HARVARD JOURNAL
on
LEGISLATION

VOLUME 30, NUMBER 2 SUMMER 1993

THIRTIETH ANNIVERSARY ISSUE

ARTICLES

STATUTE

COMMENT

RECENT DEVELOPMENTS AND LEGISLATION

PREFACE

ERWIN N. GRISWOLD*

The *Harvard Journal on Legislation* is now completing its thirtieth volume. Since I was present at the creation, the editors of this volume have kindly asked me to contribute a preface for this issue. This is not a new experience for me, since I was asked to write a preface for the first issue,[1] and I also contributed a preface to the concluding issue of Volume 20.[2]

The thirty years since a group of students came to my office in Langdell Hall and sought approval for a journal on legislation have gone by very quickly. It is sobering, though, to realize that those students are now in their mid- or late fifties, and are busy as thoroughly experienced lawyers, or, perhaps, as judges or administrators, or are engaged in other legally related activities.[3] I feel sure that, without exception, they regard their work in overseeing the first volume of the *Journal on Legislation* as an important part of their legal educational experience.

The *Journal on Legislation* has, in its thirty years, performed at least two useful functions for the legal community. First, it has contributed to the collection and then dissemination of materials associated with legislation. Second, it has made an opportunity for law review type experience available to a considerable body of students. The number of students who have had

---

* Mr. Griswold was Dean of the Harvard Law School from 1946 to 1967 and was a strong supporter of the *Journal*'s founding in 1963. He served as Solicitor General of the United States from 1967 to 1973 and is currently a partner with Jones, Day, Reavis & Pogue in Washington, D.C.

[1] Erwin N. Griswold, *Preface*, 1 HARV. J. ON LEGIS. 3 (1964).

[2] Erwin N. Griswold, *Preface: The Explosive Growth of Law Through Legislation and the Need for Legislative Scholarship*, 20 HARV. J. ON LEGIS. 267 (1984).

[3] One of these I remember particularly. Tedson J. Meyers was the founding president of the Harvard Student Legislative Research Bureau, and the leader of the group that provided the motivating force for the establishment of the *Journal*. After graduation from Law School in 1953, he served for two years as a lieutenant in the Marine Corps in Korea. He later moved to Washington, D.C., where he has had extensive experience in public service and in private practice. Among other things, he was a member of the City Council in D.C. from 1972–75 and was instrumental in drafting and obtaining the adoption of legislation for the benefit of handicapped drivers. Later, as president of the Cosmos Club, he skillfully brought about the admission of women as members, after a decade of controversy.

The *Journal*'s Board became active in the fall of 1963, and the first issue of the *Journal* was published in January, 1964. The first president of the *Journal*'s Board was Karl Vance Hart, J.D. 1964, now practicing in Ocala, Florida.

editorial and writing experience through the *Journal* during the thirty years of its life must now be several hundred.

There are some risks in all of this, and I think I see some traces of these risks in the current issue. The articles are perhaps less and less devoted to the specifics of legislation, including analysis of recently enacted statutes. They are more and more like those of any other law review. They sometimes deal in great, perhaps excessive, detail with problems that arise because of legislation, but where the legislation is primarily part of the background, sometimes little more than incidental. Ten years ago in the *Journal*, I discussed the growing importance of legislation to the development of law.[4] Legislation remains an important aspect of legal growth, and the *Journal* should attempt to retain its focus on legislation.

Another problem with the articles in many law reviews in recent years is that they are not as well written as they might be. Many articles tend to be esoteric, using ponderous language that often seems to spin off into the clouds. "Deconstruction" is often one of the favorite words, along with "normative" and other terms rarely used by lawyers. Many of these articles appear to be vehicles through which junior faculty members seek to obtain tenure, and it is currently fashionable to do this in highly theoretical terms. Rarely do such articles today contain material that might be of real help to a lawyer engaged in the active task of practicing law.[5]

I think that this issue of the *Journal* measures up very well, though perhaps with some qualifications. Two of the articles are especially appealing to me. First on the list I would put the article on "Asbestos: A Multi-Billion-Dollar Crisis," by Professors Edley and Weiler. Although this problem has long been with us, and much has been written about it, it is a massive

---

[4] *See* Griswold, *supra* note 2, at 268–71.

[5] This may indeed be a reflection of current trends in law schools that like to be called "elite." For many years, the Catalogue of the Harvard Law School contained a statement that, I believe, was formulated by Dean Pound, which said: "The School seeks as its primary purpose to prepare for the practice of the legal profession wherever the common law prevails." HARVARD LAW SCHOOL, THE LAW SCHOOL CATALOGUE 11 (1961–1962). It now reads: "The School prepares its graduates to serve as lawyers and in law-related roles. Its goal is to provide comprehensive and enlightened training for law practice, for public service at the local, state, federal and international levels, and for law teaching and legal scholarship. The School, through its faculty, students and graduates, seeks to make substantial contributions towards solving society's complex problems. Graduates of the School work in many different settings during their professional lives." HARVARD LAW SCHOOL, HARVARD LAW SCHOOL CATALOG 17 (1992–1993).

It still seems to me that a person cannot do these related tasks well unless he is first a good and well-trained lawyer.

one, both in terms of its social impact and in terms of its effect on our court system. The Edley-Weiler article gives an excellent summary of the problems, offers possible solutions, and confronts the difficulties with those solutions. Viewed abstractly, the obvious solution is to establish something analogous to workers' compensation. It was, indeed, a sort of miracle that workers' compensation, taking workers' claims out of court, became established nationwide between 1906 and 1915. The plaintiffs' lawyers were not as strong an influence on the legislatures then, and they apparently did not foresee the impact of the statutes. Legislative difficulties did arise with respect to the Safety Appliance Act[6] of 1893 and the Employer's Liability Act[7] of 1908. They both provide for dual jurisdiction, thus preserving the right to sue in the federal court, including the right to jury trial. Since that time, the strength of the plaintiffs' bar has clearly increased.[8] This is recognized by the authors, and they acknowledge that the chance of enacting a statute of the workers' compensation sort to deal with asbestos claims is slight. It seems equally hard to be optimistic about the chances of a comprehensive judicial solution. Perhaps it can be done by a straight-arm method, but our courts, understandably, are not very adept—or successful—at that.

Another article in this issue that appeals to me is the one by Professor Durchslag on "Property Tax Abatement for Low-Income Housing." As his subtitle indicates, this may be a matter "whose time may never arrive." The article is, however, a thoughtful analysis of a real problem, which might be susceptible to resolution or improvement through legislation.

Professor Dowd, the author of the article on family leave, has been a frequent contributor to the *Journal on Legislation*[9] and other Harvard publications.[10] Professor Dowd writes well and has a stimulating point of view. The problems she discusses are important ones, and their appropriate answers are far from clear. I hope that she will continue making her thoughtful contribution,

---

[6] Safety Appliance Act, ch. 196, 27 Stat. 531 (1893) (codified as amended at 45 U.S.C. §§ 1–43 (1988)).

[7] Employer's Liability Act, ch. 149, 35 Stat. 65 (1908) (codified as amended at 45 U.S.C. §§ 51–60 (1988)).

[8] The "plaintiffs' bar" is by no means the same as the "trial bar" to which the authors refer.

[9] *See* Nancy E. Dowd, *Envisioning Work and Family: A Critical Perspective on International Models*, 26 HARV. J. ON LEGIS. 311 (1989).

[10] *See* Nancy E. Dowd, *Work and Family: The Gender Paradox and the Limitations of Discrimination Analysis in Restructuring the Workplace*, 24 HARV. C.R.-C.L. L. REV. 79 (1989).

and that she may devote some of her intellectual energy to the problems raised by permeating regulation which may materially change the nature of life in this country.

From this point on, I have a little more difficulty. The article dealing with medical care systems for children by Dr. Barden, joined by a substantial list of other participants, is in many ways a pioneering sort of article. It concludes with a "model statute" that brings it more nearly in line with the original concept of the *Journal on Legislation*. It is interesting to note that "virtually all of the provisions of this model Act" have already "been enacted—unanimously—in New Jersey." Thus, the article can be of immediate use in implementing the New Jersey law and in sponsoring similar statutes in other states. My chief reservation is the fact that the proposed solution provides further governmental regulation which, at some point, can overwhelm the "freedom" that has been one of the great aspects of American life. Within my own lifetime, there once was a widespread feeling that none of these matters was appropriate for action by the federal government.

The Reitze-Needleman article on air pollution is massive, sometimes ponderous, and not always easy to follow because of the numerous acronyms. It would be more effective to divide it into separate articles, or, even better, to expand the discussion somewhat and publish it as a book—which would, of course, include a convenient but comprehensive table of acronyms. I do not want in any way to belittle the thorough scholarship of the work done by the authors. I am only concerned about the ponderous nature of what they have produced as an item for publication in the *Journal*.

Despite these observations and queries, I readily conclude that this is an excellent issue of the *Journal*, dealing with problems that were not foreseen when the *Journal* developed from an idea into reality some thirty years ago. Although the *Journal* has not managed fully to avoid the risks attendant upon modern legal scholarship, it has managed to contribute valuable information to the legal community on a topic of increasing importance: legislation. Of course, I cannot foresee the questions that will be discussed in the *Journal* after another thirty years. I have great confidence, though, that the *Journal*—building on the experience of its first thirty years—will still be making important contributions to the field of legislation and to the training of students at the Law School.

ARTICLE

FAMILY VALUES AND VALUING FAMILY: A BLUEPRINT FOR FAMILY LEAVE

NANCY E. DOWD*

In the spring of 1989, the Harvard Journal on Legislation *sponsored a symposium on legislative approaches to work and the family. Scholars and political figures from across the country gathered to discuss families, child care, and the government's role in easing the conflict between work and family. As part of that conference, the author used the examples of Sweden and France to present a comparative analysis of family-support policies from an international perspective.*

Four years later, Professor Dowd now revisits the subject of family leave, but on the domestic level. The transition to a Democratic administration has already brought changes to federal policy in this area. In this Article, she combines a history of family-leave legislation with a vision for the future. Although the new family-leave legislation is a positive step, she argues that a comprehensive family policy must include more than just parental leave. To that end, she provides insights and suggestions for how to craft a family policy that responds to the nation's needs.

The time has come for family leave. Weeks after the presidential inauguration, Congress passed and President Clinton signed the Family and Medical Leave Act of 1993.[1] But what does it look like? Over the years, family leave has steadily gained support as the substance of leave proposals has steadily shrunk.[2] Moreover, the concept of family leave was limited even before this shrinkage occurred. Most importantly, leave proposals have never included wage replacement. In addition, the proposals have offered significantly less time for leave than child-development specialists recommend for childbirth and

* Professor, University of Florida College of Law. A.B., University of Connecticut, 1971; M.A., University of Illinois, 1973; J.D., Loyola University of Chicago School of Law, 1981. This Article is dedicated to Zachary.

[1] Pub. L. No. 103-3, 107 Stat. 6. During the 1992 election campaign, family leave was used to highlight party differences. To embarrass a Republican president claiming to be committed to "family values," the predominantly Democratic Congress enacted family-leave legislation sure to be met with a presidential veto. Similar tactics had been used in the 1988 campaign.

[2] One recent poll showed nearly 60% of those polled favored family leave. Women supported leave more strongly than men (62% versus 51%), and working mothers supported leave more strongly than working fathers (66% versus 52%). Blacks supported family leave much more strongly than whites (78% versus 53%). Support shifted from a majority in favor to approximately 50% among persons whose incomes exceeded $50,000. Felicity Barringer, *In Family Leave Debate, A Profound Indifference*, N.Y. TIMES, Oct. 7, 1992, at A1, A22. For discussion of the legislative history of family-leave proposals, see *infra* notes 4–14 and accompanying text.

adoption, and have not provided leave for parents to care for a sick child with an ordinary illness that can be treated at home. The enacted leave legislation does not depart from this limited, constricted pattern. The leave provided is paltry in light of the psychological and developmental needs of family, and the social and political implications of ongoing work-family conflict.

In this Article, I sketch a different blueprint for family leave. It is an opportunity to revisit my analysis of leave and family policy included in a symposium sponsored by the *Harvard Journal on Legislation* ("JOL") in 1989.[3] My premises are that support of children is a social as well as an individual responsibility, and that children are a social as well as an individual benefit. Furthermore, meaningful support must be sensitive to context, that is, to the class differences and diversity of family structures within which children are raised. The shape and character of family support inevitably reflects our vision of families. That vision may consist of unexamined assumptions or explicit moral, political, and social goals. Attention to the connection between context and vision enlightens policy. Failure to consider it may sabotage meaningful change and simply recreate powerful hierarchies of gender, race, and class, while rendering the promise of support empty and meaningless.

