.C. § 355(b)(2)(A)(iv) (2000).

brought within the forty-five day period reference in clause (i) may be brought only in the judicial district identified pursuant to subparagraph (B)(ii).

"(D) DECLARATORY JUDGMENTS.—

"(i) If a recipient of notice under subparagraph (B)(ii) does not initiate an infringement action as specified under subparagraph (C)(i), the applicant may bring a declaratory judgment action under section 2201 of title 28, United States Code, for invalidity, noninfringement, and nonenforceability of the patents within the notice under subparagraph (B)(ii).

"(ii) The declaratory action under subparagraph (D)(i) shall be brought only after—

"(I) the applicant sends a request to the recipient requesting a covenant not to sue with respect to the patents within the notice under subparagraph (B)(ii); and

"(II) the recipient of the request indicated in subparagraph (D)(ii)(I) does not issue such covenant not to sue to the applicant within thirty days of receipt of the request in subparagraph (D)(ii)(I).

"(iii) If the recipient of the request indicated in subparagraph (D)(ii)(I) does not issue a covenant not to sue with respect to the patents within the notice under subparagraph (B)(ii), this shall be deemed a 'reasonable apprehension' as that term is used in section 2201 of title 28, United States Code."

These sections follow ALSMA with respect to the timeframe for patent infringement suits and jurisdiction.[514] However, the section differs markedly in other respects. It attempts to address the follow-on applicant's potentially limited knowledge as to whether the product infringes on a reference drug patent, and therefore the appropriateness of investing further into the marketing and sales of the product. Litigation at this stage of development is preferable if there is a question of patent infringement. If it occurs during the application phase, patent litigation resolves patent validity and infringement issues before the generic company has commercialized the product or the brand-name firm has suffered any damages.

Importantly, brand-name firms in the chemical drug markets have chosen not to sue generic firms for patent infringement, resulting in significant doubt for generic firms as to whether their drugs violate brand-name patents. Under these circumstances, if the generic firm enters the market, it does so at great risk. Such a strategy may be a "bet [the] company"[515] decision if it is later found to be in violation of brand-name patents, subjecting it to treble

---

[514] H.R. 1038 § (3)(a)(2)(k)(17)(C)–(D).

[515] *See Lifesaving Drugs Hearing*, *supra* note 179, at 10–11.

damages[516] and other penalties.[517] This, in turn, may result in inefficient decisions to enter or forego entry, both of which waste valuable resources.

Because of this concern, Congress provided in the Medicare Modernization Act that a generics firm is entitled to bring a declaratory judgment action against a brand-name firm that has refused to sue.[518] However, the Federal Circuit has effectively emasculated this provision by concluding that if brand-name firms refuse to bring suit, there is no "reasonable apprehension" on the part of the generic firm, and therefore the court has no jurisdiction.[519]

To ensure that there are efficient decisions regarding early entry of follow-on biologics, the proposal mandates that either the originator firm sue for patent infringement upon notice from the follow-on firm or issue a covenant not to sue. Otherwise, the follow-on firm may bring a declaratory judgment action to resolve any potential question or conflict regarding the listed patents in question. This allows early, reasonable certainty with respect to the patents in question. The section also expressly notes that an originator's refusal to issue a covenant not to sue is deemed a "reasonable apprehension" within the meaning of the declaratory judgment statute, thus expressly bringing the conflict within the bounds of Federal Circuit jurisdiction.[520]

Contrary to the provisions of ALSMA, which alter remedies for patent infringement and encourage litigation so as to avoid royalties as its sole remedy,[521] in this system, the resolution of the conflict will occur earlier through a substantive evaluation performed by each party. The originator

---

[516] *See* 35 U.S.C. § 287 (2000) (allowing for treble damages for patent infringement).

[517] *See, e.g.*, 35 U.S.C. §§ 283–85 (2000) (allowing damages, injunction, enhanced damages, and attorney fees).

[518] *See* 21 U.S.C. § 355(c)(3)(D)(i) (2000 and Supp. IV 2004).

[519] *See* Teva Pharms. U.S.A., Inc. v. Pfizer, Inc., 395 F.3d 1324, 1337 (Fed. Cir. 2005). *But see* MedImmune, Inc. v. Genentech, Inc., 127 S. Ct. 764 (2007) (holding that payer of royalties (under protest) under licensing agreement was not required to breach agreement prior to seeking declaratory judgment). The Court quotes the *Teva* case negatively. *See id.* at n.11. The *MedImmune* case would seem at least to clarify that payment of royalties under a protested license agreement can represent a reasonable apprehension of imminent suit, and to acknowledge federal court Article III jurisdiction in patent infringement actions for declaratory relief. However, the *MedImmune* decision may still not be helpful to generics firms hoping for guidance about whether investment and marketing expenditures in generic products that may or may not violate a brand-name firm's patent should be expended. In the latter case, there is no contact between the generic and the brand-name company: the generic firm is simply awaiting a brand-name firm lawsuit. There is no license agreement, contract, or voluntary payment of funds, under protest, as compared with *MedImmune*. Of course, the long-term consequences of this decision are unclear at this point. But at present, there are distinctions between current generics and brand-name circumstances that arguably require attention.

[520] *See* Shell Oil Co. v. Amoco Corp., 970 F.2d 885, 887 n.2 (Fed. Cir. 1992) (for Federal Circuit jurisdiction, "the defendant's conduct must have created on the part of the plaintiff a reasonable apprehension that the defendant will initiate suit if the plaintiff continues the allegedly infringing activity." (citation omitted).

[521] *See* Access to Life-Saving Medicine Act, H.R. 1038, 110th Cong. § (3)(b)(1)(6)(B) (2007).

firm has forty-five days to bring a patent infringement suit. If it does not, the follow-on firm can request a covenant not to sue. If not granted by the originator firm, a declaratory judgment action may be pursued, which will clarify the rights of the parties early in the process and before socially wasteful investments are made or infringing products are marketed. If granted, the follow-on firm has confidence in additional investments to bring the product to market, resulting in lower costs and lower consumer prices.

In attempting to ensure timely approval of follow-on applications, addressing delay tactics is important. However, the standards within ALSMA are so onerous as to preclude any challenge or efforts to challenge using civil processes.[522] Instead, simpler procedures to avoid abuse should be considered.

> **"(10) PETITIONS AND CIVIL ACTIONS REGARDING APPROVAL OF CERTAIN APPLICATIONS.—Section 505 of the Federal Food, Drug, and Cosmetic Act (21 U.S.C. 355) is amended by adding at the end the following:**
>
> **"(o) CITIZENS PETITIONS AND PETITIONS FOR STAY OF AGENCY ACTION.—With respect to any petition that seeks to have the Secretary take, or refrain from taking, any form of action relating to the approval of a comparable biological product application, the following shall apply:**
>
> **"(1) NO DELAY OF APPROVAL.—The Secretary shall not delay approval of a comparable biological product application submitted under this Act while a petition is reviewed and considered. Consideration of a petition shall be separate and apart from the review and approval of an application submitted under this Act, unless such petition review directly implicates the public health and safety associated with the comparable biological product in the application.**
>
> **"(2) TIMING OF FINAL AGENCY ACTION.—The Secretary shall take final agency action with respect to a petition within six months of receipt of that petition. The Secretary shall not extend such six-month review period, even with consent of the petitioner, including based upon the submission of comments relating to a petition or supplemental information supplied by the petitioner. If the Secretary has not taken final agency action on a petition by the date that is six months after the date of receipt of the petition, such petition shall be deemed to have been denied on such date.**
>
> **"(3) PUBLIC HEALTH ISSUES.—Notwithstanding the provisions indicated in section (o)(2), the Secretary may extend such six-month review period and delay approval of a comparable**

[522] *See id.* § (3)(a)(2)(k)(18).

biological product application when such a delay is necessary to protect the public health and welfare. If the Secretary extends the six-month review period, or delays approval of a comparable biological product, then—

"(A) The Secretary shall publish on the Internet site of the Food and Drug Administration a statement providing the reasons underlying the determination.

"(B) Not later than ten days after making the determination, the Secretary shall provide notice to the comparable biological product applicant and an opportunity for a meeting with the Commissioner of the Food and Drug Administration to discuss the determination.

"(4) VERIFICATION.—The Secretary shall not accept for review a petition unless it is signed and contains the following verification: 'I certify that, to my best knowledge and belief: (a) this petition includes all information and views upon which the petition relies; (b) this petition includes representative data and/or information known to the petitioner which are unfavorable to the petition; and (c) I have taken reasonable steps to ensure that any representative data and/or information which are unfavorable to the petition were disclosed to me. I further certify that the information upon which I have based the action requested herein first became known to the party on whose behalf this petition is filed on or about _____. I verify under penalty of perjury that the foregoing is true and correct.,' with the date of the filing of such petition inserted in the blank space.

"(5) For the purposes of this paragraph, the term 'petition' includes any request to the Secretary, without regard to whether the request is characterized as a petition.

"(6) The Secretary shall annually submit to the Congress a report that specifies the number of applications under this subsection that were approved during the preceding twelve-month period, the number of such applications whose effective dates were delayed by petitions under this subsection, and the number of days by which the applications were so delayed."

This section bases its provisions on a previous bipartisan proposal addressing potential abuses of the civil citizen petition processes.[523] This methodology is simpler than the one is ALSMA, and focuses upon the goal of ensuring that petitions are reviewed within six months, unless public health and safety issues are involved.

Critically, however, public health and safety are the focus of this proposal. It rejects ALSMA's complete ban on assessing the petition in con-

[523] *See* Lower Priced Drugs Act, S. 2300, 109th Cong. (2006).

junction with the follow-on biologic application.[524] Instead, it allows for the assessment of the petition in relation to the application if there is a public health and welfare concern.

Further, this proposal does not attempt to dictate under what circumstances and with what standards a court may or may not order the Secretary to take action or refrain from taking action on a follow-on biologics application.[525] Because of the plethora of situations that may arise with follow-on product challenges, it would be unwise to handcuff the courts in their assessments of the facts and circumstances of each case, particularly in this young and developing industry.

The proposal does adopt the ALSMA approach of requiring an accounting of the applications and relative delays associated with petitions under this section.[526] This accounting provides information that can allow alteration of the petition process if and when such challenges create policy concerns.

Finally, the safety of the biologic drug supply requires attention. To protect vulnerable patients, biologics, due to their sensitive nature and easy counterfeiting, should not be subject to alternative sourcing.

"(11) PROHIBITION AGAINST BIOLOGICAL DRUG SALES VIA IMPORTATION OR THE INTERNET.—

"(A) All biological drugs, whether they be originator biological drugs or approved comparable biological drugs, shall be excluded from any importation program and shall not be permitted to be imported, except under the provisions of the Food, Drug, and Cosmetic Act, by manufacturers and authorized distributors under section 381(d)(1) of title 21 of the United States Code.

"(B) All biological drugs, whether they be originator biological drugs or approved comparable biological drugs, shall not be subject to sales through Internet sellers."

Here, in an effort to ensure maximum safety and to make genuine biological products available to patients, the proposal prohibits all nonstandard purchases and sales of biologics. Because importation and the Internet are documented sources of counterfeit and diverted drugs,[527] these means of obtaining biologics are not permitted by this proposal.[528]

---

[524] H.R. 1038 § (3)(a)(2)(k)(18)(A)(i)(I). "Consideration of a petition shall be separate and apart from the review and approval of the application."

[525] *See id.* § (3)(a)(2)(k)(18).

[526] See *id.* § (3)(a)(2)(k)(18)(D).

[527] *See* Liang, *Fade to Black, supra* note 102, at 285.

[528] Importation legislation previously introduced has also excluded biologics from their provisions. *See, e.g.*, Pharmaceutical Market Access and Drug Safety Act, S. 334, 109th Cong. § 4 (2005,) at 112.

VII. Concluding Remarks

The biotechnology revolution has provided great benefits in the treatment of a vast array of human disease. Research and development involving the very processes of life have brought unprecedented tools for improvement of humankind's quality and quantity of life.

But the excitement and promise of biologics come with a cost. Prices for these wünder-drugs are very high, reflecting the tremendous resources required to bring a functional, safe, and effective biologic protein to the market. But more importantly, the cost of these drugs creates access issues for vulnerable patients, which simultaneously creates tremendous burdens on our health care system and weakens the ethical foundations of medical care delivery in this country.

At its moral core, access to drugs is a measure of our egalitarianism. Unfortunately, only those with the ability to afford the high costs of drugs can access them and the health and quality of life they bring. We must eradicate this inequality so that every person in this country may benefit from the promise of biotechnology.

Earlier policies, such as the HWA, have been successful in creating access to cheaper chemical drugs while also maintaining a robust incentive to innovate. However, even with this important policy, many are still without vital access to lifesaving medications.

We are now faced with similar considerations for biologic medicines. However, the science, risks, and concerns with these products are as substantively different from chemical medicines as biology is from chemistry. Given scientific information gaps, vulnerable citizens who bear the risk of policy failure, and the great potential harm, public policy must err on the side of safety. Blind application of the chemical generic drug regulatory regime to the complexities of biologics is clearly inappropriate.

We do need to reward those who take the risks of innovation, and this is particularly so in the cutting-edge field of biologics. Ensuring that the momentum of scientific progress is sustained is critical for our society's future benefit. Yet the United States needs to develop processes to guarantee access to drugs for those most in need and to promote health to the general population of this nation.

Hence, what is needed is a regulatory structure that addresses the policy analysis, challenges, and realities represented by biologics. It must create incentives to innovate—both in formulating new biologics as well as developing follow-on products—and provide flexibility to allow advances in science to be incorporated into the regulatory process. Moreover, to address the policy vacuum that ignores those most in need, the regulatory structure should also include provisions for access to these beneficial medicines by those without resources, preferably utilizing public-private-community partnerships. All of this should be built on a policy infrastruc-

ture that focuses on safety, more information, and more scrutiny rather than potentially less-than-dramatic price benefits.

This Article, through an assessment of scientific, regulatory, and policy issues, builds a proposed regime that takes into account the medical and safety complexities of biologics, the related regulatory issues, and international experience in biologics review. It also creates programs for drug access that are constructed upon existing public and private efforts to ensure access to crucial biologic products. Through an informed process of policy analysis, it seeks to ensure that the risks and benefits of follow-on biologics are suitably balanced and that access is enhanced.

We are fortunate to live in the era of biotechnology. Yet to sustain the benefits of this remarkable understanding of life's very foundations, we need to ensure that safety is emphasized while access is ensured for all. It is a challenging policy goal. But we must embrace it so that we give ourselves and our children the best chance for health as well as fulfill our potential for an inclusive society.

APPENDIX

H. R. ____

To amend the Public Health Service Act to provide for the licensing of comparable biological products, and for other purposes.

A BILL

To amend the Public Health Service Act to provide for the licensing of comparable biological products, and for other purposes.

Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled,

SECTION 1. SHORT TITLE.

This Act may be cited as the "Biologic Drug Safety and Access to Medicines Act."

SECTION 2. FINDINGS.

Congress makes the following findings:

(1) Biological drugs, also known as biologics, created through the use of recombinant DNA technologies and other biotechnology means have provided significant and important health benefits to patients, including those with the most severe and debilitating diseases such as cancer, AIDS, and hepatitis.

(2) Biologics are highly complex, and their research, development, and manufacture are uniquely characterized and distinguished from the research, development, and manufacture of chemical drugs.

(3) Because of their complexity, biologic characterization and their mechanisms of action may be much more difficult, if not impossible, to describe compared with smaller, less complex chemical drugs.

(4) Because of their complexity and size, biologics may induce severe, adverse, and unwanted immunologic reactions in patients.

(5) Biologic drugs have intellectual property patent protection, which has promoted their development and rewarded these originator companies through market exclusivity rights.

(6) Important biologic drugs have patent terms that have ended or are to end within the next several years.

(7) For the greatest benefits to inure to patients, biologics with patent protections that have ended should be subject to competition from appropriately safe, comparable biologic products, also known as follow-on biologic drugs.

(8) Many patients, particularly minorities, the uninsured, and the underinsured, cannot access biologic products at current prices.

(9) Since follow-on biologic products are more complex than traditional chemical medicines, follow-on biologic products require a

review of applications that is different from chemical medicines as well as tailored specifically for biologic molecules.

(10) Other areas of the world, including the European Union and Australia, have created regulatory systems to assess, approve, and deny follow-on biologic applications and products.

(11) Biologic drugs are generally supplied as clear fluids and are sensitive to temperature and other environmental conditions.

(12) Biologic drugs have been counterfeited and harmed patients.

(13) Biologic drugs require secure and traceable supply chains to ensure appropriate transport and authenticity, and should not be accessed using nontraditional or nonstandard supply means.

SECTION 3. DEFINITIONS.

Section 351(i) of the Public Health Service Act (42 U.S.C. 262(i)) is amended—

(1) by striking "In this section, the term biological product means" and inserting the following:

"In this section:

(1) The term 'biological product' means"; and

(2) by adding at the end the following:

"(2) The term 'comparable biological product application' means an application for a license of a biological product containing the same, or similar, active ingredient as a biological product for which a license has been approved under subsection (a). A comparable biologic application is a human drug application under section 735(1)(C) of the Federal Food, Drug, and Cosmetic Act.

"(3) The term 'reference product' under this Act means the single licensed biological product, approved under subsection (a) or subsection (k), against which a comparable biologic product is evaluated for demonstration of safety, potency, or purity.

"(4) The term 'comparable' in reference to a comparable biologic product application means the absence of clinically meaningful differences between the comparable biological product and the reference product in terms of the safety, purity, and potency associated with the clinical indication for which the comparable biologic product application is applying for approval. Such absence of clinically meaningful differences between the comparable biological product and the reference product shall be assessed based upon—

"(A) data derived from chemical, physical, and biological assays, other non-clinical laboratory studies; and

"(B) data from clinical study or studies sufficient to confirm safety, purity, and potency for appropriate conditions of use that the comparable biologic is applying for, which the reference product is licensed and intended to be used for.

"Any studies under subparagraph (B) shall be designed to avoid duplicative and unethical clinical testing.

"(5) The term 'thorough characterization' means an analysis of structural and functional features based upon appropriate analytical and functional testing sufficient to identify, and should be focused upon determining differences between the comparable biologic product and the reference biologic product relevant to safety, purity, potency, and use.

"(6) The term 'interchangeable' means that a biological product contains an active ingredient or ingredients with principle molecular structural features comparable to the reference product, and that the comparable biological product can be expected to produce the same clinical result as the reference product in any given patient in the condition or conditions of use for which both products are labeled.

"(7) The term 'process for the review of a comparable biological application' means, with respect to a comparable biological product application, the procedural activities of the Secretary, with respect to the review of human drug applications and supplements as defined in section 735(6) of the Federal Food, Drug, and Cosmetic Act, except as otherwise defined herein.

"(8) The term 'final action' means, with respect to a comparable biological product application, the Secretary's issuance on the final action date of a final action letter to the sponsor of a comparable biological product application under this Act, which may—

"(A) approve the application; or

"(B) disapprove the application and set forth in detail an enumeration of the specific deficiencies in the particular application and of the specific, enumerated actions the sponsor would be required to take in order for the sponsor to receive a final action letter that addresses the deficiencies of the application, which then may lead to approval of such application.

"Under subsection (B), addressing specific deficiencies of the particular applications and engaging in the specific enumerated action does not require the Secretary to approve the application, if other deficiencies are identified associated with the original application, or on the basis of further, submitted information.

"(9) The term 'final action date' means, with respect to a comparable biological product application, the date that is eighteen calendar months following the Secretary's acceptance of a sponsor's submission of such application, except that the final action date hereunder may be extended for such period of time as is agreed to by the Secretary and the sponsor of such application in a jointly executed written agreement that is countersigned by the Secretary and the sponsor of such application no later than thirty days prior to the final action date provided for by this subsection.

"(10) The term 'reviewing division' means the division responsible for the review of an application for approval of a biological product (including all scientific and medical matters, chemistry, manufacturing, and controls)."

SECTION 4. CATEGORIZATION OF CERTAIN BIOLOGICAL PRODUCTS.

(a) IN GENERAL.—Section 351 of the Public Health Service Act (42 U.S.C. 262) is amended—

(1) in subsection (a)(1)(A), by inserting after "biologics license" the following: ," or comparable biologics license,"; and

(2) by adding at the end the following subsection:

"(k) REGULATION OF COMPARABLE BIOLOGICAL PRODUCTS.—

"(1) BIOLOGICAL PRODUCT CATEGORIZATION.—The Secretary shall categorize all currently approved reference biological products. The Secretary shall:

"(A) utilize the following categories that shall be amended, changed, or revised from time to time, subject to subsection (C) herein:

"(I) Presumptively Well-Known Mechanism(s) of Action

"(A) Without Known Adverse Immunogenicity;

"(B) With Known Adverse Immunogenicity;

"(II) Partially Known Mechanism(s) of Action

"(A) Without Known Adverse Immunogenicity;

"(B) With Known Adverse Immunogenicity;

"(III) Unknown Mechanism(s) of Action

"(A) Without Known Adverse Immunogenicity;

"(B) With Known Adverse Immunogenicity.

"(B) allow any person to petition the Secretary to have a currently approved reference biological product moved from one category to another.

"(C) review the categories for approved reference biological products biannually to ensure substantive scientific relevance, effectiveness, and efficiency for review of comparable biological product applications."

"(2) GUIDANCE BY CATEGORY.—The Secretary shall issue guidance by categories denoted in subparagraph (1) herein on the general requirements that all comparable biological product applications must fulfill for application review. This guidance shall include information and requirements associated with comparable biological product applications, including:

"(A) pharmacokinetic studies;

"(B) pharmacodynamic studies;

"(C) pharmacokinetic/pharmacodynamic studies;

"(D) clinical comparability studies;

"(E) immunogenicity studies;

"(F) adverse reaction studies;

"(G) pharmacovigilence monitoring;

"(H) traceability methodology;

"(I) manufacturing methodology;

"(J) quality assessment methodology;

"(K) post-marketing studies; and

"(L) any other assessments necessary to evaluate the comparability, safety, purity, potency, and function of the comparable biological product.

"(3) SPECIFIED INFORMATION.—Notwithstanding the previous section requirements, the Food Drug Administration shall require the comparable biological product applicant to provide evidence of investigation of immunogenicity issues including data from investigation prior to the comparable biological product application; to monitor the clinical safety of its drug during the post approval period; and within its application, to provide for a risk specification and pharmacovigilance program plan that includes a description of potential safety issues that may be a result of differences in the manufacturing process from the originator biologic.

"(4) AUTHORITY OF THE FOOD AND DRUG ADMINISTRATION.—

"(A) The Secretary shall designate the Food and Drug Administration with the authority to review comparable biological drug applications.

"(B) The Food and Drug Administration may utilize its own agency expertise in evaluating comparable biological drug applications, and may utilize other experts for the purpose of evaluating these applications.

"(5) MEETINGS WITH APPLICANTS.—

"(A) GUIDANCE FOR SCIENTIFIC EVIDENCE.—The Food and Drug Administration, upon the written request of any applicant intending to submit a comparable biological product application, shall meet with such applicant to determine the type of valid scientific evidence that will be necessary to demonstrate for purposes of approval of an application the safety, purity, potency, and function of a comparable biological product and the conditions of use proposed by such applicant. The written request shall include a detailed description of the comparable biological product, a detailed description of the proposed conditions of use of the comparable biological product, and a proposed plan for determining whether there is a reasonable assurance of safety, purity, potency, and function of the comparable biological product. Within thirty days after such meeting, the Food and Drug Administration shall specify in writing the type of valid scientific evidence that will provide a reasonable assurance that the comparable biological product is safe, pure, potent, and functional under the conditions of use proposed by such person.

"(B) LEAST BURDENSOME MEANS.—The Food and Drug Administration shall consider, in consultation with the applicant, the least burdensome appropriate means of evaluating the comparable biological product's safety, purity, potency, and function that have a reasonable likelihood of resulting in application approval.

"(C) BINDING NATURE.—Any written response by the Food and Drug Administration under the provisions of paragraph (A), or in consultations with the applicant under the provisions of paragraph (B), regarding the parameters of valid scientific evidence of a comparable biological product, shall be reduced to writing and made part of the administrative record. Such agreement shall not be changed after collection of such scientific evidence begins, except—

"(i) with the written agreement of the sponsor or applicant; or

"(ii) pursuant to a decision, made in accordance with subparagraph (D) by the director of the reviewing division, that a substantial scientific issue essential to determining the safety, purity, potency, or function of the comparable biological product has been identified after the testing has begun. Any challenge by a comparable biological product applicant to a determination that there is a substantial scientific issue must be shown by clear and convincing evidence.

"(D) SUBSTANTIAL SCIENTIFIC ISSUE MEETING.—A decision under subparagraph (C)(ii) by the director shall be in writing and the Secretary shall provide to the sponsor or applicant an opportunity for a meeting at which the director and the applicant will be present and at which the director will document the scientific issue(s) involved.

"(E) WRITTEN DECISIONS AND FIELD AND COMPLIANCE PERSONNEL.—The written decisions of the reviewing division shall be binding upon, and may not directly or indirectly be changed by, the field or compliance division personnel unless such field or compliance division personnel demonstrate to the reviewing division why such decision should be modified.

"(F) DELAY.—The FDA may delay the reviewing division action because of the unavailability of information, or for any reason to assure the marketing of a safe, pure, potent, and functional drug.

"(6) APPROVAL OF COMPARABLE BIOLOGICAL PRODUCTS.—

"(A) The Food and Drug Administration shall review the information submitted in the comparable biological product application and any other information available, subject to subsections (B) and (C) of this paragraph.

"(B) The Food and Drug Administration, in its review under subsection (A), shall not disclose information deemed a trade secret of the reference biological drug.

"(C) The Food and Drug Administration, in its review under subsection (A), may use reference biological drug information, in-

cluding conclusions regarding safety and efficacy as well as conditions of its approval of the reference biological drug, in its assessment of the comparable biological product application.

"(D) The Food and Drug Administration shall not approve a comparable biological product application for any conditions of use relating to the reference product if the comparable biological product is not shown as comparable under terms and provisions of this section, and shall provide the applicant with a written, detailed explanation for its decision.

"(E) The Food and Drug Administration shall not approve a comparable biological product application for any conditions of use relating to the reference product unless the comparable biological product application:

"(i) shows the comparable biological product and the reference product have comparable principal molecular structural features as demonstrated by thorough categorization of the two products;

"(ii) shows that the comparable biological product is comparable to the reference product for the conditions of use prescribed, recommended, or suggested in the labeling proposed in the application;

"(iii) shows that the comparable biological product and reference product use the same mechanism(s) of action for the conditions of use prescribed, recommended, or suggested in the labeling proposed for the comparable biological product. However, this section shall not apply if the mechanism(s) of action is not known for the reference product for such conditions;

"(iv) shows that the route of administration, the dosage form, and the strength of the comparable biological product are the same as those of the reference product;

"(v) shows that the conditions of use prescribed, recommended, or suggested in the labeling proposed for the comparable biological product are limited to one or more of the same use(s) as have been previously approved for the reference product;

"(vi) shows that the inactive ingredients of the comparable biological product are not unsafe for use under the conditions prescribed, recommended, or suggested in the labeling proposed for the biological product, or the composition of the comparable biological product is not unsafe under such conditions because of the type or quantity of inactive ingredients included or the manner in which the inactive ingredients are included;

"(vii) shows that the facility in which the comparable biological product is manufactured, processed, packed, or held meets standards designed to assure that the comparable biological product continues to be safe, pure, potent, and functional; and

"(viii) the Secretary has not withdrawn or suspended the license of the reference product, for safety or effectiveness reasons, or has

published a notice of opportunity for hearing to withdraw such license for safety or effectiveness reasons, or has determined that the reference product has been withdrawn from sale for safety or effectiveness reasons; and

"(ix) the application does not contain an untrue statement of material fact; and

"in any event, shall provide the applicant with a written, detailed explanation for the decision.

"(F) If the Food and Drug Administration does not approve a comparable biological product application, within its written, detailed explanation for the decision, it shall give the applicant notice of an opportunity for a hearing before the Secretary on the question of whether the application is approvable. If the applicant elects to accept the opportunity for a hearing by written request within thirty days after such notice, such hearing shall commence not more than ninety days after the expiration of such thirty days unless the Secretary and the applicant otherwise agree. Any such hearing shall thereafter be conducted on an expedited basis and the Secretary's order thereon shall be issued within ninety days after the date fixed by the Secretary for filing final briefs. The Secretary shall take a final action on a comparable biological product application by the final action date.

"(G) NAMING OF COMPARABLE BIOLOGIC PRODUCTS.—If a comparable biological product is approved, the Food and Drug Administration shall assign the comparable biologic product a unique name. The Food and Drug Administration shall, within 180 days of the passage of this Act, work with international authorities to harmonize naming protocol to allow unique names for comparable biological products approved under this section.

"(H) INTERCHANGEABLE COMPARABLE BIOLOGICAL PRODUCTS.—

"(i) A comparable biological product applicant may request in an original application or in a supplement that the Food and Drug Administration make a determination that the comparable biological product is interchangeable with the reference product.

"(ii) If a comparable biological product applicant requests the Food and Drug Administration to make a determination that the comparable biological product is interchangeable, the Food and Drug Administration shall first determine if the comparable biological product is comparable under the provisions of this Act.

"(iii) If the Food and Drug Administration approves the comparable biological product, it will then make a determination of the interchangeability of the comparable biological product if requested under section (H)(i). The Food and Drug Administration may approve a comparable biological product for marketing before a deci-

sion is made on the interchangeability of the comparable biological product.

"(iv) If the Food and Drug Administration deems the comparable biological product interchangeable, it shall assign the comparable biological product an identifier and publish a therapeutic comparability code indicating that the comparable biological product is interchangeable with the reference product.

"(v) For interchangeable products, the comparable biological product found interchangeable with the reference product shall be named using the INN protocol with the reference product INN; however, a numerical appendage shall be made on the basis of the order in which the comparable biological product was shown interchangeable, such as INN-1 for the first comparable biological product that was found interchangeable, INN-2 for the second comparable biological product that was found interchangeable, etc.

"(vi) Interchangeable comparable biological products may include a statement upon its label indicating that it is interchangeable with the biological reference product to which the sponsor of the comparable biological product application has demonstrated comparability to the reference product for the conditions of use prescribed, recommended, or suggested in the labeling proposed for the comparable biological product.

"(7) ACCESS TO DRUGS AND DATA EXCLUSIVITY.—

"(A) The Secretary shall direct the Department of Health and Human Services Office of Minority Health to—

"(i) identify public and private low and no-cost drug programs in the United States, including those with culturally competent language translation services, and identify all state-level Offices of Minority Health;

"(ii) develop a nationwide program for low and no-cost drugs for minority and vulnerable patient populations under 400% of the federal poverty level utilizing and expanding the programs identified in section (A)(i) above, with the assistance of the Department Advisory Committee on Minority Health, state-level Offices of Minority Health, and industry members and groups, as appropriate;

"(iii) work with state governments to utilize the program developed in (A)(ii) to enroll participants into eligible health programs, including Medicaid, state children's health insurance programs, Supplemental Security Income, and state high risk insurance programs; and

"(iv) develop appropriate education, terms, and conditions of participation to ensure that access to biological drugs is provided to minority and vulnerable patient populations and that identification of any adverse reactions or events associated with these drugs are noted, reported, and disseminated.

"(B) No comparable biological product application may rely upon investigations of a reference drug application under section 351 of the Public Health Service Act performed for a successful Biological License Application that were not conducted by or for the comparable biological product applicant for approval of the comparable biological product application and for which the comparable biological product applicant has not obtained a right of reference or use from the person by or for whom the investigations were conducted, before five years from the date of the approval of the reference drug Biological License Application; and the Food and Drug Administration shall not use or consider such studies in any comparable biological product application under section (6), except a comparable biological application may be submitted and investigations used under section (6) after the expiration of four years from the date of approval of the reference drug application if it contains a certification of patent invalidity or noninfringement described in section 9(C), subject to subsection (C) of this section.

"(C) The Food and Drug Administration may consider reference drug applicant investigations without regard to the five-year exclusion noted in section (B) in a comparable biological product application under subsection (6) if the reference drug Biological License Application applicant in section (B) does not participate in the nationwide drug program described in section (A) through provision of the approved reference drug to the nationwide drug program.

"(D) For any applicant that has a comparable biological product application approved by the Food and Drug Administration under section (B) or (C), such applicant must participate in the nationwide drug program described in section (A) within eighteen months of the date of application approval.

"(E) Subject to the provisions and terms of subsection (A)(v), to obtain the data exclusivity period of five years noted in section (B), a reference drug applicant must agree to participate in the nationwide program for a minimum of fifteen years after the date of approval of its Biological License Application or until the reference drug is withdrawn from the market; and subject to the provisions and terms of section (A)(v) and section (D), the producer of the comparable biological product must agree to participate in the nationwide program for a minimum of ten years after the date of approval of its comparable biological products application or until the comparable biological product is withdrawn from the market.

"(8) MARKETING EXCLUSIVITY.—

"(a) If the Food and Drug Administration approves a comparable biological product, and the Food and Drug Administration approves such comparable biological product as the first interchangeable version of the reference drug, the Food and Drug Administra-

tion shall not approve a second interchangeable or subsequent interchangeable comparable biological product application, and no holder of a biological product license under subsection (a) shall manufacture, market, sell, or distribute a rebranded interchangeable biological that is interchangeable with the reference product, until the earlier of—

"(i) Ninety days after the first commercial marketing of the first interchangeable comparable biological product to be approved as interchangeable for that same reference product; or

"(ii) one year after—

"(I) a final court decision on all patents in suit in an action instituted under paragraph (9) against the applicant for the first approved interchangeable comparable biological product; or

"(II) the dismissal with or without prejudice of an action instituted under paragraph (9) against the applicant that submitted the application for the first approved interchangeable comparable biological product; or

"(iii) either—

"(I) Thirty-six months after approval of the first interchangeable comparable biological product if the applicant has been sued under paragraph (9) and such litigation is still ongoing within such thirty-six month period; or

"(II) one year after approval in the event that the first approved interchangeable comparable applicant has not been sued under paragraph (9).

"(b) For the purposes of this section—

"(i) Notwithstanding the foregoing provision, the applicant for a subsequent comparable biological product application that has demonstrated interchangeability with the reference product may elect, at its option, to have the product approved as a non-interchangeable comparable biological product whose approval will not be delayed by operation of this paragraph.

"(ii) A 'final court decision' means a final decision of a court from which no appeal, other than a petition to the United States Supreme Court for a writ of certiorari, has been or can be taken.

"(iii) A 'rebranded interchangeable biologic' means any rebranded interchangeable version of a reference product that the holder of the biological product license approved under subsection (a) for that reference product seeks to commence marketing, selling, or distributing, directly or indirectly, but does not include any product to be marketed, sold, or distributed by an entity eligible for exclusivity with respect to such product under this paragraph or after expiration of any exclusivity with respect to such product under this paragraph.

"(9) PATENTS.—

"(A) REQUEST FOR PATENT INFORMATION.—

"(i) IN GENERAL.—At any time, including at the initial stages of development, but no later than the date upon which the applicant files its comparable biological product application with the Secretary, an applicant or prospective comparable biological product applicant may send a written request for patent information to the holder of the approved application for the reference product. Within sixty days of receipt of such request, the holder of the approved application for the reference product shall provide to the applicant or prospective comparable biological product applicant a list of all patents owned by, or licensed to, the holder of the approved application that the application holder in good faith believes relate to the reference product, including patents that claim the approved biological product, any method of using such product, any component of such product, or any method or process of manufacturing such product or component.

"(ii) COSTS OF COMPLYING WITH REQUEST.—The application holder may demand payment not exceeding $1,000 to offset the cost of responding to the information request, subject to adjustment from time to time by the Secretary to reflect increased costs of fulfilling such requests.

"(iii) UPDATES.—For a period of two years from the date of the request for information, the holder of the approved application for the reference product shall update its response to the request for information by identifying newly issued or licensed relevant patients. The updates must be provided within thirty days of patent issuance for newly issued patents, and within thirty days of obtaining a license for newly licensed patents.

"(iv) ADDITIONAL REQUESTS.—The comparable biological product applicant may submit additional requests for patent information, subject to the requirements of this paragraph, at any time.

"(B) PATENT NOTIFICATIONS.—The comparable biological product applicant must provide a notice under this subparagraph with respect to any one or more patents provided by the holder of the reference product, if the provisions under subparagraph (B)(ii) apply. Each notice shall—

"(i) be sent to the holder of the approved application for the reference product and to the owner of the patent identified pursuant to subparagraph (A)(i);

"(ii) include a detailed statement of the factual and legal bases for the applicant's belief that the patents included in the notice are invalid, unenforceable, or will not be infringed by the commercial sale of the product for which approval is or has been sought; and

"(iii) identify the judicial district or districts in which the applicant consents to suit being brought in response to the notice.

"A comparable biological product applicant, within the comparable biological product application, shall certify that, if applicable, such notice has been sent to the parties noted in subparagraph (B)(i).

"(C) ACTION FOR INFRINGEMENT.—

"(i) TIMEFRAME FOR BRINGING ACTION.—Within forty-five days of receipt of notice described in subparagraph (B), the holder of the approved application for the reference product, or the owner of the patent, may bring an infringement action solely with respect to the patent or patents included in such notice.

"(ii) APPROPRIATE JUDICIAL DISTRICT.—Notwithstanding section 1391 of title 28, United States Code, an infringement action brought within the forty-five day period reference in clause (i) may be brought only in the judicial district identified pursuant to subparagraph (B)(ii).

"(D) DECLARATORY JUDGMENTS.—

"(i) If a recipient of notice under subparagraph (B)(ii) does not initiate an infringement action as specified under subparagraph (C)(i), the applicant may bring a declaratory judgment action under section 2201 of title 28, United States Code, for invalidity, noninfringement, and nonenforceability of the patents within the notice under subparagraph (B)(ii).

"(ii) The declaratory action under subparagraph (D)(i) shall be brought only after—

"(I) the applicant sends a request to the recipient requesting a covenant not to sue with respect to the patents within the notice under subparagraph (B)(ii); and

"(II) the recipient of the request indicated in subparagraph (D)(ii)(I) does not issue such covenant not to sue to the applicant within thirty days of receipt of the request in subparagraph (D)(ii)(I).

"(iii) If the recipient of the request indicated in subparagraph (D)(ii)(I) does not issue a covenant not to sue with respect to the patents within the notice under subparagraph (B)(ii), this shall be deemed a 'reasonable apprehension' as that term is used in section 2201 of title 28, United States Code.