But even more importantly, we must go beyond family leave. My fear about the enactment of family-leave legislation is not only that we will be satisfied with too little, but also that we will consider family leave a sufficient family policy. It is tempting to set family leave as the boundary. It is, after all, the simplest of family policy issues, dealing with time-bound work-family conflicts rather than more complex structural issues. But we accomplish very little if we ensure the initial foundation of families, or support them during extreme emergencies, only to

---

[3] Nancy E. Dowd, *Envisioning Work and Family: A Critical Perspective on International Models*, 26 HARV. J. ON LEGIS. 311 (1989) [hereinafter Dowd, *Envisioning Work and Family*]. In that article, I focused on the insights to be gained from the comparative analysis of family-support policies, specifically using the examples of Sweden and France. What has happened since then is both progression and regression on the issue of family leave, with virtually no movement toward a comprehensive family policy. This Article focuses on family leave as the likely starting point for family policy. I also emphasize once again the essential vision of family policy that was at the core of my 1989 article. *See also* Nancy E. Dowd, *Work and Family: Restructuring the Workplace*, 32 ARIZ. L. REV. 431 (1990) [hereinafter Dowd, *Restructuring*]; Nancy E. Dowd, *Work and Family: The Gender Paradox and the Limitations of Discrimination Analysis in Restructuring the Workplace*, 24 HARV. C.R.-C.L. L. REV. 79 (1989) [hereinafter Dowd, *Gender Paradox*].

undercut them by leaving intact a structure that exacerbates work-family conflicts. What is needed is both integrated policy-making and a broader scope of policy. Little has been done, for example, to coordinate family leave with the two major pieces of family policy that we do have: child-care and welfare legislation. The second part of this Article suggests connections that need to be made and also outlines the necessary expansion of family policy. The vision underlying a broader policy is critical. Determining the choices and alternatives, as well as the progression of policy, is difficult. Yet, as virtually the last advanced industrialized country to consider family policy, the United States can benefit from the structures and experiences of other countries.

It is perhaps misleading to call this Article a blueprint. In some areas I have specific recommendations, while in others, I simply raise questions and explore alternatives. In Part I, I briefly outline the legislative history of family leave, discuss the importance of the vision underlying family leave, and make specific policy recommendations. In Part II, I expand upon the recommendation that family leave must be part of a comprehensive family policy by sketching the elements of such a policy and by discussing two pieces of a comprehensive policy: child care and family income supports.

I. Family Leave

A. *A Brief History and Analysis*

Representative Patricia Schroeder (D.-Colo.) first introduced family-leave legislation in 1985.[4] The original bill provided for up to eighteen work weeks of parental leave in any two-year period, and up to twenty-six work weeks of disability leave in any one-year period.[5] The right to leave was protected by an anti-discrimination provision that provided for equitable relief, compensatory and punitive damages, and attorneys' fees and costs.[6] In addition, the legislation proposed creating a commission to recommend means to provide wage replacement for both

[4] Parental and Disability Leave Act of 1985, H.R. 2020, 99th Cong., 1st Sess..
[5] *Id.* §§ 102–103.
[6] *Id.* § 107(d).

types of leave.[7] The bill was reintroduced in subsequent sessions with little success until 1990, when election year politics propelled its passage.[8] As passed, the bill provided a combined total of twelve work weeks for parental, family, and disability leave.[9] Employees also could be required to substitute paid vacation, personal, or family leave for parental or family leave.[10] The bill's anti-retaliation provision allowed either charges under the mechanisms of the Fair Labor Standards Act or a civil action.[11] The bill provided for equitable remedies (including cease and desist orders, permanent or temporary injunctions, temporary restraining orders, or any other relief the court would deem appropriate) and monetary damages limited to lost wages and benefits, including interest on the total monetary damages, plus liquidated or consequential damages.[12] President Bush vetoed the legislation in the summer of 1990.[13] Again under the pressure of national elections, the 1991 family-leave bill, substantially similar to the 1989 bill, was passed in the fall of 1992, and vetoed by President Bush.[14]

President Clinton made family leave a legislative priority at the beginning of his administration. Passing leave legislation became symbolic of the end of legislative gridlock, and the beginning of executive-legislative cooperation. The new Democratic administration did not, therefore, reopen the leave debate, but rather characterized the legislation as a model of bipartisan consensus. The enacted legislation, therefore, was based on the previously vetoed 1990 and 1992 acts.

---

[7] *Id.* § 201.

[8] *See* 136 CONG. REC. S7979–8008 (daily ed. June 14, 1990); 136 CONG. REC. H2198–240 (daily ed. May 10, 1990).

[9] H.R. 770, 101st Cong., 1st Sess. (1989).

[10] *Id.* § 102.

[11] *See id.* § 105; Fair Labor Standards Amendments of 1989, 29 U.S.C. §§ 201, 216–217 (Supp. II 1990).

[12] H.R. 770, 101st Cong., 1st Sess. § 109 (1989).

[13] President's Message to the House of Representatives Returning Without Approval the Family and Medical Leave Act of 1990, 26 WEEKLY COMP. PRES. DOC. 1030 (June 29, 1990).

[14] S. 5, 102d Cong., 2d Sess. (1991). Compared with the 1989 version, the 1991 bill provided even fewer damages by removing even limited compensatory damages and expanding the good faith defense of employers against liquidated damages. With the exception of disability leave, the 1991 bill did not allow leave to be taken intermittently, and required that leave be used within 12 months of birth or adoption/placement. If both parents worked for the same employer, only a total of 12 weeks of leave was allowed. Unlike the 1989 bill, the 1991 bill provided for a study commission, but it was limited to an impact study. The 1991 bill did not commission a study of merging parental, family, and disability leave or of developing a broader family policy.

The Family and Medical Leave Act of 1993[15] provides aggregate annual leave of twelve weeks, which may be composed of parental leave, leave to care for a seriously ill immediate family member, and/or leave for the employee's own serious illness.[16] The Act covers employers of fifty or more employees, and an employee must have worked for twelve months for a minimum of 1250 hours (slightly more than twenty hours per week for fifty weeks) to be eligible for leave.[17] Leave is unpaid, although if an employer provides any form of paid leave the employer may require the employee to utilize that leave during the period of entitlement to family and disability leave.[18]

The Act concentrates primarily on three areas: parental leave, eldercare leave, and personal disability leave. The goal of parental leave[19] has been to eliminate the stark choice between work and family dictated by the lack of job security guarantees for parents employed in the wage labor market at the time of childbirth or adoption, or in the event of a severe medical emergency of a child. Because childbirth or adoption are more common, those events have motivated the push for parental leave. The goal of childbirth or adoption leave has been to provide for essential time to create a solid parent-child bond. In contrast, the goal in providing leave for a child's severe medical emergency has been to ensure the availability of parents to support children through crises, as well as to reduce financial and other stress for parents, by guaranteeing that they will have a job to return to when they need it most. Indeed, the tragic circumstances of childhood illnesses have provided some of the most compelling testimony of the need for such legislation.

Eldercare leave[20] closely parallels the goals of severe health care emergency leave to care for children, except that an adult

---

[15] Pub. L. No. 103-3, 107 Stat. 6.

[16] *Id.* § 102(a), 107 Stat. at 9.

[17] *Id.* § 101, 107 Stat. at 7–9.

[18] *Id.* § 102, 107 Stat. at 9–11.

[19] *See generally* Dowd, *Gender Paradox*, *supra* note 3, at 119–28; David K. Haase, *Evaluating the Desirability of Federally Mandated Parental Leave*, 22 FAM. L.Q. 341 (1988); Deborah L. Rhode, *Occupational Inequality*, 1988 DUKE L.J. 1207; Marjorie Jacobson, Note, *Pregnancy and Employment: Three Approaches to Equal Opportunity*, 68 B.U. L. REV. 1019 (1988); Stephen Keyes, Note, *Affirmative Action for Working Mothers: Does* Guerra's *Preferential Treatment Rationale Extend to Childrearing Leave Benefits?*, 60 FORDHAM L. REV. 309 (1991); Wendy S. Strimling, Comment, *The Constitutionality of State Laws Providing Employment Leave for Pregnancy: Rethinking* Geduldig *After* Cal Fed, 77 CAL. L. REV. 171 (1989); Deirdre A. Whittaker, Note, *Should We Have a National Leave Policy?: A Survey of Leave Policies, Problems and Solutions*, 34 How. L.J. 411 (1991).

[20] *See generally* Larry Polivka, *In Florida the Future Is Now: Aging Issues and*

child takes the leave in order to care for an ailing parent. Like leave for a sick child, eldercare leave is intended to ensure workers the ability to provide care over a limited time period in cases of severe medical emergency.

Personal disability or medical leave[21] is not strictly limited to family support. Rather, the goal of personal medical leave is to provide employees with job protection in the event of their own physical or mental disability.[22] Job security is not limited to those with families; it is a benefit or right of all employees covered by the legislation. Disability leave fills a significant gap in existing disability benefits: the lack of job security despite health insurance and wage replacement benefits.[23] Disability leave is, therefore, an individual benefit that crosses the line from family-support policies. Families benefit because dependent family members suffer from loss of family income due to a wage-earner's disability, but such benefits are equally or more critical to single people who may have no other income support.

Policies in the 90s, 18 FLA. ST. U. L. REV. 401 (1991); Nadine Taub, *From Parental Leaves to Nurturing Leaves*, 13 N.Y.U. REV. L. & SOC. CHANGE 381 (1984–1985); Steven K. Wisensale & Michael D. Allison, *An Analysis of 1987 State Family Leave Legislation: Implications for Caregivers of the Elderly*, 28 GERONTOLOGIST 779 (1988).

[21] *See generally* Joan Vogel, *Containing Medical and Disability Costs by Cutting Unhealthy Employees: Does Section 510 of ERISA Provide a Remedy?*, 62 NOTRE DAME L. REV. 1024 (1987). On leave for pregnancy-related disability and the issues surrounding maternity leave and the definition of equality, see generally Nancy E. Dowd, *Maternity Leave: Taking Sex Differences into Account*, 54 FORDHAM L. REV. 699 (1986); Lucinda M. Finley, *Transcending Equality Theory: A Way out of the Maternity and the Workplace Debate*, 86 COLUM. L. REV. 1118 (1986); Herma Hill Kay, *Equality and Difference: The Case of Pregnancy*, 1 BERKELEY WOMEN'S L.J. 1 (1985); Wendy W. Williams, *Equality's Riddle: Pregnancy and the Equal Treatment/Special Treatment Debate*, 13 N.Y.U. REV. L. & SOC. CHANGE 325 (1984–1985).

[22] This is where pregnancy-related leave or maternity leave fits within family leave. Folded within disability leave, maternity leave is included as part of a gender-neutral benefit, as opposed to providing for it separately. The gender-neutral approach potentially has a disparate impact on women, however, since those women disabled by non-pregnancy-related ailments may have less disability time available to them than men if they have already taken time in connection with childbirth. If a woman needs significant time to recover from an unusually difficult pregnancy or a Caesarean section, for example, then little or no disability leave may be available to her for non-pregnancy-related ailments. On the other hand, separating disability leave from parental leave is preferable to collapsing maternity leave and childrearing leave. *See infra* notes 49–52 and accompanying text.

[23] It is important to note, however, that wage replacement or payment of expenses in the form of sick leave, short-term disability, and health insurance benefits is far from universal, and for most employers is not required. Approximately two-thirds of employers have sick-leave policies. Estimates of employers who provide disability insurance for short-term disability range from 50% to 95%. Dowd, *supra* note 21, at 712 nn.56–57; *see also* U.S. BUREAU OF THE CENSUS, THE STATISTICAL ABSTRACT OF THE UNITED STATES 1992, at 417 [hereinafter STATISTICAL ABSTRACT]. Health insurance is one of the most widely offered benefits, provided by 92% of medium and large firms in 1989 and by 69% of small firms in 1990. *Id.*

Leave for any of these purposes under the 1993 Act is extremely limited. The time benefit is seriously inadequate to resolve the work-family conflicts that the legislation is designed to ameliorate or eliminate. With respect to family formation, for example, most experts agree that a minimum of six months is needed to create solid family bonds at birth or adoption.[24] The minimum length of parental leave in other countries that provide this benefit ranges from five to eighteen months.[25] The twelve-week limit of the 1993 Act is inadequate for family formation, and some employees may have considerably less leave time if they have been required to use some of their twelve-week total to cover eldercare and personal disability as well.

Even more fundamentally, the lack of wage replacement profoundly affects the impact of family leave across divisions and intersections of gender, race, and class. The lack of wage replacement has the strongest effect along class lines. Lower-income families can least afford any loss of income precisely when family expenses are increasing. But the lack of wage replacement strikes very broadly, given the predominance of dual-wage-earner families, and the essential contribution of both wage-earners to family income.[26] For most dual-wage-earner families, family leave without wage replacement is a hollow right, at most an ultimate safeguard to prevent job loss, but hardly a support structure to ensure healthy family formation.