"(10) PETITIONS AND CIVIL ACTIONS REGARDING APPROVAL OF CERTAIN APPLICATIONS.—Section 505 of the Federal Food, Drug, and Cosmetic Act (21 U.S.C. 355) is amended by adding at the end the following:

"(o) CITIZENS PETITIONS AND PETITIONS FOR STAY OF AGENCY ACTION.—With respect to any petition that seeks to have the Secretary take, or refrain from taking, any form of action relating to the approval of a comparable biological product application, the following shall apply:

"(1) NO DELAY OF APPROVAL.—The Secretary shall not delay approval of a comparable biological product application submitted un-

der this Act while a petition is reviewed and considered. Consideration of a petition shall be separate and apart from the review and approval of an application submitted under this Act, unless such petition review directly implicates the public health and safety associated with the comparable biological product in the application.

"(2) TIMING OF FINAL AGENCY ACTION.—The Secretary shall take final agency action with respect to a petition within six months of receipt of that petition. The Secretary shall not extend such six-month review period, even with consent of the petitioner, including based upon the submission of comments relating to a petition or supplemental information supplied by the petitioner. If the Secretary has not taken final agency action on a petition by the date that is six months after the date of receipt of the petition, such petition shall be deemed to have been denied on such date.

"(3) PUBLIC HEALTH ISSUES.—Notwithstanding the provisions indicated in section (o)(2), the Secretary may extend such six-month review period and delay approval of a comparable biological product application when such a delay is necessary to protect the public health and welfare. If the Secretary extends the six-month review period, or delays approval of a comparable biological product, then—

"(A) The Secretary shall publish on the Internet site of the Food and Drug Administration a statement providing the reasons underlying the determination.

"(B) Not later than ten days after making the determination, the Secretary shall provide notice to the comparable biological product applicant and an opportunity for a meeting with the Commissioner of the Food and Drug Administration to discuss the determination.

"(4) VERIFICATION.—The Secretary shall not accept for review a petition unless it is signed and contains the following verification: 'I certify that, to my best knowledge and belief: (a) this petition includes all information and views upon which the petition relies; (b) this petition includes representative data and/or information known to the petitioner which are unfavorable to the petition; and (c) I have taken reasonable steps to ensure that any representative data and/or information which are unfavorable to the petition were disclosed to me. I further certify that the information upon which I have based the action requested herein first became known to the party on whose behalf this petition is filed on or about _____. I verify under penalty of perjury that the foregoing is true and correct.,' with the date of the filing of such petition inserted in the blank space.

"(5) For the purposes of this paragraph, the term 'petition' includes any request to the Secretary, without regard to whether the request is characterized as a petition.

"(6) The Secretary shall annually submit to the Congress a report that specifies the number of applications under this subsection

that were approved during the preceding twelve-month period, the number of such applications whose effective dates were delayed by petitions under this subsection, and the number of days by which the applications were so delayed."

"(11) PROHIBITION AGAINST BIOLOGICAL DRUG SALES VIA IMPORTATION OR THE INTERNET.—

"(A) All biological drugs, whether they be originator biological drugs or approved comparable biological drugs, shall be excluded from any importation program and shall not be permitted to be imported, except under the provisions of the Food, Drug, and Cosmetic Act, by manufacturers and authorized distributors under section 381(d)(1) of title 21 of the United States Code.

"(B) All biological drugs, whether they be originator biological drugs or approved comparable biological drugs, shall not be subject to sales through Internet sellers."

RECENT DEVELOPMENTS

THE MILITARY COMMISSIONS ACT OF 2006

In response to the Supreme Court's decision in *Hamdan v. Rumsfeld*,[1] Congress passed the Military Commissions Act of 2006 ("MCA").[2] The MCA remedies some of the statutory and constitutional infirmities cited by the Court in *Hamdan*.[3] In addition, the MCA limits the liability of government officials responsible for conducting interrogations[4] and features a jurisdiction-stripping provision that prevents all courts from hearing and considering writs of habeas corpus filed by alien enemy combatants.[5] This Recent Development examines the Supreme Court's approach to the extra-territorial application of the Constitution and concludes that the jurisdiction-stripping provision of the MCA is inconsistent with the reach of constitutional guarantees as they have been defined in cases arising from the war on terror.

This Recent Development is divided into six parts. Part I summarizes the *Hamdan* Court's criticisms of pre-MCA military commissions as well as the standards to which the Court held they must adhere. Part II analyzes the MCA in light of these requirements and concludes that only some have been meet. In Part III, I examine the possible constitutional justifications for the MCA. Specifically, I discuss a recent D.C. Circuit decision in which the court upheld the jurisdiction-stripping provision on the grounds that it is consistent with both the scope of habeas corpus at common law as well as Supreme Court precedent.[6] However, an analysis of common law authority and the Supreme Court's recent treatment of this issue reveal that neither supports the D.C. Circuit's contention that Guantanamo Bay Naval Base is Constitution-free. In Part IV, I apply the contextual due process test for the extraterritorial application of the Constitution. Though the Supreme Court has not expressly rejected the test used by the D.C. Circuit, the Court has criticized it at length and has indicated that a contextual due process approach is better suited for determining the extent to which the Constitution applies at Guantanamo. The contextual due process analysis involves a balancing of factors which, when applied to the situation of Guantanamo detainees, weigh in favor of extending the constitutional rights of habeas corpus and due process. Part V explores the possible constitutional justifications for suspending the writ,

---

[1] 126 S. Ct. 2749 (2006).

[2] Pub. L. No. 109-366, 120 Stat. 2600 (2006) (codified in scattered sections of 10, 18, and 28 U.S.C.).

[3] *See infra* Parts I and II.A.

[4] *See* MCA § 6, 18 U.S.C. § 2441(d) (2006).

[5] *See* MCA § 7, 28 U.S.C. § 2241(e) (2006).

[6] Boumediene v. Bush, 476 F.3d 981 (D.C. Cir. 2007), *cert. denied*, 127 S. Ct. 1478 (2007).

and Part VI assesses the relative advantages and disadvantages of providing detainees with habeas corpus review by either legislative or judicial action.

I. IMPETUS FOR THE MCA: *HAMDAN V. RUMSFELD*

In the 2006 term, the Supreme Court granted the writ of habeas corpus of Salim Ahmed Hamdan, a Yemeni national held in custody at Guantanamo Bay.[7] Hamdan had been captured by militia forces in Afghanistan.[8] After being held for two years, Hamdan was charged with conspiracy to commit offenses triable by military commission.[9]

Justice Stevens, speaking for the Court,[10] held that the military commission convened to try Hamdan was unlawful on two grounds.[11] First, the Court found that the commission was not expressly authorized by any congressional act.[12] Because the Constitution does not provide for military commissions, the Court turned to common law as a possible source of authority.[13] Here, the Court found that the military tribunal convened to try Hamdan failed to satisfy two of the four common law preconditions for the exercise of jurisdiction.[14] First, the alleged acts committed by Hamdan did not occur in a theater of war or during the relevant conflict.[15] Second, conspiracy—the alleged offense—was not a violation of the laws of war and therefore not triable by a law-of-war military commission.[16]

---

[7] *See Hamdan*, 126 S. Ct. at 2759.

[8] *See id.*

[9] *See id.*

[10] Justices Souter, Ginsburg, and Breyer joined the majority opinion in part. *See id.* Justices Kennedy and Breyer filed concurring opinions. *See id.* Justices Scalia, Thomas, and Alito each filed dissenting opinions. Chief Justice Roberts did not participate in the decision. *See id.*

[11] *See id.*

[12] *See id.* at 2774–75. The court found that neither the Detainee Treatment Act ("DTA"), Pub. L. No. 109-148, 119 Stat. 2680 (2005) (codified at 10 U.S.C. § 948r(c)–(d)(2006)), nor the Authorization to Use Military Force ("AUMF"), Pub. L. No. 107-44, 115 Stat. 224 (2001), contained language that authorized commissions for unlawful enemy combatants.

[13] *See Hamdan*, 126 S. Ct. at 2773–75 (2006).

[14] *See id.* at 2777–79. The Court cites W. WINTHROP, MILITARY LAW AND PRECEDENTS 836–38 (rev. 2d ed. 1920), which describes four preconditions for the exercise of jurisdiction by a tribunal of the type convened to try Hamdan. *See id.* at 2777. First, a military commission only has jurisdiction over offenses committed in a theater of war. *Id.* Second, the offense charged must have been committed within the period of the war. *Id.* Third, a military commission not established pursuant to martial law or an occupation may only try enemy soldiers suspected of illegitimate warfare or other violations of the laws of war, as well as members of one's own army "who, in time of war, become chargeable with crimes or offences not cognizable, or triable, by the criminal courts or under the Articles of war." *Id.* (citations omitted). Finally, a law-of-war commission has jurisdiction to try only violations of the laws and usages of war cognizable by military tribunals as well as breaches of military orders or regulations for which offenders are not legally triable by court-martial under the Articles of war. *Id.* (citations and quotations omitted).

[15] *See Hamdan*, 126 S. Ct. at 2777–79.

[16] *See id.* at 2779.

In Article 21 of the Uniform Code of Military Justice ("UCMJ"),[17] Congress explicitly provided that military commissions shall have jurisdiction to try offenses against the laws of war.[18] While the boundaries of this body of law are unclear, the Court asserted that it at least encompasses the Geneva Conventions.[19] The military commission convened to try Hamdan did not satisfy Common Article 3 of the Geneva Conventions,[20] however, because it failed to provide "all the judicial guarantees which are recognized as indispensable by civilized peoples," a condition that the Convention imposes upon signatories.[21] Specifically, the Court noted that the tribunal could exclude a defendant and his counsel from the proceedings.[22] Furthermore, the tribunal could admit, upon a finding of reliability, hearsay, and statements gained through coercion.[23] The Court also noted that trying Hamdan for conspiracy violated Common Article 3 because "conspiracy has rarely if ever been tried in this country by military commission and does not appear in the major treaties on the law of war."[24]

II. THE MILITARY COMMISSIONS ACT

A. Congress Responds to Hamdan v. Rumsfeld

Immediately after the Supreme Court held that the military commissions convened at Guantanamo lacked statutory authority and were therefore unconstitutional, the White House proposed legislation to Congress that would remedy this problem as well as grant the Executive additional powers that the White House deemed necessary to fight the war on terror.[25] After two months of negotiations between the Bush administration and reluctant members of its own party,[26] the Senate passed the MCA on Sep-

---

[17] 10 U.S.C. § 821 (2006).

[18] 10 U.S.C. § 821.

[19] *See Hamdan*, 126 S. Ct. at 2794 ("[The Geneva Conventions] are, as the Government does not dispute, part of the law of war . . . . And compliance with the law of war is the condition upon which the authority set forth in Article 21 is granted." (citation omitted)).

[20] Geneva Convention Relative to the Treatment of Prisoners of War Art. 3, adopted Aug. 12, 1949, 6 U.S.T. 3316, 75 U.N.T.S. 135 [hereinafter Third Geneva Convention].

[21] *Hamdan*, 126 S. Ct. at 2795–98. *See also* Third Geneva Convention, *supra* note 20.

[22] *See Hamdan*, 126 S. Ct. at 2786 ("The accused and his civilian counsel may be excluded from, and precluded from ever learning what evidence was presented during, any part of the proceeding that either the Appointing Authority or the presiding officer decides to 'close.'").

[23] *See id.* at 2798.

[24] *Id.* at 2780–81.

[25] *See* Kate Zernike, *Senate Approves Broad New Rules to Try Detainees*, N.Y. TIMES, Sept. 29, 2006, at A1; Martin Kady II, *Congress Clears Detainee Bill*, CQ WEEKLY, Sept. 29, 2006.

[26] *See* Kady, *supra* note 25. Senators John W. Warner (R-Va.), chairman of the Armed Services Committee, John McCain (R-Ariz.), and Lindsey Graham (R-S.C.) derailed the Bush administration's original attempt to pass the MCA and forced the White House into nego-

tember 28, 2006.[27] The companion bill cleared the House on the following day[28] and President Bush signed the bill into law on October 17, 2006.[29]

In direct response to the *Hamdan* Court's finding that military commissions lacked statutory authorization,[30] the MCA gives the President the authority to establish military commissions for unlawful enemy combatants.[31]

The MCA also partially addresses the second defect of the military commission noted in *Hamdan*, i.e., that it did not satisfy the standards of Common Article 3 of the Geneva Conventions as incorporated into the UCMJ. The MCA declares as a matter of Congressional *ipse dixit* that the Act affords all the "'judicial guarantees which are recognized as indispensable by civilized peoples' for purposes of common Article 3 of the Geneva Conventions."[32] Despite this language, the actual set of rights granted in the MCA falls short of that standard. For example, the MCA shifts the burden of contesting hearsay evidence introduced for the purpose of establishing guilt to the defendant.[33] If the defendant cannot demonstrate that the hearsay evidence offered by the prosecutor is unreliable or lacking in probative value, it is deemed admissible.[34] The MCA also fails to address the Court's concern about the use of statements obtained through coercion.[35] The MCA provides that a statement that is the product of coercion or compulsory self-incrimination is admissible so long as a military judge finds it reliable and probative.[36] Moreover, despite

---

tiations. *See id.*

[27] S. 3930, 109th Cong. § 2 (2006). The MCA passed in the Senate by a vote of 65–34. 152 CONG. REC. S10420–31 (daily ed. Sept. 28, 2006).

[28] H.R. 1054, 109th Cong. § 2 (2006). The House version passed by a vote of 250–170. *See* 152 CONG. REC. H7925–36 (daily ed. Sept. 29, 2006).

[29] The President did not issue a signing statement for the MCA. His press secretary stated that a statement was unnecessary because the administration believed the MCA would pass constitutional muster. *See* Tony Snow, White House Press Sec'y, Press Briefing at the White House Conference Center Briefing Room (Oct. 17, 2006), *available at* http://www.whitehouse.gov/news/releases/2006/10/20061017-3.html. *See also* Sheryl Gay Stolberg, *Bush Signs Bill on Terror Prosecution*, N.Y. TIMES, Oct. 18, 2006, at A20.

[30] *See* Hamdan v. Rumsfeld, 126 S. Ct. 2749, 2774–75 (2006).

[31] MCA § 3(a)(1), 10 U.S.C. § 948b(b) (2006).

[32] MCA § 3(a)(1), 10 U.S.C. § 948b(f) (2006) (citing Third Geneva Convention, *supra* note 20).

[33] *See* MCA § 3(a)(1), 10 U.S.C. § 949a(b)(2)(E)(i)–(ii) (2006). *But cf.* Hamdi v. Rumsfeld, 542 U.S. 507, 534 (2004) (suggesting a similar burden-shifting scheme with respect to determining the status of possible enemy combatants).

[34] *See* MCA § 3(a)(1), 10 U.S.C. § 949a(b)(2)(E)(ii) (2006).

[35] *See Hamdan*, 126 S. Ct. at 2786–87 (2006).

[36] *See* MCA § 3(a)(1), 10 U.S.C § 948r(c)–(d) (stating that the articles of the UCMJ relating to compulsory self-incrimination do not apply to trial by military commissions).

The MCA distinguishes between coercion that occurred before and after the enactment of the DTA. *See id.* Because the DTA prohibits cruel, inhuman, or degrading treatment, *see* DTA § 1003, 10 U.S.C. 948r (2006), statements obtained after the DTA's enactment through interrogation methods that employed such treatment are inadmissible. *See* MCA § 3(a)(1), 10 U.S.C. § 948r(d) (2006). Statements obtained before the DTA's enactment in which the degree of coercion is disputed are deemed by the MCA to be admissible, provided that the totality of the circumstances renders the statement reliable and possessing

the Court's finding that the law of war, which the UCMJ incorporates,[37] does not recognize the charge of conspiracy, the MCA includes it as a substantive offense that may be punishable by death.[38]

Though failing to provide several procedural guarantees listed by the Court in *Hamdan*, the MCA makes substantial improvements in other areas. Detainees now have the right, with narrow exceptions,[39] to attend all sessions of the military commission, other than those for deliberation or voting.[40] In addition, detainees must be allowed to examine and respond to evidence seen by the commission.[41] If the evidence is classified, the government must supply the detainee with an unclassified summary.[42] The MCA also establishes the right to counsel by either a Judge Advocate General or civilian counsel.[43]

B. Beyond Hamdan*: Other Major Changes Enacted by the MCA*

In addition to responding to *Hamdan*, the MCA changes the liability of government officials,[44] permits indefinite detention,[45] and strips the courts of their jurisdiction to consider writs of habeas corpus filed by detainees.[46]

The MCA sets the boundaries of permissible interrogation by defining torture through a list of abuses.[47] Human rights groups have lauded the MCA for criminalizing the most extreme interrogation methods.[48] However, they also note that because the MCA defines torture in the narrowest of terms, it greatly limits the liability of government officials who mistreat detainees.[49]

---

sufficient probative value and that the interests of justice would best be served by admission of the statement into evidence. *See* MCA § 3(a)(1), 10 U.S.C. § 948r(c) (2006).

[37] The Court in *Hamdan* noted that UCMJ, Art. 21, 10 U.S.C. § 821 (2006), grants the President the authority to convene commissions on express condition that the President and those under his command comply with law of war. *See Hamdan*, 126 S. Ct. at 2774.

[38] *See* MCA § 3(a)(1), 10 U.S.C. § 950v(b)(28) (2006).

[39] *See* MCA § 3(a)(1), 10 U.S.C. § 949d(c) (2006) (stating that the accused shall be excluded when members of the military commission deliberate or vote); MCA § 3(a)(1), 10 U.S.C. § 949d(e) (2006) (stating that the accused may be excluded if he persists in conduct that requires exclusion in order to ensure the physical safety of individuals or to prevent the disruption of the proceedings).

[40] *See* MCA § 3(a)(1), 10 U.S.C. § 949d(c) (2006).

[41] *See* MCA § 3(a)(1), 10 U.S.C. § 949a(b)(1)(A) (2006).

[42] *See* MCA § 3(a)(1), 10 U.S.C. § 949d(f)(2)(A) (2006).

[43] *See* MCA § 3(a)(1), 10 U.S.C. § 949c(b)(3) (2006).

[44] *See infra* notes 47–55 and accompanying text.

[45] *See infra* notes 56–59 and accompanying text.

[46] *See infra* notes 60–62 and accompanying text.

[47] *See* MCA § 3(a)(1), 10 U.S.C. § 948r(b) (2006).

[48] *See, e.g.*, Scott Shane & Adam Liptak, *Shifting Power to a President, Bill Creates Legal Basis for Policy on Detainees*, N.Y. TIMES, Sept. 30, 2006, at A1 (quoting Elisa Massimino, Washington director of Human Rights First).

[49] *See, e.g.*, Press Release, Amnesty Int'l, Military Commissions Act of 2006—Turning Bad Policy into Bad Law (Sept. 29, 2006), *available at* http://web.amnesty.org/library/Index/engamr511542006.

Common Article 3 of the Third Geneva Convention requires that each signatory criminalize acts of torture conducted by government officials.[50] The War Crimes Act fulfills this obligation by providing a cause of action to individuals who have suffered "grave breaches" of the standards set forth in Common Article 3.[51] Under the MCA, torture as well as cruel and inhuman treatment are considered to be "grave breaches" of Common Article 3, while degrading or humiliating treatment is not.[52] Moreover, the MCA's definition of cruel and inhuman treatment is limited to extreme abuses, such as conduct that causes substantial risk of death, physical disfigurement, and organ loss or impairment.[53] The Act's definitions of specific abuses prohibited by Common Article 3 are also narrowly defined.[54] For example, the definitions of rape and sexual assault cover coerced activity, but not other forms of nonconsensual sex.[55]

Another problem cited by many critics is that the MCA does not address the possibility of indefinite detention.[56] The MCA requires that detainees be notified of the charges against them prior to trial,[57] but it contains no requirement that this trial occur within a specified period of time, or even that it occur at all.[58] The MCA explicitly affirms that there is no time constraint by proclaiming that the article of the UCMJ that guarantees defendants the right to a speedy trial does not apply to military commissions.[59]

The most criticized aspect of the MCA is section 7, the jurisdiction-stripping provision.[60] This section amends 28 U.S.C. § 2241, the statutory

---

[50] *See* Third Geneva Convention, *supra* note 20.

[51] 18 U.S.C. § 2441(a), (c)(1) (2006).

[52] *See* MCA § 6(b)(1)(B), 18 U.S.C. § 2441(d) (2006); *see also* Joanne Mariner, *The Military Commissions Act of 2006: A Short Primer*, FINDLAW'S WRIT, (Oct. 9 & 25, 2006), *available at* http://writ.news.findlaw.com/mariner/20061009.html (Part 1), and http://writ.news.findlaw.com/mariner/20061025.html (Part 2).

[53] MCA § 6, 18 U.S.C. § 2441(d)(2)(D) (2006).

[54] *See* MCA § 6, 18 U.S.C. § 2441(d) (2006).

[55] *See* MCA § 6, 18 U.S.C. § 2441(d)(1)(G)–(H) (2006); *see also* Editorial, *Rushing Off a Cliff*, N.Y. TIMES, Sept. 28, 2006 at A22 [hereinafter *Rushing Off a Cliff*].

[56] *See, e.g.*, Mariner *supra* note 52; Shane and Liptak, *supra* note 48, at A11 (quoting Bruce Ackerman, who stated that "[i]f Congress can strip courts of jurisdiction over cases because it fears their outcome, judicial independence is threatened").

[57] *See* MCA § 3(a)(1), 10 U.S.C. § 948q(b) (2006).

[58] For an analysis of the constitutionality of indefinite detention, *see* Elizabeth Sepper, *The Ties That Bind: How the Constitution Limits the CIA's Actions in the War on Terror*, 81 N.Y.U. L. REV. 1805 (2006); *see also* Hamdi v. Rumsfeld, 542 U.S. 507, 520–21 (2006) (noting that the AUMF did not authorize indefinite detention for the purpose of interrogation).

[59] *See* MCA § 3(a)(1), 10 U.S.C. § 948b(d)(1)(A) (2006).

[60] *See* MCA § 7(a), 18 U.S.C. § 2441(e) (2006); *see also, e.g.*, Mariner, *supra* note 52 (claiming that the jurisdiction-stripping provision qualifies as the single worst provision of the Act); Statement of U.S. Senator Russ Feingold In Opposition of the Military Commissions Act, (Sept. 28, 2006), *available at* http://feingold.senate.gov/~feingold/statements/06/09/20060928.htm ("Among [the MCA's] many flaws, this is the most troubling—that the legislation seeks to suspend the Great Writ of habeas corpus."). *But see* Andrew C. McCarthy, *The New Detainee Law Does Not Deny Habeas Corpus*, NAT'L REV. ONLINE

provision for habeas corpus, by adding language stating that no court has jurisdiction to hear an application for a writ of habeas corpus filed by or on behalf of an alien detained by the United States and who has been found to be an enemy combatant, "or is awaiting such determination."[61] This provision retroactively applies to all aliens detained since September 11, 2001.[62] The practical effect of these amendments is to permanently deny aliens the right to challenge before an independent legal body the reasons, if any, for their imprisonment.

Senator Arlen Specter (R-Pa.) introduced an amendment that would have removed the jurisdiction-stripping provision.[63] The amendment failed, 51 to 48.[64] Senator Specter nonetheless voted for the un-amended bill, claiming that it had several good items, "and the Court will clean it up" by striking down the jurisdiction-stripping provision.[65]

C. Reaction to the MCA and Subsequent Congressional Action

The passage of the MCA sparked impassioned rhetoric from both sides of the aisle. Republicans argued that the new rules would provide the necessary tools to fight a new kind of war and would finally allow detainees in Guantanamo Bay to be tried for their crimes.[66] Former Deputy Attorney General John Yoo defended the MCA by arguing that *Hamdan* overturned what he argues is the traditional understanding that terrorists are neither signatories nor combatants under Common Article 3 of the Geneva Convention.[67] He contended that the MCA responds to the Supreme Court's "stunning power grab" in *Hamdan* by restoring the President's command over the management of the war on terror, which presumably includes the ability to determine the legal status of those captured.[68]

These proponents of the MCA have proven to be outnumbered. Most of the MCA's provisions have been a lightning rod for criticism from across

---

(Oct. 3, 2006), http://article.nationalreview.com/?q=ywnlmjg3ywrlnmnjmtk0ndc1nze0zwi2yzblogrlnzu ("There are innumerable positives in the Military Commissions Act . . . . Among the best is Congress's refusal to grant habeas corpus rights to alien terrorists.").

[61] MCA § 7(a), 28 U.S.C. § 2241(e)(1) (2006). For a definition of term "unlawful enemy combatant," see MCA § 3(a)(1), 10 U.S.C. 948a(1)(A) (2006).

[62] MCA § 7(b).

[63] S. Amdt. 5087, 109th Cong. § 2 (2006).

[64] 152 CONG. REC. 1056 (daily ed. Sept. 28, 2006).

[65] Charles Babington & Jonathan Weisman, *Senate Approves Detainee Bill Backed By Bush*, WASH. POST, Sept. 29, 2006, at A1.

[66] *See* Zernike, *supra* note 25, at A1 (internal quotations omitted). After signing the bill, President Bush commented, "As our troops risk their lives to fight terrorism, this bill will ensure they are prepared to defeat today's enemies and address tomorrow's threats." *Id.*; *see also* Stolberg, *supra* note 29, at A20.

[67] John Yoo, *Congress to Courts: 'Get Out of the War on Terror,'* WALL ST. J., Oct. 19, 2006, at A18; *see also* McCarthy, *supra* note 60 (arguing that terrorists have no constitutional rights).

[68] Yoo, *supra* note 67, at A18.

the political spectrum.[69] Both left-leaning civil rights groups and conservative libertarians see the MCA as an unjustified and unprecedented infringement on personal freedom.[70] Commenting on the effect of the bill as whole, Senator Russ Feingold (D-Wis.) predicted that "[w]e will look back on this day as a stain on our nation's history,"[71] and a *New York Times* editorial described the bill as "our generation's version of the Alien and Sedition Acts."[72]

Two months after the bill's passage, Senator Patrick Leahy (D-Vt.) (the current chairman of the Senate Judiciary Committee) and Senator Arlen Specter (the chairman of the Committee at the time and currently the ranking minority member) moved to repeal the jurisdiction-stripping provision by introducing the Habeas Restoration Act.[73] This Act would restore to federal courts the jurisdiction that was stripped by the MCA[74] and grant aliens detained at Guantanamo Bay the right to challenge their detention and the military commission procedures through writs of habeas corpus.[75] Since the Habeas Restoration Act was introduced, five similar bills have been introduced in the current Congress.[76]

III. THE RIGHT TO HABEAS BEYOND THE WATER'S EDGE

The right to habeas corpus has long been a central tenet of American democracy.[77] Courts have not, however, consistently held that aliens de-

---

[69] *See* Matt Apuzzo, *Terror Case Shows Bush, Libertarian Rift*, WASH. POST., Dec. 13, 2006 (discussing the "strange bedfellows made by the president's anti-terrorism policies"); *see also* Letter from 609 Law Professors to Members of Congress (Sept. 26, 2006), *available at* http://www.law.harvard.edu/news/2006/09/letter.pdf (claiming that the bill "rob[s] individuals detained by the United States of the hallmark of American freedom, the right of anyone detained by the government to demand to know why and to challenge the conditions of confinement before a federal court, independent of the executive and the military").

[70] *See* Apuzzo, *supra* note 69; *see also* Press Release, ACLU, President Bush Signs Un-American Military Commissions Act (Oct. 17, 2006), *available at* http://www.aclu.org/safefree/detention/27091prs20061017.html (describing the bill as "one of the worst civil liberties measures ever enacted in American history").

[71] Statement of Senator Russ Feingold on the President Signing the Military Commissions Act (Oct. 17, 2006), *available at* http://feingold.senate.gov/~feingold/statements/06/10/20061017.htm.

[72] *Rushing Off a Cliff*, *supra* note 55, at A22.

[73] S. 4081, 109th Cong. § 2 (2006).

[74] *See* 152 CONG. REC. S11197–S11199 (daily ed. Dec. 5, 2006) (statements by Sens. Specter and Leahy).

[75] *Id.*

[76] *See* Habeas Corpus Restoration Act of 2007, S.185, 110th Cong. (2007); Habeas Corpus Restoration Act of 2007, H.R.1416, 110th Cong. (2007); Military Commissions Habeas Corpus Restoration Act of 2007, H.R.267, 110th Cong. (2007); Restoring the Constitution Act of 2007, H.R.1415, 110th Cong. (2007); Restoring the Constitution Act of 2007, S.576, 110th Cong. (2007).

[77] *See* The Federalist No. 83, at 427 (Alexander Hamilton) (Max Beloff ed., 2d ed. 1987) (citing habeas corpus, along with trial by jury, as the means by which individuals can protect themselves from the "the great engines of judicial despotism"); INS v. St. Cyr., 533 U.S. 289, 301 (2001) ("The writ has always been available to review the legality of Executive detention.") (citation omitted); *see also* Hamdi v. Rumsfeld, 542 U.S. 507, 554 (2004)

tained abroad possess this right.[78] Since the 1950s, constitutional rights have only extended to members of the U.S. body politic.[79] While the Supreme Court has not explicitly rejected the cases used to support this view, a number of recent decisions demonstrate that reliance on these cases is no longer tenable. In their place, the Court has advocated and applied a contextual due process analysis that balances the rights accorded to individuals with the burden those rights impose on the government. Applying this approach to the situation of alien detainees reveals that habeas corpus rights should be afforded to detainees imprisoned by the U.S. Government at Guantanamo.

The Supreme Court has yet to definitively answer the question of whether the constitutional writ of habeas corpus extends to those imprisoned at Guantanamo. This February, however, the United States Court of Appeals for the District of Columbia in *Boumediene v. Bush*[80] expressly indicated that no constitutional rights extend to the Guantanamo detainee petitioners.[81] Specifically, the court held that the jurisdiction-stripping provision of the MCA applies to the Guantanamo detainees' habeas petitions[82] and that the provision is not an unconstitutional suspension of the writ of habeas corpus.[83]

The D.C. Circuit's reasoning tracked the distinction between the two forms of habeas corpus—first the statutory right embodied in 28 U.S.C. § 2241 (which the jurisdiction-stripping provision amends), and second, "the privilege of the Writ of Habeas Corpus," one of the few constitutional rights enshrined by the Framers in the original Constitution of 1787.[84]

Regarding the statutory basis for habeas corpus jurisdiction, the D.C. Circuit held that the MCA clearly places Guantanamo detainees beyond the reach of 28 U.S.C. § 2241.[85] The court focused on MCA § 7(b), which states that the jurisdiction-stripping provision "shall apply to all cases, with-

---

(Scalia, J., dissenting) ("The very core of liberty secured by our Anglo-Saxon system of separated powers has been freedom from indefinite imprisonment at the will of the Executive.").

[78] *See, e.g.*, Johnson v. Eisentrager, 339 U.S. 763, 780–81 (1950) (concluding that "that no right to the writ of habeas corpus appears" for aliens detained abroad). *But see* Rasul v. Bush, 542 U.S. 466, 481 (2004) (stating that applying the habeas statute to Guantanamo detainees is consistent with the historical reach of the writ); United States v. Tiede, 86 F.R.D. 227 (U.S. Ct. Berlin 1979) (holding that constitutional protections extended to two German citizens detained in the American Sector of occupied Berlin). *See generally* Elizabeth A. Wilson, *The War on Terrorism and "The Water's Edge": Sovereignty, "Territorial Jurisdiction," and the Reach of the U.S. Constitution in the Guantanamo Detainee Litigation*, 8 U. PA. J. CONST. L. 165 (2006); James E. Pfander, *The Limits of Habeas Jurisdiction and the Global War on Terror*, 91 CORNELL L. REV. 497 (2006).

[79] For a discussion of this "membership approach" to constitutional guarantees, see *infra* notes 150–154, and accompanying text.

[80] 476 F.3d 981 (D.C. Cir. 2007), *cert. denied*, 127 S. Ct. 1478 (2007).

[81] *Id.* at 991–92.

[82] *Id.* at 988.

[83] *Id.* at 992.

[84] U.S. CONST. art. I, § 9.

[85] *See* Boumediene v. Bush, 476 F.3d at 986–88.

out exception, pending on or after the date of the enactment of this act."[86] The court found that Congress's decision to emphasize "all cases" with the redundant clarification, "without exception," speaks to its unambiguous intent on this issue.[87] Accordingly, the court declined the detainees' invitation to find exceptions to the provision.[88]

A. The First Constitutional Ground for Upholding the Jurisdiction-Stripping Provision: Common Law Precedent

Although the D.C. Circuit had ample support for its claim that Congress has denied the statutory writ to detainees, the support for its conclusions concerning the reach of the constitutional writ is less compelling. As one legal scholar has observed, constitutionally authorized habeas corpus "exists in speech and is celebrated as the 'great writ of liberty,' but it has no content because it is so rarely used."[89] Cases dealing with habeas corpus in the context of the war on terror prove no exception—*Rasul v. Bush*,[90] *Hamdi v. Rumsfeld*,[91] *Rumsfeld v. Padilla*[92] and *Hamdan v. Rumsfeld*[93] were all decided under 28 U.S.C. § 2241, not under Article I, Section 9 of the Constitution.

In the absence of clear precedent, the D.C. Circuit's analysis focused on two questions—whether the English common law writ of habeas corpus extended to non-subjects beyond the Crown's dominions,[94] and whether constitutional guarantees extend to aliens with no substantial connection to the United States.[95]

Previously, in *Al Odah v. United States*,[96] the proceeding that later became *Rasul v. Bush*, the D.C. Circuit stated that the habeas statute extended only to sovereign territories.[97] Because Guantanamo Bay is not part of the sovereign territory of the United States, the D.C. Circuit in *Al Odah* held

---

[86] MCA § 7(b), 28 U.S.C. § 2241(e)(1) (2006); *see Boumediene*, 476 F.3d at 986 (quoting the provision in full and refuting detainee petitioners' interpretations).

[87] *See Boumediene*, 476 F.3d at 987 ("It is almost as if the proponents of these words were slamming their fists on the table shouting 'When we say "all," we mean all—without exception!'").

[88] *Id.* at 987–88.

[89] CARY FEDERMAN, THE BODY AND THE STATE: HABEAS CORPUS AND AMERICAN JURISPRUDENCE 165 (Robert J. Spitzer ed., State University of New York Press 2006). For an account of the development of these two forms, see *id.* at 166 n.52.

[90] 542 U.S. 466, 473 (2004).

[91] 542 U.S. 507, 525–26 (2004).

[92] 542 U.S. 426, 434–35 (2004).

[93] 126 S. Ct. 2749, 2762–63 (2006).

[94] *See* Boumediene v. Bush, 476 F.3d 981, 988–90 (D.C. Cir. 2007), *cert. denied*, 127 S. Ct. 1478 (2007).

[95] *See id.* at 990–92.

[96] 321 F.3d 1134 (D.C. Cir. 2003), *rev'd sub nom.* Rasul v. Bush, 542 U.S. 466 (2004).

[97] *Al Odah*, 321 F.3d at 1144.

that a court does not have jurisdiction to grant habeas relief to those detained there.[98]

In *Rasul*, the Supreme Court reversed *Al Odah*,[99] and expressly repudiated the D.C. Circuit's analysis.[100] After providing an in-depth examination of common law sources,[101] the Supreme Court concluded that: "[l]ater cases confirmed that the reach of the writ depended not on formal notions of territorial sovereignty, but rather on the practical question of the exact extent and nature of the jurisdiction or dominion exercised in fact by the Crown."[102] Applying this standard to Guantanamo, a territory over which the "United States exercises plenary and exclusive jurisdiction,"[103] the Court concluded that it "is in every practical respect a United States territory."[104] Therefore, the Court held that the habeas statute confers on the District Court jurisdiction to hear petitioners' habeas corpus challenges to the legality of their detention at Guantanamo.[105]

Amazingly, the D.C. Circuit in *Boumediene* dismissed these determinations as dicta and once again focused on formal notions of territorial sovereignty.[106] This of course led the court to the same conclusions first stated in *Al Odah*. In *Boumediene*, the D.C. Circuit considered section 1005(g) of the DTA, which places Guantanamo Bay outside United States in the geographic sense, to be dispositive.[107] As a result, the court did not consider any of the factors found to be necessary by the Supreme Court in making the same determination.

Although the D.C. Circuit felt free to dismiss without explanation these determinations as dicta, it later relied on Supreme Court dictum because "firm and considered dicta . . . binds this court."[108] The D.C. Circuit gave no explanation for why it considered the exhaustive analysis and unqualified findings in *Rasul* less "firm and considered" than the dicta later respected by the court.[109]

---

[98] *Id.*

[99] *See Rasul*, 542 U.S. at 485.

[100] *Id.* at 481–82.

[101] *See id.*

[102] *Id.* at 482 (quotation marks, citation, and footnote omitted).

[103] *Id.* at 475.

[104] *Id.* at 487.

[105] *See id.* at 484.

[106] *See* Boumediene v. Bush, 476 F.3d 981, 990 (D.C. Cir. 2007), *cert. denied*, 127 S. Ct. 1478 (2007) (noting that the petitioners "point to dicta in *Rasul*, in which the Court discussed English habeas cases and the 'historical reach of the writ'" (citation omitted)).

[107] *See id.* at 992; *cf.* Al Odah v. United States, 321 F.3d 1134, 1142–44 (D.C. Cir. 2003), *rev'd sub nom.* Rasul v. Bush, 542 U.S. 466 (2004).

[108] *Boumediene*, 476 F.3d at 992 (referring to the Supreme Court's description of Johnson v. Eisentrager, 339 U.S. 763 (1950), in United States v. Verdugo-Urquidez, 494 U.S. 259, 268 (1990)).

[109] *See Rasul*, 542 U.S. at 480–82. As the dissent noted, the majority's dismissal of the Supreme Court's resolution of this issue as dictum violates the settled principle in the Circuit that "carefully considered language of the Supreme Court, even if technically dictum, generally must be treated as authoritative." *Boumediene*, 476 F.3d at 1002 (Rogers, J., dissenting) (citing Sierra Club v. EPA, 322 F.3d 718, 724 (D.C. Cir. 2003) (internal quotation

While the D.C. Circuit did not explain why *Rasul* does not apply, a recent decision by the U.S. District Court for the District of Columbia attempts to distinguish the Court's findings in *Rasul*. In the remand of *Hamdan v. Rumsfeld*,[110] Judge Robertson claimed that the fact that Guantanamo is under exclusive U.S. jurisdiction and control "was enough for the Court to conclude in Rasul that the broad scope of the habeas statute covered Guantanamo Bay detainees, but the detention facility lies outside the sovereign realm, and only U.S. citizens in such locations may claim entitlement to a *constitutionally* guaranteed writ."[111]

Judge Robertson is correct in noting that the Supreme Court's analysis of common law authority in *Rasul* concerned the scope of the habeas statute;[112] however, he does not explain why the scope of the constitutional writ involves different considerations. Because the constitutional writ "has no content because it is so rarely used,"[113] Judge Robertson's assertion that aliens under U.S. control and jurisdiction have not previously been found to possess this right speaks more to a barren precedential landscape, rather than to a settled understanding.