These economic implications have race and gender impact as well. Because a higher proportion of nonwhite populations have low incomes, leave provisions benefit white populations to a greater extent. The disproportionate income pattern of women and men virtually ensures that women will be the only employ-

---

[24] *See* THE YALE BUSH CTR. IN CHILD DEV. & SOCIAL POLICY, ADVISORY COMM. ON INFANT CARE LEAVE, RECOMMENDATIONS OF THE YALE ADVISORY COMMITTEE ON INFANT DAY CARE LEAVE, *reprinted in* Daily Lab. Rep. (BNA) No. 231, at F-1 (Dec. 2, 1985) [hereinafter YALE BUSH CTR.].

[25] Barringer, *supra* note 2, at A13. The most liberal policy is Sweden's, which provides full-time leave until the child is 18 months old, and the right to part-time work until the child is eight years old. *See* Karin Widerberg, *Reforms for Women—On Male Terms—The Example of the Swedish Legislation on Parental Leave*, 19 INT'L J. SOC. L. 27, 28 (1991).

[26] Dual-wage-earner families are the predominant family form in the workplace and constitute nearly 51% of married couple families. *See Number of Two-Earner Families Rose in Last Year, BLS Reports*, Daily Lab. Rep. (BNA) No. 147, at B-9 (July 31, 1990).

For an extended discussion of the impact of no wage replacement, see Maria O'Brien Hylton, *"Parental" Leaves and Poor Women: Paying the Price for Time Off*, 52 U. PITT. L. REV. 475 (1991).

ees who will exercise the right to leave, to minimize the loss of family income and to maximize wage replacement that they may be entitled to under elective disability policies.[27]

The lack of wage replacement also makes this a virtually meaningless right for single-parent families. Again, the leave entitlement may provide a shield against job loss, but it does not provide real family support. This has both race and gender implications as well. While single-parent families currently constitute thirty percent of all American families, this is the dominant family form among African-Americans.[28] Moreover, single-parent families in the black community are dominantly non-marital families, not families formed by divorce.[29] If leave benefits are limited to legal parents and/or by marital status, then non-marital families will suffer disproportionately as compared to single-parent families of divorce.

The gender consequence of ignoring single-parent families is the disproportionate impact on women, who are the actual heads of household of most single-parent families, regardless of the genesis of the single-parent family. The lack of meaningful leave benefits thus impacts most heavily women who, like racial minorities, continue as a group to occupy a highly disadvantageous labor market position.

The most highly disfavored group then, under the current structure of family leave, is composed of low-income single-parent black women. Those most likely to benefit from leave in its current form are upper-income white males with wage-earning spouses with lower incomes. The limitations of family leave in its current form make it a benefit most advantageous to

---

[27] *See* Dowd, *Restructuring*, *supra* note 3, at 451–56, 461–68 (discussing the effects of race and gender in the work-family relationship). This is so despite the avowed purpose of the Act to ensure equal opportunity for women. Family and Medical Leave Act of 1993, § 2(a)(5)–(6), (b)(4)–(5), 107 Stat. at 6–7.

[28] U.S. BUREAU OF THE CENSUS, HOUSEHOLD AND FAMILY CHARACTERISTICS 10 (1990) [hereinafter HOUSEHOLD AND FAMILY CHARACTERISTICS]; John Ermisch, *Demographic Aspects of the Growing Number of Lone-Parent Families*, *in* LONE-PARENT FAMILIES: THE ECONOMIC CHALLENGE 27, 29 (Elizabeth Duskin ed., OECD Social Policy Studies No. 8, 1990). The Act acknowledges the growth of single-parent families, but otherwise makes no special provision for single parents. *See* Family and Medical Leave Act of 1993, § 2(a)(1), 107 Stat. at 6–7.

[29] HOUSEHOLD AND FAMILY CHARACTERISTICS, *supra* note 28, at 10.

For an alternative model, see the District of Columbia Family Leave Act, which liberally defines family member and family relationships to include homosexual relationships, extended families, and short-term assumption of parental duties even without legal action. D.C. CODE ANN. § 36-1301 to -1317 (Supp. 1991). *See generally* Eleanor Nace, *Time Out: An Overview of the District of Columbia Family and Medical Leave Act of 1990*, WASH. LAW., Sept.–Oct. 1991, at 43, 44.

middle- and upper-income families that can afford unpaid leave, in effect heightening existing class, race, and gender advantages.

The inadequacies of family leave are not purely political in origin. Rather, I think they are tied to the vision underlying family leave. As I outlined in my earlier *JOL* article, the shape of leave policy is powerfully driven by the reasons leave is enacted and the vision of what it is to accomplish. These reasons are often strongly interrelated.[30] In the case of the countries profiled in my prior article, Sweden and France, the genesis of family policy was either the desire to draw women into the workforce, or to support families to increase the birth rate due to demographic trends. The vision was one of equality, or one of difference, both focused on women.[31]

The genesis for American policy seems somewhat different. Certainly it is, as with other countries, the shift of women into the wage labor market that has generated interest in the issue. To that extent, as with other family-leave policies, it is women-generated. But the purpose of family leave seems neither to pull more women in, nor to encourage them to stay at home to increase demographic balance. Rather, it seems to be a reflection of women's political power and economic position. Family leave has been advocated as a women's rights issue. The sheer number of women in the workforce, and their concomitant increasing presence in politics, means their concerns are now heard. The concerns of middle- and upper-class women, which are listened to in a way that the concerns of poor women and women of color are not, include the competing concerns of work and family. Furthermore, the economic position of women, as sole or significant income earners, has intensified the concern over this issue and also generated support among men, as an essential support for family incomes. The legislative history also suggests that men support gender-neutral family leave, even if weighted toward women's exercise of the right to leave, as an opportunity for fathering in a different way from traditional patriarchal models.

The vision of family leave is less clear. A cynic might argue that the vision is the reinstitutionalization of patriarchy, as well as the preservation of racial and class privilege. The consequences of the legislation in terms of gender, race, and class

---

[30] *See* Dowd, *Envisioning Work and Family*, *supra* note 3, at 339–44, 346–48.
[31] *Id.* at 316–19, 328–31.

might support such an analysis, but it hardly has been an explicit vision (with, perhaps, the exception of class). The strongest vision seems to be ensuring women's wage labor market position, while casting the policy in gender-neutral terms as a means toward the desirable goal of equality. But it is not at all clear what this equality looks like in practical terms.

The experiences of Sweden and France tell us that these goals are exceedingly difficult to attain. Even more significantly, in neither country, despite radically different visions expressed in their leave policies, has the economic position of women dramatically improved.[32] On the other hand, it is important to acknowledge that these policies have achieved other significant gains, by providing time to care for and raise children, and some economic support for families, all of which has materially improved the status of women. What is intriguing is that the support that theoretically would seem to diffuse work-family conflict and permit greater work and home equality may only have contributed toward the first goal, but not the latter, and this has been achieved primarily for women, not for men. A recent study found that women who did not have children did not do significantly better in their careers than women who did, suggesting that continuing inequality has something to do with sex discrimination against women in general, not just disadvantages resulting from attempting to balance work and family.[33]

B. *Visions and Blueprints*

1. Vision

I think part of the lesson from these experiences is to ask whether the goal of the policy is to focus on women's current status or to set a goal of what women's status should be—or both. This is complicated and difficult because one must balance generalizations about women with differences among women. Even more difficult, however, is envisioning what women's status should be. Goals often implicitly enshrine a single model. Is

[32] *Id.* at 324–28, 335–37.

[33] Jane E. Brody, *For the Professional Mother, Rewards May Outweigh Stress*, N.Y. TIMES, Dec. 9, 1992, at C16 ("[O]n the whole childless women, whether married or single, devoted more hours to their careers but earned no more than married women with children and were less satisfied with their lives."). This was a Canadian study of 1123 Canadian career women.

it possible to promote diversity, a range of combinations of work and family, rather than to mandate a shift from traditional gender roles to requiring market work, full-time, for all adults? The latter, I would suggest, values work but not family.[34]

What seems to be essential to improving women's status is to improve their ability to make real choices, rather than culturally or economically determined choices. Economic independence or self-sufficiency is essential to the power to make choices. Even where the choice is interdependence or dependence there must be the assurance of self-sufficiency as an alternative. Equality may be more easily envisioned and achieved by assuring that those formally equal have the power of self-determination.

I do not mean to suggest that our vision must be limited to women. Yet it must be focused on women given the realities of the distribution of family responsibilities, the structural discrimination of the workplace, and the economic disadvantage of women in the family and in the workplace. It is a difficult and important question, however, to consider how we can make the transition to thinking of family leave, and family issues more generally, in family terms, not in gender terms. What role should government play in changing consciousness?

A second lesson suggested by these policies is to think about family policies from a family or child perspective, rather than from an adult perspective. A family perspective suggests that the goal of these policies is to provide support to ensure that children's caretakers have structures within which to nurture their families, rather than structures that are premised on implicit conflicts between work and family. This has an economic component in that it trades long-term benefits for short-term costs, as it is apparent by all indicators that early investment in children reaps long-term developmental and other benefits that inure both to the individual and to society.[35] Children are the economic and political future, the most essential part of our society; it pays to invest in them for the health of society. The recognition of children as a social benefit and a social respon-

---

[34] The French model, which provides family support to stay-at-home parents as well as working parents, is extremely valuable in this regard. *See* Dowd, *Envisioning Work and Family*, *supra* note 3, at 332–33.

[35] *See* MARIAN WRIGHT EDELMAN, FAMILIES IN PERIL: AN AGENDA FOR SOCIAL CHANGE (1987); NATIONAL COMM'N ON CHILDREN, BEYOND RHETORIC: A NEW AMERICAN AGENDA FOR CHILDREN AND FAMILIES 372 (1991) [hereinafter BEYOND RHETORIC].

sibility is central to a vision of family policy. A corollary is the assumption that gender does not dictate parenting ability, and that to the contrary, the promotion of less gendered parenting is critical to undercutting the reproduction of patriarchy.[36]

A third suggestion is that the vision must be flexible, rather than fixed. In view of the shortcomings of existing models of American family-leave policy, certainly a primary goal must be incorporating gender, class, and race perspectives into policy. Furthermore, policy must reflect family diversity—making policy sensitive to the range of family forms. Flexibility should also be sought in devising benefit schemes that maximize the ability of individuals and their employers to tailor policies to particular situations and work environments. Flexibility should also be sought in the long-term goals of policy. We need to think in terms of progressive implementation and revision, rather than a fixed scheme.

2. A Blueprint for Family Leave

How do these amorphous visions translate into specific policy recommendations? I intend this section to be suggestive, not inclusive or final, and to illustrate the impact of vision on policy. Where I am unsure, I hope to set out the relevant issues and questions.

a. *Pay/Financing.* Pay is probably the single most important consideration in a reformed family-leave policy.[37] Wage replacement is critical to make family leave a real benefit, to present a real opportunity for family formation, and to provide real support through serious family crisis. Wage replacement must provide adequate income for a working parent to forgo income for the benefit period. Calculations can be based on mean or median wage figures. In order to minimize class differences, wage re-

[36] DOROTHY DINNERSTEIN, THE MERMAID AND THE MINOTAUR: SEXUAL ARRANGEMENTS AND HUMAN MALAISE 54, 93–94 (1976); CAROL GILLIGAN, IN A DIFFERENT VOICE: PSYCHOLOGICAL THEORY AND WOMEN'S DEVELOPMENT 7–9 (1982); Nancy Chodorow, *Mothering, Male Dominance, and Capitalism*, *in* CAPITALIST PATRIARCHY AND THE CASE FOR SOCIALIST FEMINISM 83, 84–86 (Zillah R. Eisenstein ed., 1979).

[37] Realistically, pay will probably have to be phased in, but could be done progressively, from no pay to partial pay to a full wage replacement scheme, with full wage replacement being capped at a level sufficient to support essential family income. Alternatively, the phasing-in could be by income level so that those most in need of wage replacement would be entitled to full pay. Such a scheme hopefully would encompass the special income needs of single-parent families.

placement must be sufficient to replace income totally or provide a sufficient replacement to sustain family income loss over a short period of time. An even more sophisticated delivery of benefits would permit exceptional support when need is demonstrated.[38]

In addition to the class concern that low- and middle-income families be able to afford to take leave, pay is also a gender issue. Income replacement encourages men, who generally earn more than women, to take leave. It may be that the gender issue can be more effectively addressed in the assignment of benefits, by use of a capitation policy, rather than by addressing the income differential.[39] But an alternative would be to structure income replacement so that it would be financially possible for the primary income earner to take leave. Another gender concern is that pay be sufficient to support single-parent families, who are predominantly female-headed households.

How should wage replacement be financed? The larger question is how all the expenses of family leave should be financed, including wage replacement. The experiences of other countries can be particularly useful in exploring the implications of various financing systems. Domestically, the most analogous models are the disability systems set up by states that require this coverage of all employers.[40] I believe that financing of family leave, as with all family support benefits, should be socially-based rather than employer-based. The benefits of family support are societal, not limited to employers. But even more significantly, the burdens should not be borne solely by employers.[41] If family leave is a child-oriented policy, instead of an adult-oriented policy, it should reflect the community's interest in the strength and support of families.

Moreover, the costs of providing leave apart from pay ought to be considered. As this varies significantly by industry and position, flexible, individualized support would be ideal if possible. An affirmative action model, rather than a required benefit

---

[38] Combining a set support figure with a need-based figure will result in an inability to take the full leave time.

[39] For discussion of capitation, see *infra* notes 43–52 and accompanying text.

[40] *See, e.g.*, CAL. UNEMP. INS. CODE §§ 2625–2778 (West 1972 & Supp. 1992); HAW. REV. STAT. §§ 392-1 to -101 (1976 & Supp. 1992); N.J. REV. STAT. §§ 43:21-25 to -56 (1962 & Supp. 1992); N.Y. WORK. COMP. LAW §§ 200–242 (McKinney 1965 & Supp. 1992); R.I. GEN. LAWS §§ 28-39-1 to -41-33 (1979 & Supp. 1992).

[41] *Cf.* INSTITUTE FOR WOMEN'S POLICY RESEARCH, UNNECESSARY LOSSES: COSTS TO AMERICANS OF THE LACK OF FAMILY AND MEDICAL LEAVE (1988) (calculating costs of no parental leave).

would be sensible. Where costs exceed some standard presumed by the right to leave and would result in significant loss, tax benefits to reduce losses or to permit the write-off of such losses against tax liability might be one way to cover costs while also encouraging individualized support programs that would reduce costs.

The other important perspective on costs is that it must be recognized and broadcast, as part of public education and employer education, that the costs of failing to provide leave or only providing unpaid leave are substantial. For example, the lack of family leave results in absenteeism, turnover, and retraining costs. Moreover, the potential long-term cost of unsupported families and children is substantial.[42]

b. *Time*. Reconstituting the time benefit is the essence of leave, and is nearly equal in importance to a pay provision. Time is meaningful only to the extent that it achieves the goals of family leave. Where the goal is family support during a severe family medical crisis, the conditions included, the range of time during which leave is needed, and the average time needed can be used to devise a benefit scheme that serves the greatest number of families. The benefit can also be structured to flexibly accommodate longer-term illnesses if the business can accommodate the employee for a longer period.

The needed time period can be similarly defined for birth or adoption of most infants. Absent medical crisis, which would be covered by the medical-crisis-leave provision, leave for birth or adoption requires sufficient time to establish a foundation for the parent-child relationship. Studies suggest that a minimum of three months should be provided, with a more desirable minimum of six months.[43]

As with the pay benefit, the structure of the time benefit has profound gender implications. Women and men are not similarly situated with respect to childbirth or lactation. The time structure can take those differences into account or exacerbate them. It can promote parenting regardless of gender or ensure that women most likely will be the primary or sole parents. When a child is born, the mother will be recovering from childbirth for some period of time afterwards. The standard period of recovery

[42] *See generally* BEYOND RHETORIC, *supra* note 35.
[43] *See* YALE BUSH CTR., *supra* note 24, at F-1.

is six to eight weeks, although the medical necessity of that period is questionable in terms of physical ability to work in particular jobs.[44] Sleep deprivation and other stresses associated with newborns continue longer, at least until the newborn sleeps through the night—a euphemism for sleeping six hours at night in a single block—which occurs on average by the time the child is two to three months old.[45] If the period of leave is limited to this childbirth recovery/early newborn period, and presumes one parent is on leave, then it is highly likely that the parent who will take leave will be the mother. If two parents may take leave at the same time, the father may take all or part of the same time to be at home. A longer time frame of six months, consistent with child development research, might permit both dual-parent leave and leave seriatim.[46]

If we are to change the way men can parent, the very nature of gender roles and the reproduction of those roles in the earliest role models and relationships with children, then family leave must respond to the powerful social and cultural constructions of motherhood and fatherhood that stand in the way of men parenting and that reinforce women's parenting. The justification for this is again to view family leave from the perspective of the child. Research indicates that the child needs the security of initial bonding, but also that the child can accommodate and benefit from multiple bonds.[47] Research also indicates that children of egalitarian parents do well and are better adjusted.[48]

The most radical structure to encourage men to take leave is the concept of capitation, coupled with a requirement that leave

---

[44] *See* Dowd, *supra* note 21, at 703.

[45] *See* MARIANNE R. NEIFERT, DR. MOM: A GUIDE TO BABY AND CHILD CARE 123–24 (1988) (stating that 12 a.m. to 5 a.m. is sleeping through night; even at one year, one-third to one-quarter of babies awaken during the night).

[46] As with pay provisions, expanded time provisions can be progressively incorporated into the workplace.

[47] *See, e.g.*, JOHN BOWLBY, A SECURE BASE: PARENT-CHILD ATTACHMENT AND HEALTHY HUMAN DEVELOPMENT (1988); Kara L. Boucher & Ruthann M. Macolini, *The Prenatal Rights of Unwed Fathers: A Developmental Perspective*, 20 N. C. CENT. L.J. 45, 56–61 (1992); John Moreland & Andrew I. Schwebel, *A Gender Transcendent Perspective on Fathering*, 9 THE COUNSELING PSYCHOLOGIST 45 (1981).

[48] On the child development implications of more egalitarian parenting patterns, and on the implications of single parenting when socioeconomic factors are factored out, see Jay Belsky, *Parental and Nonparental Child Care and Children's Socioemotional Development: A Decade in Review*, 52 J. MARRIAGE & FAM. 885 (1990); Michael E. Lamb, *The Changing Roles of Fathers*, *in* THE FATHER'S ROLE: APPLIED PERSPECTIVES 3 (Michael E. Lamb ed., 1986); Norma Radin & Graeme Russell, *Increased Father Participation and Child Development Outcomes*, *in* FATHERHOOD AND FAMILY POLICY 191 (Michael E. Lamb & Abraham Sagi eds., 1983).

be taken. Under such a structure, men and women would have leave and could not trade off who takes the leave period. If this is combined with a child-care system that provides little or no care for infants, then in effect parents must take leave and the division of leave entitlement would require parents to take leave seriatim. Some might object that this program would undermine individual choice and restrict personal liberty. Furthermore, it might be objected that such a requirement might contribute to domestic violence. One response, however, is to encourage change in the labor-market culture that would encourage such a pattern, permit it by the terms of leave structure, and support it when it occurs. The difficult issue here is to balance a desire for greater choice, especially by women, with a desire to reorient men's roles. Either you create greater choice presuming women's greater caretaking role, or you aim towards greater caretaking by men; but can you force them?

The time benefits in the 1993 Act are profoundly gendered for childbirth leave. By providing a maximum of twelve weeks of leave, the Act subtracts recuperation from childbirth from total leave. Because some period of disability after childbirth is unavoidable, this reinforces the role of women as primary parents, unless dual leaves are permitted. At the same time, on the disability side this costs women dearly, by using a substantial portion of their disability entitlement for pregnancy-related disability.[49]

Another desirable reform in time benefits would be to encourage maximum flexibility.[50] In many instances both employee and employer benefit from a phased return to work, or a part-time or modified schedule rather than full-time leave.[51] The issue here is whether to make part-time work a right. A right to part-time work redresses the imbalance of employee-employer power by giving more options to the less powerful employee. It may make planning by employers more difficult, however. As a result, employers may resist leave or those who take leave. Much of the desire for part-time work is due to the hostility of the workplace to combining work and family. Support of part-time

[49] In some instances the leave would be insufficient for pregnancy-related disability prior to childbirth. Dowd, *supra* note 21, at 712.

[50] This might be achieved with the use of mediation and/or arbitration to resolve conflicts, rather than imposing a minimum or maximum.

[51] Whether to provide part-time or modified work as an entitlement is more difficult. The original Schroeder family-leave bill did include such an entitlement. *See* H.R. 2020, *supra* note 4, § 105.

leave, particularly over a long period of time, grafts onto the workplace an exception for parents, usually women, rather than changing the structure. This result would be a serious possiblity with a part-time work entitlement similar to that of Sweden, which gives employees a right to a part-time schedule while children are young.[52]

A further question is how long-term leaves will be counted for purposes of other benefits, including how this time should be counted for social security and private pension purposes, for seniority, and for promotion and transfer.

c. *Expansion of benefits.* In addition to pay and time issues, family leave as currently constructed raises other issues related to the terms of the benefits provided and the extent of employee coverage. First, ordinary sick leave to care for sick children and family members is essential. The easiest means to accomplish this is to mandate that sick leave, where provided, must be permitted for the employee's own illness or for the illness of a family member. Sick leave, while a common benefit, is not legally required.[53] The unavailability of sick leave to care for children assumes that someone is available at home, not doing wage work, to care for sick children.[54]

Less essential, but certainly desirable, is personal leave for family reasons not hinging on illness or crisis, but to support children's development. Again, the easiest means to achieve this is to mandate that personal leave, where available, may be taken for child- or family-related reasons. Personal leave is even less universal, however, than sick leave.[55] The longer-term issue here is the structure of American vacations, which provide sig-

[52] Sweden provides the right to part-time work (up to six hours a day) until the child reaches age eight. *See* Widerberg, *supra* note 25, at 28.

[53] *See supra* note 23 (noting that approximately two-thirds of employers have sick-leave policies).

[54] Few child-care facilities care for sick children and even fewer facilities exist for care of sick children when parents must work. WHO CARES FOR AMERICA'S CHILDREN? CHILD CARE POLICY FOR THE 1990s 298 (Cheryl D. Hayes et al. eds., 1990) [hereinafter WHO CARES FOR AMERICA'S CHILDREN?].

[55] Paid personal leave is provided to 22% of employees in medium and large firms in private industry, and drops to half that percentage in small firms. STATISTICAL ABSTRACT, *supra* note 23, at 417. For employees of state and local governments, the percentage is 39%. *Id.* at 306. The average amount of annual leave time ranges from one to five days. BUREAU OF LABOR STATISTICS, U.S. DEP'T OF LABOR, BULLETIN NO. 2363, EMPLOYEE BENEFITS IN MEDIUM AND LARGE FIRMS, 1989, at 9 (1990); BUREAU OF LABOR STATISTICS, U.S. DEP'T OF LABOR, BULLETIN NO. 2398, EMPLOYEE BENEFITS IN STATE AND LOCAL GOVERNMENTS, 1990, at 8 (1992).

nificantly less time off in comparison to other industrialized countries.[56]

Coverage issues are even more important than either one of these expanded benefit issues. The 1993 Act only covers employers with more than fifty employees, leaving up to fifty percent of the workforce not entitled to leave.[57] State law leave requirements, which vary significantly, are not sufficient to resolve the problem. Only twenty-five states provide family leave and their coverage limits range from employers with twenty-one to fifty employees.[58] The resulting discrimination against children unfortunate enough to have parents who work for employers not covered by state or federal legislation is intolerable if a primary goal of family leave is to provide all children with a solid beginning and adequate support in the event of serious medical crisis. Universal coverage can only be accomplished in conjunction with financing and support for business expenses that is society-wide, rather than limited to the resources of individual businesses.

d. *Family diversity/enhancing the benefits—single parents.* Throughout the process of structuring family leave and more broadly, family policy, we must keep in mind the diversity of family structures that perform the function that family leave is designed to support. While we may disagree over the definition of family functions, such controversy should not stand in the way of a family-leave policy designed to provide support for family members caring for each other, a function of family so central and essential that it is indisputable. Definitions of family do become important in determining what relationships will be honored as entitlements to leave.[59] For example, limitation of

---

[56] Vacation time is linked to length of service in the United States, averaging 11 days off in the first year, 15 days after five years, 17 days after 10 years, and 24 days after 30 years. In contrast, vacation is an entitlement of uniform duration in European countries, averaging five weeks. *U.S. Employees Work More and Play Less*, PR Newswire, Sept. 16, 1992, *available in* LEXIS, Nexis Library, PRNEWS file.

[57] Dowd, *supra* note 21, at 711.

[58] Thomas C. Goeldner & Beverly L. Nelson-Glade, *The Wisconsin Family and Medical Leave Act*, WIS. LAW., Apr. 1993, at 21–22; *Employers Report Few Problems Adjusting to State Parental Leave Laws, Study Finds*, *in* Daily Lab. Rep. (BNA) No. 100, at A-7 (May 23, 1991); Michael J. Langan & Richard G. Gisonny, *Family Leave Proposals and Existing State Laws*, 4 BENEFITS L.J. 289 (1991).