In the absence of a developed constitutional habeas doctrine, a possible touchstone would be the scope of the writ at common law. Indeed, it is well settled that "[i]n construing any act of legislation, whether a statute enacted by the legislature, or a constitution established by the people as the supreme law of the land, regard is to be had . . . to the condition and to the history of the law as previously existing."[114] Therefore, the Supreme Court's analysis of the scope of the common law writ in *Rasul* is relevant to both statutory and constitutional habeas corpus.

There is another possible argument for the irrelevance of *Rasul*, suggested by the D.C. Circuit's approach to the issue, which conforms to the analysis put forth by Justice Scalia in his *Rasul* dissent.[115] Given that the composition of the Court has changed since *Rasul* was decided in 2004, the D.C. Circuit's decision to hold fast to its prior determinations may have been motivated by the prediction—or hope—that *Rasul* no longer commands a majority. The D.C. Circuit did not openly express its view of *Rasul*, though it tipped its hand when it gently reminded a reviewing

---

marks omitted)).

[110] 464 F.Supp. 2d 9 (D.D.C. 2006).

[111] *Id.* at 18 (emphasis in original) (citation omitted).

[112] See *Rasul*, 542 U.S. at 473.

[113] *See* Federman, *supra* note 89, at 165.

[114] U.S. v. Wong Kim Ark, 169 U.S. 649, 653–54 (1898). *See, e.g.*, Hamdan v. Rumsfeld, 126 S. Ct. 2749, 2773–75 (2006) (turning to common law as a source of authority in the absence of an applicable constitutional provision).

[115] *See* Rasul v. Bush, 542 U.S. at 502–05 (Scalia J., dissenting) (surveying common law sources in support of the proposition that the writ of habeas corpus did not issue to aliens in foreign lands).

court that "[t]he determination of sovereignty over an area is for the legislative and executive departments."[116]

This April, the Supreme Court denied the petition for writs of certiorari filed by the detainees in *Boumediene v. Bush* and a companion case, *Al Odah v. US*.[117] Justices Stevens and Kennedy wrote jointly to explain that the denial was rooted in the Court's practice of requiring exhaustion of available remedies and of avoiding constitutional questions.[118] The Justices were careful to observe, "as always, denial of certiorari does not constitute an expression of any opinion of the merits."[119] The decision to deny certiorari may not in itself have been a reflection of the Court's view of the merits; however, the same cannot be said for the citation to *Rasul* following this statement.[120] By citing the pages in *Rasul* which contain the conclusions that the D.C. Circuit dismissed as dicta in order to reach a contrary result, the Court indicated that these conclusions are not so easily disregarded. In dissent, Justice Breyer, joined by Justices Souter and Ginsburg, was not as oblique. He noted, "petitioners plausibly argue that the lower court's reasoning is contrary to this Court's precedent."[121]

Commentators speculated that the array of votes in the denial of certiorari indicates that the Justices who wanted to hear the cases had failed to attract the swing vote of Justice Kennedy.[122] Although Justice Stevens could have provided the fourth vote for a grant of certiorari, he appears to have opted to join with Justice Kennedy to salvage some prospect of ultimate relief for the detainees.[123]

Because the reasoning in *Boumediene* may persuade a majority of the Court if and when it decides this issue, it is worth examining the D.C. Circuit's claim that the writ did not extend to non-subjects beyond the Crown's dominions. The D.C. Circuit's examination of common law followed both a negative and positive line of inquiry. First, it noted that there is no common law antecedent for the proposition that aliens detained in non-sovereign territory have habeas rights.[124] Second, it cited several treatises confirming that the writ did not extend to people beyond the Crown's dominions.[125]

---

[116] Boumediene v. Bush, 476 F.3d 981, 992 (D.C. Cir. 2007), *cert. denied*, 127 S. Ct. 1478 (2007) (quotations and citation omitted).

[117] Boumediene v. Bush, 127 S. Ct. 1478 (2007).

[118] *See id.* at 1478.

[119] *Id.* (citing *Rasul*, 542 U.S. at 480–81).

[120] *Id.* at 1479

[121] *Id. See also id.* at 1480 (noting that the D.C. Circuit rejected petitioners' arguments under Supreme Court precedent that fundamental rights afforded by the Constitution extend to Guantanamo).

[122] Posting of Lyle Denniston to SCOTUSblog, http://www.scotusblog.com/movabletype/archives/2007/04/court_denies_ha.html (Apr. 2, 2007 10:03 EST).

[123] *Id.*

[124] Boumediene v. Bush, 476 F.3d 981, 988–89 (D.C. Cir. 2007), *cert. denied*, 127 S. Ct. 1478 (2007).

[125] *See id.* at 989–90.

The detainees had put forth two cases in which late eighteenth century English courts heard the habeas petitions of aliens who were captured and detained in areas beyond the Crown's dominion.[126] However, the court noted that those aliens were eventually transported to England, and it was there that they filed their habeas petitions.[127] Therefore, the court observed, "[n]one of these cases involved an alien outside of the territory of the sovereign."[128]

As the majority in *Boumediene* emphasized, none of the cases cited provides a clear answer as to how this issue would have been resolved at common law.[129] Therefore, a finding based only on this negative inference would be unjustified. The majority may have found grounds for distinguishing the cases put forth by the detainees, but as Judge Rogers countered in her dissent, "[the majority] can point to no case where an English court has refused to exercise habeas jurisdiction because the enemy being held, while under the control of the Crown, was not within the Crown's dominions."[130] She further noted that "[t]he paucity of direct precedent is a consequence of the unique confluence of events that defines the situation of these detainees and not a commentary on the reach of the writ at common law."[131]

After concluding that the detainees cited no case in which the English common law writ extended beyond the Crown's dominions, the D.C. Circuit commented that "[o]ur review shows the contrary."[132] This review is foremost rooted in William Holdsworth's *A History of English Law*, which the majority cited four times.[133] In support of the claim that the writ did not extend to those held in extraterritorial detention, the court cited Holdsworth for the proposition that "[e]ven British citizens imprisoned in 'remote islands, garrisons, and other places' were 'prevent[ed] from the benefit of the law.'"[134]

The unedited text casts light on the majority's sleight of hand. The full sentence from Holdsworth's treatise reads: "Some of these defects in the writ of *Habeas Corpus* were illustrated by the arbitrary proceedings of Clarendon, who, in 1667, was accused of sending persons to remote islands, garrisons, and other places, thereby preventing them from the benefit of the law . . . ."[135] In the impeachment trial of the Earl of Clarendon,[136]

---

[126] *See id.* at 988–89 (discussing Three Spanish Sailors, 96 Eng. Rep. 775 (C.P. 1779) and Rex v. Schiever, 97 Eng. Rep. 551 (K.B. 1759)).

[127] *See id.*

[128] *Id.* at 989.

[129] *See id.* at 989.

[130] *Boumediene*, 476 F.3d at 1000 (Rogers, J., dissenting).

[131] *Id. at* 1000–01 (Rogers, J., dissenting).

[132] *Id.* at 989.

[133] *Id.* at 989–90.

[134] *Id.* at 989.

[135] 9 WILLIAM HOLDSWORTH, A HISTORY OF ENGLISH LAW 116 (1982 ed.) (emphasis in original).

[136] Proceedings in the House of Commons, touching the impeachment of Edward late

the State Trials Court addressed this "defect," which it considered one of Clarendon's "many Great Crimes."[137] The court held that this attempt to evade the law through extraterritorial detention was evidence of his "Intent to draw Scandal and Contempt upon His Majesty's Person."[138] For this and other offenses, the court found Clarendon guilty of high treason and banished him from the Kingdom.[139]

The State Trials Court never passed on whether Clarendon's practice of transporting persons to remote islands actually placed them beyond the benefit of the law. One thing, however, is certain—the plea Clarendon had written in his defense of deeds was found to be so repugnant that the Court ordered it burned by the Hangman so that "it should not live."[140]

As the dissent points out, there are two other instances in which the majority masked problems in its analysis by omitting the unseemly parts of a sentence that are inconsistent with its reasoning.[141] And in the instances where there was nothing salvageable, the court ignored the text entirely. Repeatedly, the court asserted that no case had held that aliens imprisoned outside sovereign territory could file writs of habeas corpus.[142] The dissent cited four such cases that directly contradict this claim,[143] but the majority did not factor these cases into its analysis. For example, a number of cases establish that the writ was issued to non-subjects in India, well before England recognized its sovereignty over the territory.[144]

The dissent's findings are also consistent with the Supreme Court's examination of common law sources on this issue. In *Rasul*, the Court cited several common law authorities that also contradict the *Boumediene* ma-

---

Earl of Clarendon, Lord High Chancellor of England, (1667), 2 Compleat Collection of State-Tryals, 1 (1719 ed.).

[137] *Id.* at 1.

[138] *Id.* at 26.

[139] *See id.*

[140] *Id.*

[141] *See* Boumediene v. Bush, 476 F.3d 981, 988 (D.C. Cir. 2007), *cert. denied*, 127 S. Ct. 1478 (2007). The majority quoted INS v. St. Cyr, 533 U.S. 289, 301 (2001), for the proposition that the Suspension Clause protects the writ "as it existed in 1789." *Id.* The dissent noted that "The court oddly chooses to ignore [how the writ has developed since 1789] by truncating its reference to *St. Cyr*, without comment, and omitting the qualifier 'at the absolute minimum.'" See *Boumediene*, 476 F.3d at 1000 n.5 (Rogers, J., dissenting).

The majority also quoted Lord Mansfield in support of its claim that the writ did not run "to foreign dominions." *Id.* at 989 (quoting 2 Burr. at 856, 97 Eng. Rep. at 599). Lord Mansfield limited this claim, however, to only those lands "which belong to a prince who succeeds to the throne of England." *See Boumediene*, 476 F.3d at 1002 (Rogers, J., dissenting) (quoting 97 Eng. Rep. at 599–600). "Through the use of ellipsis marks," the dissent commented, "the court excises the qualification and concludes that the writ does not extend '[t]o foreign dominions.'" *Id.* at 1002 (Rogers, J., dissenting).

[142] *See Boumediene*, 476 F.3d at 989–91.

[143] *See id.* at 1003–04 (Rogers, J., dissenting).

[144] *See id.* (citing 16 B.N. Pandey, THE INTRODUCTION OF ENGLISH LAW INTO INDIA, 112, 149, 151 (1967); Rex v. Mitter, 1 Indian Dec. 1008 (Calcutta S.C. 1781); Rex v. Hastings, 1 Indian Dec. 1005, 1007 (Calcutta S.C. 1775) (opinion of Chambers, J.); *id.* at 1007 (opinion of Impey, C.J.); *see also* Kal Raustiala, *The Geography of Justice*, 73 FORDHAM L. REV. 2501, 2530 n.156 (2005)).

jority's claim and concluded that "[a]pplication of the habeas statute to persons detained at the base is consistent with the historical reach of the writ of habeas corpus."[145]

Moreover, there is another line of common law precedent, unmentioned by the D.C. Circuit, which is crucial to answering how this situation would have been dealt with at common law. "At its historical core, the writ of habeas corpus has served as a means of reviewing the legality of Executive detention," the Supreme Court has explained, "and it is in that context that its protections have been strongest."[146] Most habeas corpus petitions today seek post-conviction relief.[147] These petitioners have been through state and possibly federal court-—often several times—and have been found guilty beyond a reasonable doubt after having been afforded all of the attendant procedural safeguards.[148] Executive detention lacks these hallmarks of due process. Blackstone emphatically emphasized the danger inherent in this type of detention:

> To bereave a man of life, or by violence to confiscate his estate, without accusation or trial, would be so gross and notorious an act of despotism, as must at once convey the alarm of tyranny throughout the whole kingdom. But confinement of the person, by secretly hurrying him to gaol, where his sufferings are unknown or forgotten; is a less public, a less striking, and therefore a more dangerous engine of arbitrary government.[149]

As the Court's statements in *Rasul* and these common law sources demonstrate, the D.C. Circuit's decision in *Boumediene* to hold fast to its determination first made in *Al Odah* flies in the face of both binding precedent and common law practice.

B. The Second Constitutional Ground for Upholding the Jurisdiction-Stripping Provision: The Membership Approach

In addition to its common law analysis, the D.C. Circuit stated that alien detainees imprisoned abroad are not entitled to any constitutional habeas rights because "[p]recedent in this court and the Supreme Court holds that the Constitution does not confer rights on aliens without property or presence within the United States."[150] The court frequently cited *Johnson*

---

[145] Rasul v. Bush, 542 U.S. 466, 480–81 (2004).

[146] INS v. St. Cyr, 533 U.S. 289, 301 (2001) (footnote and citations omitted).

[147] *See* Erwin Chemerinsky, FEDERAL JURISDICTION § 15.2 (4th ed. 2003).

[148] *See Id.*

[149] 1 W. BLACKSTONE, COMMENTARIES ON THE LAWS OF ENGLAND 131–32 (1765).

[150] Boumediene v. Bush, 476 F.3d 981, 991 (D.C. Cir. 2007), *cert. denied*, 127 S. Ct. 1478 (2007).

v. Eisentrager[151] and *United States v. Verdugo-Urquidez*[152] for this proposition.[153] These cases establish a "membership approach" to Constitution, whereby its protections extend only to those whose significant voluntary contacts with the United States have created a "substantial connection with our country."[154]

Before discussing the Guantanamo detainees' relation to this analysis, it should be noted that the D.C. District Court has suggested that the jurisdiction-stripping provision may fail the membership approach in some circumstances. In the *Hamdan* remand, Judge Robertson asserted, "[i]f and to the extent that the MCA makes the writ unavailable to anyone who is constitutionally entitled to it, it must be unconstitutional."[155] The MCA removes the courts' jurisdiction over any habeas writ by an alien detained as an enemy combatant or is awaiting a determination as to whether he is an enemy combatant.[156] The MCA defines enemy combatants as not just aliens captured on the battlefield, but as any non-citizen suspected of one or more of the substantive offenses listed in the MCA—offenses that include materially supporting terrorism, spying, conspiring or attempting to commit one of the listed offenses, and serving as an accessory after the fact for any of the listed offenses.[157] Therefore, the jurisdiction-stripping provision applies to aliens abroad as well as to any of the twelve million lawful permanent resident aliens of the United States who may be suspected of committing a listed offense.[158] The Supreme Court has repeatedly asserted that depriving permanent resident aliens, many of whom have established a substantial connection with the United States, of habeas rights violates the Suspension Clause.[159]

A more difficult question arises in the context of non-citizen detainees held at Guantanamo Bay, an area that is not a U.S. sovereign territory, but is under exclusive U.S. control and jurisdiction.[160] While the Supreme

---

[151] 339 U.S. 763 (1950).

[152] 494 U.S. 259 (1990).

[153] *See Boumediene*, 476 F.3d at 990–91.

[154] *Verdugo-Urquidez*, 494 U.S. at 260. For a discussion of the history of the membership approach and its relevance today, see generally Wilson, *supra* note 78. In his concurrence in *Verdugo-Urquidez*, Justice Kennedy quoted Justice Story's objection to this approach. Justice Story noted that the "'difficulty' in describing the Constitution as a 'compact between the people of each state, and all the people of the other states, is that the Constitution itself contains no such expression, and no such designation of parties.'" *Verdugo-Urquidez*, 494 U.S. at 276 (Kennedy, J., concurring) (*quoting* 1 JOSEPH STORY, COMMENTARIES ON THE CONSTITUTION OF THE UNITED STATES § 365, at 335 (Cambridge, Brown, Shattuck & Co. 1833)).

[155] 464 F.Supp. 2d at 16.

[156] *See* MCA § 7(a), 28 U.S.C. § 2241(e)(1) (2006).

[157] MCA § 3(a)(1), 10 U.S.C. § 950v(b) (2006).

[158] MCA § 7(a), 28 U.S.C. § 2241(e)(1); *see* 152 CONG. REC. S11199 (daily ed. Dec. 5, 2006) (statement of Sen. Leahy).

[159] *See, e.g.*, INS v. St. Cyr., 533 U.S. 289, 290 (2001) ("The Constitution's Suspension Clause, which protects the privilege of the habeas corpus writ, unquestionably requires some judicial intervention in deportation cases." (citation omitted)).

[160] *See* Rasul v. Bush, 542 U.S. 466, 487 (2004) (Kennedy, J., concurring).

Court has not overruled the membership approach established in *Eisentrager* and *Verdugo-Urquidez*, it has asserted that this approach has no relevance to the current system of military commissions at Guantanamo.[161] In *Eisentrager*, German nationals in custody of the United States Army in Germany filed habeas petitions after a military commission convicted them.[162] Justice Jackson, speaking for the Court, held that an alien's lawful presence in the United States gave him certain rights that become more extensive when he declares his intention to become a citizen.[163] Accordingly, the Court held that "no right to the writ of habeas corpus appears" in the case of enemy aliens who at no relevant time were within U.S. sovereign territory.[164]

In *Rasul*, the Court distinguished the situation in *Eisentrager* from that at Guantanamo. Unlike the *Eisentrager* petitioners, the Guantanamo detainees in *Rasul*

> deny that they have engaged in or plotted acts of aggression against this country; they have never been afforded access to any tribunal, much less charged with and convicted of wrongdoing; and for more than two years they have been imprisoned in territory over which the United States exercises exclusive jurisdiction and control.[165]

Most strikingly, the Court in *Rasul* suggested that detainees possess constitutional rights, which necessarily rejects the membership approach of *Eisentrager*, and relied on precedent that advocates a case-by-case approach to recognizing constitutional guarantees. The clearest support for this argument is the much discussed *Rasul* footnote 15,[166] in which the Court stated:

---

[161] *See id.* at 476 (noting that "petitioners in these cases differ from the Eisentrager detainees in important respects").

[162] Johnson v. Eisentrager, 339 U.S. 763, 765–66 (1950).

[163] *See id.* at 770.

[164] *Id.* at 781.

[165] Rasul v. Bush, 542 U.S. 466, 476 (2004).

[166] *Id.* at 483, n.15. For conflicting interpretations of the implications of footnote 15, see Khalid v. Bush, 355 F.Supp. 2d. 311, 314 (D.D.C. 2005) (stating that *Rasul* footnote 15 is limited to the question presented, i.e., jurisdiction, and therefore only stands for the proposition that petitioners had met the pleading requirements for habeas); *In re* Guantanamo Detainee Cases, 355 F.Supp. 2d 443, 463 (D.D.C. 2005) (concluding that the Court's reference to Justice Kennedy's concurring opinion requires consideration of the precedent that it discusses); *see also* Wilson, *supra* note 78, at 196–202 (explaining why *In re Guantanamo Detainee Cases* has the better interpretation of Rasul footnote 15 than that provided in *Khalid*); Alan Tauber, *Ninety Miles From Freedom? The Constitutional Rights Of The Guantanamo Bay Detainees*, 18 ST. THOMAS L. REV. 77, 78–84 (2005) (noting that "while Judge Leon [in *Khalid*] ignored the rich case history of the previous century, Judge Green [in *In re Guantanamo Detainee Cases*] gave appropriate attention to the Court's previous pronouncements").

> Petitioners' allegations—that, although they have engaged neither in combat nor in acts of terrorism against the United States, they have been held in executive detention for more than two years . . . without access to counsel and without being charged with any wrongdoing—unquestionably describe custody in violation of the Constitution or laws or treaties of the United States.[167]

The Court cited Justice Kennedy's concurring opinion in *United States v. Verdugo-Urquidez* "and cases cited therein" for support.[168] Justice Kennedy's concurrence in *Verdugo-Urquidez* rejected the membership approach followed by the majority in that case and advocated a contextual due process analysis.[169] Quoted at length in Justice Kennedy's concurring opinion is Justice Harlan's *Reid v. Covert* concurrence,[170] in which Justice Harlan stated that although the Constitution is always and everywhere applicable, not all provisions apply under all circumstances.[171] Justice Harlan's approach rejects an all-or-nothing view of constitutional guarantees and instead "examines the conditions and considerations that would make adherence to a specific guarantee altogether impracticable and anomalous."[172]

Justice Harlan's flexible approach to the Due Process Clause is not in keeping with the Court's traditional resistance to picking and choosing among Constitutional guarantees.[173] The Court has been sensitive, however, to the unique demands of the war on terror. The Court not only stated its willingness to be flexible in *Rasul* footnote 15, but also applied a contextual due process approach in *Hamdi*.[174] The plurality in *Hamdi* engaged in an analysis similar to the one advocated by Justice Kennedy in his *Verdugo-Urquidez* concurrence and concluded that certain "essential constitutional promises may not be eroded,"[175] while others may be displaced in a proceeding tailored to alleviate the burden on an executive branch dealing with an ongoing military conflict.[176] For these reasons, there is strong

---

[167] 542 U.S. at 483, n.15 (internal quotations and citation omitted).

[168] *Id.*

[169] United States v. Verdugo-Urquidez, 494 U.S. 259, 277–78 (1990). (Kennedy J., concurring).

[170] 354 U.S. 1, 74 (1957) (Harlan J., concurring).

[171] *See id.*

[172] *Verdugo-Urquidez*, 494 U.S. at 277–78 (Kennedy J., concurring) (citing *Reid*, 354 U.S. at 74 (Harlan J., concurring)).

[173] *See, e.g.*, *Reid*, 354 U.S. at 8–9 (finding "no warrant, in logic or otherwise, for picking and choosing among the remarkable collection of 'Thou shalt nots' which were explicitly fastened on all departments and agencies of the Federal Government by the Constitution and its Amendments").

[174] Hamdi v. Rumsfeld, 542 U.S. 507, 533–34 (2004).

[175] *See id.* at 533 ("We therefore hold that a citizen-detainee seeking to challenge his classification as an enemy combatant must receive notice of the factual basis for his classification, and a fair opportunity to rebut the Government's factual assertions before a neutral decisionmaker.").

[176] *See id.* at 533–34 (stating that exigent circumstances render hearsay admissible and allow the burden of proof to shift to the defendant "once the Government puts forth credi-

support for the view that, as Judge Green of the D.C. District Court concluded, the Supreme Court has set "an implicit, if not express, mandate to uphold the existence of fundamental rights" of aliens detained abroad in areas under United States control and jurisdiction.[177]

In *Boumediene*, the D.C. Circuit acknowledged that the Supreme Court in *Rasul* cast doubt on the continuing vitality of *Eisentrager*.[178] However, it noted that "absent an explicit statement by the Court that it intended to overrule Eisentrager's constitutional holding, that holding is binding on this court."[179] This exercise of judicial restraint underscores a question lingering since *Rasul* was decided in 2004. Unless the Supreme Court clearly disposes of *Eisentrager*, lower courts will continue to enforce what has implicitly been cast aside as a dead letter.

Several commentators have counseled against any further diminishing of the membership approach. They fault the approach advocated in *Rasul* footnote 15 for "open[ing] the courthouse doors to enemy fighters in wartime for the first time in American history."[180] The Court is ill-equipped to make war policy, they have argued, and the MCA reflects a contrary vision as articulated by the proper branch of government.[181] While the membership approach supports the practical effects of the jurisdiction-stripping provision, an argument that attempts to cast this approach as an unwavering constant is misguided.

The membership approach began to erode during the first half of the twentieth century. In *The Insular Cases*,[182] a group of cases decided after the conclusion of the Spanish-American War in 1898, the Supreme Court focused on the relationship between the territory at issue and the United States and held that "fundamental" constitutional rights apply to citizens and aliens alike in non-sovereign territories governed by the United States.[183]

This line of reasoning, which *Eisentrager* cuts off, is consistent with common law tradition. The right to the English common law writ was not limited to subjects or to those with a substantial connection with England.[184]

---

ble evidence that the habeas petitioner meets the enemy-combatant criteria").

[177] *In re* Guantanamo Detainee Cases, 355 F.Supp.2d 443, 461 (D.D.C. 2005).

[178] *See* Boumediene v. Bush, 476 F.3d 981, 1011 (D.C. Cir. 2007), *cert. denied*, 127 S. Ct. 1478 (2007) (citing Rasul v. Bush, 542 U.S. 466, 475–79 (2004)).

[179] *Id.* (citations omitted).

[180] *See* McCarthy, *supra* note 60.

[181] *See* Yoo, *supra* note 67, at A18.

[182] Downes v. Bidwell (The Insular Cases), 182 U.S. 244, 287 (1901). In incorporated territories destined for statehood constitutional rights would apply. *See id.* at 343. In the unincorporated territories, by contrast, only "fundamental" constitutional rights would apply of their own force and courts would determine on an objective basis what relationship Congress had created with the territory. *See id.* at 268. The Insular Cases thus established that the applicability of the Constitution in U.S. territories not destined for statehood was decided on a case-by-case basis, taking into account the particular provision at issue and the nature of the relationship that Congress had established with the particular territory. *See* Wilson, *supra* note 78, at 169 n.19.

[183] *Downes*, 182 U.S. at 268.

[184] *See* Gerald L. Neuman, *Habeas Corpus, Executive Detention, and the Removal of*

In *Rex v. Schiever*,[185] one of the cases cited by the habeas petitioners in *Boumediene*, a French privateer took Schiever, a Swedish citizen, prisoner while at sea.[186] The privateer then transported Schiever to Liverpool and had him imprisoned there.[187] Schiever filed a writ of habeas corpus, which a court accepted and decided on the merits.[188] As the dissent in *Boumediene* noted, Schiever surely was not voluntarily brought into England.[189] While Schiever had habeas rights at common law in England, he would not today in the United States under the membership approach because "involuntary [presence] is not the sort to indicate any substantial connection with our country."[190] Similarly, in *Somerset v. Stewart*,[191] a slave from Africa was brought into England by way of Virginia.[192] The Court of the King's Bench accepted the slave's habeas petition and freed him on the ground that slavery did not exist in England.[193]

The membership approach of *Eisentrager* and *Verdugo-Urquidez* is not only inconsistent with common law, but it is also ill-equipped to deal with the issues raised by the war on terror. The detainee petitioners in *Boumediene* have been held for more than five years and have yet to be found guilty of any crime, whether by a federal court or military commission.[194] In denying those imprisoned an opportunity to challenge the basis of their detention or conditions of confinement, the detainees have chosen instead to speak through more desperate measures such as suicide, rioting and hunger strikes.[195] The lawyers for those detained complain that the mental health of their clients' is deteriorating.[196] These circumstances, which raise a number of constitutional concerns, were not before the Court in *Eisentrager*.

Most importantly, the membership approach is incompatible with the long-term security of the United States with respect to the war on terror. Recently, the CIA reported to Congress that Al Qaeda's numbers and influence have grown significantly despite the efforts of the United States government.[197] The CIA attributes this to propaganda that capitalizes on anti-American sentiment.[198] A recent FBI report culled from accounts by

---

Aliens, 98 COLUM. L. REV. 961, 989–90 (1998).

[185] 97 Eng. Rep. 551 (K.B. 1759).

[186] *Id.*

[187] *Id.* at 552.

[188] *Id.*

[189] *See* Boumediene v. Bush, 476 F.3d 981, 1001 (D.C. Cir. 2007), *cert. denied*, 127 S. Ct. 1478 (2007) (Rogers, J., dissenting).

[190] United States v. Verdugo-Urquidez, 494 U.S. 259, 271 (1990).

[191] 98 Eng. Rep. 499 (K.B. 1772); *see also* Neuman, *supra* note 184, at 989–90.

[192] 98 Eng. Rep. at 499.

[193] *Id.*

[194] *See* Boumediene v. Bush, 127 S. Ct. 1478, 1479 (2007) (Breyer, J., dissenting).

[195] *See* Tim Golden and Margot Williams, *Guantanamo Detainees Stage Hunger Strike Despite Force-Feeding Policy*, N.Y. TIMES, Apr. 9, 2007, at A12.

[196] *See id.*

[197] *See* Dafna Linzer and Walter Pincus, *Taliban, Al-Qaeda Resurge In Afghanistan, CIA Says*, WASH. POST, Nov. 16, 2006, at A22.

[198] *Id.*

Guantanamo employees and FBI agents catalogs a litany of abuses at the base. In addition to waterboarding and aggressive interrogation tactics, the report also lists accounts of sexual taunts, "baptizing" detainees, and wrapping a bearded detainee's head in duct tape to prevent him from praying aloud.[199] While Congress has made sure that courts are powerless to address these abuses, judgment of a different sort is still being rendered. Al Qaeda's successful recruitment efforts ensure that these degrading acts do much to endanger United States servicemen and allied forces.

The effects of the membership approach do not play out in a vacuum. A ruling that forbids those imprisoned a basic right to challenge the conditions of their confinement will continue to have a profound impact on our interests both in the United States and abroad.

IV. CONTEXTUAL DUE PROCESS: THE "IMPRACTICABLE AND ANOMALOUS" INQUIRY

Congress and the Supreme Court may each ultimately reconsider the challenges to the jurisdiction-stripping provision of the MCA. Thus, an examination of how the jurisdiction-stripping provision fares in a contextual due process analysis is useful for both forums. If the Supreme Court examines this provision, it is possible that the Court will follow through on its attempt in *Rasul* to disavow *Eisentrager*. In addition, Congress's deliberations will likely focus on competing considerations of individual rights and national security-—the same considerations that guide the contextual due process approach.

In his *Verdugo-Urquidez* concurrence, Justice Kennedy identified three considerations for a contextual due process analysis: (1) the nature of the government action; (2) the nature of the relationship between the United States and the territory at issue; and (3) the nature of the particular right at issue, all taken in light of the specific facts and circumstances of the case.[200]

A. Contextual Due Process—First Inquiry: The Nature of the Government Action

The nature of the government's action prong of the contextual due process analysis implicates both separation of powers and national security concerns. By removing the courts as a check on executive action, the jurisdiction-stripping provision of the MCA enables the President to meet the challenges of the war on terror without judicially imposed constraints. In an editorial on the MCA, John Yoo stressed that while the risk of detain-

---

[199] *See* Dan Eggen, *FBI Reports Duct-Taping, 'Baptizing' at Guantanamo*, WASH. POST, Jan. 3, 2007, at A1.

[200] *See* United States v. Verdugo-Urquidez, 494 U.S. 259, 278 (1990) (Kennedy, J., concurring).

ing the innocent is greater when the enemy fights covertly, so is the risk of releasing the dangerous.[201] If the United States is unable to prevent an attack, the responsibility will fall squarely upon the Executive, not upon the judicial branch.

Critics suspicious of an unchecked executive branch respond by noting that a habeas corpus proceeding is not a *de novo* review of the merits.[202] The purpose is not to second-guess the military commission's finding that the trial counsel satisfied its burden of proof; rather, a habeas proceeding would ensure that the commission had proper jurisdiction over the detainee, and that fair consideration had been given to the petitioner's claims in a proceeding that provided an opportunity to contest the factual basis for his detention.[203] Given their scope, habeas corpus proceedings are more appropriately characterized as a circumspect review than as a byproduct of "swashbuckling assertions of judicial supremacy."[204]

Another potential problem with habeas corpus review lies in the burden it imposes on those fighting the war on terror. As the Government contended in *Hamdi*, military officers may be unnecessarily and dangerously distracted by litigation half a world away.[205] In addition, the Government asserted that allowing discovery of military operations would intrude upon the sensitive secrets of national defense and result in a futile search for evidence buried under the rubble of war.[206]

The MCA requires that the defendant have the right to examine and respond to evidence admitted against him.[207] There are exceptions for information that would threaten national security or would reveal military sources, methods or activities,[208] but not for evidence lost in the "rubble of war."[209] If trial counsel for the Government is unable to find such evidence, he is still required to present other evidence that is sufficient to persuade the military commission of guilt beyond a reasonable doubt.[210]

However, the administrative burden and the distraction of a proceeding in a district hundreds of miles from Guantanamo are significant. Guantanamo Bay is not within a federal judicial district and therefore military

---

[201] *See* Yoo, *supra* note 67.

[202] *See generally* C.J.S. HabeasCorp § 419 (2007).

[203] *See id; see also* Hamdi v. Rumsfeld, 542 U.S. 507, 536 (2004).

[204] Yoo, *supra* note 66, at A18.

[205] *See Hamdi*, 542 U.S. at 531–32.

[206] *See id.*

[207] MCA § 3(a)(1), 10 U.S.C. § 950b(b)(1) (2006).

[208] MCA § 3(a)(1), 10 U.S.C. § 949d(f)(1)–(2) (2006); *see also* MCA § 3(a)(1), 10 U.S.C. § 949d(f)(2)(A)(iii) (2006) (requiring the Government to provide a non-classified summary or otherwise disclose the statement of relevant facts that the classified information would tend to prove if it wants to claim either of these exceptions); *see also* Mariner, *supra* note 52 (claiming that the likely impact of protecting sources, methods, and activities will be that any inquiry into the CIA's allegedly abusive interrogation practices will be barred).

[209] *Hamdi*, 542 U.S. at 532.

[210] *See* MCA § 3(a)(1), 10 U.S.C. § 949l(c)(4) (2006).

officials would need to participate in a proceeding in one of the ninety-four federal districts.[211] Moreover, if the personal jurisdiction rules for constitutional habeas corpus mirror those of the statutory writ, there is a possibility that the detainees would be able to choose any of these ninety-four districts.[212]

B. Contextual Due Process—Second Inquiry: The Relationship Between the United States and Guantanamo

The second factor in Justice Kennedy's analysis weighs in favor of recognizing habeas corpus rights for Guantanamo detainees. Judge Robertson in the *Hamdan* remand distinguished the Court's determination in *Rasul* that "Guantanamo Bay is in every practical respect a United States territory"[213] on the grounds that this finding turned on the language of 28 U.S.C. § 2241, the statutory provision for habeas corpus.[214] Because the doctrine of constitutional habeas corpus does not make a similar distinction, Judge Robertson found that the Court's determination had no relevance to an analysis of the reach of the constitutional writ.[215]

If a court adopts the "impracticable and anomalous" balancing test, the fact that a territory with Guantanamo's status has not previously been found to be within the ambit of the constitutional writ would no longer settle the issue. Rather, this test requires an examination of the "conditions and considerations" that characterize the nature of the relationship between the United States and Guantanamo.[216] By its terms, this detailed assessment requires that a court go beyond assessments of sovereignty and attempt to understand the connection, if any, that exists between the two territories. This is precisely what the Supreme Court did in *Rasul* when it decided the issue under 28 U.S.C. § 2241.[217] For this reason, the Court's determination that a non-sovereign territory that is under exclusive U.S. jurisdiction and control "is in every practical respect a United States territory"[218] speaks directly to the second prong of this inquiry.

---

[211] In habeas challenges to present physical confinement, the default rule is that the proper respondent is the warden of the facility where the prisoner is being held. Rumsfeld v. Padilla, 542 U.S. 426, 435 (2004). Although federal district courts have been granted jurisdiction over non-sovereign territories, see *e.g.*, United States v. Tiede, 86 F.R.D. 227 (U.S. Ct. Berlin 1979), a federal district has not been created for Guantanamo, nor has Congress given an existing district jurisdiction over the base.

[212] *See* Rasul v. Bush, 542 U.S. 466, 506 (2004) (Scalia, J., dissenting); *cf. Padilla*, 542 U.S. at 452–53 (Kennedy, J. concurring) (proposing that courts limit jurisdiction to the forum with the most immediate connection to the named custodian). *See generally* Pfander, *supra* note 78, at 506 (comparing Justices Scalia and Kennedy's interpretations).

[213] *Rasul*, 542 U.S. at 487.

[214] *See* Hamdan v. Rumsfeld, 464 F.Supp.2d 9, 16 (D.D.C. 2006).

[215] *See id.*

[216] *See* United States v. Verdugo-Urquidez, 494 U.S. 259, 277 (1990) (Kennedy, J., concurring) (quoting Reid v. Covert, 354 U.S. 1, 74 (1957) (Harlan, J., concurring)).

[217] *See Rasul*, 542 U.S. at 480–81.

[218] *Id.* at 487.

C. Contextual Due Process—Third Inquiry: The Right at Issue

The nature of the right at issue also militates in favor of allowing detainees to challenge the legality of their detention. Judge Green of the D.C. District Court described the right at stake when deciding the habeas petitions filed in the wake of *Rasul*:

> Short of the death penalty, life imprisonment is the ultimate deprivation of liberty, and the uncertainty of whether the war on terror—and thus the period of incarceration—will last a lifetime may be even worse than if the detainees had been tried, convicted, and definitively sentenced to a fixed term.[219]

The fundamental purpose of habeas corpus is to ensure that individuals are protected from this deprivation by providing independent legal review of the grounds for detention.[220] Today, this basic protection is still at the heart of constitutionally authorized habeas corpus. "[U]nless Congress acts to suspend it," wrote Justice O'Connor for the plurality in *Hamdi*, "the Great Writ of habeas corpus allows the Judicial Branch to play a necessary role in maintaining this delicate balance of governance, serving as an important judicial check on the Executive's discretion in the realm of detentions."[221] As this claim suggests, denial of habeas corpus review not only implicates the Suspension Clause, but also the structural checks imposed by the Constitution as a whole.[222]

Independent legal review of executive detention is especially needed when the government has foreclosed virtually every opportunity for detainees to obtain judicial review. The MCA states that "no alien unlawful enemy combatant subject to trial by military commission under this chapter may invoke the Geneva Conventions as a source of rights."[223] Therefore, any claim based on these rights cannot be brought in "any habeas corpus or other civil action or proceeding . . . ."[224] In addition, the MCA drastically limits a detainee's ability to claim under the War Crimes Act that he has been a victim of "grave breaches" of the Geneva Conventions.[225] The MCA

---

[219] *In re* Guantanamo Detainee Cases, 355 F.Supp. 2d 443, 465–66 (D.D.C. 2005); *see also* Hamdi v. Rumsfeld, 542 U.S. 507, 529–30 (2004) ("We have always been careful not to minimize the importance and fundamental nature of the individual's right to liberty, and we will not do so today.") (citations and quotations omitted).

[220] *See supra* notes 146–149 and accompanying text.

[221] *Hamdi*, 542 U.S. at 536.

[222] *See* Swain v. Pressley, 430 U.S. 372, 386 (1977) (Burger, C.J., concurring) ("A doctrine that allowed transfer of the historic habeas jurisdiction to an Art. I court could raise separation-of-powers questions, since the traditional Great Writ was largely a remedy against executive detention.")

[223] 10 U.S.C. 948b(g) (2006).

[224] MCA § 5(a).

[225] *See* MCA § 6(b)(1)(B), 18 U.S.C. § 2441(d) (2006); *see also supra* notes 47–55 and accompanying text.

also designates the President as the ultimate interpreter of the Geneva Conventions,[226] thereby granting him the discretion to decide whether all other instances of mistreatment should be considered grave breaches actionable under the War Crimes Act.[227]

Together with the jurisdiction-stripping provision of the MCA, these provisions effectively silence detainees. At common law and for most of our history, habeas corpus has ensured that fundamental rights and liberties are not a matter of executive grace. The President's ability to detain foreigners indefinitely without any real legal challenge represents a stark break with the traditional limits placed on executive power.