[59] *See generally* Katharine T. Bartlett, *Re-Expressing Parenthood*, 98 YALE L.J. 293 (1988); Katharine T. Bartlett, *Rethinking Parenthood as an Exclusive Status: The Need for Legal Alternatives When the Premise of the Nuclear Family Has Failed*, 70 VA. L. REV. 879 (1984); Martha Fineman, *Dominant Discourse, Professional Language, and Legal Change in Child Custody Decisionmaking*, 101 HARV. L. REV. 727 (1988).

leave benefits to legal parents, as that term is presently defined, will exclude step-parents, gay and lesbian co-parents, extended family parents, and psychological parents.[60]

In addition to defining who is a parent, we must also keep in mind the composition of the family that we presume will support these functions. Most leave discussions have implicitly presumed a two-parent, heterosexual couple. The significant growth in single-parent families suggests that this family form should be kept in mind as well. Single-parent families constitute thirty percent of families, and two-thirds of all children will spend some time in a single-parent family before reaching age eighteen.[61] A child-centered family-leave benefit, designed to insure bonding with at least one parental figure, might therefore need to provide a longer benefit to a single parent.

One could argue that using a single parent as a model is not unrealistic, even for children in two-parent families. The bulk of child care and elder care is performed by one adult, usually a woman. The likelihood that the household will shift from this de facto, single-parent, female-dominant model to de jure single parenthood is high, given the high divorce rate (fifty percent) and the predominance of women as sole or primary custodial parents.[62] By this reasoning, the most significant difference between two-parent and single-parent families is resources, not differences in family responsibilities. Support is therefore needed for single parents in both circumstances.

The differential in income between dual-parent and single-parent families does suggest that family-leave policy should incorporate supplemental benefits for single-parent families. This would ensure sufficient support to children, parents, and families regardless of family form. The difficult issues here are whether such a structure supports or deters parental involvement by non-resident parents, or, perhaps, whether it supports co-parenting over a primary parent structure. Some will object

---

[60] *But see* D.C. CODE ANN. §§ 36-1301 to -1317 (Supp. 1991) (defining family relationships to include homosexual relationships, extended families, and short-term assumption of parental duties). In the Schroeder draft, the definition of "parent" would have encompassed de facto parents. H.R. 2020, *supra* note 4, § 101(9)(B).

[61] *See* HOUSEHOLD AND FAMILY CHARACTERISTICS, *supra* note 28, at 10 (listing demographic data on number of single-parent families); Ermisch, *supra* note 28, at 27, 47; Arnold J. Norton & Paul C. Glick, *One Parent Families: A Social and Economic Profile*, 35 FAM. REL. 9, 16 (1986).

[62] Over half of all first marriages and remarriages end in divorce. STATISTICAL ABSTRACT, *supra* note 23, at 90. Overwhelmingly, the children of divorced or unmarried parents are raised by their mothers. *Id.* at 55.

that single parents ought not to be supported because that will value or encourage that particular family form over two-parent families.[63]

It is important to remember that the purpose of leave is to support children. We should not punish children for the families into which they are born. We also need to realistically consider the short-term versus long-term consequences of non-support or inadequate support. Single-parent families are more likely to be poor, and children of poor families are more likely to experience health, education, and employment risks. If the goal of family leave is to strengthen families, then single-parent families arguably ought to be the primary focus of family leave, particularly with respect to reforming leave to include adequate wage replacement and adequate time for family development or family support during medical crisis.

e. *Nondiscrimination and anti-retaliation.* Family leave currently protects employees who choose to take leave by sanctioning failure to provide leave, discrimination against those who exercise their rights, or retaliation for taking leave.[64] The sanctions include money damages for loss of pay and benefits, or in some instances the cost of services for dependent care, plus equitable relief, attorneys' fees, and costs.[65] In addition, double damages are provided when an employer has not acted in good faith. The remedies do not, however, include punitive damages. It certainly can be argued that emotional harms are the central injury that would result from denial of leave rights. Furthermore, denial of leave rights under circumstances where punitive damages would ordinarily be justified certainly merits the availability of these damages for injury to family relationships. Limited damages ignore the character of the harm and devalue the significance of the harm.

Even a full panoply of remedies, however, should be viewed as a deterrent of last resort. A more effective approach to support leave policies would be an affirmative action model, which should include two parts. First, a modified affirmative action

---

[63] *See* Elizabeth Duskin, *Overview*, *in* LONE PARENT FAMILIES: THE ECONOMIC CHALLENGE, *supra* note 28, at 9–26 (discussing implicit ambivalence toward single-parent families in European countries); Nancy E. Dowd, Single Parents and the Legal Ideology of Family (1993) (unpublished manuscript, on file with author) (discussing the stigmatization of single-parent families).

[64] Family and Medical Leave Act of 1993, § 105, 107 Stat. at 14.

[65] *Id.* § 107, 107 Stat. at 15–17.

model should be applied to federal contractors. The modification should be that any reporting, plan, or sanction requirements would expire for employers who demonstrate, over a five- or ten-year period, consistent use of family leave without negative consequences. Second, the model should include affirmative support for implementing family-leave and other family-support policies. Whether in the form of tax incentives, information banks, or other strategies, the government's role ought to be to encourage innovation and flexibility around core benefit standards.[66]

C. *Summary*

In summary, family leave should provide adequate time for family support during family formation, for eldercare or care of other family members, and for individual disability. Leave should include time for family formation at birth or adoption, sick-leave time for ordinary illnesses of children, leave for extraordinary or emergency medical crises of children, and leave for educational or other functions. With respect to other adult family members for whom the employee is responsible, leave should be provided for medical or other emergencies that require either care or arrangement for care. In addition, sick leave may be necessary where caregivers are temporarily ill and the employee cannot arrange substitute care. Family-support needs should determine the time allowed for leave in different circumstances, rather than a single fixed period. In order to achieve real family support, leave must be paid, with the level of pay taking into account the dual wage-earner and single-parent aspects of family structure. In addition, the entitlement must be sensitive to gender roles in the family and in the workplace, because these roles may affect the use of leave. Public education and public relations must educate employees and employers on the value and importance of leave, and the protection available to employees who choose to take leave. Affirmative action should be used to support measures that maximize flexibility, choice, and long-term perspectives. The availability of a civil

[66] The statute expressly states that it does not preempt more expansive leave provided under state law or a collective bargaining agreement. *Id.* §§ 401–403, 107 Stat. at 26. It also permits agreements between employers and employees to modify parental leave from absolute leave to intermittent or part-time leave. *Id.* § 102, 107 Stat. at 9. There are otherwise no affirmative supports for more expansive leave.

action with the full panoply of legal and equitable remedies will provide additional incentive to reform.[67]

This policy must be coordinated with a more comprehensive family-support policy. For family leave to be more than an isolated tinkering with workplace norms, it is critical that family leave be seen in conjunction with other family-support policies. The existing child-care and welfare policies need to be coordinated with family-leave policy. More broadly, a comprehensive approach to family support is needed that would go beyond coordination to the creation of a coordinated family-support policy. The changes needed in the workplace are nothing less than revolutionary. Family leave is the easy piece—that is, carving out a limited time away from work to put family first. Ongoing structural change will require much more. The next part briefly outlines some of the areas that must be addressed, and then discusses child care and family allowances in more detail.

II. FAMILY POLICY

Family leave is only a small piece of family policy. This is apparent when you dissect work-family conflict under existing family and work structures. Those structures remain fundamentally premised on the separation of work and family and the presumption of full-time or primary work in one sphere or the other, assigned by gender. Furthermore, work is valued and prioritized above family, although ideologically and culturally family is sometimes placed on a pedestal. Because economic survival is no longer family-based, but rather market- or government-based,[68] family must support work, rather than work supporting the family. The relationship between the family and the market or government is hierarchical. The valuing of wage work is replicated by hierarchy within the family where paid market work is valued over unpaid family work.[69] This ignores

[67] This package does not include health benefits, because I presume that universal health insurance, in some form, is imminent.

[68] *See* MARY ANN GLENDON, THE NEW FAMILY AND THE NEW PROPERTY (1981); Dowd, *Restructuring*, *supra* note 3, at 433–37; Charles A. Reich, *The New Property*, 73 YALE L.J. 733 (1964).

[69] *See, e.g.*, Myra Marx Ferree, *Beyond Separate Spheres: Feminism and Family Research*, 52 J. MARRIAGE & FAM. 866, 877–79 (1990); Hilary Graham, *Being Poor: Perceptions and Coping Strategies of Lone Mothers*, *in* GIVE AND TAKE IN FAMILIES: STUDIES IN RESOURCE DISTRIBUTION 56, 59, 62 (Julia Brannen & Gail Wilson eds., 1987).

the interdependence of family and work, as well as the dependence of those within the family on family wage-earners for income, and the dependence of wage-earners for non-market support.

The results of this structure include conflicts of time, including not enough hours to do both family and wage work; overloading of family and child-care work on women; the inability to control and use time; clashes between occupational and family life cycles; psychological role conflict and role strain between family and work roles; dissonance between workplace culture and valuing family, particularly for men; pervasive sex stereotyping and sex segregation in the workplace tied in part to gendered family role stereotypes; and ideological conflict created by the ideology of individualism, choice, and autonomy, which justifies individual responsibility for family and thereby obscures structural features that create conflict.[70] The aspects of the workplace that contribute to work-family conflict are based upon a male-centered stereotype of the worker's parental role as being primarily economic, coupled with the separation of work and family concerns. Thus, time schedules are remarkably inflexible, occupational demands are highest when family demands are also likely to be high, fringe benefits and required benefits facilitate separation, not connection, of work and family, and essential family support benefits like maternity leave, parental leave, personal leave, disability leave, and sick leave are not legally required, nor is health insurance.[71]

Given these layers of conflict, one of the issues for work-family policy is to consider the complex ways in which these tensions connect and the impact of implementing policy in a particular area on the broader perspective. Family policy, like most radical reforms, is a progressive process, not an overnight revolution. As the most blatant obstacles are confronted, the next layer of issues may be more subtle and sophisticated, necessitating a more sustained, long-term approach and representing a more radical change in the existing structure.

Family leave is arguably a first generation issue, dealing with circumstances in which the conflict between work and family results in job loss as a price for necessary family care, or the

---

[70] Dowd, *Gender Paradox*, *supra* note 3, at 84–100.

[71] *Id.* at 100–09; *see, e.g.*, MASS. GEN. L. ch. 118F, §§ 1–20 (1991).

sacrifice of family foundations to retain a job.[72] It is a serious issue meriting immediate attention. It also is an issue that is relatively easy to resolve; the legislative choices are relatively easy to map. On the other hand, reforming workplace culture to eliminate the impact of work-family concerns in order to reduce job segregation, or reconfiguring family roles, both to redistribute family work and change family roles, are far more difficult tasks. Ensuring true equality of opportunity for children is similarly challenging.

Family policy is the affirmative support of families for the benefit of the political and social community. The United States has no family policy per se, by name or in substance. The elements of family policy common in other countries include paid, job-protected maternity leave; parental leave; universal health insurance; nearly universal preschool child care; family income support in the form of tax benefits or family allowances; and paid, job-protected disability leave.[73] Family policy should also be attentive to such issues as the impact of leave on pension rights; the structure of property settlement and ongoing income transfers at divorce designed to value family and caregiving as well as wage work, and whether they take into account long-term as well as immediate effects of family work for advancement in and income from wage work; reevaluating the educational structure (daily timing, vacations, yearly timing), including consideration of year-round schools as well as child-care and/or after-school programs for school-age children; devising ways to value family work, including affording essential benefits such as health care regardless of employment in the wage labor market, and considering whether it is possible to value stay-at-home parents without reconstituting patriarchy; and structuring affirmative support for innovative efforts to reorient work and family culture, and engage in effective public education and support for private efforts for change.

Family policy must proceed in at least two phases. First, the concept of family policy ought to be considered in conjunction with leave policy. The 1993 Act includes a study commission, but the focus of the commission is exclusively on the impact of

---

[72] This characterization may embody a particular class bias. For poor families, income policies might be more critical than leave.

[73] Dowd, *Envisioning Work and Family*, *supra* note 3, at 316.

leave, not its connection to a broader family policy.[74] The experience of other countries suggests that the most important elements to coordinate are child-care and income support policies. It is important, then, to consider family leave in conjunction with existing child-care and welfare policy.

Second, the study and implementation of a more expansive family policy should begin. A broader policy should include reform and expansion of child-care and welfare systems, as well as changes in part-time work, educational systems, housing, support for innovative business restructuring, children's rights, and family income benefits (pay and taxes), for example. What is needed is a central vision, and policies that would achieve this vision. Meanwhile, it is essential to coordinate what we have in order to make the vision a reality.

This is a vast canvas. I will sketch only two of the critical pieces, child care and family allowances.

A. *Child Care*

Child-care policy is an obvious area that needs coordination with family leave. Yet child-care and family-leave legislation have evolved separately, with separate sets of advocates, each espousing separate agendas. Child-care policy has prioritized poor families, although indirect subsidy of child care through the tax system primarily benefits middle- and upper-class families.