V. POSSIBLE JUSTIFICATIONS FOR DENYING THE RIGHT TO HABEAS CORPUS REVIEW

Even if detainees are constitutionally entitled to petition through writs of habeas corpus, Congress may still have been justified in denying them this right. The Suspension Clause allows Congress to suspend habeas corpus when "in Cases of Rebellion or Invasion the public Safety may require it."[228] Yet the language of the MCA does not speak of suspending the writ, nor does it mention either of these predicates. Moreover, the legislative history of the MCA also makes clear that Congress did not intend to suspend the constitutional writ, but rather to redefine the scope of statutory habeas corpus review.[229]

As an alternative to citing a case of rebellion or invasion, Congress may satisfy the Suspension Clause by either proving a judicial remedy whose scope is "commensurate with habeas corpus," or by preserving access to the writ in cases where the new remedy proves "inadequate or ineffective."[230]

Proponents of the jurisdiction-stripping provision contend that the Combatant Status Review Tribunal ("CSRT") provides an adequate and effective substitute for a constitutionally mandated habeas remedy.[231] Be-

---

[226] MCA § 6(a)(3) establishes that "the President has the authority for the United States to interpret the meaning and application of the Geneva Conventions and to promulgate higher standards and administrative regulations for violations of treaty obligations which are not grave breaches of the Geneva Conventions."

[227] *See id.*

[228] U.S. CONST. art. I, § 9.

[229] *See, e.g.*, 152 CONG. REC. S10368 (daily ed. Sept. 28, 2006) ("Fact No. 3, uncontested. We do not have a rebellion or an invasion."); 152 CONG. REC. H7548 (daily ed. Sept. 27, 2006) (statement of Rep. Sensenbrenner) (stating that the MCA did not suspend the Great Writ, but rather redefined the statutory writ); *see also* Hamdan v. Rumsfeld, 464 F.Supp.2d 9, 16 (D.D.C. 2006) ("Neither rebellion nor invasion was occurring at the time the MCA was enacted."); Hamdi v. Rumsfeld, 542 U.S. 507, 575 (2004) (Scalia, J., dissenting) (insisting that the Suspension Clause, "which carefully circumscribes the conditions under which the writ can be withheld, would be a sham if it could be evaded by congressional prescription . . .").

[230] Swain v. Pressley, 430 U.S. 372, 381–382 (1977).

[231] *See, e.g.*, McCarthy, *supra* note 60 ("It's not the name of the remedy that counts;

fore a Guantanamo detainee is tried by a military commission for the offenses charged, he must first be found by a CSRT to be an unlawful enemy combatant.[232]

Yet, it is doubtful that a CSRT would be found to be "commensurate with habeas corpus."[233] The Department of Defense guidelines require that tribunal members determine the status of a detainee by a preponderance of the evidence.[234] They also create a rebuttable presumption that the Government's evidence is genuine and accurate.[235] Furthermore, detainees before a CSRT have no right to counsel. While detainees are assisted by a "personal representative," that representative need not be an attorney, does not have a duty of loyalty to the detainee, and is prohibited from accessing any classified information used against the detainee.[236] Another problem, cited by Justice Breyer in his dissent from the Supreme Court's denial of certiorari in *Boumediene v. Bush*, is that "procedural infirmities cannot be corrected by review under the DTA [because it] provides for no augmentation of the record on appeal and . . . will provide no remedy for any constitutional violation."[237]

An additional concern, which the *Hamdan* Court raised with regard to military commissions but is equally relevant to CSRTs, is that these tribunals lack structural insulation from military influence.[238] Unlike Article I judges, CSRT tribunal members are military officers under the jurisdiction of the Secretary of Defense. They are bound by whatever policies or guidelines the Secretary of Defense issues, and their promotions and demotions are determined by officials under his command.[239]

---

it's the substance. The DTA gives the detainee exactly what habeas provides."). The Detainee Treatment Act and the Department of Defense Guidelines authorized by the Act establish Combatant Status Review Tribunals for Guantanamo detainees. *See* DTA § 1005(a)(1)(A); *see also* Memorandum from Deputy Secretary of Defense Gordon England re: Implementation of Combatant Status Review Tribunal Procedures for Enemy Combatants Detained at U.S. Naval Base Guantanamo Bay, Cuba (July 14, 2006), *available at* http://www.defenselink.mil/news/Aug2006/d20060809CSRT Procedures.pdf [hereinafter CSRT Procedures].

[232] *See* CSRT Procedures, *supra* note 231, at Enclosure (1).

[233] *Swain*, 430 U.S. at 381; *see* 152 CONG. REC. S11198 (daily ed. Dec. 5, 2006) (Statements of Sen. Specter) (claiming that one of the reasons that CRSTs are not an adequate and effective substitute for habeas corpus is because they are not adversarial, but consist of a one-sided interrogation of the detainee by the tribunal members).

[234] *See* CSRT Procedures, *supra* note 231, at Enclosure (1).

[235] *Id.*

[236] *See id.* at Enclosure (3); *see also* Brief for Professors of Constitutional Law and Federal Jurisdiction as Amici Curiae Advocating Denial of Motion to Dismiss (Reversal) at 20–25, Al-Marri v. Wright, No. 06-7427 (4th Cir. Dec. 12, 2006) [hereinafter Al-Marri Amicus Brief].

[237] 127 S. Ct. 1478, 1481 (2007) (Breyer, J., dissenting) (citing DTA § 1005(e)(2)(C)); *see also* Boumediene v. Bush, 476 F.3d 981, 1005 (D.C. Cir. 2007) (Rogers, J., dissenting).

[238] *See* Hamdan v. Rumsfeld, 126 S. Ct. 2749, 2771 (2006).

[239] *See* CSRT Procedures, *supra* note 231, at Enclosure (1) ("Each Tribunal shall be composed of a panel of three neutral commissioned officers of the U.S. Armed Forces . . . ."); *see also* Federman, *supra* note 89, at 170.

The Supreme Court has held that structural insulation is not a necessary condition for due process; however, CSRTs raise a number of constitutional concerns not present in the type of hearing previously assessed by the Court. In *Goldberg v. Kelly*,[240] the Court held that an administrative hearing to determine whether welfare benefits should be terminated satisfies procedural due process so long as the person presiding over the hearing did not participate in making the determination under review.[241] CSRTs satisfy this principle—the military official who charged the detainee with being an enemy combatant does not serve as a member of the Tribunal.[242] However, it is unclear whether this rule, which was crafted for administrative hearings where public assistance benefits may be terminated, is suitable for military tribunals where the loss of physical liberty and life are at stake.

CSRTs also raise constitutional concerns not at issue in *Goldberg*. The Court in *Goldberg* asserted that constitutional restraints apply to withdrawal of public assistance benefits.[243] The agency hearing allowed welfare recipients to contest this withdrawal.[244] Military detention, by contrast, raises multiple constitutional issues, such as cruel and unusual punishment and denial of due process.[245] Unlike the administrative agency hearing in *Goldberg*, which provided a forum for all relevant constitutional claims, CSRTs are not designed to adjudicate the constitutional claims that military detention implicates.[246]

Proponents of the view that CSRTs are an adequate substitute for habeas review claim that detainees have a right to challenge their imprisonment in federal court[247] because the Detainee Treatment Act provides detainees with a right to appeal CSRT decisions to the D.C. Circuit.[248] However, the narrow scope of review renders this possible substitute "inadequate or ineffective" under the *Swain* standard.[249] Senator Kyl (R-Ariz.) has explained that "[t]he only thing the DTA asks the courts to do is check that the record of the CSRT hearings reflect that the military has used its own rules."[250] As twenty-nine professors of constitutional law and

---

[240] 397 U.S. 254 (1970).

[241] *See id.* at 271.

[242] *See* CSRT Procedures, *supra* note 231, at Enclosure (1) (stating that "none of the officers appointed [to the Tribunal] shall have been involved in the apprehension, detention, interrogation, or previous determination of status of the detainee other than the CSRT process").

[243] *See* 397 U.S. at 262.

[244] *Id.*

[245] *See supra* notes 194, 199 and accompanying text.

[246] *See* CSRT Procedures, *supra* note 231, at Enclosure (1) (stating that the purpose and function of CSRTs is to determine whether each detainee meets the criteria to be designated as an unlawful enemy combatant).

[247] *See* McCarthy, *supra* note 60.

[248] DTA § 1005(e)(2) (2005).

[249] Swain v. Pressley, 430 U.S. 372, 382 (1977).

[250] 152 CONG. REC. S10271 (daily ed. Sept. 27, 2006) (Sen. Kyl).

federal jurisdiction have argued in amicus: "On this reading, the D.C. Circuit cannot examine the central question that has always been raised on habeas corpus even for enemy aliens and prisoners of war—the lawfulness of the prisoner's detention."[251]

As a practical matter, the fact that the DTA provides for D.C. Circuit review of constitutional infirmities in a CSRT is irrelevant because, as Justice Breyer notes, "the lower court has already rendered that provision a nullity."[252] Claiming that detainees can litigate their constitutional claims in a forum that proclaims they possess no constitutional rights is cold comfort to detainees who currently believe the only way their voices can be heard is through suicide, rioting and hunger strikes.[253]

VI. LEGISLATIVE AND JUDICIAL REMEDIES

Through its reconsiderations of the jurisdiction-stripping provision,[254] the 110th Congress has demonstrated that it is dedicated to the same constitutional issues that concern the Supreme Court. And unlike the Court, Congress is equipped with institutional resources that would allow it to fully appreciate the impact of any changes to the MCA. Realistically, however, Congress is currently not in a position to change the jurisdiction-stripping provision. Despite significant support behind the bills that seek to repeal the jurisdiction-stripping provision, it is unlikely that proponents could muster a two-thirds majority in both houses to override an expected veto from President Bush.[255]

The other alternative, wholesale invalidation of the provision by the Supreme Court, would lead to a number of practical difficulties. Because of the sensitive nature of the operations that led to their arrest, many detainees can only be prosecuted with classified information. Federal prosecutors and defense attorneys agree that the procedures currently available in federal court for dealing with classified information are inadequate. For the defense attorney, secret evidence raises significant concerns as to whether defense counsel can fulfill his ethical responsibility as a diligent, competent, and zealous advocate for his client.[256] Without the ability to share certain information with his client, a defense attorney often cannot conduct an adequate investigation and prepare a defense.[257] Moreover, shielding a defendant from the evidence used to prosecute him, even if his attorney can

---

[251] *See* Al-Marri Amicus Brief, *supra* note 236, at 24.

[252] Boumediene v. Bush, 127 S. Ct. 1478, 1480 (2007) (Breyer, J., dissenting) (referring to DTA § 1005(e)(2)(C)(ii).

[253] *See* Golden and Williams, *supra* note 195, at A12.

[254] *See supra* notes 73–76, and accompanying text.

[255] Editorial, *Pawn in Guantanamo's Game*, BOSTON GLOBE, Mar. 11, 2007, at 8D.

[256] Ellen Yaroshefsky, *Secret Evidence Is Slowly Eroding the Adversary System: CIPA and FISA in the Courts*, 34 HOFSTRA L. REV. 1063, 1064 (2006).

[257] *Id.* at 1074.

review it, has the potential to run afoul of both the Due Process and Confrontation Clause.[258]

A former federal prosecutor contends that procedures currently in place make it extraordinarily difficult for the Government to prosecute an alleged terrorist held in Guantanamo Bay in federal court.[259] Although there is statutory law in place that protects classified information from unwarranted disclosure, such as the Classified Information Procedures Act ("CIPA"),[260] the Government's ability to present its case fully and completely, and to provide the defense the evidence to which it is constitutionally entitled, is limited by the Government's desire not to declassify otherwise relevant information at trial.[261] Typically, the Government declassifies only that material which it must under CIPA, and seeks alternatives to declassification—or refuses to disclose facts altogether, if need be—which can lead to sanction and perhaps even dismissal by the Court under the Statute.[262] A dismissal of all or part of the case, an acquittal, or a hung jury can result because the CIPA substitutes do not permit the Government to develop a complete narrative by presenting all the relevant information in context.[263]

Secret evidence is currently seeping slowly into federal criminal prosecutions.[264] Striking down the jurisdiction-stripping provision would greatly increase the number of cases handled by federal courts that involve secret evidence. While this would address one constitutional concern, a failure to tailor procedures to deal with the problem of secret evidence would exacerbate several others.

The contextual due process approach offers a third possibility, one that is designed to balance individual and national interests. There is no constitutional barrier to applying a contextual due process approach in this context. Because there is no developed doctrine for constitutional habeas corpus,[265] there appear to be no constitutionally mandated procedural requirements aside from some form of judicial review. Therefore, new procedures that effectively deal with the problems plaguing trials in federal court can be developed so long as they do not erode essential constitutional promises.[266] This freedom allows the Court to set a constitutional floor that

---

[258] *See id.* at 1066. *See also* Kiareldeen v. Reno, 71 F.Supp.2d 402 (D.N.J. 1999) (holding that the Immigration and Naturalization Service's use of secret evidence, both at the alien's bond hearing and throughout his removal proceedings, violated the alien's right to due process).

[259] Telephone Interview with Michael D. Ricciuti, former Assistant United States Attorney and Chief of the Anti-Terrorism and National Security Section, U.S. Attorney's Office (Apr. 6, 2007) (Mr. Ricciuti's views were his own; he was not speaking on behalf of the Department of Justice) [hereinafter Michael D. Ricciuti Interview].

[260] 18 U.S.C. app. 3 §§ 1–16 (2000).

[261] Michael D. Ricciuti Interview, *supra* note 259.

[262] *Id.*

[263] *Id.*

[264] *See* Yaroshefsky, *supra* note 256, at 1064.

[265] *See* Federman, *supra* note 89, at 165.

[266] *See* Hamdi v. Rumsfeld, 542 U.S. 507, 533 (2004). Through the Rules Enabling

would sufficiently address the "'risk of an erroneous deprivation' of a detainee's liberty interest while eliminating certain procedures that have questionable additional value in light of the burden on the Government."[267]

CONCLUSION

Underlying this debate is the widely held sentiment that terrorists do not have, nor deserve, constitutional rights.[268] This notion, however, fails to address the reality that Guantanamo is home not only to terrorists such as September 11 mastermind Khalid Shaikh Mohammed, but also to people who the Government has admitted are a far cry from "the worst of the worst."[269] According to data published by the Department of Defense, the majority of Guantanamo detainees have been determined not to have committed hostile acts against the United States or its coalition allies, and only eight percent of the detainees have been characterized as Al Qaeda fighters.[270]

The contextual due process approach, as well as our common law tradition, requires that the wrongfully imprisoned have a fair opportunity to challenge the legality of their detention before a neutral decisionmaker. Habeas corpus review may impose a significant administrative burden on the Government. However, the conditions and considerations that guide the extraterritorial application of the Constitution ultimately weigh in favor of allowing courts to consider the writs filed by individuals detained by the United States Government. This determination is not only consistent with the Supreme Court's requirements under due process, but also with the broader view held by several Justices that the Constitution both protects individual rights and defines the limits of governmental action.[271]

Act, 28 U.S.C. §§ 2071–77 (1990), Congress has granted the Supreme Court the authority to create procedural rules. 28 U.S.C. §2072(a), (b) (1990). Other legislation, however, limits the Court's ability to sufficiently address this problem. As discussed in the text accompanying note 260, the Classified Information Procedures Act, 18 U.S.C. app. 3 §§ 1–16, regulates the use and admissibility of secret evidence. Therefore, any solution would likely require the attention of Congress as well as the Supreme Court.

[267] *Id.* at 534 (quoting Mathews v. Eldridge, 424 U.S. 319, 335 (1976)).

[268] *See* Yoo, *supra* note 67, at A18.

[269] *See* Editorial, *They Came For the Chicken Farmer,* N.Y. TIMES, Mar. 6, 2006, at A22 (quoting former Secretary of Defense Donald Rumsfeld's description of Guantanamo's detainee population).

[270] *See* MARK DENBEAUX & JOSHUA DENBEAUX, ESQ., REPORT ON GUANTANAMO DETAINEES: A PROFILE OF 517 DETAINEES THROUGH ANALYSIS OF DEPARTMENT OF DEFENSE DATA (Feb. 2006), *available at* http://law.shu.edu/aaafinal.pdf; *see also* CTR. FOR CONSTITUTIONAL RIGHTS, FACES OF GUANTANAMO: GUANTANAMO'S MANY WRONGLY IMPRISONED, http://ccr-ny.org/v2/reports/docs/faces_of_Guantanamo.pdf (profiling a number of Guantanamo detainees and the questionable bases for their detention); Editorial, *They Came For the Chicken Farmer, supra* note 269 (describing a Pakistani chicken farmer who was accused of being the Taliban's deputy foreign minister and imprisoned at Guantanamo Bay due a similarity in the spelling of his name and that of the foreign minister).

[271] *See, e.g.*, United States v. Verdugo-Urquidez, 494 U.S. 259, 277 (1990) (Kennedy, J., concurring) ("I take it to be correct, as the plurality opinion in Reid v. Covert sets forth,

The Court's approach to cases concerning the war on terror strongly favors this conception. While decided on different grounds, *Hamdi*, *Padilla* and *Hamdan* all emphatically reject the Government's contention that Guantanamo detainees held exist in a legal black hole where the Constitution and Geneva Conventions do not shine. The Supreme Court's reluctance to accept this interpretation is also deeply rooted in our common law tradition of disfavoring lawless enclaves; a tradition that has existed even with respect to individuals who are not subjects or citizens of the sovereign.[272]

—*Daniel Michael*[*]

that the Government may act only as the Constitution authorizes, whether the actions in question are foreign or domestic."); Hamdi v. Rumsfeld, 542 U.S. 507, 556 (2004) (Scalia, J., dissenting) ("The gist of the Due Process Clause, as understood at the founding and since, was to force the Government to follow those common-law procedures traditionally deemed necessary before depriving a person of life, liberty, or property.").

[272] *See* Pfander, *supra* note 78, at 512.

[*] J.D., Harvard Law School, 2007; B.A., New York University, 1998. I thank Professor Martha Minow for her comments on earlier drafts of this Recent Development.

THE MISPLACED ROLE OF IDENTITY THEFT IN TRIGGERING PUBLIC NOTICE OF DATABASE BREACHES

> Don't let anybody kid you that this is about privacy. It's about power and autonomy.[1]

In 1973, the U.S. Department of Health, Education, and Welfare ("HEW"), in response to public concern about the increased collection and storage of personal information, presented a report ("HEW Report") that ambitiously recommended the creation of a "Code of Fair Information Practice."[2] The HEW Report promoted several data usage principles, including the right of individuals to limit the use of personal information to the purpose for which it was collected, and the requirement that organizations ensure the accuracy and prevent the misuse of any personal data that they collect and use.[3] The HEW Report helped lay the foundation for passage of the landmark Privacy Act of 1974.[4]

Several decades later, it appears that the impact of the HEW Report was transient. In 2005, alarming newspaper headlines announced the theft of millions of records of sensitive, personal information.[5] Database security breaches had compromised the personal privacy of hundreds of thousands of Americans, and the federal government lacked a regulatory framework for database businesses that would prevent such breaches.[6] Immediately after reports of the security breaches became public, lawmakers began drafting bills consistent with the principles of the HEW Report. These bills emphasized prevention; they focused on ensuring that industry controlled personal information more stringently before accidents occurred.[7] Yet, as the months

---

[1] Stephen Pounds, *Identity Complex*, PALM BEACH POST, Apr. 10, 2005, at 1F (quoting Robert O'Harrow, Jr.).

[2] *See* U.S. DEP'T OF HEALTH, EDUC. & WELFARE, SEC'Y'S ADVISORY COMM. ON AUTOMATED PERSONAL DATA SYS., *Records, Computers, and the Rights of Citizens* 41–42 (1973) [hereinafter HEW REPORT], *available at* http://www.aspe.hhs.gov/datacncl/1973privacy/tocprefacemembers.htm; *see also* DANIEL J. SOLOVE ET AL., PRIVACY, INFORMATION, AND TECHNOLOGY 145 (2006) (discussing the influence of the HEW Report).

[3] *See* HEW REPORT, *supra* note 2, at 41–42.

[4] 5 U.S.C. § 552a (2000). In particular, the Act requires that agencies inform people of "the principal purpose or purposes for which the information is intended to be used" when their information is collected. *Id.* at § 552a(e)(3)(B). *See also* SOLOVE ET AL., *supra* note 2, at 147.

[5] See, *e.g.*, Morey Elizabeth Barnes, *Falling Short of the Mark: The United States Response to the European Union's Data Privacy Directive*, 27 NW. J. INT'L L. & BUS. 171, 171 (2006); Robert O'Harrow, Jr., *ID Data Conned From Firm*, WASH. POST, Feb. 17, 2005, at E1; Gary Rivlin & Tom Zeller, Jr., *Purloined Lives*, N.Y. TIMES, Mar. 17, 2005, at C2.

[6] *See* Robert O'Harrow, Jr., *ChoicePoint Data Cache Became a Powder Keg*, WASH. POST, Mar. 5, 2005, at A1.

[7] *See, e.g.*, Personal Data Privacy and Security Act of 2005, S. 1332, 109th Cong. (2005) (seeking "to prevent and mitigate identity theft; to ensure policy; and to enhance . . . protections against security breaches, fraudulent access, and misuse of personally identifiable information"); Information Protection and Security Act, S. 500, 109th Cong. (2005) (regu-

passed, and more details about the database breaches became public, lawmakers began introducing bills that diverged from the HEW Report's guiding principles. Under these later proposals, only certain types of information were deemed sensitive,[8] and the notification requirement became a business decision,[9] based on whether the potential for identity theft existed.[10]

This Recent Development is not a call to arms for broad privacy regulations[11] or an appeal for the elusive right to be left alone.[12] Rather, this Recent Development simply argues that consumers should not have to pay for costs they did not incur. Recent congressional efforts in the information privacy context have amounted to a disservice to the American public because they do not assign value to non-pecuniary costs and they premise notification on the narrow problem of identity theft.

Part I of this Recent Development briefly describes the database industry and the state of some important privacy laws fitfully used to regulate these businesses and similar organizations. Part II describes common features and salient differences between several different bills introduced by the 109th Congress. Finally, Part III makes three recommendations: (1) database businesses should not have sole discretion to determine when to notify either the public or the government; (2) the 110th Congress should sever the misleading metric of identity theft from the criteria governing when business should notify the public in the event personal information has been compromised; and (3) because the compromise of personal information can have more than financial implications, the definition of protectible personal information should be broader.

I. A GROWING INDUSTRY AND ITS IMPERFECT REGULATION

Database businesses control billions of records containing information about American citizens.[13] These companies use technological acu-

---

lating "information brokers and protecting individual rights with respect to personally identifiable information").

[8] *See, e.g.*, H.R. 4127 § 5(7)A (2005) (proposing an amendment to the Fair Credit Reporting Act and procedures for "nationwide notice" in the event of a security breach).

[9] "Breach of security" is defined as a significant risk of identity theft, a determination made by the business. *See, e.g.*, Financial Data Protection Act of 2005, H.R. 3997, 109th Cong. § 630(b)(3) (2005); H.R. 4127, 109th Cong. (2005).

[10] *See H.R. 3997, Financial Data Protection Act of 2005: Hearing Before the H. Subcomm. on Financial Institutions and Consumer Credit of the Comm. on Financial Serv.*, 109th Cong. 101 (2005) [hereinafter *H.R. 3997 Hearing*] (statement of Oliver Ireland, Financial Serv. Coordinating Council) (defining identity theft as when "a criminal uses personal identifying information related to another person . . . to open a new account in that person's name").

[11] *See, e.g.*, DANIEL J. SOLOVE & MARC ROTENBERG, INFORMATION PRIVACY LAW 688 (2d ed. 2003) (discussing the global trends of data privacy); *see also* DAVID BRIN, THE TRANSPARENT SOCIETY 8–23 (1998) (stating that "it will prove quite impossible to legislate away the new surveillance tools and databases. They are here to stay.").

[12] *See, e.g.*, Samuel D. Warren & Louis D. Brandeis, *The Right to Privacy*, 4 HARV. L. REV. 193, 205 (1890).

[13] *See* Pounds, *supra* note 1, at 1F (discussing the various sizes of the database compa-

men to efficiently access and aggregate public,[14] publicly available,[15] and non-public information[16] and profit from society's expanding use of this data.[17] This industry has seen its profits soar as businesses and government[18] have come to rely on its services[19]—a result that is hardly surprising, given that the rapid data aggregation, digitization, and dissemination of information facilitated by database businesses have profoundly increased their clients' efficiency and ability to deliver services.[20] The use of databases reassures the Boy Scouts that their scoutmasters are not paroled sex offenders, instills patients with confidence about the competency of their medical providers, and gives law enforcement an edge in securing the homeland.[21] Readily accessible information enables businesses to extend credit for mort-

---

nies, including ChoicePoint, which has 19 billion documents, and Acxiom Corp., which claims that its consumer database covers 95% of U.S. households).

[14] "Public information" includes personal data that individuals submit to the government and is generally available to the public. *See Protecting Consumers' Data: Policy Issues Raised by ChoicePoint: Hearing Before the Subcomm. on Commerce, Trade, and Consumer Protection of the H. Comm. on Energy and Commerce*, 109th Cong. 20–21 (2005) [hereinafter *Protecting Consumers' Data Hearing*] (statement of Deborah Platt Majoras, Chairman, FTC).

[15] "Publicly available information" consists of facts about an individual that might be found in non-governmental sources, either in print or online. *Id.*

[16] "Non-public information," the most sensitive category, includes items found on applications submitted to obtain credit, secure employment, or obtain insurance. *Id.*

[17] *See* Pounds, *supra* note 1, at 1F (describing entrepreneur Hank Asher's creation of a type of parallel processing "where hundreds of computers break down information requests and conduct searches simultaneously in different databases, in the end pulling together divergent blocks of material in seconds" and thereby "learns more about that person as it scours one database and uses that information to expand its search in another database"). Asher's technology remains the "core" of ChoicePoint and Reed Elsevier, PLC., parent company of LexisNexis. *Id.*

[18] *See* Gov't Accountability Office ("GAO"), GAO-06-609T, PERSONAL INFORMATION: AGENCIES AND RESELLERS VARY IN PROVIDING PRIVACY PROTECTIONS 11 (2006) (stating that in 2005 the Social Security Administration and the Departments of Justice, Homeland Security, and State had approximately $30 million in contractual arrangements with information resellers); *see also* Glenn R. Simpson, *Big Brother-in-Law: If the FBI Hopes to Get the Goods on You, It May Ask ChoicePoint*, WALL ST. J., Apr. 13, 2001, at A1 (documenting the government's increasing reliance on database businesses).

[19] *See Identity Theft: Recent Developments Involving the Security of Sensitive Consumer Information: Hearing Before the S. Comm. on Banking, Housing, and Urban Affairs*, 109th Cong. 56 (1st Sess. 2005) [hereinafter *Identity Theft Hearing*] (statement of Don McGuffey, Vice President, ChoicePoint) (noting that ChoicePoint provides services to more than 7000 federal, state, and local law enforcement agencies); ROBERT O'HARROW, JR., NO PLACE TO HIDE 34, 145 (2005) (stating that Acxiom is "a billion-dollar player in the data industry, with details about nearly every adult in the United States" and that ChoicePoint has more than 250 total terabytes of data regarding the lives of nearly every adult in America); Daniel J. Solove, *Access and Aggregation: Public Records, Privacy and the Constitution*, 86 MINN. L. REV. 1137, 1151 (2002) (discussing the increasing flow of information from the private sector to the public).

[20] *See* ALAN CHARLES RAUL, PRIVACY AND THE DIGITAL STATE: BALANCING PUBLIC INFORMATION AND PERSONAL PRIVACY 43 (2002) (giving examples of this efficiency).

[21] *See generally* DEREK V. SMITH, RISK REVOLUTION (2004) (providing extensive treatment of the benefits). Smith is the CEO of ChoicePoint. ChoicePoint, Executive Management, http://www.choicepoint.com/about/executive.html#dvs (last visited Apr. 17, 2007).

gages less expensively,[22] enhances public health surveillance,[23] and even assists in finding missing children.[24]

But the industry's comprehensive infiltration of our lives is not costless; as evidenced by the widespread impact of the recent database breaches, weaknesses in industries that constitute society-wide systems can have profound consequences.[25] Database giant ChoicePoint exemplified these dangers in February 2005 when it mistakenly disclosed the personal information of 145,000 Americans to scam artists and failed to inform the public for three months.[26] Revelations soon surfaced that other database businesses, as well as financial institutions and merchants, had also disclosed sensitive data.[27] By the end of 2005, over 50 million individuals' personal information had been compromised.[28]

The financial implications of such breaches are significant. The Federal Trade Commission ("FTC") estimates that the wrongful use of data, some of which can be traced to poor data security practices,[29] costs businesses and consumers $55 billion annually.[30] Data insecurity also imposes non-pecuniary losses by impairing citizens' ability to participate meaningfully in society:[31] inaccurate or misused data can restrict an individual's ability to secure employment, obtain a mortgage, or purchase a car.[32] It can

---

[22] *See* FRED CATE, PRIVACY IN PERSPECTIVE, at xiii (2001).

[23] *See* Lawrence O. Gostin et al., *Balancing Communal Goods and Personal Privacy Under a National Health Informational Privacy Rule*, 46 ST. LOUIS U. L.J. 5, 9 (2002).

[24] *See* SMITH, *supra* note 21, at 8.

[25] *See* James P. Nehf, *Recognizing the Societal Value in Informational Privacy*, 78 WASH. L. REV. 1, 81 (2003) (discussing how other society-wide systems, from capital markets to banking, benefit from effective oversight).

[26] *See* Tom Zeller, *Data Security Laws Seem Likely, so Consumers and Businesses Vie to Shape Them*, N.Y. TIMES, Nov. 1, 2005, at C3; *see also* ACLU, THE CHOICEPOINT ID THEFT CASE: WHAT IT MEANS (2005), *available at* http://www.aclu.org/privacy/consumer/15301leg20050310.html (describing ChoicePoint's initial reaction to the disclosure of data).

[27] *See* David Lazarus, *Shifting Sands in Data Leak*, S.F. CHRON., Feb. 25, 2005, at C1 (documenting how ChoicePoint discovered the breach in October but waited until January to notify consumers); *see also* Privacy Rights Clearinghouse, A Chronology of Data Breaches, http://www.privacyrights.org/ar/ChronDataBreaches.htm (last visited Apr. 16, 2007) (providing a continuously updated list of data breaches since the ChoicePoint incident in February 2005).

[28] *See H.R. 3997 Hearing*, *supra* note 10, at 5 (statement of Rep. Michael Oxley (R-Ohio), Chairman, H. Subcomm. on Financial Institutions and Consumer Credit); Dennis D. Hirsch, *Protecting the Inner Environment: What Privacy Regulation Can Learn from Environmental Law*, 41 GA. L. REV. 1, 19 (2006).

[29] Information security breaches come in many forms. Hacks into computer systems are one variety. Other breaches result from misplaced disks, stolen laptops, losses or thefts while data is in transit, and the activities of rogue employees. *See* Fred H. Cate, *Bank Secrecy Act, Security Breaches, Electronic Commerce, and Identity Theft*, 60 CONSUMER FIN. L.Q. REP. 344, 345 (2006).

[30] *See Protecting Consumers' Data Hearing*, *supra* note 14, at 2 (statement of Rep. Cliff Stearns (R-Fla.)).

[31] *See* Paul M. Schwartz, *Privacy and Democracy in Cyberspace*, 52 VAND. L. REV. 1609, 1653 (1999) (suggesting that meaningful democratic deliberation is threatened by information processing that is not secure).

[32] *See, e.g.*, ROBERT GELLMAN, PRIVACY, CONSUMERS, AND COSTS: HOW THE LACK OF PRIVACY COSTS CONSUMERS AND WHY BUSINESS STUDIES OF PRIVACY COSTS ARE BIASED

even lead to wrongful arrests.[33] Reacting to the threat of security breaches, individuals may choose not to participate in activities that require them to reveal personal data, from registering to vote to engaging in efficiency-enhancing activities like e-commerce and internet banking.[34]

The U.S. Constitution does not explicitly guarantee privacy.[35] Nevertheless, as state governments and other entities have increasingly opted to sell personal information,[36] Congress has responded with legislation intended to curtail the practice.[37] Congress's piecemeal, reactive approach, however, has left gaps, the result of which has been that the emerging private database industry is largely unregulated.[38]

For example, the Fair Credit Reporting Act of 1970 ("FCRA"),[39] which limits how credit-reporting agencies use personal data, applies to database businesses only to the extent that they engage in consumer services similar to those of credit-reporting agencies.[40] The Privacy Act of 1974 ("Privacy Act"),[41] which regulates the use of personal information by federal agencies and their private contractors, does not reach the activities of privately created databases.[42] The Gramm-Leach-Bliley Act ("GLBA"),[43] which gives individuals the right to prevent financial institutions from shar-

---

AND INCOMPLETE 25–28 (2002), *available at* http://www.epic.org/reports/dmfprivacy.pdf.

[33] *See* GAO, GAO-02-363, IDENTITY THEFT: PREVALENCE AND COST APPEAR TO BE GROWING 56 (2002) (reporting that identity theft has led to wrongful criminal investigations, arrests, or convictions in almost 13,000 complaints to the FTC).

[34] *See* UCLA CTR. FOR COMMC'N POLICY, THE UCLA INTERNET REPORT: SURVEYING THE DIGITAL FUTURE 48–53 (2003) [hereinafter UCLA REPORT], *available at* http://www.forbes.com/fdc/mediaresourcecenter/UCLA03.pdf (noting that privacy concerns are the primary worry for online shoppers). *See generally* Paul M. Schwartz, *Privacy and Participation: Personal Information and Public Sector Regulation in the United States*, 80 IOWA L. REV. 553, 560 (1995) (discussing how data collection "creates a potential for suppressing a capacity for free choice: the more that is known about an individual, the easier it is to force his obedience").

[35] *See* SOLOVE, *supra* note 11, at 20.

[36] *See* SOLOVE, *supra* note 19, at 1143–45.

[37] *See, e.g.*, Daniel J. Solove & Chris Jay Hoofnagle, *A Model Regime of Privacy Protection* 3 (George Washington Univ. Law Sch. Pub. Law Research, Paper No. 132, 2005), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=699701.

[38] *See* ANGIE A. WELBORN, INFORMATION BROKERS: FEDERAL AND STATE LAWS (2005); *see also* DANIEL J. SOLOVE, THE DIGITAL PERSON 71–72 (2004) (quoting Joel Reidenberg, *Privacy in the Information Economy: A Fortress or Frontier for Individual Rights?*, 44 FED. COMM. L.J. 195 (1992) (observing that the "sectoral" nature of privacy laws "derives from the traditional American fear of government intervention in private activities and the reluctance to broadly regulate industry").

[39] 15 U.S.C. § 1681a (2000).

[40] *See* GAO, *supra* note 18, at 6.

[41] *See* 5 U.S.C. § 552a (2000).

[42] *See id.*; *see also* NATHAN BROOKS, DATA BROKERS: BACKGROUND AND INDUSTRY OVERVIEW 2 (2005); Chris Hoofnagle, *Big Brother's Little Helpers: How ChoicePoint and Other Commercial Data Brokers Collect and Package Your Data for Law Enforcement*, 29 N.C. J. INT'L L. & COM. REG. 595, 623 (2004); Robert Gellman, *Does Privacy Law Work?*, *in* TECHNOLOGY AND PRIVACY: THE NEW LANDSCAPE 195–201 (Philip E. Agre & Marc Rotenberg eds., 1997) (discussing the loopholes in the Privacy Act and the difficulty of using the Privacy Act to obtain damages).

[43] 15 U.S.C. §§ 6801–6809 (2000).

ing their data with third parties, also does not apply to privately managed databases.[44]

Other congressional measures have sought to restrict information captured from areas as diverse as driving records and medical information.[45] But as the security breaches in 2005 demonstrated, from a consumer protection standpoint a need still exists for legislation that specifically addresses the database industry.[46]

II. THE 109TH CONGRESS JUMPS ON THE BANDWAGON

Notwithstanding industry opposition, as news of the database breaches became public in 2005,[47] members of Congress from both political parties[48] called for hearings and began drafting legislation.[49] California's 2003 law,[50] which arguably forced ChoicePoint and at least sixty companies to issue nationwide notifications,[51] served as a model for many of these early bills.[52]

While these bills varied greatly in their specifics, they generally shared at least three core elements with respect to the notification decision. First, each bill defined the applicable entity.[53] This definition, because it delineates the scope of the law, is particularly important for consumers, given the wide variety of institutions that collect and aggregate personal information. Possible institutions include universities, online retailers, and even churches.[54] Indeed, it is the interpretation of the phrases "credit-

---

[44] *See id.* § 6802.

[45] *See, e.g.*, Driver's Privacy Protection Act ("DPPA") of 1994, 18 U.S.C. § 2721 (2000) (applying only to state motor vehicles departments and their employees); Health Insurance Portability and Accountability Act ("HIPAA") of 1996, 42 U.S.C. § 1320d (2000) (applying to health plans, health care clearinghouses, and health care providers). *See generally* THE PRIVACY LAW SOURCEBOOK 2004 (Marc Rotenberg ed., 2005) (discussing DPPA and HIPAA).

[46] *See Protecting Consumers' Data Hearing*, *supra* note 17, at 1–4 (statement of Rep. Cliff Stearns (R-Fla.)) (noting that the current regulatory regime is ineffective because of its uncoordinated nature).

[47] *See, e.g.*, Michele Heller, *Debate Starts on Legislative Response*, AM. BANKER, Oct. 11, 2005, at 1 (reporting the heavy resistance of business groups to regulation).

[48] *See H.R. 3997 Hearing*, *supra* note 10, at 1.

[49] *See, e.g.*, *Assessing Data Security: Preventing Breaches and Protecting Sensitive Information: Hearing Before the H. Comm. on Fin. Serv.*, 109th Cong. (2005).

[50] *See* S. 1386, 2002 Leg., Reg. Sess. (Cal. 2002) (codified as amended at CAL. CIV. CODE §§ 1798.29, .82 (West Supp. 2006)).

[51] *See Securing Consumers' Data: Options Following Security Breaches: Hearing Before the Subcomm. on Commerce, Trade, and Consumer Protection of the H. Comm. on Energy and Commerce*, 109th Cong. 5 (2005) [hereinafter *Securing Consumers' Data Hearing*] (statement of Rep. Jan Schakowsky (D-Ill.)); Grant Gross, *Congress Looks to Pass Data Breach Law*, INFOWORLD, Sept. 2, 2005, http://www.infoworld.com/article/05/09/02/HNcongressdata_1.html.

[52] *See* Cal. S. 1386; David Leit & Matthew Atlas, *Data Privacy Regulations: A Patchwork for Potential Problems*, N.J. LAW., May 22, 2006, at A6.

[53] *See, e.g.*, S. 751, 109th Cong. § 2 (2005).