In contrast to the long struggle to enact leave legislation, a child-care bill was enacted by Congress in 1990.[75] The legislation combined tax credits and grant funding to support expanded

---

[74] Family and Medical Leave Act of 1993, §§ 301 - 306, 107 Stat. at 23–25. Either the commission's agenda should be revised, or a second legislative or executive commission or task force should be authorized to consider coordination of leave with other family policies.

[75] Omnibus Budget Reconciliation Act of 1990, Pub. L. No. 101-508, §§ 5081–5082, 104 Stat. 1388-1, 1388-233 to -250. This was the first major child-care legislation since 1971, when President Nixon vetoed a measure to provide child-care programs out of fear that it would encourage women to work outside of the home. *See* 1990 CONG. Q. ALMANAC 547 (citing 1971 CONG. Q. ALMANAC 504).

The other significant child-care legislation enacted since 1988 was the child-care provisions of the welfare reform overhaul of 1988. The Family Support Act of 1988, Pub. L. No. 100-485, § 301, 102 Stat. 2343, 2382, required states to provide child-care and medical benefits to families leaving welfare for jobs, along with training programs to move welfare mothers from welfare to work. *See* 1989 CONG. Q. ALMANAC 224 (citing 1988 CONG. Q. ALMANAC 347). The states have found it difficult to provide these child-care services.

child-care facilities and assistance with the cost of child care for low-income families. The tax credit expanded the earned income tax credit for working, low-income families, with additional credit for children under age one. This credit is available as long as one person in the household works, regardless of whether income tax is paid. In the event no tax is due, a refund will be given.[76] The funding provision is designed to be used in several ways, including providing direct subsidies for the poorest families to pay for child care; funding expanded care for low-income families not poor enough for welfare; and improving the quality of care.[77] The legislation represents a compromise among several legislative initiatives that began in 1988. One issue that pervaded the political process concerned whether to mandate the quality of care by imposing federal health and safety standards on child-care providers.[78] Some argued that quality standards would make it harder to maintain and expand available child care. There was also debate on whether to subsidize (in money or vouchers) poor families' child-care expenses, provide tax credits for child-care and insurance costs, increase funding of Head Start to reach a greater proportion of preschoolers, and fund after-school programs for school-age children in addition to preschool programs.[79]

Although geared to providing child-care support to low-income families, this package serves only a fraction of those families. The tax credit, for example, requires low-income families to file a tax return. Without massive public education on the availability of the credit, many low-income families will not utilize it. Furthermore, with a maximum credit of $2,013 by 1994, the amount of care that can be bought with this credit is far below the estimated $6,000 per year that child care actually costs.[80] The dependent-care tax credit, by contrast, is the largest single program of child-care assistance. However, because it provides no refund, it is useless to families with insufficient income to pay tax.

---

[76] Act of November 5, 1990, Pub. L. No. 101-508, 1990 U.S.C.C.A.N. (104 Stat.) 1388-233.

[77] *Id.*

[78] This was identified in the ABC bill (Act for Better Child Care). H.R. 3660, 100th Cong., 1st Sess. (1988).

[79] 1990 CONG. Q. ALMANAC 547–51.

[80] Women who work full-time and require 41 hours or more of child care pay, on average, $99 per week for child care. Child care constitutes about 10% of family expenses. *Nationwide Child Care Survey*, Business Wire, Dec. 7, 1992, *available in* LEXIS, Nexis Library, Business Wire File.

The funding programs attempt to address a broad range of child-care issues without fully addressing any of them.[81] For instance, available funding for latchkey children would serve only 300,000 of the over ten million latchkey children.[82] Child-care funding for families who needed such care to stay at work and off welfare was limited to $300 million.[83] Head Start funding, although enormously increased due to the popularity of the program and because of the political battle between Congress and the Bush administration to claim credit for supporting the program, still serves less than half of the eligible children.[84]

It is striking that child care and family leave have not been considered together. Child care is essential for parents to return to wage work. In addition, child care provides essential backup care for sick children, again allowing their parents to be able to go to work.

The interconnection also relates to how family leave should be structured now, as opposed to how it might be structured in view of a better child-care system. For instance, wage replacement during leave is further mandated by the immediate, significant expense of infant care.[85] On the other hand, it may make more sense developmentally, as well as financially, both to extend the financial support and the time frame for leave. Then we could concentrate child-care resources on programs for children over one year old, which are less expensive.

We are quite a distance from closing the gap between the need and the reality of quality child care. The hesitancy to develop a national child-care policy coordinated with family leave is tied to the individualistic ideology that so permeates the family-policy area. It is connected strongly to ideas of appropriate mothering.[86] Finally, it may stem from an aversion to scrutinizing the structure of existing child care. Gender and class interests may be subtle brakes on evaluation of the need and cost of

---

[81] Under the legislation, 75% of funding is to be used by the states to provide child-care services directly to families with incomes below 75% of the state median income, or increase the availability or quality of child care. Three-fourths of the remaining money is to be spent either on preschool, and/or on before- or after-school care for school-age children, 20% on training and salaries, and the last 5% on either activity. 1990 CONG. Q. ALMANAC 551.

[82] *Id.*

[83] 1991 CONG. Q. ALMANAC 507.

[84] *Id.* at 510.

[85] WHO CARES FOR AMERICA'S CHILDREN?, *supra* note 54, at 297–98.

[86] *See, e.g.*, Martha L. Fineman, *Images of Mothers in Poverty Discourses*, 1991 DUKE L.J. 274.

quality universal child care because child care depends so strongly on underpaid, mostly female, wage workers.

Beyond these ideological hurdles, we need to address other issues. Should family policy be centralized or decentralized? How can diversity in the provision of care be maximized without hurting quality? Should child care be connected solely to wage work? As with family leave, gender, race, and class concerns must be kept in the forefront, particularly where the structural issues are framed as "choice" versus "equality." The primary focus, however, must always remain what system produces the best quality care for all children.

B. *Family Income Support*

In contrast to child-care policy, which requires coordination, integration, and then expansion of current programs to meet the goal of providing universal quality care, family income benefits require an entirely new piece of family policy. We could model our system on the European and Scandinavian family allowance plans. Family allowances are a two-tier concept. The most important feature is that income benefits are provided regardless of need, in recognition of the additional expense of children for all households. The objection of over-inclusiveness is countered by the destigmatization of family benefits and the deemphasizing of class differences.[87] Under some family income support schemes, additional need-based income benefits are provided. Benefits targeted for single-parent families are the newest piece of this need-based tier. The tendency has been to support single parents without promoting such families. The additional benefits, which include supplemental income, housing allowances (with preferences for low-income housing), and child-care subsidies, make single parents far better supported in Europe than in the United States. Of course, single-parent families are still significantly disadvantaged compared to two-parent families.[88] These systems attempt the delicate balance between support and promotion of single parenthood.

---

[87] This might also promote adoption by reimbursing initial adoption expenses. Conceptually, however, such a scheme might better be considered along with pregnancy expenses. Another issue raised by the adoption scenario is the adoption of older children and the different needs that presents for family foundation.

[88] *See generally* LONE PARENT FAMILIES: THE ECONOMIC CHALLENGE, *supra* note 28.

Income support in the United States has been limited to need-based welfare and unemployment assistance, and it has been riddled with inconsistency and counterproductive policies. We do a bad job of supporting poor families. We do an even worse job of supporting poor children. We need to do for the very young what we have done for the very old: raise their living standard to a minimum acceptable level for all.[89] Family allowances, or some other form of broad income support, are a means to that end.

III. CONCLUSION

The development of a comprehensive family policy is an area where the models of other countries are particularly useful. As one of the last countries to engage in policy and restructuring in this area, the United States can learn from the experiences of other countries. One thing that experience has made clear is the nagging inequality of women despite family-support policies.[90] Two recent studies suggest some of the complexities of this problem.

One study debunks the myth that it is women's child-care and marital responsibilities, their "choices," that are responsible for their status in the wage labor market. The study found that women do not earn substantially more if they devote themselves exclusively to their careers, as opposed to trying to combine career and family.[91] The study also found that women report less satisfaction with their lives if they do not have children, and that the least satisfied of the women were those who were unmarried and without children.[92] This study suggests the presence of deep structural and cultural discrimination in the wage labor market regardless of whether women engage in their tra-

[89] It has been easier to improve living standards for the old than for the young, and that difference unfortunately has been a source of conflict instead of learning. *See* Robert Atchley, *Fallacies of 'Geezer Bashing'*, CHRISTIAN SCI. MONITOR, Sept. 28, 1992, at 18; Jean Dietz, *Old and Young Must Unite to Reach Goals, Aging Group Is Told*, BOSTON GLOBE, Mar. 29, 1992, at B39.

[90] Dowd, *Envisioning Work and Family*, *supra* note 3, at 337–44.

[91] *See* Brody, *supra* note 33, at C16.

[92] It is interesting that unmarried, childless women were compared to men who presumably are married and have children. This suggests that women can only succeed in a male world if they do not have marital and/or child-care responsibilities, a telling commentary both on female and male constructions of gender roles in the work-family context.

ditional gender role of sole or primary childrearing.[93] It also demonstrates how that discrimination reinforces the value of traditional roles and denies even the satisfaction of economic equality to women who most closely adhere to the male standard.

Another recent study points to the continuing reality of women's care for children, as well as the deep-seated assumption that they should do so. The study reported that children benefit from being at home with at least one parent for the first year of life. This was demonstrated by differential learning patterns emerging in preschool, specifically children's ability to relate geographic items.[94] What is most notable in the reporting of this study is that it spoke only in terms of mothers. The comparison was day care to mother care, and the implications of the study were presented only for mothers.[95] There is much research to support the value of parent-child bonding during infancy.[96] But the thrust of this study echoes the concerted, although largely unsuccessful, effort to attack day care over the past several decades as increasing percentages of women returned to work when their children were infants.[97] Women's preeminent role as parents, and the continuing marginalization of fathers, is strongly reflected in this study.

Stereotypes are hard to change. Amelioration of the problem might mean trying to undercut the economic disadvantages of women's family roles. If, however, we fail to go beyond economics, we miss very real harms. We also may only end up reinforcing the disadvantages by compensating for them, rather than attacking them. How we see work and family is critical to our construction of both gender roles and the structures that we create that presume those roles. Our gender stereotypes are the foundation of the continuing inequality of women and the limited family role of men. This marginalization is magnified when combined with inequities of class and race.

---

[93] *See* Vicki Schultz, *Telling Stories About Women and Work: Judicial Interpretations of Sex Segregation in the Workplace in Title VII Cases Raising the Lack of Interest Argument*, 103 HARV. L. REV. 1750, 1815–16 (1990) (arguing that the structure of the workplace encourages women to choose those traditional female jobs in which they have generally excelled rather than to attempt to enter into non-traditional careers).

[94] Erik Eckholm, *Learning if Infants Are Hurt when Mothers Go to Work*, N.Y. TIMES, Oct. 6, 1992, at A1, A21.

[95] *Id.*

[96] *See supra* note 47 and accompanying text.

[97] *See* Eckholm, *supra* note 94, at A21.

Do we want to confront the deeply patriarchal model of family, work, and work-family connections? To do so requires a vision that recognizes that we must get beyond a gendered arrangement built on inequality, separation, and dominance. Rethinking gender in family and work is a revolutionary task. The greatest challenge will be to get beyond gender and focus more clearly on the kinds of family and work relationships and structures we want to have as part of the foundation for the society we wish to be.

ARTICLE

PROPERTY TAX ABATEMENT FOR LOW-INCOME HOUSING: AN IDEA WHOSE TIME MAY NEVER ARRIVE

MELVYN R. DURCHSLAG*

Substandard and overcrowded urban housing conditions in the United States have been a problem in need of a legislative solution for many years. Twenty years ago, in Volume 11 of the Harvard Journal on Legislation, *Norman Alpert argued that property tax abatement might stimulate private, for-profit investment in low- and moderate-income housing. Under Mr. Alpert's market-oriented proposal, deferred reassessment of improved property would attract private capital into the construction or rehabilitation of housing units for low-income individuals and families.*

In this Article, as part of the Journal's *Thirtieth Anniversary Issue, Professor Durchslag revisits the property tax abatement issue. Drawing on empirical studies, Professor Durchslag argues that, though plausible in theory, property tax abatement does not in reality generate a significant increase in the supply of low-income housing because of several economic and political factors. Furthermore, he argues that tax abatement proposals have had the additional negative effect of indirectly impeding movement toward greater public responsibility for low-income housing. Professor Durchslag concludes that legislatures should abandon attempts to attract private capital into the low-income housing market through the use of tax incentives. Instead, he argues, the public sector ought to be the low-income housing supplier of choice.*

Twenty years ago, Norman Alpert wrote that property tax abatement offered promise as an incentive to profit-motivated housing investors to either construct new housing units or rehabilitate existing units for low-income individuals and families.[1] The theory is simple. Property taxes represent a significant

* Professor of Law, Case Western Reserve University School of Law. B.S., Northwestern University, 1962; J.D., Northwestern University, 1965. I appreciate the critical comments offered by my colleagues, William Marshall, Andrew Morriss, and William T. (Tom) Bogart.