[54] *See, e.g.*, BROOKS, *supra* note 42, at 2.

reporting agency," "GLBA-institution," and "government agency" that have traditionally kept databases outside of regulatory purview.[55] Yet unbounded terms are not necessarily the panacea; overinclusive definitions may ignore the disparate abilities and practices of different institutions. Conversely, legislation aimed strictly at database businesses may overlook certain industries, such as health maintenance organizations and financial institutions, whose poor data practices can also put sensitive information at risk.

Second, the recent database notification bills set forth certain tests or triggers to guide the public notification process.[56] California's Notice of Breach Security Act,[57] for example, mandates notification whenever personal information is "reasonably believed to have been acquired by an unauthorized person," regardless of whether there is evidence that an unauthorized person actually took or used any personal information.[58] On its face, California's law does not require an actual breach, nor does it require a factual finding of identity theft.[59] As discussed below, although early federal bills followed this model, later proposals have deviated significantly from it.

Third, these recent bills set forth what types of information must be protected, what instructions must be contained in notices of breaches sent to consumers, who must be notified (consumers, government authorities, or both), and how these individuals must be notified (by mail, print publication, or online posting).[60]

The earliest bills imposed the strictest requirements on the database industry. For example, under a bill sponsored by Senator Dianne Feinstein (D-Cal.) in January 2005,[61] a breach involving the unauthorized acquisition of and access to personal information would be sufficient to trigger the consumer notice requirement.[62] Two months later, in the wake of the cas-

---

[55] *See* Hoofnagle, *supra* note 42, at 622–27.

[56] *See, e.g.*, S. 1789, 109th Cong. § 421 (2005) (designating when business are to give notice).

[57] Cal. S. 1386.

[58] *See id.*; *see also* Tyler Paetkau & Roxanne Torabian-Bashardoust, *California Deals with Identity Theft*, BUS. L. TODAY, May-June 2004, at 37, *available at* http://www.abanet.org/buslaw/blt/2004-05-06/bashardoust.shtml.

[59] Cal. S. 1386.

[60] *See, e.g.*, S. 1789 (setting guidelines for notice procedures and the contents of such notices).

[61] *See* S. 115, 109th Cong. (2005).

[62] *See id.* § 2(2). A similar bill introduced the same day, the Privacy Act of 2005, would have prohibited a commercial entity from collecting personal information and then disclosing or selling it to non-affiliated parties unless notice was given to the individual whose personal information was at issue and gave that individual an opportunity to restrict the disclosure or sale. *See* S. 116, 109th Cong. § 101 (2005). A later version of this bill, introduced in April 2005, S. 751, 109th Cong. (2005), received several endorsements from the online financial services industry. *See* Press Release, Senator Dianne Feinstein, E-Loan and ING Direct Endorse Feinstein Identity Theft Legislation (June 7, 2005), *available at* http://feinstein.senate.gov/05releases/r-idtheft-endor.pdf; *see also* S. 1332, 109th Cong. § 3 (2005) (advancing the same disclosure requirement).

cading security breach announcements, Senator Bill Nelson (D-Fla.) introduced a bill that focused on the preventative aspects of database security. [63] The bill directed the FTC to promulgate rules regulating the security procedures of information brokers.[64] Senator Nelson's bill defined "personally identifiable information" as any information used "to identify a person or cause harm to such person."[65] That bill did not explicitly link the definition of "harm" to identity theft.[66]

Senator Arlen Specter (R-Pa.) introduced one of the most expansive bills.[67] Under his proposal, business entities would only earn an exemption from individual notice requirements if a joint consultation with state and federal authorities concluded that there was a "de minimis risk of harm" to sensitive, personally identifiable information compromised by a security breach."[68] The bill also required data brokers to create preventative programs[69] and contemplated the government leveraging its considerable market strength to change the business practices of databases[70] by having the General Services Administration review all government contracts to ensure that the best data practices were followed. One implication of this procedure was that financial penalties and the loss of government business would result if the industry did not meet specific benchmarks.[71]

Subsequent Senate bills significantly narrowed the circumstances under which data brokers would be required to notify individuals.[72] These bills listed discrete types of "sensitive personal information," which seemed designed to preclude an expansive interpretation of the term.[73] Unlike their predecessors, these bills conflated the risk of harm with the risk of identity theft.[74] Their provisions stated that breaches of security were material only when the affected business "establish[ed] a reasonable basis to conclude that a significant risk of identity theft to an individual exist[ed]."[75]

---

[63] *See* Information Protection and Security Act, S. 500, 109th Cong. (2005).

[64] *Id.* §§ 2, 3.

[65] *See id.*

[66] *Id.* S. 768, 109th Cong. (2005); H.R. 4127, 109th Cong. (2005).

[67] S. 1332, 109th Cong. (2005).

[68] *See id.* § 424.

[69] *See id.* § 402.

[70] *See id.* § 601.

[71] *See id.* § 601.

[72] *See, e.g.*, S. 500, 109th Cong. (2005) (providing a comprehensive definition not limited to personal identities).

[73] *See, e.g.*, S. 1326, 109th Cong. § 2 (2005) (defining the term as including only an individual's first and last name, address or telephone number, as well as other identifying numbers, but not including any other description or grouping of individuals); Identity Theft Protection Act, S. 1408, 109th Cong. § 10 (2005) (defining "sensitive personal information" as including a name, address, or telephone number combined with at least one other "data element" defined in the section).

[74] *See* S. 1408. While this bill purported to expand the definition of "personal information" and to permit the FTC to modify the definition of harm through the rulemaking process, see *id.* § 10, it also made notification contingent on whether the defined entity determined that there was a reasonable risk of identity theft, see *id.* § 3.

[75] Notification of Risk to Personal Data Act, S. 1326, 109th Cong. § 2 (2005).

In addition to imposing the higher threshold of "significant risk,"[76] the bills contained a sizable exemption from the notification requirements: no notification would be required if the compromised business concluded that no risk of identity theft existed or if the business had a security program "reasonably designed to block unauthorized transactions before they are charged to an individual's account."[77]

The House also circulated bills tying notification requirements to identity theft.[78] The most prominent of these bills, which gathered considerable momentum in the waning days of the 109th Congress, were the Financial Data Protection Act ("H.R. 3997")[79] and the Data Accountability and Trust Act ("H.R. 4127" or "DATA").[80] Factions that supported and opposed these measures quickly emerged.[81]

The original version of H.R. 3997 required public notification of a breach with respect to "sensitive financial identity information"[82] if such information "has been or is reasonably likely to be misused in a manner causing substantial harm or inconvenience against the consumers to whom such information relates to commit identity theft."[83] Critics of the bill insisted that the "substantial harm" test was too stringent.[84] In response, an amended version altered the test, requiring only that "any person . . . following the discovery of a breach of security of the system maintained by such person that contains such data . . . notify each individual that their personal information was acquired by an unauthorized person" and notify

---

[76] *See Securing Consumers' Data Hearing*, *supra* note 51, at 45 (statement of Rep. Jan. Schakowsky (D-Ill.)) (finding the bar for notification and materiality quite high, and expressing puzzlement over a data broker's testimony that the "unauthorized access of information by a former employee does not constitute a significant risk"); *see also Securing Electronic Personal Data: Striking a Balance Between Privacy and Commercial and Governmental Use: Hearing Before the H. Comm. on the Judiciary*, 109th Cong. 14 (2005) [hereinafter *Securing Electronic Personal Data Hearing*] (statement of Chris Swecker, Assistant Dir., FBI).

[77] S. 1326, 109th Cong. § 2 (2005).

[78] *See, e.g.*, H.R. 3997, 109th Cong. (2005); H.R. 4127, 109th Cong. (2005).

[79] H.R. 3997, 109th Cong. (2005).

[80] H.R. 4127, 109th Cong. (2005).

[81] *See, e.g.*, Press Release, Consumers Union et al., Re: Do Not Bring H.R. 3997, the Financial Services Data "Security" Bill, to the Floor (July 21, 2006), *available at* http://www.consumersunion.org/pdf/HR3997-Floor.pdf; Privacy Rights Clearinghouse, *Congressional Update: Bad ID Theft Bill Will be Considered This Week*, PRC'S PRIVACY UPDATE, July 24, 2006, http://www.privacyrights.org/newsletter/060710.htm; Letter from Ed Mierzwinski, Dir. of Consumer Prot., U.S. Pub. Int. Res. Group, to Members, H. Subcomm. on Commerce, Trade, and Consumer Prot. (Nov. 2, 2005), *available at* http://www.consumersunion.org/pub/core_financial_services/002828.html.

[82] *See* H.R. REP. No. 109-454, pt. 1, at 15 (2005) (defining such information as "the first and last name, the address, or the telephone number of a consumer," in combination with a Social Security number, driver's license number, taxpayer numbers, or unique biometric data).

[83] H.R. 3997, 109th Cong. § 630(e)(1)(C) (2005) (as introduced in H.R.). This bill is identical to the Financial Data Protection Act of 2005, S. 2169, 109th Cong. (2005).

[84] H.R. REP. No. 109-454, pt. 1 at 78 (explaining concerns about the substantial harm test).

the FTC.[85] Not only did this amendment seem to eliminate the substantial harm test, it also contained a new provision that provided an exemption from all notification requirements if the covered entity determined that "no reasonable risk of identity theft, fraud, or other unlawful conduct" existed.[86] This exception effectively swallowed whatever remained of the facially broad notification rule.

Representative Cliff Stearns (R-Fla.) introduced the second prominent bill, H.R. 4127, whose varying levels of preemption and enforcement received mixed reviews from industry members and consumers.[87] In its amended form, the bill required private companies to notify "consumers and certain authorities whenever there is a breach in the security of a consumer's personal information" and "to investigate and take steps to repair the breach."[88] As reported in the House, a "breach of security" is defined as:

> the unauthorized acquisition of data in electronic form containing personal information that establishes a reasonable basis to conclude that there is a significant risk of identity theft to the individual to whom the personal information relates.[89]

The bill's opponents roundly criticized the "significant risk of identity theft" test in the original version[90]; reflecting a compromise with this contingent, the amended version flatly states that security breaches require businesses to notify consumers.[91] Like H.R. 3997, the amended version of H.R. 4127 provides a similarly broad exemption for businesses if the covered entity determines that "there is no reasonable risk of identity theft, fraud, or other unlawful conduct."[92]

III. SUGGESTED LEGISLATIVE GUIDEPOSTS

As these recent legislative efforts suggest, lawmakers from both parties have constricted the debate on database notification to questions about

---

[85] H.R. 3997, § 3(a) (as reported in H.R.).

[86] H.R. 3997 (as reported in House); *see also* H.R. REP. No. 109-454, pt. 1, at 21–24 (discussing proposed amendments to § 3 of the bill).

[87] Posting of Ed Mierzwinski to U.S. PIRG Consumer Blog, Cutting the Privacy Baby in Half, http://www.uspirg.org/html/consumer/archives/2005/11/index.html (Nov. 3, 2005, 6:30 PM).

[88] *See* H. REP. No. 109-453, pt. 3, at 4 (2005).

[89] *See* H.R. 4127, 109th Cong. § 5 (2005) (as introduced in H.R.).

[90] *See* Press Release, H. Comm. on Energy and Commerce Democrats, Democrats Reject Weak Data Security Bill (Nov. 3, 2005), *available at* http://energycommerce.house.gov/Press_109/109nr24.shtml. *But see H.R. 3997 Hearing*, *supra* note 10, at 29.

[91] *See* H. REP. No. 109-453, pt. 2, 4 (2005) (Any person engaged in interstate commerce . . . shall, following the discovery of a breach of security of the system . . . notify each individual [and] notify the Commission).

[92] *See* H.R. 4127, § 3 (as introduced in H.R.) (second reported); *see also* H.R. 3997.

what should be the appropriate burden for proving identity theft. Unfortunately, this limited debate has not considered or attempted to institute incentive structures that might allow better policing of the database industry before breaches occur. As the next round of negotiations, hearings and legislative proposals begin, any legislation regulating the database industry should, at a minimum, require that the notification decision be a product of a joint consultation with the government; not allow any covered entity to predicate the notification decision on the occurrence (or nonoccurrence) of identity theft; and maintain a flexible, expansive definition of personal information. While such legislation could be directed at many industries, at this time there is a particularly acute need for the database industry due to the combination of minimal transparency and the indirect market relationship that exists between database businesses and the people whose personal information it acquires and aggregates.

A. The Notification Decision Should Not Be Left to Industry Alone

When businesses classify which threats warrant public notification by conducting an internal risk assessment, they are essentially self-regulating—particularly if these decisions are not subject to external review or veto. Supporters of self-regulation emphasize its flexibility, arguing that it leads to "a more tailored balance between information uses and privacy than privacy laws do."[93] According to this argument, state and federal laws are not well-suited to overseeing an increasingly complex and ever-changing industry.[94] Opponents of strict regulation therefore favor at most a low federal floor that would allow companies to develop their own policies for notification.[95]

These opponents also argue that strong regulatory measures are unnecessary because database businesses have significant, market-based incentives to prevent the misuse of information.[96] Such opponents argue that companies will suffer reputational damage if they fail to notify victims of breaches and that the market will accordingly favor those companies that do.[97]

---

[93] CATE, *supra* note 22, at 26; *see also* GAO, *supra* note 18, at 22 (noting that many agencies currently rely on the users of the information to self-police).

[94] *See* Robert W. Hahn & Anne Layne-Farrar, *The Benefits and Costs of Online Privacy Legislation*, 54 ADMIN. L. REV. 85, 111 (2002) (discussing the position of "information flow advocates" that privacy regulations would quickly become obsolete as the Internet changes).

[95] *See, e.g.*, Solove & Hoofnagle, *supra* note 37, at 8 (noting the weaknesses of the self-regulatory rules adopted by database businesses).

[96] *See H.R. 3997 Hearing*, *supra* note 10, at 12 (statement of Rep. Jeb Hensarling (R-Tex.)) (noting that Federal Reserve Chairman Alan Greenspan believed that markets and self-regulation work in the database industry context); S. Kasim Razvi, *To What Extent Should State Legislatures Regulate Business Practices as a Means of Preventing Identity Theft?*, 15 ALB. L.J. SCI. & TECH. 639, 642 (2005); *see also* CATE, *supra* note 22, at 24–25 (giving examples of company privacy policies).

[97] *See* THOMAS M. LENARD & PAUL H. RUBIN, PROGRESS AND FREEDOM FOUNDATION,

Finally, opponents of strong regulatory oversight argue identity thieves are the true danger—after all, database businesses do not steal identities, identity thieves do. Regulatory regimes divert attention from the real culprits—identity thieves—when they target business practices.[98] The government, according to this view, best protects citizens by making government-issued identification harder to obtain,[99] facilitating corrections of erroneous public reports,[100] and increasing the pressure to report[101] and prosecute identity thefts.[102] This perception of the "real" problem is also evidenced by the behavior of database businesses. For example, after the security breach in 2005, ChoicePoint framed the crisis as "crimes committed against ChoicePoint" and "fraud against the company" when it began notifying consumers three months after their personal information was first put at risk.[103]

The position favoring little or no government regulation suffers from several shortcomings, particularly as applied to the database industry.

First, increased government enforcement of criminal measures and increased government enforcement of regulatory measures are not mutually exclusive—a crew may plug the leaks in a boat while continuing to bail out the water. Furthermore, this position overlooks the practical difficulty of prosecuting identity theft.[104] One study has found that identity thieves have less than a 1 in 700 chance of being caught.[105] While this dreary rate does not increase the culpability of database businesses, it should put legislators on notice that more effective solutions are needed. The difficulty of tracing a personal information disclosure to the culprit of the disclosure,

---

AN ECONOMIC ANALYSIS OF NOTIFICATION REQUIREMENTS FOR DATA SECURITY BREACHES 5 (Emory L. & Econ. Res. Paper No. 05-12, 2005), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=765845.

[98] *See id.*

[99] *See* CATE, *supra* note 22, at 65–66.

[100] *See id.*

[101] *See id.*

[102] *See H.R. 3997 Hearing*, *supra* note 10, at 118 (statement of Karl F. Kaufmann, Partner, Sidley Austin Brown & Wood L.L.P. on behalf of the U.S. Chamber of Commerce) (discussing prosecutions of identity theft); *see also* CATE, *supra* note 22, at 65–66. This reasoning previously led Congress to pass the Identity Theft Penalty Enhancement Act in 1998, which raised penalties for certain related crimes by two years and for terrorism using false identification by five years. *See* Pub. L. No. 105-318, 112 Stat. 3010 (1998) (codified as amended at 18 U.S.C. § 1028 (2000)); Gary M. Victor, *Identity Theft, Its Environment and Proposals for Change*, 18 LOY. CONSUMER L. REV. 273, 296–97 (2006).

[103] *See* Lazarus, *supra* note 27, at C1.

[104] *See* Brandon McKelvey, *Financial Institutions' Duty of Confidentiality to Keep Customer's Personal Information Secure from the Threat of Identity Theft*, 34 U.C. DAVIS L. REV. 1077, 1111–12 (2001) (stating that prosecutors have refrained from acting on identity theft cases due to the difficulty, time and expense associated with a successful prosecution).

[105] *See, e.g.*, Stephen Mihm, *Dumpster-Diving for Your Identity*, N.Y. TIMES MAG., Dec. 21, 2003, § 6 at 42 (using the estimate of an analyst from Gartner Inc., a research company that advises financial institutions on security issues, and noting that some in the industry estimate that "it's more on the level of one out of a thousand"); Press Release, Gartner Says Identity Theft Is Up Nearly 80 Percent, http://www.gartner.com /5_about/press_releases/pr21july2003a.jsp (last visited Apr. 17, 2007).

be it an identity thief or a lax security policy,[106] should also underscore the need for federal regulation.

Second, market-driven self-regulation does not correct for companies' countervailing incentives to withhold effective notice in the event of a breach. Under a regime of self-regulation, database businesses are not necessarily forced to shoulder the full costs of fraud.[107] The absence of a direct market relationship between consumers and database businesses assures that any reputational consequences associated with security breaches will be at best an imperfect deterrent. For example, a national survey of consumers found that in the aftermath of security breaches, 40% of consumers considered discontinuing their relationship with a compromised business and that 19% had already done so.[108] But for many database businesses, that market relationship is simply not present because database businesses do not rely on the consumer to obtain information or for their revenues.

These characteristics of the database industry make any form of self-regulation potentially problematic. It is true that a forced disclosure requirement for security breaches, as provided for in some of the early proposals considered by the 109th Congress, might motivate businesses to assume some of the costs associated with increasing security standards ex ante.[109] However, sanctions for breaches could also work to consumers' and regulators' disadvantages by creating an incentive not to disclose.

An examination of other industries that retain personal information enables further evaluation of whether market-based self-regulation would be effective.[110] Many credit-reporting companies like Experian and TransUnion do not engage directly with consumers, so they need not fear losing business if credit information is lost. These industries therefore lack the market-based incentives contemplated by self-regulation proponents.[111]

---

[106] *See* James P. Nehf, *Incomparability and the Passive Virtues of Ad Hoc Privacy Policy*, 76 U. COLO. L. REV. 1, 27–28 (2005) (discussing the difficulties individuals face in discovering when a breach of their information has occurred, as well as the problems associated with tracing a breach to a particular cause).

[107] Database businesses do not suffer the repercussions experienced by the victims of identity theft, which include high interest rates, fraudulent debts, and "endless nightly threatening calls from collection agencies." BOB SULLIVAN, YOUR EVIL TWIN: BEHIND THE IDENTITY THEFT EPIDEMIC 36 (2004).

[108] *See* PONEMON INSTITUTE, L.L.C., NATIONAL SURVEY ON DATA SECURITY BREACH NOTIFICATION 3, 10 (2005), *available at* http://www.whitecase.com/files/FileControl/863d572d-cde3-4e33-903c-37eaba537060/7483b893-e478-44a4-8fedf49aa917d8cf/Presentation/File/Security_Breach_Survey%5b1%5d. pdf [hereinafter PONEMON, NATIONAL SURVEY].

[109] *See* CAL. OFFICE OF PRIVACY PROT., RECOMMENDED PRACTICES ON NOTICE OF SECURITY BREACH INVOLVING PERSONAL INFORMATION 6 (2007), *available at* http://www.privacy.ca.gov/recommendations/secbreach.pdf (reporting that compliance with the law's mandates led to a synthesis of best practices, the adoption of new security measures and data retention policies, and the encouragement of some entities to exit the sensitive personal information business altogether).

[110] *See* Jeff Sovern, *The Jewel of Their Souls: Preventing Identity Theft Through Loss Allocation Rules*, 64 U. PITT. L. REV. 343, 362–65 (2003) (discussing the incentive problem in the credit industry).

[111] *See id.*; *see also Transcript of FTC Public Workshop, Information Flows: The Costs*

Indeed, Congress passed the FCRA in part to respond to this incentive problem.[112] The GLBA was similarly designed to motivate financial institutions to maintain secure databases.[113] Even financial institutions, which bear the brunt of identity theft losses because they are responsible for the monetary loss accumulated under the actual fraud,[114] are not necessarily incentivized to self-regulate because they can pass much of the cost associated with imperfect security measures onto consumers.[115]

Given these considerations, the current lack of interest in ex ante federal oversight and joint consultation with public officials appears to be a curious departure from ordinary government practice.[116] Unregulated database businesses, like other businesses, operate according to their economic self-interest; assigning complete discretion to them would likely lead to profit-motivated decisions not to notify the public even when the company has a clear obligation to do so.[117] Under the legislation proposed by the 109th Congress, affected businesses could almost always insist that with the knowledge available to them at the time, they could not have concluded there was a risk of harm.[118] This amounts to, as one journalist bluntly

---

and Benefits to Consumers and Businesses of the Collection and Use of Consumer Information 17 (June 18, 2003) (statement of Charles Morgan, CEO, Acxiom Corp.) (stating that this database business does not conduct business directly with customers).

[112] *See* 15 U.S.C. § 1681 (2000); SOLOVE, *supra* note 2, at 257.

[113] *See* H.R. REP. No. 106-434, at 245 (1999), *reprinted in* 1999 U.S.C.C.A.N. 1, 245.

[114] *See* 15 U.S.C. § 1643 (2000); Berg, *supra* note 105, at 6.

[115] *See* Bob Sullivan, *Instant Credit Means Instant Identity Theft*, MSNBC.COM, May 25, 2005, http://www.msnbc.msn. com/id/6762127/.

[116] *See Securing Electronic Personal Data*, *supra* note 76, at 26 (statement of Sen. Charles Schumer (D-N.Y.)) (discussing how potential database regulations would constrain data merchants in similar ways to regulations that apply to banks and credit bureaus); Satish M. Kini & James T. Shreve, *Notice Requirements: Common Themes and Differences in the Regulatory and Legislative Responses to Data Security Breaches*, 10 N.C. BANKING INST. 87, 89–90 (2006) (discussing obligations of financial institutions to safeguard personal information).

[117] See A. Michael Polinsky & Steven Shavell, *Mandatory Versus Voluntary Disclosure of Product Risks* 2 (John M. Olin Ctr. for Law, Econ., and Bus., Discussion Paper No. 564, 2006) (giving examples of how the business motive to suppress negative information in various industries is tempered by disclosure requirements and liability for non disclosure); *see also* Kenneth M. Dreifach, *Data Privacy, Web Security, and Attorney General Enforcement*, *in* PLI'S SIXTH ANNUAL INSTITUTE ON PRIVACY LAW: DATA PROTECTION–THE CONVERGENCE OF PRIVACY & SECURITY 401, 407 (PLI Pats., Copyrights, Trademarks, and Literary Prop., Course Handbook Series No. 828, 2005) (noting that ChoicePoint's stock plunged nearly 20% in the months following disclosure of the breach).

[118] See H.R. 3997, 109th Cong. §2(b)(1) (2005) (stating that an investigation is required

> [w]henever any consumer reporter determines or becomes aware of information that would reasonably indicate that a breach of data security has or may have occurred or is reasonably likely to be about to occur . . . the consumer reporter shall immediately conduct a reasonable investigation to . . . (A) assess the nature and scope of the potential breach; (B) identify the sensitive financial personal information involved; and (C) determine if the potential breach is reasonably likely to result in substantial harm or inconvenience to any consumer to whom the information relates

put it, a situation in which "the company that didn't protect your data in the first place gets to decide if a breach is significant enough that you need to know about it."[119] Victims of the breach would then be forced to rebut the company's internal conclusion.[120] The cost in time and money needed to overcome this steep asymmetrical information burden would probably prevent some consumers from obtaining a remedy.[121]

Considering the stake that the government and private sector actors have in the database industry, a more reasonable approach would be to ensure that at least one stakeholder, the government, is involved in the risk assessment process.[122] Database businesses may balk at this arrangement, although such an arrangement could ultimately be to their advantage: should harm occur after the government and the business make a joint decision not

---

and that only "if a consumer reporter determines after commencing an investigation . . . that a potential breach of data security may result in substantial harm or inconvenience to any consumer to whom the sensitive financial personal information involved in such potential breach relates," are they obligated to notify); *Addressing Measures to Enhance the Operation of the Fair Credit Reporting Act: Before S. Comm. on Banking, Housing, & Urban Affairs*, 106th Cong. 294–312 (2003) (testimony of Edmund M. Mierzwinski, Consumer Program Dir., U.S. Pub. Int. Res. Grp.), *available at* http://banking.senate.gov/index.cfm?Fuseaction=Hearings.Detail&HearingID=56; *see also* Paetkau et al., *supra* note 58, at 37 (noting similar interpretation issues with regards to the California law, including the meaning of "reasonably believe" and when exactly a company would be on inquiry notice of a breach that would raise a duty to investigate). The recent federal bills further complicate matters by specifically tying the notification requirement to the burden of establishing risk of identity theft. *See, e.g.*, H.R. 3997, 109th Cong. § 2(e) (2005) (requiring notice to consumers only when sensitive information "has been or is reasonably likely to be misused in a manner causing substantial harm or inconvenience against the consumers to whom such information relates to commit identity theft," or to be misused "in a manner causing substantial harm or inconvenience against consumers to whom such information relates to make fraudulent transactions on such consumers' financial accounts").

[119] *See* Loren Steffy, *Identity Theft Legislation Provides Easy Way Out*, HOUSTON CHRON., Nov. 9, 2005, in Bus., at 1.

[120] *See, e.g.*, S. 1326, 109th Cong. § 2(2)(B) (2005) (allowing the entity to not classify an incident as a breach of system security if it concludes "after conducting a reasonable investigation, that there is not a significant risk of identity theft to an individual"). An average victim and potential plaintiff would most likely find it exceedingly difficult to wade through the technical and managerial decisions that led to a failure to notify and meet the burden of proving that such an investigation was unreasonable. Of course, identity theft victims can only sue if the bill grants a private cause of action, which many did not. *See, e.g.*, S. 1408, 109th Cong. § 5 (2005); *Securing Consumers' Data Hearing*, *supra* note 51, at 34 (noting that the letter that ChoicePoint sends out, which recommends that people review their credit reports and continue to check them for unusual activity, is akin to saying, "[w]e've had a spill, now you go and protect yourself").

[121] *See, e.g.*, OSCAR H. GANDY, JR., THE PANOPTIC SORT: A POLITICAL ECONOMY OF PERSONAL INFORMATION 206–07 (1993) (discussing the practical difficulty arising from the number of "potential providers of privacy invasions," which cumulatively make identity theft too large and costly for an individual to deal with the implications alone).

[122] *See H.R. 3997 Hearing*, *supra* note 10, at 67 (statement of Julie Brill, Assistant Att'y Gen., Vermont) (stating that the "breached entity should be required to consult with law enforcement and receive an affirmative response that there is no risk of harm or misuse of personal information from the breach").

to notify the public, their partnership with the government could insulate the business from liability.[123]

Further, the nature of the industry provides part of the rationale for including the government. First, the "commodity" at issue makes database businesses unique; each piece of personal information has special significance for an actual individual. Second, due to the high transaction costs, consumers cannot bargain effectively in order to achieve the efficient outcome. For these reasons, government regulation that imposes liability on the database businesses, which are the least cost avoiders, is justified. Limiting the discretion of industry when it comes to avoiding notification of the public is therefore a precondition to effective oversight, because it ensures that the task of adjusting behavior falls on the party that can accomplish it at the lowest cost.[124]

B. Notification Decisions Premised on the Risk of Identity Theft Miscalculate that Risk and Ignore Non-pecuniary Harms

If Congress creates specific joint consultation procedures, it should expand the conditions that would trigger mandatory public notification. Recent legislative proposals that explicitly tie the decision to notify to the risk of identity theft are flawed because they mistakenly presume that the risk of identity theft can be rapidly and accurately measured, and overly discount (or ignore) non-pecuniary concerns.

A typical counterargument to this position is framed as follows: "[t]he benefits of a notification requirement consist of the reduction in the costs associated with identity theft."[125] To test this claim, a study found that the benefits of notification ranged from $7.50 to $10 per individual whose personal information was compromised.[126] Since only 2% of database breach victims actually experience fraud (and had very limited legal responsibility for the effects of that fraud),[127] a "well-designed notification program"

---

[123] The nuclear industry provides an extreme, but effective, example. Federal regulations require immediate notification of the Nuclear Regulatory Commission ("NRC") and other agencies if there is a radioactive leak, see 10 C.F.R. § 20.2202 (2000), but grant exemption from initiating further measures if the NRC deems the incident to be not hazardous. *See* 10 C.F.R. § 20.2301 (2000).

[124] *See generally* Robert S. Pindyck & Daniel L. Rubinfeld, MICROECONOMICS 621–55 (5th ed. 2001) (discussing the role of government in ensuring economic efficiency). Alternatively, regulations might set punitive damages far above expected liability. This would be preferable where, as here, the sources of harm are difficult to identify and victims are reluctant to bring suit. *See* ALAN J. AUERBACH & MARTIN FELDSTEIN, HANDBOOK OF PUBLIC ECONOMICS 1675–76 (2002).

[125] LENARD, *supra* note 97, at 12.

[126] *See* Posting of David Canton to eLegal Canton, Reporting Data Loss Debatable, http://www.canton.elegal.ca/archives/2005/09/reporting_data.html#trackbacks (Sept. 12, 2005, 7:50 AM) (discussing the study where those benefits were reported).

[127] *See* Liz Pulliam Weston, The Hysteria Over Identity Theft, MSN Money, http://articles.moneycentral.msn.com/Banking/FinancialPrivacy/TheHysteriaOverIdentityTheft.aspx?page=all (last visited Apr. 17, 2007). Other studies seem to confirm that many of the

would probably only eliminate about 10% to 20% of expected costs to consumers while increasing indirect costs both to consumers and to sectors of the economy that depend on the free flow of information, due to the likelihood of consumer overreaction to notifications.[128] The shortcoming of premising notification on a the traditional cost-benefit analysis is that this approach threatens to obscure problems that actually exist because the analysis omits important considerations[129]—indeed, one need not fall back on arguments about human dignity and personal autonomy to make this point, but rather simply look at what factors are not considered.[130]

First, the current method of dividing the total financial sum stolen due to identity theft by the number of people affected by fraud only (as of the date of the study) de-emphasizes the significant losses to those parties actually affected. Current estimates suggest that the approximately ten million identity theft victims a year[131] spend an average of 600 hours and $1,400 each fixing their credit.[132]

Second, identity theft is characterized by low reporting rates. The Federal Trade Commission recently found that a sizeable 62% of identity theft victims did not contact law enforcement; in those cases, no report was taken.[133] Underreporting thus skews the perception of the true breadth of the problem, as well as evaluation of the limited problem considered by the cost-benefit analysis.

Third, while losses from identity theft may be mitigated if discovered quickly,[134] many victims are unaware of the fraud until it is too late. For example, if a thief opens an entirely new account in an individual's name,[135]

---

recent breaches did not result in identity theft. *See, e.g.*, Steve Lohr, *Surging Losses, but Few Victims in Data Breaches*, N.Y. TIMES, Sept. 27, 2006, at G1; *see also H.R. 3997 Hearing*, *supra* note 10, at 128 (statement of ID Analytics Corp.); Press Release, ID Analytics, Inc., ID Analytics' First-Ever National Data Breach Analysis Shows the Rate of Misuse of Breached Identities May Be Lower than Anticipated (Dec. 8, 2005), *available at* http://www.idanalytics.com/news_and_events/20051208.htm.

[128] LENARD, *supra* note 97, at 12.

[129] *See H.R. 3997 Hearing*, *supra* note 10, at 44 (statement of Rep. Gary L. Ackerman (D-N.Y.)) (stating that protections for security breaches should be triggered without requiring a precondition of financial fraud).

[130] *See* Nehf, *supra* note 106, at 3 (stating that privacy advocates favor the rhetoric of fundamental rights but have trouble identifying and quantifying the costs of data proliferation and the benefits of data protection). Cost-benefit analyses can underestimate the benefits when the costs of a policy change are more easily quantified than its benefits and when the comparison must attempt to measure unquantifiable concepts, like the "value" of personal privacy. *See id.* at 30.

[131] *See Identity Theft Hearing*, *supra* note 19, at 8 (statement of Deborah Platt Majoras, Chairman, FTC).

[132] This figure does not even include remaining debt from account fraud that victims may have to pay themselves. IDENTITY THEFT RES. CTR., FACTS AND STATISTICS (2003), http://www.idtheftcenter.org/facts.shtml.

[133] FTC, Consumer Fraud and Identity Theft Complaint Data 14, Feb. 2007, *available at* http://www.consumer.gov/sentinel/pubs/Top10Fraud2006.pdf.

[134] *See* GAO, GAO-06-833T, PRIVACY: PREVENTING AND RESPONDING TO IMPROPER DISCLOSURES OF PERSONAL INFORMATION 12–13 (June 2006).

[135] *See* Anthony White, *The Recognition of a Negligence Cause of Action for Victims of*

and the victim receives neither physical nor electronic statements, the person would most likely be dependent on their free annual credit report to learn of the fraud.[136] If the criminal alters fundamental information about the victim (if, for instance the criminal behavior results in an arrest under the assumed identity), this misinformation may be reintroduced into the public records system during subsequent data gathering. Since the database businesses are private companies and have no legal obligation at present to share such information with each other, a victim's success in correcting one database business's report would not necessarily correct errors in other databases.[137] Such consumers would need to be extraordinary vigilant to correct information at every database company.

Additionally, these situations also require consideration of the special problem of potential harm. Under many state laws, for example, plaintiffs are often unable to sue if they did not sustain actual injury.[138] Yet data spills "are generally irreversible and their effects indeterminate . . . [C]ompanies themselves can never be sure when (if ever) a spill is truly 'cleaned up.'"[139] Many plaintiffs are thus likely to have an enduring risk of identity theft, although their ability to sue is probably not similarly enduring.[140] As this discussion suggests, the diverse possible motives of data breachers, the inability to know ex ante the extent of the information they will possess, and the difficulty in forecasting if and when the breachers will use that information, showcase the fallacy of forecasts that omit future harm as a variable.[141]

Identity Theft: Someone Stole My Identity, Now Who Is Going to Pay for It?, 88 MARQ. L. REV. 847, 852 (2005) (describing "true name fraud," where accounts opened fraudulently under addresses other than the victim's are not discovered until the victim makes a major purchase).

[136] *See Protecting Consumers' Data Hearing*, *supra* note 17, at 59 (statement of Derek Smith, CEO, ChoicePoint) (testifying that ChoicePoint only provides one year of credit monitoring). Members of Congress have expressed concern that this limited monitoring window could fail to protect consumers victimized farther in the future. *See id.*; *see also* John Leland & Tom Zeller, Jr., *Technology and Easy Credit Give Identity Thieves an Edge*, N.Y. TIMES, May 30, 2006, at A1 (giving an overview of the problem).

[137] *See generally* O'HARROW, *supra* note 19, at 37–50 (discussing the existence of many different databases); Erin M. Shoudt, *Identity Theft: Victims "Cry Out" for Reform*, 52 AM. U. L. REV. 339, 366–67 (2002) (documenting the difficulty of recognizing fraudulent activity).

[138] *See* Forbes v. Wells Fargo Bank, N.A., 420 F. Supp.2d 1018, 1020 (D. Minn. 2006) (holding that in an action for damages based on stolen personal information "the threat of future harm, not yet realized, will not satisfy the damage requirement" if no other harm is alleged or proved) (internal citations omitted); PROSSER AND KEETON ON TORTS 143 (William L. Prosser et al. eds, 5th ed. 2001) (stating that "the threat of future harm" is not enough to satisfy the elements of cause of action); SEVENTH ANNUAL INSTITUTE ON PRIVACY LAW: EVOLVING LAWS AND PRACTICES IN A SECURITY-DRIVEN WORLD 33 (Francoise Gilbert et al. eds., 2006) (giving examples of how courts have been reluctant to assess damages in this area).

[139] *See* Dreifach, *supra* note 117, at 406.

[140] For example, the FTC estimates that 32% of people find out about the identity theft over a year later. *See* FTC, IDENTITY THEFT VICTIM COMPLAINT DATA, fig. 8, http://www.ftc.gov/bcp/edu/microsites/idtheft/downloads/clearinghouse_2006.pdf.

[141] *See Data Security: The Discussion Draft of Data Protection Legislation; Hearing Before the Subcomm. on Commerce, Trade, and Consumer Protection of the H. Comm. on Energy and Commerce*, 109th Cong. 29, 29–30 (2005) [hereinafter *Data Security Hearing*]

Beyond actual or potential pecuniary harm, database breaches can have serious repercussions that have nothing to do with stealing identities or the associated economic loss, at least when using the traditional definition of economic loss. A security breach that discloses substantial health information could lead to embarrassment or reputational harm, with only subtle, largely incalculable effects on future economic well-being.[142] Putting people's personal information at risk could also have demoralizing emotional consequences, including feelings of defilement, shame, and heightened fears of financial insecurity.[143] The revelation of personal medical history may have a very indirect effect on one's financial identity, but the repercussions of it can be enormous.[144] Moreover, in at least one incident, poor database security methods led to the death of a disclosure victim.[145]

Recent bills have not adequately addressed these concerns. Under a regulatory regime like H.R. 4127, individuals entitled to notification by database businesses would still experience difficulty determining how much personal information had been taken because the legislation's proposed language does not require fact-finding by the database businesses to determine exactly what personal information has been compromised.[146] Further, unless faced with stringent requirements, database companies are unlikely to err on the side of generous disclosure; instead, recent events suggest they will tack close to the existing legal standard. For example, after the security breaches in 2005,[147] ChoicePoint sent generic letters to victims describing the types of information that might have been taken, despite the company's admitted ability to give a particularized assessment to each victim

---

(statement by Chris Hoofnagle, President and Executive Dir., Electronic Privacy Information Center) (discussing how "security breaches may be motivated by a number of crimes unrelated to attempted identity theft").

[142] *See* Mike Wereschagin, *Medical ID Theft Leads to Lengthy Recovery*, PITT. TRIB. REV., Oct. 24, 2006, *available at* http://www.pittsburghlive.com/x/pittsburghtrib/s_476326.html (discussing the harms attendant to medical identity theft).

[143] *See* IDENTITY THEFT RESEARCH CTR., IDENTITY THEFT—THE AFTERMATH 2003: A COMPREHENSIVE STUDY TO UNDERSTAND THE IMPACT OF IDENTITY THEFT ON KNOWN VICTIMS AS WELL AS RECOMMENDATIONS FOR REFORM 35–39 (2003), *available at* http://www.idtheftcenter.org/idaftermath.pdf (describing specific emotional responses of identity theft victims); *see also* McKelvey, *supra* note 104, at 1087 (discussing health problems).

[144] *See, e.g.*, Standards for Privacy of Individual Identifiable Health Information, 65 Fed. Reg. 82,462, 82,464 (Dec. 28, 2000) (codified at 45 C.F.R. pts. 160, 164) (acknowledging that medical privacy is a fundamental right different form "ordinary economic good[s]").

[145] *See* Remsburg v. Docusearch, Inc., 816 A.2d 1001 (N.H. 2003). (finding the database business liable for an improper disclosure of personal information that aided in the stalking and murder of a victim).

[146] *See* H.R. 4127, 109th Cong. (2005) (as introduced in H.R.) (requiring only "[a] description of the nature and types of information and accounts as appropriate that were, or are reasonably believed to have been, subject to the breach of data security").

[147] *See supra* Part I.

of this disclosure.[148] In defense of this approach, ChoicePoint argued that its notification strategy was actually for consumers' "own benefit."[149]

Basing decisions on narrowly defined conceptions of imminent economic harm shifts the costs of less quantifiable perils to consumers. Accordingly, legislation that fails to account for the risk of such harms falls far short of recognizing the true extent of the dangers imposed by data security breaches.[150]

Supporters of linking the consumer notification trigger to the risk of identity theft make the valid point that a more sensitive trigger could lead to over-notification.[151] They further argue that over-notification may desensitize victims (causing them to ignore all threats),[152] or may lead to needless business expenditures undertaken to placate consumers who seek unnecessary card and account replacements in response to false alarms.[153]

But this presumes that the total costs associated with these outlays from overnotification exceed the potential pitfalls of undernotification. While a lower threshold for notification may result in some short-term losses, these costs could be addressed through higher fees for consumers.[154] Whether these increases are significant enough to outweigh their ability to mitigate the extent of overall loss achieved by warning victims is an empirical question. An increase in mass notification also signals to potential identity thieves that the information is being monitored, and reduces the expected benefits from theft, potentially prompting some to refrain from committing the crime.

A preferable provision for triggering notification, which should deter fraud and better account for non-pecuniary risks, would state that notifica-

---

[148] *See Securing Consumers' Data Hearing*, *supra* note 51, at 34 (giving an example of the generalized letter notifying a customer of a security breach).

[149] *See Protecting Consumers' Data Hearing*, *supra* note 30, at 60 (statement of Derek Smith, CEO, ChoicePoint) (positing that ChoicePoint needs to be careful about disseminating information about what data was lost into the public domain).

[150] *Cf. Data Security Hearing*, *infra* note 153, at 12 (statement of Rep. Ed Markey (D-Mass.)) (stating that the identity theft test is "murky" and proposing a more expansive definition).

[151] *See Protecting Consumers' Data Hearing*, *supra* note 17, at 56 (testimony of Kurt P. Sanford, President and CEO, U.S. Lexis Nexis).

[152] *See id.*; *see also H.R. 3997 Hearing*, *supra* note 10, at 6 (statement of Rep. Barney Frank (D-Mass.)) (cynically noting the industry's hesitation to promptly notify disclosure victims "because of a very new-found concern for the capacity of people's mailboxes"); Press Release, Info. Tech. Ass'n. of Am., Statement of the Information Technology Association of America to the Virginia Joint Commission on Technology & Science on HB 2721, The Personal Information Privacy Act (Aug. 3, 2005) [hereinafter ITAA Press Release] (on file with author) ("[L]egislation should not result in excessive notifications that inure consumers to breach threats because they are unable to distinguish between defensive notifications, and notifications that alert to real risks of misappropriated data.").

[153] *See Enhancing Data Security: The Regulators' Perspective: Hearing Before the Subcomm. on Financial Inst. and Consumer Credit of the H. Comm. on Financial Services*, 109th Cong. 9 (2005) [hereinafter *Enhancing Data Security Hearing*] (statement of Lydia B. Parnes, Dir., Bureau of Consumer Protection, FTC).

[154] *See* Sullivan, *supra* note 115.

tion of security breaches is required unless there is *no* risk of harm.[155] Determining whether or not a risk of harm exists for these purposes would, again, be a product of the consultation between state or federal authorities and would occur before a business could proceed to notify the public at large.[156]

Current legislation also places undue emphasis on the ex post effects of database breaches. Regardless of whether businesses err on the side of under- or over-notification, that result does not remedy or prevent the breach. Assuming that breaches are a net negative to consumers and the economy, the question then becomes how best to incentivize the database industry to prevent the breach from occurring. While rigorous notification requirements might provide some incentive to these businesses to increase preventative measures because of reputational costs, a more direct solution might be civil or criminal liability. Providing a private right of action to individuals harmed by the breach would probably provide sufficient incentive for the businesses to implement the rigorous procedures necessary to make breaches an extreme rarity.[157]

C. The Definition of Personal Information Must Be Flexible

The architect of the technology used by companies like ChoicePoint and LexisNexis once observed, "I believe only 2% of the data has been collected and 98% hasn't even been thought up yet."[158] The potential for misuse of stored biometric and genetic information,[159] largely ignored in the current policy debates, highlights the shortcomings of premising notification requirements solely on the probability of financial loss by disclosure victims. Consideration of the untapped potential of the database industry should fundamentally transform how lawmakers view its regulation.

In other contexts, opponents of regulations have argued that regulators should use narrow definitions, because more expansive ones might burden society with restrictions on information that no one has a reasonable privacy interest in.[160] The use of biometric data, however, differs from tradi-

---

[155] *See H.R. 3997 Hearing*, *supra* note 10, at 67, 72 (statement of Julie Brill, Assistant Att'y Gen., Vermont on behalf of Nat'l Assoc. of Att'ys Gen.).

[156] See *supra* Part III.A.

[157] *See* Sovern, *supra* note 110, at 375, 384 (discussing how loss allocation optimally puts the burden on the business, which has more power to prevent the loss than does the consumer).

[158] Pounds, *supra* note 1.

[159] *See Securing Consumers' Data Hearing*, *supra* note 51, at 35 (statement of Daniel Solove, Professor, George Washington School of Law); *see also* Smith, *supra* note 21, at 13 (noting the expanding use of biometrics and forensic DNA analysis); John Schwartz, *For Sale in Iceland: A Nation's Genetic Code; Deal with Research Firm Highlights Conflicting Views of Progress, Privacy, and Ethics*, WASH. POST, Jan. 12, 1999, at A1 (recounting how Iceland "has decided to become the first country in the world to sell the rights to the entire population's genetic code to a biotechnology company").

[160] Fred H. Cate, *Principles for Protecting Privacy*, 22 CATO J. 54 (2002).

tional data in several important respects. While individuals can always obtain a new Social Security number, they cannot be issued new fingerprints, DNA, retinas, or faces. Some scholars predict that personal genomic databases will someday stockpile an individual's complete biological profile.[161] If so, this stored information may eliminate the use of identification cards, by making biometric information accessible for confirmatory identification.[162] Indeed, this prediction is already coming true on a limited scale: federal agencies are reissuing identification cards to their employees that contain biometric indicators,[163] and children in Glasgow, Scotland, scan their thumbprints to check out library books.[164] But as this practice is expanded to other applications, the fraudulent capture of these immutable characteristics becomes more likely.

Some states have already recognized the dangers of these practices and are crafting legislation in response.[165] At the very least, new federal legislation must be tailored to anticipate the widespread use of biometric information, or must at least give sufficient leeway to a regulating body (such as the FTC) to adapt its regulations to new data collection practices.[166]

While the United States has chosen a reactive, targeted approach to establishing fair information practices, many other nations have enacted comprehensive "data protection" laws, applicable to public and private sectors.[167] The existence of such comprehensive data protection laws demonstrates that it is possible to enact legislation designed to deter the occurrence of security breaches and mitigate the aftermath of security breaches when they occur. When enacting such any new data protection legislation, Congress should ensure that a broad definition of "personal information" is included in the legislation.

---

[161] *See* Nehf, *supra* note 25, at 27–30 (discussing how advances in genomics are creating DNA databases, palm and retinal scanners, and other advanced collections techniques are enabling the use of biometric signatures in everyday life); *see also* JAMES CANTON, THE EXTREME FUTURE 235 (2006) (envisioning that personal genomic databases, where DNA records are stored, will be in existence by 2020).

[162] *See* AMITAI ETZIONI, THE LIMITS OF PRIVACY 116 (1999).

[163] *See* Press Release, White House, Homeland Security Presidential Directive/Hspd-12 (Aug. 27, 2004), *available at* http://www.whitehouse.gov/news/releases/2004/08/20040827-8.html (describing the parameters of the new identification initiative); Griff Witte, *Unlocking Fingerprints: Plan for Enhanced Federal IDs Could Open Door to a Biometrics Boom*, WASH. POST, Aug. 28, 2006, at D1.

[164] *See* Martyn McLaughlin, *Civil Rights Row Over School Fingerprints: Pupils Asked for Thumb Image to Check Out Library Books*, HERALD (GLASGOW), Sept. 12, 2006, at 9.

[165] *See* Nehf, *supra* note 25, at 27 n.101; *see, e.g.*, Cal. Civ. Code § 56.17 (West 2002); N.J. Stat. Ann. § 10:5-45 (West 2002).

[166] *See* S. 768, 109th Cong. § 2 (2005) (giving the FTC discretion to include any type of information it deems appropriate); *see also* H.R. 958, 110th Cong. § 5 (2007).

[167] *See* Joel R. Reidenberg, *Setting Standards for Fair Information Practice in the U.S. Private Sector*, 80 IOWA L. REV. 497, 500 (1995); *see also* EPIC & PRIVACY INT'L, PRIVACY AND HUMAN RIGHTS (2004) (providing an analysis of privacy laws around the world).

III. CONCLUSION

How the new Congress will ultimately respond to this problem is difficult to predict.[168] At the very least, it appears that the new Congress has not forgotten about database protection. In January 2007, Senator Dianne Feinstein (D-Cal.) introduced legislation that requires consumer notification if personal information has been, or is reasonably believed to have been, accessed or acquired."[169] By not strictly tying notification to identity theft, but rather to unauthorized access and acquisition, the bill places the burden on the business to demonstrate that its notification procedures comply with the proposed legislation's requirements.[170] While the bill does provide a safe harbor if the database business conducts a risk assessment and concludes there is not a significant risk, the business still must alert the Secret Service of its intention to use the exemption and receive clearance.[171] Most importantly, however, the term "security breach" is broadly defined as any unauthorized access to personal information, which itself is defined to include biometric indicators.[172]

In March 2007, a Senate bill was introduced that would assess fines and criminal penalties for databrokers who willfully conceal security breaches.[173] Meanwhile, a bipartisan group of House members have reintroduced a bill resembling the problematic proposals of the 109th Congress in almost every respect, including name: the Data Accountability and Trust Act.[174] It again gives an exemption for businesses if they determine "that there is no reasonable risk of identity theft, fraud, or other unlawful conduct."[175] The country awaits the outcome of this renewed debate.

The practices of database businesses are of particular concern because the businesses' actions (or inactions) affect the well-being of parties who are not part of a transactional relationship. The currently favored cost-benefit analysis that links security breaches to identity theft obscures the central policy issue of what actual rights citizens should have over the whereabouts and release of their personal information.[176] It also leaves unanswered the question of whether database businesses should internalize some of the costs they produce when information is inadvertently disclosed. The analysis further fails to consider the very real possibility that this is a situation characterized by high transaction costs that arguably justifies shifting liabil-

---

[168] However, some recent undertakings support the view that there will be less rancorous this time around. *See* Jon Swartz, *Lawmakers get less combative on data-breach bills*, USA TODAY, Mar. 1, 2007, at 5B.

[169] S. 239, 110th Cong. § 2(a) (2007).

[170] *Id.* § 2(c).

[171] *Id.* § 3(b).

[172] *Id.*

[173] S. 495, 110th Cong. § 103 (2007).

[174] H.R. 958, 110th Cong. (2007).

[175] *Id.* § 3(f)(1).

[176] *See supra* Part III.

ity to database businesses, the least-cost avoiders.[177] This position also foolishly gives database businesses the discretion to notify the public as they wish, without requiring a full reckoning of the intangible repercussions of personal data misuse and accidental disclosure. Whereas database businesses might lose market value, consumers can lose their jobs,[178] their right to vote, and even their freedom.[179] The speed at which institutions have begun collecting new types of personal data,[180] particularly biometric indicators, raises these stakes. The idea that organizations can collect data on immutable identity indicators should fundamentally alter the basic policy calculus underlying these legislative proposals. The 110th Congress should take action with these concerns in mind.

—*Brendan St. Amant*[*]

[177] *See id.*

[178] *See* Information Protection and Security Act, S. 500, 109th Cong. § 2 (2005).

[179] *See* GAO, *supra* note 33, at 8.

[180] *See supra* Part III.C.

[*] J.D. and MPP Candidate, Harvard Law School and the Kennedy School of Government, Class of 2008; B.A., Cornell University, 2000. The author would like to express his appreciation to Professor John Palfrey for his assistance.

THE EVOLUTION OF THE DISPARATE IMPACT THEORY OF TITLE VII: A HYPOTHETICAL CASE STUDY

The evolution of the disparate impact theory of Title VII of the Civil Rights Act of 1964 has been remarkable. When disparate impact first developed into a cognizable claim for relief from employment discrimination, it appeared that employers would have to meet a stringent standard before legally partaking in activities that unduly impacted groups protected by Title VII. This allowed individuals protected under Title VII and disparately impacted by a facially neutral employment policy to seek relief from the courts even if they would be unable to prove intentional discrimination by the employer. If a plaintiff succeeded in showing that an employment policy created a disparate impact on a protected class, the plaintiff would prevail unless the employer could demonstrate that the policy arose out of business necessity and was related to job performance. Even if the employer met this burden, the plaintiff could still succeed by demonstrating that other selection devices with a less discriminatory impact would also serve the employer's legitimate interests. Disparate impact relief was intended to supplement the relief afforded by the disparate treatment theory, under which a plaintiff could recover damages if she could prove that an employer intentionally discriminated based on a characteristic protected under Title VII.

However, the effectiveness of the disparate impact theory has dissipated over the years, and today there is serious debate over how the current Supreme Court will view disparate impact protection. The Supreme Court has not decided a Title VII disparate impact case since the passage of the 1991 Civil Rights Act (the "Act"), at which time Congress attempted to strengthen the disparate impact protections that had been weakened by the Supreme Court's interpretation of the doctrine. Thus, it is difficult to predict the precise effect the Act will have on future litigation.

Part I of this Recent Development introduces the two ways in which an employee can challenge Title VII employment discrimination in court: the disparate treatment doctrine and the disparate impact doctrine. Part II examines the erosion of the disparate impact theory and predicts that in order to prove business necessity before the current Supreme Court, an employer will only have to demonstrate that an employment policy serves an important business purpose. Part III traces the evolution of the disparate impact doctrine through the example of a currently disputed employment practice. Finally, Part IV highlights how the recent Court has overstepped clear congressional intent and argues that it is essential for Congress to monitor future Court decisions and demand the faithful application of the laws it passes.

I. METHODS OF CHALLENGING AN EMPLOYMENT PRACTICE UNDER TITLE VII

A. A Brief Introduction to Disparate Treatment Theory

Disparate treatment claims arise under Title VII when employers allegedly treat applicants or employees differently because of their membership in a protected class. The central issue is whether the employer's actions were motivated by discriminatory intent, which may be proven by direct, indirect, or circumstantial evidence.[1] Further, the plaintiff may be able to prove intentional discrimination even if the employer had a mixed motive for the employment action.[2] Both individuals and classes can make disparate treatment claims.

B. A Brief Introduction to Disparate Impact Theory

Although it may be difficult to prove disparate treatment, the Supreme Court has held that Title VII does not require proof of intentional discrimination for a successful claim: plaintiffs can also recover by claiming disparate *impact*.[3] In order to establish a prima facie case of disparate impact, the Court originally required the plaintiff to demonstrate only that an employment policy impacted members of a group protected by Title VII in a discriminatory pattern.[4] Once the plaintiff had demonstrated a

---

[1] *See* McDonnell Douglas Corp. v. Green, 411 U.S. 792, 802 (1973) (holding that an employee had presented a prima facie case of racial discrimination by showing that he belonged to a racial minority, was rejected for a job for which the employer knew he was qualified, and that the job thereafter remained open); Dep't of Water & Power v. Manhart, 435 U.S. 702, 711 (1978) (holding that the plaintiffs established a prima facie case of discrimination by demonstrating that an employer charged all female employees higher retirement fund premiums than it charged to males); Hazelwood School Dist. v. United States, 433 U.S. 299, 308–12 (1977) (holding that in order to establish a prima facie case of discrimination when relying on statistical analysis, the plaintiff must first define the labor market and the relevant segment of the population qualified for the job at issue, and then establish that there is a statistically significant difference between the composition of the relevant segment of the employer's workforce and the composition of the qualified population in the defined labor market).

[2] *See* Civil Rights Act of 1991 § 108(a)–(b), 42 U.S.C. § 2000e-2(m), 5(g)(2)(B) (2000).

[3] *See* Griggs v. Duke Power Co., 401 U.S. 424, 431 (1971) ("[G]ood intent or absence of discriminatory intent does not redeem employment procedures or testing mechanisms that operate as 'built-in headwinds' for minority groups . . . . Congress directed the thrust of [Title VII] to the *consequences* of employment practices, not simply the motivation.").

[4] *See id.* at 430–31 (holding that plaintiffs established a prima facie case of disparate impact by demonstrating that an employer's high school diploma requirement and general intelligence test adversely impacted African Americans because 34% of white males, but only 12% of African Americans had completed high school, and 58% of white applicants passed the employment tests compared with only 6% of African Americans); *see also* Dothard v. Rawlinson, 433 U.S. 321, 329–30 (1977) (holding that plaintiffs established a prima facie case of disparate impact by showing that women 14 years of age or older composed 52.75% of the Alabama population and 36.89% of the labor force, but only 12.9% of

prima facie case, the burden shifted to the defendant to demonstrate the business necessity and job relatedness of the challenged employment policy.[5] If the employer could not meet this burden, the Court would find in favor of the plaintiff. Even if the employer did meet this burden, the plaintiff could then show that the employer's justification was a pretext by demonstrating that other tests or selection devices, without a similar discriminatory effect, could "also serve the employer's legitimate interest in efficient and trustworthy workmanship."[6] The period during which the aforementioned standards were law is described in this Recent Development as the "*Griggs-Albemarle-Dothard*" or "*Griggs*" era.

In the "modern era"[7] of disparate impact doctrine, the plaintiff will likely have a more difficult burden of establishing a prima facie case of disparate impact. The plaintiff will need to identify the actual hiring practice that resulted in the disparity.[8] Further, the plaintiff will have the difficult task of demonstrating how the policy has a disparate impact on a protected class.[9] Once the plaintiff has established a prima facie case, the employer will face a more lenient standard in its effort to rebut the prima facie case, enabling many employers to escape liability. While the "official" burden on the employer will be to demonstrate that the employment requirements are "job related for the position in question and consistent with business necessity,"[10] the Court is likely to interpret these requirements loosely such that the employer will probably only have to demonstrate that the policy serves an important business purpose to meet the requirement that the policy be "consistent with business necessity."[11]

II. The Evolution of the Disparate Impact Doctrine

A. *The Disparate Impact Theory in the* Griggs-Albemarle-Dothard *Era*

The Supreme Court first accepted disparate impact as a claim for relief in *Griggs v. Duke Power Co.*[12] In *Griggs,* the Court held that requiring an employee to have a high school education and to pass a screening test violated Title VII because it had a disparate impact on African Americans

---

the correctional counselor positions, and that the 5'2" requirement in conjunction with the 120 pound weight requirement would exclude 41.13% of the female population while excluding less than 1% of the male population).

[5] *See Griggs*, 401 U.S. at 431.

[6] Albemarle Paper Co. v. Moody, 422 U.S. 405, 425 (1975).

[7] For purposes of this discussion, the "modern era" is defined as the time since the passage of the Civil Rights Act of 1991.

[8] *See* Civil Rights Act of 1991 § 105(a), 42 U.S.C. § 2000e-2(k)(1)(A)(i) (2000) (stating that a complaining party must show an employer has a "particular employment practice" that causes a disparate impact).

[9] *See infra* text accompanying notes 61–66 and 72–77.

[10] 42 U.S.C. 2000e-2(k)(1)(A)(i).

[11] *See infra* text accompanying notes 81–88.

[12] 401 U.S. 424 (1971).

that the employer was unable to justify.[13] The Court held that even though the employer had no "'intention to discriminate against Negro employees.' . . . absence of discriminatory intent does not redeem employment procedures or testing mechanisms that operate as 'built-in headwinds' for minority groups and are unrelated to measuring job capability."[14] Thus, a primary question for the Court became what standard to apply to determine whether or not an employment policy produced a disparate impact on a group protected by Title VII.

In *Griggs*, the plaintiff made out a prima facie case of disparate impact by showing that the employer's hiring policy impacted African Americans in a discriminatory pattern.[15] The burden then shifted to the defendant to justify the disputed practice. The Court stated that "the touchstone [of defendant's burden] is business necessity."[16] This language was quite significant because it appeared that the Court was adopting the strict business necessity language previously used by lower courts.[17] If so, business necessity required the demonstration that the employment practice was necessary for the "safe and efficient operation of the business."[18] The Court further held that "if an employment practice which operates to exclude Negroes cannot be shown to be related to job performance, the practice is prohibited."[19] Therefore, within one opinion, the Court set out two distinct standards: business necessity and job relatedness.

Although the *Griggs* opinion was not wholly clear regarding the exact burden the employer would need to meet, it appeared that the Court would require the defendant to demonstrate both that the employment policy was job-related and justified by business necessity. If the employer failed to meet either of these burdens, it could not prevail.

---

[13] *See id.* at 430–32.

[14] *Id.* at 432 (citations omitted).

[15] *See id.* at 430–31 (stating that an employer's high school diploma requirement and general intelligence test adversely impacted African Americans because 34% of white males but only 12% of African Americans had completed high school, and 58% of white applicants passed the employment tests compared with only 6% of African Americans).

[16] *Id.* at 431.

[17] In a case involving a no-transfer policy that disparately impacted African Americans, the Tenth Circuit held that:

> [t]he remedial nature of Title VII requires the adoption of the business necessity test. If employers or unions could pursue, upon a showing of mere rationality, neutral policies which have the effect of perpetuating past discrimination, the value of the principles developed in [earlier] cases would be eroded. When a policy is demonstrated to have discriminatory effects, it can be justified only by a showing that it is necessary to the safe and efficient operation of the business.

Jones v. Lee Way Motor Freight Co., 431 F.2d 245, 249 (10th Cir. 1970). *Jones* followed the holding in *Local 189 United Papermakers & Paperworkers v. United States*, 416 F.2d 980, 989 (5th Cir. 1969), and was decided almost one year prior to *Griggs*.

[18] *Jones*, 431 F.2d at 249.

[19] *Griggs*, 401 U.S. at 431.

Ultimately, *Griggs* held that the employer at issue had failed to show that its policy requiring a high school diploma and a screening test had a "demonstrable relationship to successful performance of the jobs for which it was used."[20] According to the Court, the policy had been adopted "without meaningful study of [its] relationship to job-performance ability."[21] The employer even admitted that it had created the policy out of an undemonstrated belief that its requirements would improve the overall quality of the workforce.[22] Giving great deference to the Equal Opportunity Employment Commission ("EEOC") Guidelines, the Court held that the relationship of the employment practice to job performance must be shown through the use of "data demonstrating that the test is predictive of or significantly correlated with important elements of work behavior that comprise or are relevant to the jobs for which candidates are being evaluated."[23] Since the employer in *Griggs* had failed to meet this burden, the plaintiff prevailed.

The Court clarified the job-relatedness prong of the employer's burden in *Albemarle Paper Co. v. Moody*.[24] In *Albemarle*, African American employees at a large paper mill brought a class action suit alleging that the application of two general ability screening tests resulted in African Americans being disproportionately assigned to lower-paying, less-skilled job lines within the plant.[25] The Court held that the employer's screening mechanisms[26] violated Title VII because they had a disparate impact on African Americans and the employer could not prove a "manifest relationship to the employment in question."[27] Although the employer in *Albemarle* conducted a validation study on the efficacy of the employment screening tests, the Court found that the half-day study did not sufficiently demonstrate the required relationship between the policy and job performance.[28] The Court noted three reasons why the validation study did not demonstrate the necessary link: (1) its results were not statistically significant; (2) it compared test results with subjective supervisoral rankings elicited without proper care; and (3) it was conducted on an unrepresentative group of employees.[29] In reaching this conclusion, the Court continued to give "great deference" to the EEOC Guidelines.[30]

Thus, the *Albemarle* Court confirmed the exacting burden an employer asserting a Title VII defense would need to meet: testing policies that have a

---

[20] *Id.*

[21] *Id.*

[22] *See id.*

[23] *Id.* at 434 n.9 (quoting 29 C.F.R. § 1607.4 (1970)).

[24] 422 U.S. 405 (1975).

[25] *Id.* at 427–29.

[26] The screening mechanism of most concern in *Albemarle* was the employer's use of employment tests that were racially discriminatory in effect, though not necessarily in intent. *See id.* at 425–28.

[27] *Id.* at 425.

[28] *See id.* at 429–31.

[29] *See id.* at 430–35.

[30] *See id.* at 430–31.

discriminatory impact on a group protected by Title VII would be "impermissible unless shown, by professionally acceptable methods, to be 'predictive of or significantly correlated with important elements of work behavior which comprise or are relevant to the job or jobs for which candidates are being evaluated.'"[31] To accomplish this, the EEOC Guidelines recognized three methods that an employer could use to validate an employment test: criterion validation, content validation, and construct validation.[32]

The *Albemarle* Court then articulated a third step used to evaluate claims of disparate impact. If the defendant were to meet its burden of proof in demonstrating "job relatedness," the plaintiff could still prove pretext by showing "that other tests or selection devices, without a similarly undesirable racial effect, would also serve the employer's legitimate interest in 'efficient and trustworthy workmanship.'"[33]

Thus, after *Griggs* and *Albemarle* there were two distinct scenarios in which a plaintiff would win a case using a disparate impact challenge. The plaintiff would win if she could make out a prima facie case of disparate impact and the defendant could not demonstrate both the job relatedness and the business necessity of the employment policy. Second, even if the employer did meet this burden, the plaintiff would still win if she could show that other employment policies would similarly serve the employer's interests without causing an undesirable racial effect.

In *Dothard v. Rawlinson*,[34] the Court clarified the plaintiff's burden in establishing a prima facie case and continued to interpret the disparate impact theory as placing a demanding burden on the employer. In *Dothard*, the lower court found that women fourteen years of age or older compose 52.75% of the Alabama population and 36.89% of the labor force, but only 12.9% of the correctional counselor positions.[35] Furthermore, the court

---

[31] *Id.* at 431 (quoting 29 C.F.R. § 1607.4(c) (1970)). "Professionally acceptable methods" are methods such as those described in the Standards for Educational and Psychological Tests prepared by a joint committee of the American Psychological Association, the American Educational Research Association, and the National Council on Measurement in Education, and standard textbooks and journals in the field of personnel selection. *See* 29 C.F.R. § 1607.5 (2005).

[32] *See* 29 C.F.R. § 1607.5A. Criterion validation requires the selection of some criterion of performance on the job that includes the most important aspects of the job. Then, the person performing the validation seeks to correlate good performance according to that criterion with good performance on the test or other selection device. *See* 29 C.F.R. § 1607.14B. Content validation requires "that the behavior(s) demonstrated in the selection procedure are a representative sample of the behavior(s) of the job in question or that the selection procedure provides a representative sample of the work product of the job." See 29 C.F.R. § 1607.14C. Construct validation looks to abilities that are generally useful in a variety of different work behaviors and attempts to replicate those abilities in a test. *See* 29 C.F.R. § 1607.14D.

[33] *Albemarle*, 422 U.S. at 425 (quoting McDonnell Douglas Corp. v. Green, 411 U.S. 792, 801 (1973)).

[34] 433 U.S. 321 (1977).

[35] *See id.* at 329–30.

showed that the combination of the employer's height and weight requirements would exclude 41.13% of the female population but less than 1% of the male population in the United States.[36] In finding that this evidence met the plaintiffs' burden of establishing a prima facie case, the Court clarified that "the plaintiffs in such a case as this are not required to exhaust every possible source of evidence, if the evidence actually presented on its face conspicuously demonstrates a job requirement's grossly discriminatory impact."[37] The Court concluded that it would be up to the employer to "adduce countervailing evidence of his own" if the employer believes that the plaintiff's statistics are deficient.[38]

The *Dothard* majority further held that the employer failed to meet its burden of demonstrating either job relatedness or business necessity. The Court found that the employer had not demonstrated job relatedness because it had not established a satisfactory correlation between the height and weight requirements and the requisite amount of strength believed to be essential to good job performance.[39] Moreover, the Court held that regardless of whether the defendants had met their burden of demonstrating job relatedness, the employer had not demonstrated the other necessary prong: business necessity.[40]

The *Dothard* Court adopted the holding of the Tenth Circuit in *Jones v. Lee Way Motor Freight Co.*,[41] stating that "a discriminatory employment practice must be shown to be necessary to safe and efficient job performance to survive a Title VII challenge."[42] In *Dothard*, because the employer "failed to offer evidence of any kind in specific justification" of the necessity of the height and weight requirements as opposed to a strength requirement, the employer failed to demonstrate that the policy was necessary for safe and efficient job performance.[43]

The *Dothard* majority noted that in enacting Title VII, "Congress required 'the removal of artificial, arbitrary, and unnecessary barriers to employment.'"[44] Such removal required that an employer be held to an exacting standard. In the *Griggs-Albemarle-Dothard* era, an employer could not legally create neutral employment policies with a disparate impact on a protected group unless it could establish that the policy was both job-related and necessary to the business. Even then, an employer could still be found to have violated Title VII if the plaintiff could show that the

---

[36] *See id.* The employer had a policy against hiring individuals below 5'2" tall, as well as individuals weighing less than 120 pounds. *See id.*

[37] *Id.* at 331.

[38] *Id.*

[39] *Dothard*, 433 U.S. at 331.

[40] *See id.* at 331–32.

[41] 431 F.2d 245, 249 (10th Cir. 1970).

[42] *Dothard*, 433 U.S. at 329–30 n.14.

[43] *Id.* at 331.

[44] *Id.* at 328 (quoting Griggs v. Duke Power Co., 401 U.S. 424, 431 (1971)).

employer could have used other tests with less discriminatory impact to achieve the same purpose.

Further, although the Court allowed exceptions for business necessity and job relatedness, several circuit courts held that some job-related qualities were inherently discriminatory and therefore fell under disparate treatment. In *Gerdom v. Continental Airlines*,[45] the Ninth Circuit held that an airline's policy of requiring flight hostesses to comply with strict weight requirements as a condition of employment could not be justified by passengers' tastes for thin women, since the justification was discriminatory on its face.[46]

Similarly, in *Diaz v. Pan Am Airways*,[47] another case involving disparate treatment, the Fifth Circuit stated that "customer preference may be taken into account only when it is based on the company's inability to perform the primary function or service it offers."[48] Accordingly, the court held that because the employer had not demonstrated that "all or substantially all men have been shown to be inadequate" as flight attendants the employer's policy of excluding all men violated Title VII.[49]

Although the latter two cases were decided under the disparate treatment doctrine, they suggest that the lower courts would have been equally unhappy with an employer's rationale of customers' discriminatory tastes to rebut a plaintiff's prima face case under the disparate impact doctrine.

B. Washington v. Davis *and* New York City Transit Authority v. Beazer*: A Bridge Between the* Griggs *Era and the Modern Era*

One of the first signs that the disparate impact standard might not be as strict as originally believed came from *Washington v. Davis*.[50] Although this case was decided a year before *Dothard,* some of its analysis and dicta became fertile ground for jurists seeking to undercut the reach of Title VII.[51] While *Davis* interpreted a statute that was only applicable to the District of Columbia, the Court went out of its way to address matters significant to Title VII. Specifically, the Court focused on the methods of validation necessary to properly demonstrate job relatedness.[52]

The Supreme Court held in *Davis* that the Washington, D.C., Police Department successfully demonstrated the job relatedness of the civil ser-

---

[45] 692 F.2d 602 (9th Cir. 1982).

[46] *Id.* at 608–09.

[47] 442 F.2d 385 (5th Cir. 1971).

[48] *Id.* at 389.

[49] *Id.* at 388 (internal quotation marks omitted).

[50] 426 U.S. 229 (1976).

[51] *See* Robert Belton, *Title VII at Forty: A Brief Look at the Birth, Death, and Resurrection of the Disparate Impact Theory of Discrimination*, 22 HOFSTRA LAB. & EMP. L.J. 431, 463–64 (2005).

[52] The Court did not address the business necessity prong of Title VII.

vice test that it used to screen applicants for jobs in the Department.[53] African American applicants scored significantly lower on this examination than Caucasian applicants, leading the plaintiffs to argue that the test had a disparate impact on African American applicants.[54] The plaintiffs' position appeared to be strengthened when the Court recognized that the screening test had not been shown to have any correlation with an applicant's ability to perform the job of a police officer; rather, it had been shown only to be correlated with how an applicant would perform in the police academy.[55] Such a test would seemingly have violated the EEOC Guidelines mandating that any test must be "predictive of or significantly correlated with important elements of work behavior which comprise or are relevant to the job or jobs for which candidates are being evaluated," which is the language that the Court accepted in both *Griggs* and *Albemarle*.[56] Yet, the Court held in *Davis* that the test could be validated by demonstrating its positive relationship to performance in the policy training academy, and that such a ruling was a perfectly "sensible construction of the job-relatedness requirement."[57]

If the Court had wanted to hold the D.C. Police Department to an exacting validation requirement, as it had done in prior disparate impact cases under Title VII,[58] the Court could have required that the Department meet the EEOC Guidelines for criterion validation by designing a screening test that could be correlated with an individual's performance as a police officer. Instead, the *Davis* Court decided that the test, although not proven to have any firm link to performance as a police officer, met the job relatedness requirement.[59] This holding suggested that the strict job relatedness standard that the Court had required in earlier Title VII cases might no longer be necessary. Accordingly, if the Court were to take the same approach in a purely Title VII context,[60] it would severely erode the job relatedness requirement.

---

[53] *See Davis*, 426 U.S. at 229.

[54] *Id.* at 235.

[55] *Id.* at 250–51 & n.17.

[56] *See* Griggs v. Duke Power Co., 401 U.S. 424, 433–34 & n.9 (1971) (citing 29 C.F.R. § 1607.4(c) (1970)); Albemarle Paper Co. v. Moody, 422 U.S. 405, 431 (1975) (citing 29 C.F.R. § 1607.4(c) (1970)).

[57] *Davis*, 426 U.S. at 250–51.

[58] *See, e.g.*, *Albemarle Paper Co.*, 422 U.S. at 431–36 (holding that Albemarle's validation study did not meet the validation requirement because it was not properly designed to demonstrate a link between performance on the test and performance on the job).

[59] *Davis*, 426 U.S. at 229.

[60] It is reasonable to believe that the Court would take a similar approach in a Title VII context because the Court in *Davis* demonstrated its liberal interpretation of the EEOC statute mandating that the test must be "predictive of or significantly correlated with important elements of work behavior which comprise or are relevant to the job or jobs for which candidates are being evaluated," and that interpretation would affect Title VII cases as well. *See* 29 C.F.R. § 1607.4(c) (1970).

The Supreme Court further clarified its disparate impact jurisprudence in *New York Transit Authority v. Beazer*,[61] a Title VII case. The Court increased the plaintiff's burden of establishing a prima facie case while lowering the defendant's burden of demonstrating business necessity.[62]

In *Beazer*, the New York City Transit Authority refused to employ recovering heroin addicts currently being treated with the drug methadone.[63] Although the plaintiffs believed they had compelling statistical evidence of discriminatory employment practices, the Court disagreed. The plaintiffs were able to demonstrate that 81% of the employees referred to the Transit Authority for suspected violations of its drug policy were African American or Hispanic. Yet, the Court noted that these statistics indicated little about the racial composition of the employees suspected of or dismissed for using methadone.[64] The Court further found that the plaintiffs' statistics demonstrating that 62% to 65% of all methadone users were African American or Hispanic also failed to demonstrate disparate impact because the plaintiffs could not demonstrate how many of these individuals actually worked for the Transit Authority.[65]

The Court was clearly troubled that the plaintiffs were not able to demonstrate that actual Transit Authority employees suffered a disparate impact from the employment policy.[66] However, upon the facts available to the plaintiffs in *Beazer*, such a demonstration seemed very difficult to make and placed a much greater initial burden on the plaintiffs than the Court had required in the past. While such a burden may certainly be reasonable, it was nonetheless a seemingly large change in Court policy from the *Griggs* era.

The Court then addressed the next step in the disparate impact inquiry—the business necessity prong—suggesting that even if the plaintiffs had established a prima facie case of disparate impact, the employer still would have succeeded in showing the business necessity of its policy by demonstrating that it served "the general objectives of safety and efficiency."[67] The Court believed that the employer met this burden by demonstrating that safety and efficiency required the exclusion of persons taking methadone from all safety-sensitive positions, and that those goals were "significantly served by . . . [the employment policy] as it applies to

---

[61] 440 U.S. 568 (1979).

[62] *Id.* at 585.

[63] Methadone is a well-tested, commonly used medication that is safe and effective for the treatment of narcotic withdrawal and dependence. By occupying the dopamine receptors of the brain, thereby reducing the cravings associated with opiate use, methadone is a stabilizing factor that helps addicts to discontinue their heroin use. *See* ERIN BROEKHUYSEN, DRUG POLICY INFO. CLEARINGHOUSE, METHADONE 1 (Apr. 2000), http://www.whitehousedrugpolicy.gov/publications/pdf/ncj175678.pdf.

[64] *See Beazer*, 440 U.S. at 584–85.

[65] *Id.* at 585–86.

[66] *See id.*

[67] *Id.* at 592.

all methadone users, including those who are seeking employment in non-safety-sensitive positions."[68]

Thus, the Court suggested that it would allow the employer to exclude methadone users from non-safety-sensitive positions because it helped the Transit Authority to exclude methadone users from safety-related positions, while at the same time addressing cost and administrative concerns.[69] While it is possible that the policy served the "general objectives of safety and efficiency,"[70] it is doubtful that it was "necessary for the safe and efficient operation of the business," as the Court required employers to demonstrate during the *Griggs-Dothard-Albemarle* era.[71] Regardless, *Beazer* opened the door for employers to develop policies that were over-inclusive in their scope and discriminatory in their effect. The days of narrow tailoring of employment policies required in earlier cases appeared to be over.