[1] Norman Alpert, *Property Tax Abatement: An Incentive for Low Income Housing*, 11 HARV. J. ON LEGIS. 1 (1973). Property tax abatement, as Alpert uses the term and as it is generally understood, means the temporary non-recognition of an increase in value and thus relief from ad valorem taxes resulting from improvements to a particular piece of property. *Id.* at 2. This is distinguished from tax exemption, which ordinarily describes a more permanent removal of all or part of the value of a particular piece of property from tax liability. The most common forms of tax exemption include property used for charitable or religious purposes (full tax exemption) and the homestead exemption for property used as a primary residence (partial dollar amount exception).

portion of the gross rents received by urban housing suppliers.[2] If current rents maximize profitability, any increase in costs must, if the enterprise is to continue, be followed either by an increase in revenues (a rent increase) or a reduction in expenditures.[3] Upgrading the quality of low-income housing thus presents a dilemma. On the one hand, a significant investment in property improvements will almost certainly increase real estate taxes and thus increase costs.[4] On the other hand, the property owner's ability to increase rents in order to maintain the previous (optimal) rate of return is limited. Low-income tenants simply cannot increase the amount they pay for rent.[5] Indeed, any rent increase would likely displace low-income persons in favor of those with higher incomes who can afford to purchase the higher-quality units.

Other than maintaining the status quo, the property owner's only remaining option is to reduce her non-improvement-related expenses, such as maintenance. But that would be shortsighted at best and, more realistically, pure folly: it would hardly be rational for one to invest a substantial sum of money in improvements only to consciously reduce the value of those improvements by failing to maintain them. While this analysis is admittedly over-simplistic, it demonstrates the lack of "market" incentives for improving housing for low-income families. Some governmental intervention, then, is necessary to supply those incentives.

One form that intervention can take is to ensure that a property owner's costs are not increased as a result of the improve-

---

[2] The consensus seems to be that property taxes represent between 20 and 25% of gross rents. *See* DICK NETZER, ECONOMICS OF THE PROPERTY TAX 85 (1966) (25%); Alpert, *supra* note 1, at 1 (21%).

[3] Cost reductions most likely will take the form of deferred maintenance. Property taxes and debt service payments are not only fixed costs but their non-payment will eventually result in loss of the property through foreclosure. The property owner could reduce expenses by shifting some of them, such as utility costs, to the tenant. From the tenant's perspective, this would have the same impact as a rent increase and would be successful only if the demand for low-income housing were sufficiently inelastic.

[4] *See* Alpert, *supra* note 1, at 2. Netzer suggests that the increase in real estate taxes resulting from substantial improvements is geometric rather than linear, thus requiring a significantly greater increase in rents than might be justified by the value of the improvement. *See* NETZER, *supra* note 2, at 83.

[5] Current statistics indicate that most low-income persons spend between 35 and 50% of their income on housing. *See, e.g.*, National Ass'n of State Univs. & Land Grant Colleges Urban Hous. Working Group, *A Statement of Appropriate Private and Public Responses to Urban Housing Needs*, 36 WASH. U. J. URB. & CONTEMP. L. 63, 64 (1989) [hereinafter Urban Housing Working Group]; Lawrence B. Simons, *Overview: Housing Options for the 1990s*, 6 YALE L. & POL'Y REV. 259, 268–70 (1988).

ments by deferring the increased property taxes which result from improvement-related increases in property values. Alpert explained the incentive in the following way:

> A landlord calculating the value of a property tax exemption expects that his assessment would normally be increased by the amount of his rehabilitation expenditure . . . . The annual savings thus would equal the tax rate times the increase in value, adjusted for the local assessment ratio. If the owner of a building discovers that his after tax return will increase because of a reassessment deferment on an improvement, not only does his already contemplated improvement investment become more profitable, but additional improvement that would otherwise be unprofitable may become profitable. A tax abatement limited to housing improvement for low and moderate income residents thus should induce additional capital expenditure and provide more housing.[6]

Alpert's argument that a property owner should account for any real estate tax savings on improvements in her investment return is probably correct. Unfortunately, what little empirical evidence exists demonstrates either that this has not occurred or, if it has, the effect on the supply of housing for low- and moderate-income households has been negligible. Part I of this Article describes the findings of three studies, one in New York and two in Missouri, that, in whole or in part, attempted to assess real estate tax abatement as an incentive for producing or maintaining housing for low-income persons. In that Part, I also suggest two general reasons why the data show that abatements have not been successful. In Parts II and III I suggest more specifically why it is unlikely that real estate tax abatements will be successful in encouraging housing for low-income persons. Part II focuses on the disjunction between economic theory and economic reality, while Part III focuses on the factors which encourage local politicians to prefer commercial and industrial redevelopment to housing redevelopment. I conclude by suggesting that the public sector, not the private sector, ought to be the low-income housing supplier of choice, thus making tax abatement for this purpose unnecessary.

I. The Inducement that Never Worked

In 1976, Sternlieb, Roistacher, and Hughes conducted a study of the three housing tax abatement programs in New York City,

[6] Alpert, *supra* note 1, at 5–6.

the Mitchell Lama,[7] J51,[8] and 421[9] programs. The authors concluded that there was little if any *housing* benefit to low- and moderate-income persons as a result of these programs, considered either separately or combined. The only benefit was a general economic stimulus to the New York City economy produced by an increased circulation of dollars, a benefit which did not inure, at least directly, to low- and moderate-income persons.[10] Generally, the housing consumers who benefited were those in the high- to middle-income ranges.[11] More specifically, the authors found that the J51 program had been used to rehabilitate some 284,000 units, predominantly in Manhattan and, within Manhattan, largely "in the prospering areas of the upper east side, the upper west side, and Chelsea."[12] Moreover, even after the tax roll-backs provided by the program, rents in privately financed (as opposed to F.H.A. financed), newly rehabilitated units increased by approximately 242%. This "sharp increase in rents . . . suggests that there is a substantial difference in the income of tenants occupying the units before and after rehabilitation."[13]

The experience with the Mitchell Lama program was even more disappointing. Despite (1) a thirty-year tax abatement of up to 50% of the full assessed value of the property or 100% of the increased value brought about by the improvements, (2) a long-term, below-market interest rate mortgage, and (3) an occupant population which "on average [had] higher incomes and lower rent-income ratios than other city renters," the program experienced a high incidence of mortgage and even property tax default.[14]

---

[7] N.Y. PRIV. HOUS. FIN. LAW §§ 10–37 (McKinney 1991).

[8] The J51 program is codified in the New York City Administrative Code. For an explanation of its provisions, see GEORGE STERNLIEB ET AL., TAX SUBSIDIES AND HOUSING INVESTMENT: A FISCAL COST-BENEFIT ANALYSIS 12–15 (1976).

[9] N.Y. REAL PROP. TAX LAW § 421-a (McKinney 1991 & Supp. 1992).

[10] STERNLIEB ET AL., *supra* note 8, at 20.

[11] *Id.* at 19.

[12] *Id.* at 14–15.

[13] *Id.* at 15.

[14] *See id.* at 16–17. The third program evaluated by the study, the 421 program, encourages new construction. N.Y. REAL PROP. TAX LAW § 421-a2(a)(1) (McKinney Supp. 1992); STERNLIEB ET AL., *supra* note 8, at 9–10. It has not been of any direct benefit to low- and moderate-income persons. *Id.* at 19, 21. These persons could benefit indirectly through "filtering," the process by which standard units are opened for low-income families because of vacancies created by more affluent families moving to the newer housing. That is the theory, at least. However, the authors note that "any benefits from filtration may be offset by displacement of existing tenants because of demolition or major rehabilitation." *Id.* at 22. *See generally* HOUSING IN AMERICA: PROBLEMS AND

It may be difficult to generalize from the Sternlieb study to communities outside New York, largely because of rent control and its disincentive to owners contemplating the substantial rehabilitation of older structures. The value of the tax abatement incentive would have to be significant given the disincentive produced by a controlled rent structure. In addition, even tax abatement in a rent controlled community such as New York City might be seen by housing investors and owners as a mere "pay-back" for the market incentives rent control destroys, rather than as an incentive.

Yet the inducement also failed to increase the supply of units for low-income people in a city that did not have rent control. A 1980 study of the so-called 353 program[15] in St. Louis concluded that the early use of the program was primarily for downtown commercial redevelopment.[16] At the time of the study, some housing activity was attributed to the tax abatement program, but it was small compared to the amount of commercial and industrial redevelopment.[17] Moreover, much of the housing produced was not for low- and moderate-income persons. While over one-half of the units (564 out of 925) were constructed with federal "Section 8"[18] subsidies, approximately

---

PERSPECTIVES 161–203 (Roger Montgomery & Daniel R. Mandelker eds., 2d ed. 1979) (discussing all aspects of filtering); Anthony Downs, *Are Subsidies the Best Answer for Housing Low and Moderate Income Households?*, 4 URB. LAW. 405, 409–12 (1972) (arguing that filtering concentrates the social problems attendant to "slum" areas).

[15] The program is named after the relevant section of the Missouri statutes, MO. REV. STAT. § 353.010–.180 (1986 & Supp. 1991). The 353 program provides a real estate tax abatement to an urban redevelopment corporation for up to ten years, provided that the use of the property is consistent with the approved redevelopment plan for the area. *Id.* § 353.110. The abatement is with respect to all improvements on the land as determined in the calendar year immediately preceding the year in which the corporation acquired title to the land. *Id.*

[16] DANIEL R. MANDELKER ET AL., REVIVING CITIES WITH TAX ABATEMENT 27 (1980).

[17] It is difficult to extrapolate the exact figures from the data provided by tables 5 and 6, *id.* at 29, 31, because some of the projects noted in table 6 combine residential rehabilitation with commercial and industrial redevelopment. Simply comparing those projects which were exclusively residential to the total, they represent only about 15.5% of the total private investment made under the 353 program in the period between 1961 and 1981. This figure is obtained by calculating total completed and anticipated private investment, both in and outside the Central Business District ($602,564,000) and dividing that into the total completed and contemplated private investment in projects *devoted solely* to residential occupancy ($93,100,000).

[18] 42 U.S.C. § 1437f (1988 & Supp. II 1990). The Section 8 program was enacted as part of the Housing and Community Development Act of 1974, Pub. L. No. 93-383, 88 Stat. 633 (codified as amended in scattered sections of 42 U.S.C. and other titles), as a primarily "demand-side" subsidy which supplemented the difference between the market rate rent and 30% of a tenant's monthly adjusted income. 24 C.F.R. § 813.107(a) (1992). The Section 8 program has all but been repealed. *See generally* John R. Nolan, *Reexamining Federal Housing Programs in a Time of Fiscal Austerity: The Trend Toward*

seventy-seven percent of those were for the elderly or the handicapped,[19] politically favored groups with respect to whom the federal inducements are the greatest.[20] Approximately thirty percent of the total residential units were either "market rate" apartment units or condominiums.[21]

Finally, a more recent study of the 353 program concluded that while the program was originally conceived as focusing on housing, that focus was not primarily intended to directly benefit low- and moderate-income persons. Rather, housing was a means to remove the blight which accompanies slums and to attract back to the city those middle- and upper-income people who had fled to the suburbs.[22] Like the 421 program in New York City, the 353 abatement program in Missouri is an economic development program, not a poverty program; housing for the "impoverished" is left to public housing.[23]

The results of these studies should not be surprising. First and most obviously, real estate tax abatements themselves are not deep enough to lower rents sufficiently to benefit low-income persons. Using the St. Louis study's assumptions that tax abatement would reduce rents by twenty dollars a month[24] and that a tenant should spend no more than twenty-five percent of her gross earnings on rent,[25] the difference in annual income required to occupy the unit after tax abatement compared to that

---

Block Grants and Housing Allowances, 14 URB. LAW. 249 (1982); Simons, *supra* note 5, at 264–68.

[19] MANDELKER ET AL., *supra* note 16, at 28.

[20] Two examples of favored treatment for the elderly and handicapped stand out. First, the Housing Act of 1961 permitted HUD to pay up to an additional $120 per unit for any unit occupied by a "senior citizen" when necessary to ensure the financial solvency of the project. *See* Lawrence M. Friedman, *Public Housing and the Poor, in* HOUSING URBAN AMERICA 454 (Jon Pynoos et al. eds., 1973). Second, the Section 236 program included within the definition of "family" "individuals" 62 years or older or handicapped. 12 U.S.C. § 1715z-1(j)(2)(B) (1988). Moreover, "[a]t least 20 per centum of the total amount of contracts for assistance payments" must be made with respect to elderly or handicapped "families." *Id.* § 1715z-1(i)(4).