During this transition period, strict business necessity no longer appeared to be the standard under which the Court would examine an employer's policy. Rather, the Court was more willing than ever to defer to the employer's expertise when examining facially neutral employment policies.

C. The Meaning of the Disparate Impact Theory Today

The disparate impact doctrine suffered another setback in *Wards Cove Packing Co. v. Antonio*.[72] *Wards Cove* involved minority fishermen who believed the hiring practices of the company for which they worked systematically excluded them from coveted managerial jobs.[73] The Court held that the plaintiffs had not established a prima facie case of disparate impact, even though minorities were severely underrepresented within the jobs at issue.[74] Although non-whites comprised only 17% of the new hires for medical jobs and 15% of the new hires for office worker positions, the Court recognized that the plaintiffs had not established the number of qualified non-whites that actually applied for these positions.[75] The Court held that in order to establish a prima facie case of disparate impact, the plaintiffs would have the burden of identifying the specific employment practice thought to be responsible for the disparate impact.[76] Once identi-

---

[68] *Id.* at 587 n.31.

[69] *Beazer*, 440 U.S. at 590–92 & n.33.

[70] *Id.* at 592.

[71] *See* Albemarle Paper Co. v. Moody, 422 U.S. 405, 411 (1975); Griggs v. Power Co., 401 U.S. 424 (1971); Dothard v. Rawlinson, 433 U.S. 321 (1977).

[72] 490 U.S. 642 (1989).

[73] *See id.* at 646–48.

[74] *See id.* at 652–55.

[75] *Id.* at 652.

[76] The Civil Rights Act of 1991 provides that if the elements of an employer's decision-making process cannot be analyzed separately, the entire process can be treated as a single employment practice. *See* 42 U.S.C. § 2000e-2(k)(1)(B)(i) (2000).

fied, plaintiffs then had to demonstrate specifically how the employment practice disparately impacted a group protected by Title VII.[77]

If the plaintiffs could meet their burden of establishing a prima facie case of disparate impact on remand, the Court held that the burden would shift to the employer to produce evidence from which a reasonable inference could be drawn that the employer had a justification for its business practice:

> The touchstone of this inquiry is a reasoned review of the employer's justification for his use of the challenged practice . . . there is no requirement that the challenged practice be "essential" or "indispensable" to the employer's business for it to pass muster: this degree of scrutiny would be almost impossible for most employers to meet.[78]

With this holding, the Court adopted a "reasoned review" standard of an employer's practices, marking a clear departure from the *Griggs* era when no disparate impact would be accepted unless the employer actually demonstrated that the policy was necessary to the business.[79] Assuming that most employers could design an employment policy that would pass a reasonableness review, this new standard would be problematic for those employees disparately affected by such policies.

The Court then dealt another blow to the *Wards Cove* plaintiffs. While stating that the plaintiffs might still prevail by proving that a different employment practice with less disparate impact could have been used by their employer, the Court held that courts should defer to the employer's judgment about the effectiveness of these alternative practices.[80] Thus, after *Wards Cove*, if an employer claimed that the challenged practice was more effective than the plaintiffs' alternative, the courts would defer to the employer's expertise unless the claim appeared wholly meritless. As demonstrated, *Wards Cove* dealt a significant setback to Title VII plaintiffs.

Immediately following *Wards Cove*, in reaction to the decision, Congress passed the Civil Rights Act of 1991.[81] The Act codified *Wards Cove*'s statistical requirements for establishing a prima facie case of disparate impact.[82] Accordingly, plaintiffs continue to bear the reasonable burden of demonstrating how each particular employment practice causes a disparate impact on members of a protected class.[83] Nevertheless, Congress

---

[77] *See Wards Cove Packing Co.*, 490 U.S. at 656–57.

[78] *Id.* at 659.

[79] *See, e.g.*, Griggs v. Duke Power Co., 401 U.S. 424 (1971); Albemarle Paper Co. v. Moody, 422 U.S. 405 (1975); Dothard v. Rawlinson, 433 U.S. 321 (1977).

[80] *See Wards Cove Packing Co.*, 490 U.S. at 660–61.

[81] Pub. L. No. 102-166, 105 Stat. 1071 (1991) (codified in scattered sections of 42 U.S.C.).

[82] *See id.* § 105(a), 42 U.S.C. § 2000e-2(k)(1)(A)(i) (2000).

[83] *See supra* text accompanying notes 72–77.

expressly rejected *Wards Cove*'s "reasoned review" standard for evaluating employment policies allegedly responsible for a disparate impact. Specifically, the Act states in its preamble that it is meant "to codify the concepts of 'business necessity' and 'job related' enunciated by the Supreme Court in *Griggs v. Duke Power Co.* and in the other Supreme Court decisions prior to *Wards Cove Packing Co. v. Antonio*."[84] The Act requires that defendants "demonstrate that the challenged practice is job related for the position in question and *consistent with business necessity*."[85] They must do so with both the burden of production and persuasion.[86] Accordingly, Congress expressly overruled aspects of *Wards Cove* by prohibiting an employer from creating a disparate impact on a protected class of workers unless the employer could prove that the disparate impact was caused by a policy that was both job-related and consistent with business necessity.

However, congressional ambiguity has complicated the interpretation of "consistent with business necessity." Because Congress did not precisely state the standard it sought to adopt by referring to pre-*Wards Cove* decisions, the Act gave jurists suspicious of the disparate impact theory and weary of applying a strict business necessity standard room to maneuver in creating new variations of the business justification doctrine. For instance, once an employer demonstrates job relatedness,[87] it is possible (and likely consistent with the Act as drafted, although inconsistent with its intent) that defendants will only have to show that a challenged practice does not entirely preclude a protected class from employment.[88]

Congressional ambiguity regarding the pretext stage of disparate impact litigation adds to the uncertain outlook of the doctrine. Recall that the Court in *Wards Cove* held that courts were to defer to the employer's judgment about the effectiveness of different practices.[89] In the Act, Congress appeared to reject this holding. Congress required that the plaintiffs' ability to demonstrate pretext—namely by demonstrating an alternative employment practice with less of an adverse impact—meet the standards existing before *Wards Cove*.[90]

However, some academic commentators note that the Court in *Wards Cove* was quoting from *Albemarle* in holding that plaintiffs could prevail by proving that an alternative employment practice with less disparate

---

[84] Civil Rights Act of 1991 § 3(2), 42 U.S.C. § 1981 note.

[85] *Id.* § 105(a), 42 U.S.C. § 2000e-2(k)(1)(A)(i) (emphasis added).

[86] *Id.*

[87] The Act also failed to require validation. However, because Congress intended to restore disparate impact protection to it pre-*Wards Cove* status, the Court should require validation.

[88] *See* GEORGE A. RUTHERGLEN & JOHN J. DONOHUE III, EMPLOYMENT DISCRIMINATION, LAW AND THEORY, 219 (1st ed. 2005).

[89] *See* Wards Cove Packing Co. v. Antonio, 490 U.S. 642, 661 (1989).

[90] Civil Rights Act of 1991 § 3(2), 42 U.S.C. § 1981 note.

impact could have been used instead of the employer's practice.[91] The *Wards Cove* decision only added that courts should defer to the employer's judgment about the effectiveness of different employment practices.[92] *Albemarle*, like most disparate impact cases, did not discuss at any length the requirements for the proof of pretext.[93] Hence, because the Civil Rights Act of 1991 did not overrule *Albemarle*, the effect of the Act on this stage of the disparate impact analysis remains unclear.

D. Predicting the Position of the Current Court

Accordingly, as the Supreme Court has yet to decide a Title VII disparate impact case in the modern era, it is difficult to predict the precise effect the Act will have on future litigation. However, while the Court has not examined the meaning of "consistent with business necessity" in the Title VII context since *Wards Cove*, it has examined the phrase in the context of the Americans with Disabilities Act of 1990 ("ADA").[94] Such cases suggest that the Court prefers an interpretation of "consistent with business necessity" that is closer to the Court's reasoning in *Davis, Beazer,* and *Wards Cove* than the disparate impact cases decided in the *Griggs-Albemarle-Dothard* era.[95]

In *Chevron U.S.A. v. Echazabal*,[96] an employer refused to hire an applicant with liver damage caused by hepatitis C, which the employer's doctors said would be aggravated by exposure to toxins at the employer's refinery.[97] The Supreme Court held that the ADA "creates an affirmative defense for action under a qualification standard shown to be job-related for the position in question and . . . consistent with business necessity."[98] Thus, an employment practice requiring that an individual not pose a direct threat to himself would be consistent with business necessity if the employer determined that the individual could not perform the job safely with reasonable accommodation.[99] Hence, the Court suggested that paternalism is consistent with business necessity, as is reducing the chances of incurring liability due to OSHA violations.[100] Such a standard appears, at least upon initial examination, to be lenient towards the employer.[101]

---

[91] *See* RUTHERGLEN & DONOHUE, *supra* note 88, at 219.

[92] *Id.*

[93] According to Professors Rutherglen and Donohue, this is because most disparate impact cases are resolved during one of the first two stages of proof. *Id.* at 219–20.

[94] 42 U.S.C. §§ 12101–12213 (2000).

[95] *See* RUTHERGLEN & DONOHUE, *supra* note 88, at 219.

[96] 536 U.S. 73 (2002).

[97] *Id.* at 76.

[98] *Id.* at 78 (quoting 42 U.S.C. §§ 12112(b)(6), 12113(a) (2000)) (internal quotation marks omitted).

[99] *Id.* at 78, 87.

[100] *Id.* at 84–86.

[101] However, under the facts of *Echazabal*, it can be argued that the employer's motivation was to save the potential employees from danger. Therefore, it is arguable that a mate-

The question of the status of the Title VII disparate impact doctrine must wait for the Court to grant certiorari in another Title VII disparate impact case. The Supreme Court has yet to interpret the "job related for the position in question and consistent with business necessity" standard adopted by the Act in 1991. Additional cases provide insight into the jurisprudence of the Court, although not specifically its opinion on business necessity in the context of Title VII.[102] Yet, as suggested, the Court made an ideological shift away from its *Griggs* era disparate impact jurisprudence even before the appointment of Chief Justice Roberts in 2005 and Justice Alito in 2006.[103] A review of the new Justices' voting records and writings regarding disparate impact cases prior to their appointments to the Supreme Court suggests that both would view the theory of disparate impact in much the same manner as did the Justices who they replaced.[104]

This analysis advances the hypothesis that the current state of the disparate impact doctrine in general, and the business necessity theory in particular, is closer to the holdings of *Davis, Beazer, and Wards Cove* than the disparate impact cases decided in the *Griggs-Albemarle-Dothard* era. Moreover, the lower courts have given the Supreme Court little ground to stand upon in interpreting the meaning of the 1991 Act. Most courts of appeals that have applied the Act's standard to a Title VII challenge have done so with very little analysis. For example, in *Fitzpatrick v. City of At-*

---

rial distinction could be drawn when analyzing future cases that do not involve such physical danger.

[102] *See* Smith v. City of Jackson, 544 U.S. 228, 241 (2005) (holding that even though older workers suffered from significant disparate impact under the Age Discrimination in Employment Act of 1967, the complaint did not identify any specific test, requirement, or practice within a pay plan that had an adverse impact on older workers); U.S. Airways v. Barnett, 535 U.S. 391, 406 (2002) (holding that the ADA does not require an employer to assign an employee, disabled on the job, to a position in violation of an established seniority system); Se. Cmty. Coll. v. Davis, 442 U.S. 397, 405 (1979) (holding that the Rehabilitation Act of 1973 did not require an educational institution to lower its standards or abstain from requiring reasonable physical qualifications for admission to a clinical training program).

[103] *See supra* Part II.B–C. Chief Justice John Roberts replaced Chief Justice William Rehnquist. Justice Samuel Alito replaced Justice Sandra Day O'Connor.

[104] *See, e.g.*, Michael Foreman, Director of Employment Discrimination Project, Address at Harvard Law School (Mar. 2, 2006) (stating that Justice Alito decided in favor of the employer during his time on the Court of Appeals in line with how the other conservative Justices on the Supreme Court would have been thought to vote, and that Justice Roberts, while having authored few opinions on employment discrimination while a Judge on the Court of Appeals for the District of Columbia, is expected to have a conservative interpretation of Title VII and the disparate impact doctrine); Sheridan v. E. I. Dupont de Nemours & Co., 100 F.3d 1061, 1078 (3d Cir. 1996) (Alito, J., dissenting) (rejecting a majority holding that in the context of employment discrimination, "when the plaintiff has made out a prima facie case and has offered enough evidence to support a finding that the explanation was pretextual, a defense motion for summary judgment or judgment as a matter of law must always be denied"); NATIONAL PARTNERSHIP FOR WOMEN AND FAMILIES, JOHN ROBERTS' RECORD ON ISSUES IMPORTANT TO WOMEN AND FAMILIES 8–9 (2005), *available at* http://www.nationalpartnership.org/site/DocServer/RobertsReportAug05.pdf (describing memoranda written by Roberts that were critical of disparate impact analysis).

lanta[105] the Eleventh Circuit held that a regulation requiring all firefighters to be clean shaven met the standard of business necessity because it was demonstrably necessary to meet an "important business goal."[106] In *Sondel v. Northwest Airlines*, a case in which the employer had a 5'2" height requirement for certain positions of employment, the Eighth Circuit held that the employer was required to show a "compelling need" for the 5'2" height restriction and "the lack of an effective alternative policy that would not produce a similar impact."[107] And in *Bradley v. Pizzaco of Nebraska, Inc.*, the Eighth Circuit applied the EEOC Guidelines as flexible recommendations instead of binding law, ultimately holding that the defendant must prove that there is a "compelling need" for the disputed practice based on prior circuit decisions, instead of on the EEOC guidelines themselves.[108]

The Court cannot wait forever to accept another Title VII disparate impact case. And when a disparate impact case involving the interpretation of "consistent with business necessity" is next decided by the Supreme Court, it can be expected to follow the line of reasoning it developed in *Davis* and *Beazer*. This hypothesis is consistent with the general thrust of the *Davis-Beazer-Wards Cove* trilogy and recent ADA cases described above. If the Court does follow this path, such a decision might only require an employer to demonstrate an important reason for the policy, along the lines of the standard used by the Eleventh Circuit in *Fitzpatrick v. City of Atlanta*.[109]

If such a standard is adopted, it would likely take the following form: first, the plaintiffs would need to meet their burden of establishing a prima facie case of disparate impact; second, the employer would have the opportunity to begin to rebut the plaintiff's case by demonstrating the job-relatedness of its policy. If it could demonstrate that the policy is job related, it would then need to demonstrate that the policy is "consistent with business necessity." However, demonstrating that a policy is "consistent with business necessity" may only require the employer to demonstrate that its policy serves an important business goal.[110] This standard would be a difficult hurdle for most plaintiffs to overcome, obliging courts to rule in favor of the employer in an overwhelming majority of disparate impact cases in which the employer can demonstrate job-relatedness.

---

[105] 2 F.3d 1112 (11th Cir. 1993).

[106] *Id.* at 1118–20. The Eleventh Circuit acknowledged that African American men suffer disproportionately from a medical condition which prevents them from shaving and assumed, without deciding, that the plaintiffs had made out a prima facie case of disparate impact. *Id.* at 1114, 1118. The defendant's showing that bearded firefighters could not safely wear respirator masks satisfied the important business goal requirement. *Id.* at 1119–20.

[107] 56 F.3d 934, 940 n.10 (8th Cir. 1995).

[108] 7 F.3d 795, 797–98 (8th Cir. 1993); *see* RUTHERGLEN, EMPLOYMENT DISCRIMINATION LAW: VISIONS OF EQUALITY IN THEORY AND DOCTRINE, 87 (1st ed. 2001).

[109] 2 F.3d 1112, 1117–18 (11th Cir. 1993).

[110] *See supra* Part II.C.

III. The Likely Judicial Interpretation of a Mandatory Shaving Policy Through the Eyes of a Current Employment Dispute

For clarity and illustrative purposes, the evolution of the disparate impact doctrine will be traced through the eyes of a current controversial employment practice: the requirement that employees shave despite the disparate impact the policy may have on African Americans because of their unique susceptibility to pseudofolliculitis barbae.

Pseudofollicultis barbae has an estimated incidence rate of fifty percent or higher in African American men and is relatively rare in Caucasian men.[111] No other segment of the population is believed to be significantly affected. The condition causes shaved facial hair to grow back into the skin, causing inflammation, infection, and noticeable facial bumps. The only treatment known to be completely effective for all patients is to abstain from shaving.[112] Thus, a requirement that all men shave in order to gain or continue employment is likely to have a discriminatory impact on African American men in comparison with other segments of the population. Affected men would be forced to choose between leaving a position of employment with an employer that has a strict no-beard policy or facing harmful medical consequences.

A. Challenging a Mandatory Shaving Policy Under Title VII

A plaintiff could challenge a mandatory shaving policy by either claiming disparate treatment or disparate impact. In order to prove disparate treatment, a plaintiff must show that the employer actually intended to treat the plaintiff, or his class, in an adverse manner due to plaintiff's protected status.[113] In order to prove disparate impact, a plaintiff would have the lesser burden of demonstrating that a policy of the employer had the effect of harming the plaintiff or his class.[114] Thus, a challenge to a policy that barred all employees from wearing a beard in the workplace would likely fail under a disparate treatment challenge because the employee would have to prove that the employer instituted the shaving policy in order to exclude African Americans from the workplace. Conversely, under early interpretations of disparate impact, a plaintiff would likely have succeeded in making out a prima facie case because the employee would

---

[111] *See* Hermelita Winter et al., *An Unusual Ala12Thr Polymorphism in the 1A - Helical Segment of the Companion Layer-Specific Keratin K6hf: Evidence for a Risk Factor in the Etiology of the Common Hair Disorder Pseudofolliculitis Barbae*, 122 J. Investigative Dermatology 652, 652, 654 (2004).

[112] *See* Thomas G. Greidanus & Beth Honl, *Pseudofolliculitis of the Beard*, E-Medicine.com, http://www.emedicine.com/derm/topic354.htm (last visited Apr. 15, 2007).

[113] *See supra* Part I.A.

[114] *See supra* Part I.B.

only have to demonstrate that the policy disproportionately harmed African Americans in practice.

Because the Supreme Court has not decided a Title VII disparate impact case since the passage of the Civil Rights Act of 1991, which effectively overruled important parts of the last disparate impact case heard by the Court in 1989,[115] it is difficult to predict with complete accuracy whether a disparate impact challenge to this issue would succeed today. Nonetheless, the analysis in Part II of this Recent Development suggests that it will be more difficult for plaintiffs to establish a prima facie case of disparate impact and easier for an employer to establish job relatedness. It also appears the Court will be more sympathetic to an employer's business necessity defense than it has been in years past.

B. Challenging a Universal No-Beard Policy Under a Disparate Treatment Theory

The Supreme Court is unlikely to view the hypothetical employment policy requiring all men to shave as a case of disparate treatment. An analogy to the Equal Protection Clause analysis in *Geduldig v. Aiello*[116] sheds light on the manner in which the Court would handle such a claim. In *Geduldig*, the Supreme Court held that distinctions based on pregnancy are not unconstitutional under the Equal Protection Clause because not all women are pregnant.[117] Congress passed Title VII to shield protected classes of individuals from discrimination, and Title VII protects women. However, the Court in *Geduldig* held that while many women may be adversely affected by pregnancy, any adverse employment action based on pregnancy could only be said to be targeting pregnant people, and not women in general.[118] Although the policy in *Geduldig* did not specifically target women, women are the only sex capable of getting pregnant. By creating a rule that discriminates against pregnant people, an employer is essentially targeting women.

Similarly, by creating a rule banning beards in the workplace, an employer is essentially banning up to 50% of African American men from employment, while having very little impact on other Americans. African Americans can be separated into two distinct categories: those that have pseudofolliculitis and those that do not. Unless Title VII also covers protected individuals or subclasses of individuals adversely impacted by an employer's policy, an employer could potentially eliminate half of the Afri-

---

[115] *See* Wards Cove Packing Co. v. Antonio, 490 U.S. 642, 652 (1989).

[116] 417 U.S. 484 (1974).

[117] *See id.* at 496–97 n.20 ("The program divides potential recipients into two groups—pregnant women and nonpregnant persons. While the first group is exclusively female, the second includes members of both sexes.").

[118] *See id.* ("While it is true that only women can become pregnant, it does not follow that every legislative classification concerning pregnancy is a sex-based classification . . . .").

can American employee pool by creating a "neutral" policy that prohibited beards in the workplace. Thus, even though *Geduldig* was decided on Equal Protection grounds, its extensive discussion of the consequences of creating a rule for a subgroup that has an impact on a group protected by Title VII, strongly suggests that the Court would not interpret a universal no-beard policy as disparate treatment by an employer.

C. The Business Necessity of a Mandatory Shaving Policy in the Griggs-Albemarle-Dothard *Era*

The employee policy in the pseudofolliculitis example would have been unlikely to survive in the *Griggs-Albemarle-Dothard* era because it would have violated the disparate impact doctrine. The *Griggs-Albemarle-Dothard* era Court would almost certainly have viewed a policy that required all male employees to shave in order to hold a position of employment with heavy suspicion if that policy prevented up to fifty percent of African American men from obtaining employment.[119] If the plaintiff succeeded in establishing a prima facie case of disparate impact, the employer would have had the opportunity to demonstrate the business necessity of the policy.[120] But if the employer had the policy simply because it believed that beardless men looked more professional, or that customers might prefer beardless men, or out of a suspicion that beardless men did better work, there is strong evidence from the *Griggs*-era decisions to believe that the Court would have had no problem disposing of the case in favor of the plaintiffs.[121]

But what if the employer had the policy because she had statistical evidence showing that for the position in question, beardless men resulted in higher customer orders, and indeed, greater profits? What would happen if a pizza delivery company had conducted a study that demonstrated that beardless drivers resulted in greater consumer orders because customers had a distaste for facial hair? The *Griggs*-era Court still would have been likely to find in favor of the plaintiffs because it is unlikely that an employer in the *Griggs* era would have been able to establish the business necessity

---

[119] *See supra* Part II.A.

[120] *See supra* text accompanying notes 15–18. Not all employment polices can be validated for job relatedness. An example of such a policy is the no-beard requirement described above. In such a case, there is no ability or factor that the employer could test. The employee either has a beard or he does not. Thus, it is unlikely that a validation requirement would be mandated in such a case.

[121] *See* Griggs v. Duke Power Co., 401 U.S. 424, 432 (1971) ("Congress has directed the thrust of [Title VII] to the *consequences* of employment practices, not simply the motivation.") (emphasis added); *cf.* Gerdom v. Continental Airlines, 692 F.2d 602 (9th Cir. 1982) (holding that customer preference is not good rebuttal evidence for an employer in a disparate treatment case, which can be easily analogized to the disparate impact cases as discussed above); Diaz v. Pan Am Airways, 442 F.2d 385 (5th Cir. 1971) (holding that customer preference is not a valid legal excuse for the employer in disparate treatment cases).

of a policy that disparately impacted a protected class in order to cater to customer preference.[122]

It is unlikely that the *Griggs*-era Court would have accepted customer preference as a justification for creating a disparate impact on a group protected by Title VII.[123] This is evidenced by the fact that the Court had already suggested that increased costs or decreased profits alone would not be enough to justify the business necessity of an employment policy that has a discriminatory impact on a protected class. Recall that the *Dothard* Court required that the Alabama prison system develop a strength test rather than use a height and weight requirement.[124] There should be little doubt that creating a test to assess the strength of all incoming applicants would be costlier to implement than a height and weight requirement. Such a test might require special equipment or specially trained personnel to monitor and administer the examination. However, this issue was never addressed in the Court's analysis. The overriding concern was eliminating employment policies that created a disparate impact on any class of individuals protected by Title VII, and not the impact a non-discriminatory policy might have on the employer in terms of cost.[125]

Therefore, the *Griggs*-era Court would almost certainly have found that requiring all deliverymen to shave for the purpose of increasing profits and customer satisfaction was a violation of Title VII. The analysis of *Griggs*, *Dothard*, *Albemarle*, *Gerdom*, and *Diaz*, as discussed above, leaves little room for argument on this subject. Therefore, the employer would have to articulate a different rationale for business necessity, such as the fact that long hair could be unsanitary. Even if the employer were successful in justifying the business necessity prong of the *Griggs* test, the plaintiff could still have won by demonstrating that the employer had a less discriminatory alternative available; for example, a policy that allowed African Americans to trim their facial hair. This approach would serve the employer's legitimate interest in sanitary and well-groomed deliverymen while creating less of a disparate impact on African Americans. Thus, a mandatory shaving policy would have been unlikely to be upheld by the *Griggs* era Court.

---

[122] *See Griggs*, 401 U.S. at 431 ("If an employment practice which operates to exclude Negroes cannot be shown to be related to *job performance*, the practice is prohibited.") (emphasis added); Dothard v. Rawlinson, 433 U.S. 321, 331 n.14 (1977) ("a discriminatory employment practice must be shown to be necessary to safe and efficient job performance to survive a Title VII challenge.").

[123] *See Gerdom*, 692 F.2d at 609 ("It has long been established in the airline industry that passengers' preference for attendants who conform to a traditional image cannot justify discriminatory airline hiring policies."); *Diaz*, 442 F.2d at 389 ("[C]ustomer preference may be taken into account only when it is based on the company's inability to perform the primary function or service it offers.").

[124] *See supra* text accompanying notes 34–43.

[125] *See supra* notes 39–44 and accompanying text.

D. The Business Necessity of a Mandatory Shaving Policy in the Modern Era

In 1993, in *Bradley v. Pizzaco of Nebraska, Inc.*,[126] the Eighth Circuit Court of Appeals addressed a case in which the employer required all of its male employees to shave, despite the fact that some employees were unable to do so because of pseudofolliculitis barbae.[127] Finding that Domino's Pizza was unable to satisfactorily demonstrate that the policy was important to its business, the Eighth Circuit enjoined the no-beard policy because it created a disparate impact on African Americans.[128] Paul D. Black, Domino's vice president for operations, argued that Domino's policy was motivated by the belief that it is "common sense that the better our people look the better our sales will be."[129] Additionally, Black cited a survey "indicating that up to twenty percent of customers would 'have a negative reaction' to a delivery person wearing a beard."[130] Nevertheless, the court found it significant that "the survey [made] no showing that customers would order less pizza in the absence of a strictly enforced no-beard rule."[131] Thus, the court held that Black's arguments were "largely speculative and conclusory," and that "[s]uch testimony, without more, [did] not prove the business necessity of maintaining the strict no-beard policy."[132]

The Court's analysis in *Bradley* turned on Domino's failure to conduct a study demonstrating the shaving policy's effect on sales. As it is plausible that customers would prefer beardless deliverymen, and would order less from restaurants that used bearded deliverymen, the *Bradley* court left open the possibility that if Domino's had in fact demonstrated that a no-beard policy increased sales, and conducted statistically significant studies, it might have prevailed.[133]

Of course, the 1993 decision in *Bradley* does not settle how the current Supreme Court would handle such a case. As discussed above, in cases concerning disparate treatment, customer preference is not an excuse for discrimination.[134] However, in the case of a facially neutral policy, the modern Court certainly could conclude that catering to customer prefer-

---

[126] 7 F.3d 795 (8th Cir. 1993).

[127] *See id.* at 796.

[128] *See id.* at 799.

[129] *Id.* at 798 (internal quotation marks omitted).

[130] *Id.*

[131] *Bradley*, 7 F.3d at 799.

[132] *Id.* at 798.

[133] *See id.* at 798–99.

[134] *See* Diaz v. Pan Am Airways, 442 F.2d 385, 389 (5th Cir. 1971) ("[I]t would be totally anomalous if we were to allow preferences and prejudices of the customers to determine whether sex discrimination was valid. Indeed, it was, to a large extent, these very prejudices the Act was meant to overcome . . . ."); Gerdom v. Continental Airlines, 692 F.2d 602, 609 (9th Cir. 1982) ("[G]ender-based discrimination cannot be upheld on the basis of customer preferences unrelated to abilities to perform the job . . . .").

ence for the purpose of increasing profits is "consistent with business necessity," and therefore plaintiffs would not entitled to relief under a disparate impact theory. Although the Court has never gone quite this far, the most recent Title VII disparate impact decisions of *Davis*, *Beazer*, and *Wards Cove*, and the recent ADA opinions cited above have progressively required more of the plaintiff and less of the employer.[135]

Although the Civil Rights Act of 1991 attempted to reverse some of this erosion,[136] a majority of the current Supreme Court could easily read "consistent with business necessity" as requiring nothing more than an important business reason for engaging in a facially neutral policy, even if it does have a disparate impact on a group protected by Title VII. Based on the makeup of the current Court and the ambiguous statute Congress has provided, this Recent Development predicts that the Court would interpret "consistent with business necessity" in this, or a similar, manner.[137]

IV. Policy Issues and Conclusion

The disparate impact theory of Title VII has evolved since the *Griggs* era. The key question is how much. It has become progressively more difficult for plaintiffs to establish a prima facie case of disparate impact; it has become easier for defendants to demonstrate the business necessity and job relatedness of their employment policies; and it has seemingly become easier for the same defendants to avoid a finding of pretext in the third stage of the disparate impact analysis. Congress attempted to remedy some of this erosion through the Civil Rights Act of 1991. If the plaintiffs meet the burden of establishing a prima facie case of disparate impact, the Act shifts the burden of production and persuasion to the employer to demonstrate that its employment policy was "business related" and "consistent with business necessity." However, at the same time, the Supreme Court has grown suspicious of the business necessity theory, as demonstrated by many of its non-Title VII decisions discussed above. Thus, because of poor drafting of the Act, it would be unsurprising if the Court interpreted the meaning of "consistent with business necessity" to require only an important reason for the policy, even though it is clear that Congress intended the stronger standard cited above. If this theory is correct, and an employer is able to validate its test or policy and demonstrate that the policy has an important business purpose, the Court may hold that this is enough for the employer to meet its burden.

When the next generation of disparate impact cases reaches the Court, Congress will certainly have an opportunity to clarify its position on dis-

---

[135] *See* discussion *supra* Parts II.B–D.

[136] *See* Civil Rights Act of 1991 § 2(2), 42 U.S.C. § 1981 note (2000) ("[T]he decision of the Supreme Court in Wards Cove . . . has weakened the scope and effectiveness of Federal civil rights protections . . . .").

[137] *See* discussion *supra* Part II.C.

parate impact litigation. Congress might believe that the interpretation of the business necessity standard that this Recent Development predicts is appropriate.[138] However, if Congress wants to return to the interpretation of business necessity espoused in the *Griggs* era,[139] it should be careful to explicitly require that an employment policy be necessary for the operation of the business and then validated to prove job relatedness according to the EEOC Guidelines.

It is essential to remember that Congress has the sole power to decide this issue as long as its prescriptions are constitutional. Thus, while there is every indication that the Court might ignore the purpose of the Civil Rights Act of 1991, such action would be wrong because it is duty bound to follow the will of Congress. It is clear that Congress intended employers to prove that their policies are mandated by strict business necessity in order to prevail over a disparate impact claim, but the poorly drafted language used in the Civil Rights Act of 1991 leaves many legal loopholes to be exploited by those who feel business necessity is too strict a standard by which to hold employers.

Most notable in this controversy is Congress' lack of effort to clarify an issue that is of great legal importance. There is no excuse for the lack of action or clarification. Congress should not wait for the Court to decide this issue, because such a decision may be years away. In the meantime, the rights that the Civil Rights Act of 1991 bestowed upon employees are slowly being stripped away by the lower courts. This issue needs to be clarified by Congress soon, or else it will be assumed that Congress agrees with the interpretation being espoused by the courts in regard to disparate impact in general and Title VII in particular. Until Congress acts, employers will likely be able to demonstrate that their policies are consistent with business necessity by showing that the policies merely serve an important business purpose.

—*William Gordon*[*]

---

[138] This Recent Development predicts that the Court will adopt a business necessity standard that requires an employer to demonstrate that an employment policy that disparately affects a class protected by Title VII serves an "important business purpose." Such a standard would be less demanding on the employer than the business necessity standard the Court appeared to use in the *Griggs-Albemarle-Dothard* era, which demanded that "a discriminatory employment practice must be shown to be necessary to safe and efficient job performance to survive a Title VII challenge." Dothard v. Rawlinson, 433 U.S. 321, 331 n.14 (1977).

[139] *See supra* note 138.

[*] J.D., Harvard Law School, 2007; B.A, Pomona College, 2002.

THE PENSION PROTECTION ACT OF 2006: REFORMING THE DEFINED BENEFIT PENSION SYSTEM

The recent spate of corporate bankruptcies has caused workers to shoulder many of the burdens of corporate mismanagement,[1] ineffective government regulation, and the inevitable byproducts of a new economic reality.[2] One of the few silver linings for workers as they face these events comes in the form of the Pension Benefit Guaranty Corporation ("PBGC"), the federal government's insurance program for workers' defined benefit pensions. While companies themselves are not required to honor pension promises in bankruptcy, the PBGC, with limited exceptions, does guarantee those promises to workers.[3]

This Recent Development evaluates the 2006 reform of the pension system in light of the structure and financial situation of the existing pension insurance system. The results of this analysis suggest that this reform should be viewed as a successful effort to promote the fiscal solvency of the pension system because the structural changes in the pension system did significantly more to increase the financial solvency of the system than special interests did to diminish it.

Part I provides an overview of the contemporary design of the federal government's defined benefit pension regulation system, with a particular focus on the PBGC. Part II summarizes the legislative history of the Pension Protection Act of 2006 ("Protection Act"),[4] beginning with a brief discussion of previous reforms. Part III concludes by arguing that, while imperfect, the Pension Protection Act has been a step in the right direction for the long-term solvency of the system and therefore the protection of workers' defined benefit pensions.

I. THE BASIC STRUCTURE OF THE DEFINED BENEFIT PENSION SYSTEM

Some structural and financial background is helpful for understanding the most recent round of reform.[5] The PBGC, as well as other laws in

---

[1] *See, e.g.*, BRIAN CRUVER, ANATOMY OF GREED: THE UNSHREDDED TRUTH FROM AN ENRON INSIDER 25 (2002) (detailing the mismanagement and demise of Enron).

[2] *See, e.g.*, Caroline Daniel, *Stricken Airlines Seek Shelter in a Storm*, FIN. TIMES, Oct. 19, 2004, at 1 (pointing out the advantages and inevitability of Chapter 11 bankruptcy for airlines as well as for companies in many other industries).

[3] *See* CONG. BUDGET OFFICE ["CBO"], A GUIDE TO UNDERSTANDING THE PENSION BENEFIT GUARANTY CORPORATION 1 (2005), *available at* http://www.cbo.gov/ftpdocs/66xx/doc6657/09-23-GuideToPBGC.pdf [hereinafter CBO GUIDE TO UNDERSTANDING].

[4] Pension Protection Act of 2006, Pub. L. No. 109-280, 120 Stat. 780 (codified as amended in scattered sections of 26 U.S.C. and 29 U.S.C.).

[5] This Recent Development will not revisit previous work describing the history of the federal government's pension regulation system but will instead provide a broad overview of details relevant to understanding the Pension Protection Act. *See* Kathryn J. Kennedy, *Pension Funding Reform: It's Time to Get the Rules Right (Part 1)*, TAX NOTES, Aug. 22, 2005 at 907 [hereinafter Kennedy *Part I*]; Kathryn J. Kennedy, *Pension Funding Reform: It's Time to Get the Rules Right (Part 2)*, TAX NOTES, Aug. 29, 2005 at 1039 [hereinafter

this area, focus only on defined benefit pension plans.[6] Before the advent of the Employee Retirement Income Security Act of 1974 ("ERISA"),[7] employers were not required to pre-fund their pension plans.[8] In 1974, President Gerald Ford signed ERISA, thereby enacting minimum funding standards to ensure a degree of pension plan pre-funding.[9] ERISA also created a federal government guarantee for workers in the event of plan insolvency.[10]

ERISA mandates that corporations calculate the yearly costs of current obligations to retirees and any additional costs derived from previously unpaid yearly costs.[11] Along with such estimates of obligations, ERISA requires actuaries to report a "T account" for the plan, which compares the plan's yearly debits and credits.[12] If the credits (employer contributions to the plan) are equal to or greater than the debits (yearly accrued plan costs), the plan is deemed to be in compliance with the minimum funding standards.[13]

These funding rules have led to a simple system. All plans have a funding standard account ("FSA") such that if plan assets equal the present value of liabilities then the FSA is 0.[14] The FSA for each plan changes based upon the normal accrual of benefits, investment fluctuations, and structural changes.[15] If the FSA is equal to or greater than 0, the employer is not required to contribute.[16] Deficit reduction contributions ("DRCs") are required when the value of a plan's assets compared to the value of its liabilities (the funding ratio) falls below 90%.[17]

ERISA also set up a system designed to guarantee a certain level of benefits in the event of plan insolvency.[18] The PBGC, the cornerstone of

---

Kennedy *Part II*].

[6] Defined benefit plans commit an employer to providing a specific benefit at a particular retirement age as a life annuity for the plan participant. *See* 29 U.S.C. § 1002(35) (2006).

[7] Pub. L. No. 93-406, 88 Stat. 829 (1974) (codified as amended in scattered sections of 26 U.S.C. and 29 U.S.C.).

[8] *See* CBO GUIDE TO UNDERSTANDING, *supra* note 3, at 2; Kennedy *Part I*, *supra* note 5, at 910.

[9] *See* 29 U.S.C. §§ 1001(a)–(c) (2000); Kennedy *Part I*, *supra* note 5, at 911.

[10] *See* 29 U.S.C. §§ 1001(a)–(c); Kennedy *Part I*, *supra* note 5, at 911.

[11] *See* 26 U.S.C. § 412(b) (2006).

[12] *See* 29 U.S.C. § 1082(a)(1) (2006); 26 C.F.R. § 1.412(a)(1) (2006).

[13] *See, e.g.*, 26 U.S.C. § 302(b)(2); 29 U.S.C. § 1082(b)(2); 26 C.F.R. § 1.412(b)(2). If debits are greater than credits, the company is required to pay a 10% excise tax on the plan, plus an additional 100% tax if there is not a correction made within a specified correction period. *See* 26 U.S.C. § 302(a)(2); 29 U.S.C. § 1082(a)(2); 26 C.F.R. § 1.412(a). However, if the credits are greater than the debits, then the balance is carried over into the next year. *See* 26 U.S.C. § 302(b)(3); 29 U.S.C. § 1082(b)(3); 26 C.F.R. § 1.412(b)(3).

[14] *See* 29 U.S.C. § 1082(b); CBO GUIDE TO UNDERSTANDING, *supra* note 3, at 5.

[15] *See* 29 U.S.C. § 1082(b)(2); CBO GUIDE TO UNDERSTANDING, *supra* note 3, at 5.

[16] *See id.*

[17] *See* 29 U.S.C. § 1082(d)(2); CBO GUIDE TO UNDERSTANDING, *supra* note 3, at 6.

[18] At the time of ERISA's creation, the PBGC was largely an afterthought, which partly explains why the PBGC is not structured around sound financial principles. *See* Julie Kosterlitz, *Risking a Major Loss*, NAT'L J., May 27, 2006, at 10 (paraphrasing the Center

this system, maintains two legally distinct programs: one for single-employer plans and one for multi-employer plans.[19] The PBGC primarily covers single-employer plans, which it insures, and by design devotes significantly less coverage to multi-employer plans, to which it provides loans when necessary.[20]

A major component of the PBGC's insurance system is its termination structure. Initially, ERISA allowed employers to opt-out of the FSA system whenever they chose and hence transfer all of their liabilities to the PBGC.[21] In 1987, Congress revised this system to create a standard termination mechanism that allows companies to voluntarily end their plans only if plan assets exceed plan liabilities.[22] Barring a situation in which plan assets exceed plan liabilities, the PBGC will only assume a employer's obligations if: the employer is petitioning for bankruptcy or insolvency, the employer is unable to pay its debts when due and will be unable to continue business without termination, or the cost of the employer's plan has become unreasonably burdensome because of a decline in the employer's workforce.[23]

In order to fund the PBGC, ERISA initially required companies to pay an annual $1 per-participant premium ("the basic premium").[24] Congress has since increased the basic premium to $19 per participant per year.[25] The Pension Protection Act of 1987[26] added a second tier premium ("the variable premium") for plans with unfunded vested benefits, which was initially set at $6 for each $1,000 of unfunded vested benefits, with a cap of $50.[27] In 1994, Congress removed the cap and changed the variable requirement to $9 for each $1,000 of unfunded vested benefits.[28] If an employer does not meet its premium obligations to the PBGC, the PBGC can place a lien on the employer's assets.[29]

Aside from the premiums collected from employers, the PBGC also acquires assets from terminated plans.[30] Terminated assets are part of the

---

on Federal Financial Institutions' President, Douglas Elliott).

[19] *See* 29 U.S.C. § 1082; CBO GUIDE TO UNDERSTANDING, *supra* note 3, at 8.

[20] *See* CBO GUIDE TO UNDERSTANDING, *supra* note 3, at 8.

[21] *See* Employee Retirement Income Security Act of 1974, Pub. L. No. 93-406, § 4041(a), 88 Stat. 829, 1020 (current version at 29 U.S.C. § 1341(a) (2006)) (failing to require companies to meet any conditions before opting out of their plans).

[22] *See* 29 U.S.C. § 1341(b)(1)(D).

[23] *See* 29 U.S.C. § 1341(c)(2)(B).

[24] *See* CBO GUIDE TO UNDERSTANDING, *supra* note 3, at 8.

[25] *See* Omnibus Budget Reconciliation Act of 1990, Pub. L. No. 101-508, § 12021(b)(1), 140 Stat. 1388; CBO GUIDE TO UNDERSTANDING, *supra* note 3, at 8.

[26] Pub. L. No. 100-203, 101 Stat. 1330 (1987) (codified as amended in scattered sections of 42 U.S.C.).

[27] *See id.* at § 9331(a).

[28] *See* Retirement Protection Act of 1994, Pub. L. No. 103-465, § 774(a), 108 Stat. 4809, 5045; CBO GUIDE TO UNDERSTANDING, *supra* note 3, at 9.

[29] *See* 29 U.S.C. § 1082(f) (2006); CBO GUIDE TO UNDERSTANDING, *supra* note 3, at 3.

[30] *See* 29 U.S.C. § 1305(b)(1)(C); CBO GUIDE TO UNDERSTANDING, *supra* note 3, at

off-budget PBGC trust fund, which can be invested in most investment vehicles.[31] While the PBGC itself is not backed by the full faith and credit of the federal government, it does have a $100 million line of credit from the United States Treasury.[32] Finally, the PBGC does not pay out all employer-provided benefits. Instead, it only guarantees benefits up to a maximum annual pension.[33]

II. THE LEGISLATIVE HISTORY OF RECENT DEFINED BENEFIT PENSION SYSTEM REFORM

PBGC reform began to gain momentum in 2003, culminating in two separate bills: the Pension Funding Equity Act of 2004 ("Equity Act")[34] and the Pension Protection Act of 2006.[35] During the 1990s, strong equity markets inflated the value of pension plan assets around the country, reducing the required minimum contributions by employers.[36] Unfortunately, the economic problems of the early 2000s deflated plan assets while simultaneously decreasing the interest rates used for funding purposes.[37] This confluence of events, which caused massive underfunding in plans across the country, has been dubbed the "perfect storm."[38] The PBGC was without the financial reserves necessary to provide relief should many of the plan obligations become PBGC liabilities.[39]

The Congressional Budget Office ("CBO") projected that benefits paid would grow from about $4 billion in 2005 to approximately $10 billion in 2015.[40] According to this projection, the PBGC would exhaust its on-budget surpluses around 2013 and would need to use a larger percentage of its trust fund to meet benefit obligations.[41] Additionally, the Center on Federal Financial Institutions ("COFFIs") predicted in 2004 that overall PBGC funds would be depleted by 2021.[42]

11.

[31] *See* CBO GUIDE TO UNDERSTANDING, *supra* note 3, at 11.

[32] *See id.*

[33] *See* 29 U.S.C. § 1322(b)(3); CBO GUIDE TO UNDERSTANDING, *supra* note 3, at 12. In 2005, the maximum PBGC award was a $45,614 lifetime annuity beginning at age 65. *See id.*

[34] Pub. L. No. 108-218, 118 Stat. 596 (2004) (codified as amended in scattered sections of 26 U.S.C. and 29 U.S.C.).

[35] Pub. L. No. 109-280, 120 Stat. 780 (2006) (codified as amended in scattered sections of 26 U.S.C. and 29 U.S.C.).

[36] *See, e.g.*, Kennedy *Part II*, *supra* note 5, at 1040.

[37] *See id.*

[38] *The Pension Underfunding Crisis: How Effective Have Reforms Been?: Hearing Before the H. Comm. on Education and the Workforce*, 108th Cong. 2 (2003) (statement of Rep. John A. Boehner, Chairman, H. Comm. on Education and the Workforce).

[39] *See id.*

[40] CBO GUIDE TO UNDERSTANDING, *supra* note 3, at 18.

[41] *See id.* at 17.

[42] *See* COFFI, PBGC: UPDATED CASH FLOW MODEL FROM COFFI 2 (2004), *available at* http://www.coffi.org/pubs/PBGC%20Updated%20Cash%20Flow%20Model%20from%20COFFI.pdf. In addition to these cash flow estimates, accrual accounting estimates showed

The Bush administration began legislative debate on this problem by releasing its reform proposal on July 8, 2003.[43] The administration's plan was primarily concerned with establishing funding requirements based upon a realistic measurement of risk.[44] The proposal suggested structural modifications, which included changing the discount rate used in calculating plan liabilities by replacing the thirty-year Treasury bond rate with a yield curve based upon investment-grade corporate bonds, as well as creating a classification for "at-risk" plans.[45]

The bill that emerged from conference committee and that was eventually signed by President Bush on April 10, 2004,[46] represented a compromise amongst provisions and principles.[47] The final incarnation of the Equity Act included a temporary, three-year requirement to use a yield curve based upon investment grade corporate bonds.[48] The bill also included an exception that waived the deficit reduction contributions requirement for three years for applicable employer plans that were not required to make a deficit reduction contribution in 2000.[49] This exemption maintained the specific business exceptions—originally added in the full Senate—allowing airline, steel, and other companies that apply to waive substantial parts of their deficit reduction contributions in the first two years.[50] The final bill did not address the administration's desire for a separate classification for "at-risk" plans.

As the problems facing the PBGC continued to grow, however, legislators returned to reform yet again. The administration released a second

---

the PBGC to be in even worse financial shape. *See id.* Accrual accounting evaluates the net financial position of a system by comparing the system's assets to the present value of its liabilities. *See* CBO GUIDE TO UNDERSTANDING, *supra* note 3, at 14. The accrual model predicted $1.7 billion per year in additional claims for PGBC over the next ten years, which translates into a median projected accumulated deficit of $26.9 billion in 2014. *See* PBGC, 2004 ANNUAL REPORT 12 (2004), *available at* http://www.pbgc.gov/docs/2004_annual_report.pdf [hereinafter PBGC 2004 ANNUAL REPORT]. The CBO model estimated an $86.7 billion accumulated deficit in 2014 from a $63.7 billion deficit. *See* CBO, THE RISK EXPOSURE OF THE PENSION BENEFIT GUARANTY CORPORATION, at viii (2005), *available at* http://www.cbo.gov/ftpdocs/66xx/doc6646/09-15-PBGC.pdf.

[43] *See* Press Release, Office of Pub. Affairs, U.S. Dep't of Treasury, The Administration Proposal to Improve the Accuracy and Transparency of Pension Information (July 8, 2003), *available at* http://www.treasury.gov/press/releases/js529.htm.

[44] *See id.*

[45] *See id.* "At-risk" plans would be required to make larger contributions to their plans. *See infra* notes 107–113 and accompanying text; *see also* GOV'T ACCOUNTABILITY OFFICE ("GAO"), COMMERCIAL AVIATION: BANKRUPTCY AND PENSION PROBLEMS ARE SYMPTOMS OF UNDERLYING STRUCTURAL ISSUES 45 (2005), *available at* http://www.gao.gov/new.items/d05945.pdf [hereinafter COMMERCIAL AVIATION] (graphing the historical yields of investment grade corporate bonds and thirty-year Treasury bonds).

[46] *See* Denise Marois, *Pension Bill Gives Airlines Breathing Room*, AVIATION DAILY, Apr. 13, 2004, at 3.

[47] Pension Funding Equity Act of 2004, Pub. L. No. 108-218, 118 Stat. 596 (codified as amended in scattered sections of 26 U.S.C. and 29 U.S.C.).

[48] *See* 29 U.S.C. 1082(b)(5)(B)(ii)(II) (2006).

[49] *See* 29 U.S.C. 1082(d)(12).

[50] *See id.*

reform proposal in early 2005 that focused again on PBGC risk reduction.[51] The proposal raised per-participant premiums to thirty dollars from nineteen dollars and indexed such premiums to wage growth for the first time.[52] It defined ongoing liabilities as the full present value of obligations based upon an AA investment grade corporate bond yield curve.[53] Credit balances arise when a pension plan sponsor makes a contribution in one year that is higher than the minimum required in that year. The administration sought to entirely eliminate credit balances to ensure that pension plans would not reduce their payments based upon an unrealistically high value of previous payments.[54] In an attempt to more closely match payments with potential liabilities, the proposal defined "at-risk" plans as those that were attached to non-investment grade rated firms.[55]

House members began the consideration of PBGC reform when the House Committee on Education and the Workforce passed a related bill on June 30, 2005.[56] The bill only restricted the use of credit balances for those plans funded at under 80% of liabilities.[57] In another significant departure from the administration's proposal, the committee's bill defined "at-risk" plans as those plans funded at under 60% of liabilities.[58]

The House Committee on Ways and Means' bill included an increase in per-participant premiums to thirty dollars but declined to index future increases to wage growth.[59] It also imposed more severe penalties for plans that do not meet basic funding obligations.[60] Finally, the bill followed the House Committee on Education and the Workforce bill's structure with reference to "at-risk" plans.[61] Following the administration's proposal, the Committee on Ways and Means' bill also disallowed the use of credit balances entirely.[62]

Less than a month after the House Education and Workforce Committee passed its bill, the Senate Finance Committee began the legislative process

---

[51] *See* Elaine Chao, U.S. Sec'y of Labor, Protecting the Retirement Security of America's Workers: The President's Plan for Reforming Private Defined Benefit Pension Plans (Jan. 10, 2005), *available at* http://www.dol.gov/_sec/media/speeches/20050110_retirement.htm.

[52] *See* EMPLOYMENT BENEFITS AND SEC. ADMIN., U.S. DEP'T OF LABOR, STRENGTHENING FUNDING FOR SINGLE-EMPLOYER PENSION PLANS 2 (2005), *available at* http://www.dol.gov/ebsa/pdf/sepproposal2.pdf.

[53] *See id.* at 16.

[54] *See id.* at 12.

[55] *See id.* at 14–15.

[56] Pension Protection Act of 2005, H.R. 2830, 108th Cong. (as reported by H. Comm. on Education and the Workforce, Sept. 22, 2005).

[57] *See id.* § 102(f).

[58] *See id.* § 102(i)(3).

[59] *See* Pension Protection Act of 2005, H.R. 2830, 109th Cong. § 401 (as reported by H. Comm. on Ways and Means, Dec. 6, 2005).

[60] *See id.* § 102(f).

[61] *See id.* § 122.

[62] *See id.* § 431(c)(6)(B)(ii).

in the Senate.[63] The bill also deviated from the administration's principles in a few key areas. Specifically, the bill did not reduce the use of credit balances,[64] instead taking an even stronger stance on the "at-risk" issue by defining an "at-risk" plan as one attached to a non-investment grade company.[65] The committee bill also reduced the amortization period to seven years, though it created an exception for airline companies that allowed them to amortize their payments over fourteen years.[66] Amortization periods, which create a cushion for obtaining full funding, had previously been set at thirty years.[67]

By contrast, a proposal by the Senate Committee on Health, Education, Labor, and Pensions ("HELP") increased the per-participant premium to thirty dollars but did not index premiums to future wage growth.[68] The HELP proposal did not change the credit balance system. The credit balance system was allowing companies to continue to rely on artificially high estimates of previous payments to the plan funds to reduce current payments.[69]

After the passage of bills by both the Committee on Finance and HELP by early September, the process stalled in the full Senate.[70] The lag initially occurred because Senator Charles Grassley (R-Iowa), the chairman of the Committee on Finance, and Senator Michael Enzi (R-Wyo.), the chairman of HELP, could not agree on what version of the pension bill to bring to the floor of the Senate.[71] By September 28, however, the Senators had agreed to bring a compromise bill to the floor.[72] Since the bill included credit rating provisions, Senators Mike DeWine (R-Ohio) and Barbara Mikulski (D-Md.), with the backing of business groups, put a hold on the legislation.[73] The Senate could not consider the bill until the Senators eventually dropped their hold on the bill.

When the full Senate passed the bill on November 16, 2005, it departed from the principles set forth by the administration.[74] The bill fol-

---

[63] *See* National Employee Savings and Trust Equity Guarantee Act of 2005, S. 1953, 109th Cong. (as reported by S. Comm. on Fin., Nov. 2, 2005).

[64] *See id.*

[65] *See id.* § 430(f).

[66] *See id.* §§ 430(c), 334(d)(3)(B).

[67] *See* Employee Retirement Income Security Act of 1974, Pub. L. No. 93-406, § 302(b)(2)(B)(ii), 88 Stat. 829, 869 (current version at 29 U.S.C. (2006)).

[68] *See* Press Release, Senate Health, Education, Labor, and Pensions Committee, Senators Kennedy and Enzi Propose Bipartisan Pensions Reform to Provide Retirement Security for Millions of Americans (Sept. 8, 2006), *available at* http://help.senate.gov/Min_press/2005_09_08_b.pdf.

[69] *See id.*

[70] *See Provisions of Capitol Hill Pension Reform Agreement*, MAIN WIRE, Sept. 28, 2005 (on file with author).

[71] *See id.*

[72] *See id.*

[73] *See* Jerry Geisel, *Pension Plan Funding Reform Takes Tortuous Path*, BUS. INSURANCE, Dec. 26, 2005, at 14.

[74] Pension Security and Transparency Act of 2005, S. 1783, 109th Cong. (as passed by

lowed the lead of the other proposals by increasing per-participant premiums to thirty dollars, but it did not index premiums to future wage growth.[75] The Senate bill required full funding of liabilities based on an investment-grade corporate bond curve.[76] In a compromise between the administration's proposal of dropping credit balances altogether and leaving the credit balances intact, the bill required that they be valued at market value.[77] "At-risk" plans were defined as those attached to non-investment grade firms and funded at less than 93% of plan liabilities.[78] Like HELP's proposal, the Senate bill reduced the amortization period to seven years but allowed airlines to amortize over twenty years.[79]

The final House bill, passed 294-132 on December 15, 2005, closely resembled the bill put forward by the Committee on Ways and Means.[80] The only two differences in the final bill were that it indexed per-participant premiums to wage growth and allowed for an exception in contributions for interstate bus companies.[81]

In the midst of the PBGC reform debate, Congress dealt with a related bill, which was part of the budget reconciliation procedures debate. The Deficit Reduction Act of 2005[82] bears an important relationship to the PBGC reform debate. The bill imposed additional requirements on the PBGC by increasing per-participant premiums to thirty dollars, with future premiums indexed to wage growth.[83]

The final pension bill did not change the premiums set by the Deficit Reduction Act.[84] As a result of contentious conference committee negotiations over issues such as at-risk plans and credit balances, it was not until August 3, 2006 that the Senate finally passed its last version of pension reform.[85] The President had threatened to veto any bill he believed would not restore the solvency of the defined benefit pension system, but

---

Senate, Nov. 16, 2005).

[75] *See id.* § 401.

[76] *See id.* § 303(h)(2).

[77] *See id.* § 304.

[78] *See id.* § 303.

[79] *See id.* § 403(d)(3)(B).

[80] Pension Protection Act of 2005, H.R. 2830, 109th Cong. (as passed by House, Dec. 15, 2005).

[81] *See id.* §§ 401, 121(a).

[82] Pub. L. No. 109-171, 120 Stat. 4 (codified as amended at 29 U.S.C.A. § 1306(a)(3)(A)(i) (West 2006)).

[83] *See* 29 U.S.C § 1306(a)(3).

[84] *See* William J. Miner, *Defined Benefit Landscape Changes*, BUS. INSURANCE, Oct. 2, 2006, at 10.

[85] *See* Deborah Solomon, *Pension Measure to Enact Changes Over Several Years*, WALL ST. J., Aug. 5, 2006, at A4. The final few months of negotiations focused on somewhat unrelated issues such as the estate tax and the extension of other tax provisions. *See* Jim Abrams, *Congress Divide Jeopardizes Pension Bill*, ASSOCIATED PRESS, July 28, 2006, *available at* http://www.sfgate.com/cgi-bin/article.cgi?f=/n/a/2006/07/28/national/w003151D53.DTL#sections.

he signed the final version of the Pension Protection Act passed by the House and Senate.[86]

The particulars of the final bill will be discussed below in relation to the success and failure of various stakeholders in the reform process. The centerpiece of the bill, in accordance with almost all of the relevant proposals, was to increase the funding requirements to 100% over the next seven years.[87]

III. THE RESULTS OF RECENT REFORM

While there is a temptation to focus on the victories of special interests in the preceding legislative history, the overall positive fiscal outcome of the bill for the PBGC should not be ignored. Although some commentators argued that the final bill risked doing more harm than good,[88] many suggested that the bill might be a net positive fiscal development, despite also alluding to the significant inroads made by various special interest groups.[89] Yet the lack of clarification in these accounts as to why the bill should be seen as a fiscal success suggests the need for a more comprehensive analysis of specific victories achieved by special interests and the relationship of those victories to the overall changes made by the bill. This Part will discuss the aspects of the legislation that were successes for special interests before arguing that, in the structural context of the PBGC, these changes did not overwhelm the positive fiscal developments of the reform.

The success of special interests can be partially attributed to the temporary pragmatic alignment of labor special interests with business special interests. From the initial stages of the reform process, unions were reported to be aligning themselves with business interests out of a desire to preserve pension plans.[90] Business and labor interests tried to alter the pro-

---

[86] *See* Peter Baker, *Bush Signs Sweeping Revision of Pension Law*, WASH. POST, Aug. 18, 2006, at D1.

[87] 29 U.S.C. 1082.

[88] *See, e.g.*, Editorial, *The Pension Piñata*, N.Y. TIMES, Aug. 13, 2006, at 9 (arguing that the final pension bill was primarily composed of special interest accommodations); *see also* Editorial, *Pension Reforms Welcomed*, BUS. INS., Aug. 7, 2006, at 8 (declining to speculate as to the long-term effect of the bill); Editorial, *Tougher Pension Rules Will Strengthen the System*, BUS. INS., Sept. 18, 2006, at A8 (suggesting that it is likely the bill will have a positive effect); Editorial, *Pension Reform Adds Cushions but No Guarantees*, USA TODAY, Aug. 17, 2006, at 10A (arguing that the bill will be a modest improvement).

[89] *See, e.g.*, Bruce E. Davis, *New Pension Law: Mostly Good for Steel, Other Retirees*, MORNING CALL (Allentown, Pa.), Aug. 9, 2006, at A9 (praising the bill while discussing the concessions to airlines); *see also* Editorial, *Time Running Out to Enact Pension Bill*, MIAMI HERALD, July 25, 2006, (urging President Bush to veto any bill with airline relief); Editorial, *Full Funding for Promised Pensions*, CINCINNATI INQUIRER, Aug. 4, 2006, at 6B (focusing on the airline provisions while discussing the House bill); Editorial, *The Pension Endgame: The Airlines Win, Again*, WASH. POST, Aug. 2, 2006, at A14 (suggesting that the House bill's concessions to airlines might overwhelm the positive fiscal effects of the bill).

[90] *See* James A. Klein, *Uniting to Repair Pensions*, WASH. TIMES, Sept. 1, 2003, at A16

posals designed to improve the financial health of the PBGC, especially those related to funding structure, because any such proposal would increase the burden on company-operated pension plans.[91] Put another way, these special interests fought to reduce the premiums paid by employers for their pension plan insurance.[92] This strange coalition emerged because businesses were concerned about the effect of new obligations on company balance sheets, and unions were concerned about the risk that companies would terminate their pension plans because of the new requirements.[93]

An important example of the power of the business lobby was the decision of Senators DeWine and Mikulski to put a hold on the 2006 bill in the Senate.[94] While DeWine and Mikulski ultimately relented, they did so only after extracting a promise from members of the congressional leadership that they could raise the credit balance issue in conference.[95] The business lobby was particularly determined in resisting the credit rating issue, because many companies, such as General Motors, would fall into an "at-risk" category if junk bond status qualified them for such a designation.[96]

The yield curve issue in the reform discussions focused on plan risk based upon the specifics of each individual company. During the second round of reform, critics of using bond yields expressed concern that they were inappropriate because bonds are tied to particular end dates while pensions have no particular expiration date.[97] The business lobby success-

---

(pointing out the shared interest of business and labor in the continuation of pension plans). In a show of unity between labor and business on this issue, United Auto Workers representatives came out against the administration's proposal alongside representatives from the Chamber of Commerce. *See* Stephanie Kirchgaessner, *Bush Pension Reform Faces Heavy Opposition*, FIN. TIMES, Apr. 27, 2005, at 8.

[91] *See* Klein, *supra* note 90.

[92] *See Plans Proliferate to Reform Federal Pension Insurer*, BESTWIRE, June 8, 2005 (on file with author).

[93] While such a coalition did exist, it did not include the entire labor movement, as some sectors of the movement focused on the long-run fiscal stability of the PBGC and advocated increased funding. *See Pitched Debate Looms Over Pension Reform: Legislation Includes Key Provision for Retail Fund Industry*, MONEY MGMT. EXECUTIVE, Dec. 12, 2005, at 1. One possible reason for this split is that higher wage earners lose more with PBGC funded plans because of the PBGC cap on annuities. This may incline those workers to support stricter plan solvency requirements. *See* COMMERCIAL AVIATION, *supra* note 45, at 54–55.

[94] There are also reasons to suspect that, because of potentially conflicting interests, the business community may not have always acted as a cohesive lobby. For instance, businesses with underfunded pension plans do not want increased funding rules, as such rules would increase obligations on the business. *See* Letter from James A. Klein, President, American Benefits Council, to Pension Bill Conferees (Mar. 20, 2006), *available at* http://www.americanbenefitscouncil.org/documents/confereeletter032006.pdf [hereinafter Council Letter]. On the other hand, businesses with fully funded plans may see virtue in forcing underfunded businesses to increase the solvency of the overall system in a way that causes fewer plans to put obligations on the PBGC. Otherwise, tax increases in the form of premiums or general taxes, needed to meet plan obligations, would negatively affect fully funded businesses.

[95] *See* Geisel, *supra* note 73, at 14.

[96] *See* Stephanie Kirchgaessner, *Legislation Strains Cosy Links Between Business and Politics*, FIN. TIMES, Nov. 21, 2005, at 6.

[97] *See* COFFI, PBGC: A Yield Curve Primer 6 (2004), *available at* http://www.coffi.

fully limited the move from Treasury rates to yield curves in the first round of reform,[98] yet they failed to do so in the second round.[99]

In addition to disagreeing about whether or not to have bond yields at all, legislators also differed on what type of bond yields to use. Those favoring a higher standard of bonds argued that using better quality bonds would hold plans to a stricter fiscal standard, which would be appropriate given the importance of pension obligations.[100] Business lobbyists argued that a higher standard would increase contribution requirements, thereby risking harm to companies and their plans.[101] In the end, legislators sided with the business lobby and expanded the yield curves used to include investment grade corporate bonds, not just AA or better corporate bonds.[102]

One of the biggest areas of contention was the use of credit balances. Those arguing for the elimination or restriction of credit balances claimed that allowing such credit when the value of the actual assets has declined distorts the funding ratio of any given plan.[103] Those in favor of maintaining credit balances argued that companies should be encouraged to make contributions greater than the required minimum and be rewarded for doing so.[104] Business lobbyists saw this as an important issue, because credit balances reduce the amount of present funding obligations for companies.[105] Ultimately, credit balances were restricted but not eliminated.[106]

The second set of major issues dealt with in the PBGC reform debate was whether or not to create a separate category for "at-risk" plans and how to define such plans. Designating plans as "at-risk" would provide a longer-term solution to some of the PBGC's problems by requiring plans that were more likely to be in financial difficulty to make larger contributions to their plans.[107] Creating an "at-risk" status was controversial because big business argued that requiring additional payments from the most finan-

---

org/pubs/Primer %20on%20Yield%20Curve%205.pdf.

[98] *See* Pension Funding Equity Act of 2004, Pub. L. No. 108-218, § 101(a)(1), 118 Stat. 596 (codified as amended at 29 U.S.C. § 1082 (2004)).

[99] *See* Pension Protection Act of 2006, Pub. L. No. 109-280, § 101(b), 120 Stat. 780 (codified as amended at 29 U.S.C § 1082 (2006)).

[100] *See, e.g.*, Press Release, U.S. Dep't of the Treasury, Prepared Remarks to the D.C. Bar Association (Mar. 7, 2006) (Remarks of Mark J. Warshawsky, Assistant Sec'y of the Treasury).

[101] *See, e.g.*, Council Letter, *supra* note 94, at 1.

[102] *See* Pension Protection Act of 2006, § 303.

[103] *See, e.g.*, Press Release, U.S. Dep't of the Treasury, *supra* note 100.

[104] *See, e.g.*, Council Letter, *supra* note 94, at 1–2.

[105] *See id.*

[106] *See* Pension Protection Act of 2006 § 303 (eliminating double counting of credit balances by preventing credit balances from being applied to the under-funding and contribution calculations, indexing credit balances to actual returns, and removing the assumption of a specific interest rate without reference to current market conditions).

[107] *See, e.g.*, Mark Schoeff Jr., *Parties United on Increasing PBGC Premiums,* WORKFORCE MGMT., Nov. 7, 2005, at 10.

cially vulnerable companies risked pushing those companies over the edge or, at a minimum, causing them to exit the defined benefit system.[108]

While questions remained as to whether or not any "at-risk" category was acceptable when the second round of PBGC reform began, the debate shifted to the use of credit ratings in determining "at-risk" plans. Some argued that a plan should be defined as "at-risk" if the company had a non-investment grade credit rating.[109] The proponents of such a rule claimed that, while not perfectly correlated, a bad credit rating is an empirically good measure of such high-risk plans.[110] In opposing the use of credit ratings, the business lobby argued that there was minimal direct relevance of credit ratings to the funding status of a plan and that not all firms are rated.[111]

Business also argued that the potential economic problems of creating an "at-risk" definition would be exacerbated by the use of credit ratings.[112] For example, the American Benefit Council argued that credit ratings would penalize companies in cyclical industries, as such companies are likely to have poor credit ratings at precisely the times when they are most economically vulnerable.[113] In the final version of the bill, the conference committee dropped references to "non-investment grade" status in the definition of "at-risk"—a big win for business and aligned labor lobbyists.[114]

Another issue was what to do when a plan was not fully funded. Penalizing plans for immediate rule changes or short-term market fluctuations that decrease plan assets is questionable, because pension obligations are long-term obligations. One argument for a substantially reduced amortization period was that longer periods do not functionally require underfunded plans to make up the difference in plan funding before their plans terminate, anyway.[115] Those in favor of allowing economically troubled companies to maintain liquidity asserted that if funding rules were to be strengthened, then companies needed a long time to make up such shortfalls.[116] In the end, the amortization period was reduced to seven years: a compromise between requiring immediate funding and allowing for amor-

---

[108] *See id.*

[109] *See, e.g.*, Press Release, U.S. Dep't of the Treasury, Prepared Remarks to the American Bankers Association (June 28, 2005) (Remarks of The Hon. Mark W. Warshawsky).

[110] *See id.*

[111] *See e.g.*, Kate Laughlin, *Pension Reform May Penalize HY Companies*, HIGH YIELD REP., May 15, 2006, at 1.

[112] *See* Council Letter, *supra* note 94, at 1.

[113] *See id.*

[114] *See* Pension Protection Act of 2006, Pub. L. No. 109-280, § 303, 120 Stat. 780 (codified as amended at 29 U.S.C § 1083 (2006)) (defining a plan as "at-risk" if it is less than 80% funded on a normal basis and less than 70% funded on the stricter "at-risk" basis).

[115] *See Solvency of the Pension Benefit Guaranty Corporation—Current Financial Condition and Potential Risks: Hearing Before the S. Comm. on the Budget*, 109th Cong. 4 (June 15, 2005) (testimony of Bradley D. Belt, Executive Director, PBGC).

[116] *See, e.g.*, Elana Schor, *Airlines Send Out Pension Distress Call*, THE HILL, June 9, 2005, at 12 (discussing amortization arguments in the airline context).

tization over a long time period, which would have made the funding structure irrelevant.[117]

The politically contentious issue of whether or not to create funding exceptions for certain industries or particular companies turned on intuitions opposite those surrounding the "at-risk" debate. Some reasoned that giving particular exceptions was necessary because those industries or companies were financially vulnerable, such that requiring full deficit reduction contributions risked causing those plans to exit the system.[118] The arguments on the other side were twofold: drawing the line to determine which industries or companies should be exempted was impossible;[119] and vulnerable companies are precisely the ones that should be required to pay the full contribution because they are at a greater risk of default.[120]

The airline lobby was effective in achieving gains in the PBGC reform process. In the 2004 reform bill, airlines were the primary force behind the funding relief section, which was expanded in the final version to cover all businesses.[121] At one point in the process, the airlines garnered a bill which gave them a twenty-five-year amortization period. In the final bill—a compromise that strongly favored the airline lobby—aviation plans that were no longer accruing additional benefits were given a seventeen-year amortization period while other airlines were held to the same seven-year period as other companies.[122]

Although the foregoing discussion indicates that special interest groups were successful in achieving many of their stated goals in the reform process, these successes do not overwhelm the positive fiscal solvency aspect of the recent PBGC reform. While these special interest victories may act as examples of the political clout of different lobbies, this Recent Development argues that despite their influence, the eventual reform was a dramatic net positive for the fiscal solvency of the PBGC system. A detailed accounting comparison of the stages of the most recent reform reveals that the fiscal effect of the various changes obtained by special interest groups was minor with respect to the overall fiscal effect achieved.

While initially plausible, many facts suggest that business and union lobbyists did not play as big of a role in the reform process as an initial account might suggest. To adequately analyze the impact of such lobbying groups, one must focus on the context of their efforts. To be sure, any business with a pension plan would have initially decried almost every element

---

[117] *See* Pension Protection Act of 2006 § 303(c).

[118] *See* Schor, *supra* note 116, at 12.

[119] *See* Press Release, COFFI, Pension Reform: Issues for Conference Committee 3 (Feb. 26, 2006), *available at* http://www.coffi.org/pubs/Pension%20Reform%20Conference%20Committee%20Issues.pdf.

[120] *See id.*

[121] *See, e.g.*, *Funding Relief Fails to Clear Senate*, FACTS ON FILE: WORLD NEWS DIGEST WITH INDEX, Dec. 31, 2003, at 1038.

[122] *See* Pension Protection Act of 2006 § 430.

of the administration's reform proposal.[123] For example, in the first round of PBGC reform, businesses came out against all of the particular provisions of the administration's proposal, including those changing the yield curve, only to later relent and focus their efforts on the funding relief provisions.[124]

Interest groups did not change the portion of the PBGC reform debate involving the raising of per-participant premiums. Policymakers seemed to widely recognize that because these premiums had not been raised since 1991, they needed to be increased substantially.[125] By increasing PBGC's revenue base, raising the base premium provided legislators with the easiest and most effective way of improving the PBGC's financial position. In essence, such a rise in the premium would constitute a tax increase on employers since these monies would not fund employers' own plans but would be counted as on-budget revenues in the overall federal budget.[126] In the end, the increase to thirty dollars per participant, indexed to wage growth in the Deficit Reduction Act, was left unchanged and remains the law.[127] The failure of special interests to alter these rates indicates that they were primarily fighting at the margins.

This account of the influence of special interests on the recent round of reform is confirmed by detailed accounting estimates of the fiscal effect of the various proposals, bills, and final legislation. One way to evaluate the fiscal effect of special interests is to compare the fiscal solvency estimates at the beginning of the legislative process and at the end of the legislative process. As described above, the administration's initial proposal represented a strong fiscal solvency approach. The administration's proposal can be contrasted with the final House and Senate bills that had been altered during the legislative process involving special interests.

COFFI estimates indicate an insignificant difference between these reform proposals measured by the ultimate reduction in the PBGC's long-term deficit. COFFI estimated that the administration's proposal would have reduced the price of a bailout from $92 billion to $45 billion, the House bill would have reduced the price of a bailout to $49 billion and the Senate bill would have had an outcome comparable to the House bill.[128]

---

[123] This seems to be an instance where the power of an executive agency, in this case the PBGC, and the President's general role in the budgetary process created an incentive for the administration to resist business lobbies. *See* Elizabeth Garrett, *Accountability and Restraint: The Federal Budget Process and the Line Item Veto Act* 20 CARDOZO L. REV. 871, 916–19 (1999) (discussing the interbranch tension in the line item veto context).

[124] *See* Emily Heil, *Faced with Hill Resistance, Airlines Back Off Pension Fix*, NAT'L J.'S CONGRESS DAILY, Aug. 11, 2003, at 1.

[125] *See, e.g.*, Jerry Geisel, *Pension Reform Bill Clears Senate Hurdle, but Struggle Looms*, BUS. INS., Nov. 21, 2005, at 1.

[126] *See* CBO GUIDE TO UNDERSTANDING, *supra* note 3, at 11.

[127] *See* 29 U.S.C. § 1306(a)(3)(A) (2006); *see also* In Brief, BUS. INS., Oct. 30, 2006, at 1.

[128] Press Release, COFFI, PBGC: Legislation May Not Restore Solvency 1 (June 20, 2005), *available at* http://www.coffi.org/pubs/PBGC%20Legislation%20May%20Not%20

The $4 billion difference between the estimates of the administration's initial proposal and the estimates of the House bill does not demonstrate a large interest group effect on the overall fiscal solvency of the PBGC in the legislative process. This also makes it clear that industry-specific provisions included in the Senate bill and not in the House bill did not make a very large difference in the ultimate amount of contributions required by the system. Because these reductions in bailouts necessarily came from greater funding on the part of businesses, interest groups may have made a small difference, but they certainly were not in control of the policy outcome.

Further evidence of the limited effect of lobbying comes from the changes in contribution requirements. The PBGC estimates that the administration's proposals would have generated $1 trillion in contributions over the next ten years or 110% of what previous law would have provided.[129] By comparison, the PBGC estimates that both the House and Senate bills would have led to about $843 billion in funding over the next ten years, equivalent to 92% of the funding that would have been provided by the previous law.[130] Pointing to the difference between 100% and 92%, one could argue that these numbers indicate the significant effect of business lobbies in reducing required plan contributions over the course of the legislative process.

However, such an argument fails upon further investigation. This is because a notable difference arises with respect to the transition periods of each proposal. This is something which lobbyists have largely ignored, at least in public, but which creates a large difference because of the shorter (ten-year) window assumed by the PBGC estimates.[131] This is true because of the phase-ins contained in the House and Senate bills. The COFFI estimates above provide a preferable longer-term picture, ten years and beyond, and should be viewed as the definitive estimates. Some argue that projecting past ten years is pointless because there is likely to be another round of legislation in the next decade; however, this round of legislation will comprise the baseline from which future rounds of legislation depart.

Restore%20Solvency.pdf; Press Release, COFFI, PBGC: Senate Finance Reform Bill 1 (July 25, 2005), *available at* http://www.coffi.org/pubs/PBGC%20Senate%20Finance%20Bill.pdf.

[129] Pension Benefit Guaranty Corporation, The Impact of Pension Reform Proposals on Claims Against the Pension Insurance Program, Losses to Participants, and Contributions 8 (2005), *available at* http://www.pbgc.gov/docs/impact_of_reform_proposals_1005.pdf.

[130] *See id.*

[131] *See id.* at 15.

IV. CONCLUSION

Many accounts of the pension reform process have focused on the legislative maneuvers of various interest groups. These descriptions fail to explain the effect of such legislation on key stakeholders. By focusing on the structural changes in recent reform, this analysis highlights the relative ineffectiveness of major lobbyists in achieving their stated goals.[132] In particular, the adjustment of per-participant PBGC premiums reflects a serious change in PBGC finances, the fiscal importance of which outweighs the small changes in funding requirements for specific companies represented by interest groups. Workers, under pressure due to recent corporate bankruptcies, should see the passage of the Pension Protection Act as a substantial improvement in the protection of their defined benefit pension plans.

—Daniel B. Klaff[*]

---

[132] *See supra* Part III.

[*] J.D. Candidate, Harvard Law School, Class of 2008. B.S., Cornell University School of Industrial and Labor Relations, 2004. I wish to thank Richard Fallon, Daniel Halperin, Howell Jackson, Christine Jolls, Daryl Levinson, Martha Minow, Joseph Singer, Matthew Stephenson, Adrian Vermeule, and Elizabeth Warren for their generous feedback and support. I would also like to acknowledge support from the Harvard Law School Summer Academic Fellowship program.