[21] MANDELKER ET AL., *supra* note 16, at 28.

[22] Michael M. Shultz & F. Rebecca Sapp, *Urban Redevelopment and the Elimination of Blight: A Case Study of Missouri's Chapter 353*, 37 WASH. U. J. URB. & CONTEMP. L. 1, 42–47 (1990).

[23] *Id.* at 45–46. The authors also argue that federal programs have shifted their focus from blight removal and lower-income housing to economic development. *Id.* at 34. This is evidenced by Community Development Block Grants, 42 U.S.C. §§ 5301–5317 (1988 & Supp. II 1990), and Urban Development Action Grants, 42 U.S.C. § 5318 (1988 & Supp. II 1990). *See* Shultz & Sapp, *supra* note 22, at 34.

[24] MANDELKER ET AL., *supra* note 16, at 57.

[25] *Id.* at 57 n.42.

required before tax abatement is only $960.[26] For these reasons, Alpert correctly concluded that tax abatements need to be combined with other subsidies such as land write-downs, mortgage subsidies, and rent subsidies in order to make housing affordable for persons with low and moderate incomes.[27] Some land write-downs are funded locally through tax increment bonds[28] and a number of states have housing finance agencies empowered to provide mortgages for low-income housing at below-market interest rates.[29] But the subsidies which have been deep enough to put standard housing within the reach of lower-income households have, in the past, come from the federal government, and these have all but vanished.[30]

The evaporation of sufficient federal housing subsidies does not, however, fully explain why tax abatement programs have stimulated precious little private investment in housing for low- and moderate-income persons. The plain fact is that after decades of debate it has never been empirically established that real estate tax abatements play any significant role in a potential developer's decision to improve her property. Indeed, despite the theory, much of the empirical evidence suggests that they

---

[26] Because costs are different, the income difference is somewhat more, $2,168, for single-family homeowners. *Id.*

[27] Alpert, *supra* note 1, at 7.

[28] Tax increment bonds operate similarly to tax abatement, except that the abated taxes are in effect paid back by the developer over the life of the bonds. Acquisition and site improvement costs can be financed by the issuance of bonds, the amortization of which is funded by separating the real estate into two assessed value segments, that which existed before the acquisition (and presumed resale to the redeveloper) and that which exists after the improvements are made by the redeveloper. The redeveloper pays the real estate taxes based on the first value and pays the increment between the first and second values into a fund which the city uses to amortize the bonds.

The advantage to the city is that it will eventually recoup all or part of its acquisition and site improvement costs while still continuing to collect real estate taxes on the property. The advantages to the redeveloper are that (1) it pays the acquisition costs over a period of time at an interest rate below that at which it could borrow from a commercial lender, and (2) it can deduct from its federal adjusted gross income the full amount of its payments to the city, whether the money goes into the general fund as real estate taxes or into a special fund to pay the bond interest and principal. *See generally* Jonathan M. Davidson, *Tax Increment Financing as a Tool for Community Development*, 56 U. DET. J. URB. L. 405 (1979).

[29] *See, e.g.*, ILL. REV. STAT. ch. 67½, paras. 301–304 (1992); MASS. GEN. L. ch. 23A app., §§ 1-1 to 2-18 (1981 & Supp. 1992); PA. CONS. STAT. § 1680.101–.603a (1977 & Supp. 1992). For an analysis of the impact of federal tax policy on the capital acquisition efforts of state housing finance agencies, see Trevor W. Nagel & Walter J. St. Onge III, *Housing Bonds and Tax Reform: The Perils of a Partial Analysis of Low-Income Housing Programs*, 6 YALE L. & POL'Y REV. 287 (1988).

[30] *See* Simons, *supra* note 5, at 260; Urban Housing Working Group, *supra* note 5, at 64–65, 69.

play little, if any, role at all.[31] As I will demonstrate below, however, even if tax abatements were the incentive their proponents claim,[32] the inducement would not be in the direction of housing for low- and moderate-income persons.

II. Economic Theory and Economic Reality

Three economic factors restrain real estate tax abatements from steering private investment to housing for low- and moderate-income persons. First, the price of housing for low- and moderate-income persons is not necessarily profit maximizing, and may be determined in a very different manner than Alpert seemed to posit as he traced the rational investment policies of the typical urban, inner-city property owner.[33] Rather than continuing to *invest* in her property up to the point when "the ratio of rental return to cost is one," an owner of property in what Alpert described as "'deteriorated' or 'redevelopment'" areas[34] is equally likely to *disinvest* by deferring investment or maintenance in order to capture or recapture as much cash as possible before the neighborhood "goes to pot," often staying just one step ahead of the housing code inspector. This behavior, ordinarily known as "milking," quite literally liquidates a property's value much as a corporation might liquidate its fixed

---

[31] *See* Abt Assocs., U.S. Dep't Hous. & Urban Dev., Property Tax Relief Programs for the Elderly: Final Report 50 (1975) (reporting inability to conclude that property tax relief correlated with occupant's decision whether or not to expend resources on maintenance or repair); Norman Krumholz, *Equity and Local Economic Development*, 5 Econ. Dev. Q. 291, 294 (1991); Note, *Can State and Local Tax Incentives and Other Contributions Stimulate Economic Development?*, 44 Tax Law. 285 (1990). Some of the relevant literature is collected *id.* at 285 n.2. The Mandelker, Feder, and Collins study of property tax abatement in St. Louis reached essentially the same conclusion: at best, real estate tax considerations might be important at the margin, but even that conclusion may depend upon how the question is framed. *See* Mandelker et al., *supra* note 16, at 33–38. And that study, unlike the Krumholz article, is favorable to the use of real property tax abatements as a development incentive.

[32] *Cf.* Sternlieb et al., *supra* note 8, at 57 (concluding that approximately 50% of the privately financed units rehabilitated under the J51 program in New York City were probably induced by the tax abatement incentive).

[33] *See* Alpert, *supra* note 1, at 5. "When the ratio of rental return to cost is one, optimum quality of maintenance and capital investment for a particular building has been achieved. Further expenditure . . . would not be economical, just as failing to spend enough to reach that optimum would not be profit maximizing." *Id.*

[34] *Id.* at 28.

assets by converting them to cash.[35] The result is that the price of much slum housing is more than is justified by the quality of housing services provided. To a property owner engaged in "milking," a real estate tax abatement which results in a nine percent cost savings[36] is unlikely to prompt her to alter that policy in favor of positive investment.

A real estate tax abatement might prompt investment if the tax abatement program (1) were focused on a particular neighborhood so that the obsolescence of surrounding properties would not be capitalized negatively into the value of the new or rehabilitated property and (2) were available only to property owners who would be unlikely to adopt a policy of positive disinvestment.[37] This, I assume, is one reason some states have limited tax abatements to not-for-profit or limited dividend redevelopment corporations acting pursuant to a governmentally approved neighborhood redevelopment plan.[38]

Even a tax abatement program focused in this way, however, may not produce the desired results because of the second and third economic factors working against private investment in low- and moderate-income housing. The second factor is that property in deteriorating neighborhoods tends to be over-assessed relative to its value, whereas property in improving neighborhoods tends to be under-assessed relative to its value.[39] There are a number of explanations for this phenomenon: (1) a time lag between changes in property values and when those changes are reflected (if ever) in the assessed value of the property; (2) the relative political influence of those who live in more affluent neighborhoods versus those who live in deteriorating

---

[35] Milking is different, certainly in degree if not in kind, from what Alpert described as a property owner's "disinvestment" which occurs as an expected result of filtering. *Id.* at 4. The latter is not an attempt to liquidate the value of a building. It is merely an attempt to adjust expenditures to match the reduction in gross rents which result from lower-income occupants.

[36] This figure is derived from a comparison of the cost figures of a proposed residential building in St. Louis with and without the 353 abatement. *See* MANDELKER ET AL., *supra* note 16, at 56.

[37] There is arguably some question about whether a property owner who has decided to disinvest will pass on to her tenants any savings from tax abatement, particularly in those communities where housing vacancy rates are low and where the demand for housing thus tends to be inelastic. It is this concern which has prompted some legislatures to couple tax abatement with rent controls. *See, e.g.*, N.J. STAT. ANN. § 55:14E-8 (West 1989) (repealed 1991); N.Y. PRIV. HOUS. FIN. LAW § 31.

[38] *See, e.g.*, MICH. COMP. LAWS ANN. § 125.912 (1986); N.J. STAT. ANN. § 55:14E-11.

[39] *See generally* Kenneth K. Barr, *Property Tax Assessment Discrimination Against Low Income Neighborhoods*, 13 URB. LAW. 333 (1981).

neighborhoods; and (3) most significantly for present purposes, the different assessment ratios for single-family dwellings (which predominate in more stable neighborhoods) and multi-family tenements (which are more common in less stable neighborhoods).[40] If one assumes that property owners know of these differential assessment practices and that these practices actually influence investment decisions,[41] it is hard to imagine that real estate tax abatement will have much impact in inducing investment or reinvestment in neighborhoods with concentrated low-income housing in which property values are likely to be in decline. At best, the abatement, by helping to bridge the gap between the assessment/sales price ratios in the two neighborhoods, may remove a *disincentive* to investment in deteriorating neighborhoods.[42]

That leads to the third reason why tax abatements have not worked as an economic incentive to investment in low- and moderate-income housing. James Heilbrun's early study of the subject noted that although abatement of the typical American real estate tax (i.e., one which taxes both the value of land and its improvements) will have no impact on the replacement of old structures with new, it can encourage rehabilitation so that the "standing stock" will be adapted to "market conditions."[43] The problem is that the market is just not conducive to investment in housing for lower-income persons, at least not investment in standard housing.[44] Real estate taxes do nothing to improve that market, but even their total absence is insufficient to create a market that does not otherwise exist. Indeed, the Sternlieb study ends up supporting tax abatement for *newly constructed* housing, not because of the shelter it produces, but rather because the loss of real estate tax revenue caused by abatement is justified by the multiplier effect of the increased economic activity resulting from the construction of new units and the attraction of upper-income residents into New York City, who will pay other taxes to the city and shop in the city's stores.[45]

---

[40] *See id.* at 333–41, 382–400.

[41] *But cf. supra* notes 31–32 and accompanying text.

[42] MANDELKER ET AL., *supra* note 16, at 26.

[43] JAMES HEILBRUN, REAL ESTATE TAXES AND URBAN HOUSING 101 (1966).

[44] Alpert recognized this to be a significant impediment. His solution was to regulate the property owner who receives the tax abatement to ensure that she uses the abatement only to house low-income persons. *See* Alpert, *supra* note 1, at 15–17.

[45] STERNLIEB ET AL., *supra* note 8, at 3–6, 54. Their conclusion with respect to the

III. THE POLITICS OF MAXIMIZING PUBLIC INVESTMENT RETURNS

The Sternlieb study and, to a lesser degree, the St. Louis study demonstrate that market forces will drive those who supply housing to seek out the upper-income end of the demand scale and there is little that tax abatements can do to deter that movement. Politics, on the other hand, will tend to drive real estate tax abatements from housing to other forms of development which are more visible and can contribute more significantly to the tax base after the abatement period ends.

There are several factors which underlie political decisions about where and how public resources are expended. First, real estate tax abatement has never commanded widespread support even among public officials responsible for administering those programs.[46] There are many reasons for this, but most boil down to the once fashionable policy of "trickle-down economics." The direct beneficiary of the subsidy[47] is not the intended beneficiary, in this case the housing consumer. Rather it is assumed that by providing a benefit to the investor, that benefit will "trickle down" to the intended beneficiary in the form of lower rents and/or higher quality housing services. This policy of giving to A so that B might benefit is viewed by many as at best

---

J51 rehabilitation program, which is not designed to attract higher-income persons back to the city, is quite the opposite—that program is "less likely . . . to pay back its initial subsidy." *Id.* at 6.

[46] *See, e.g.*, Krumholz, *supra* note 31. Professor Krumholz was for many years the Director of City Planning for the city of Cleveland, Ohio.

Even the highly touted "enterprise zone" legislation, which exists in some 38 states and has been proposed at the federal level, is in trouble in some quarters. In Ohio, for example, the current enterprise zone legislation, OHIO REV. CODE ANN. § 5709.61–.66 (Baldwin 1991 & Supp. 1992), is due to expire on December 31, 1993. In fact, last December the legislation was extended through 1993 only after a difficult fight in the Ohio legislature, particularly the Ohio House. The major objection is that wealthier suburbs are "cleaning up" at the expense of distressed central cities; the rich are getting richer while the poor continue to languish. Timothy Heider, *Enterprising Abatements Draw Criticism*, CLEV. PLAIN DEALER, Dec. 14, 1992, at B1. Moreover, the Ohio Senate Economic Development Committee has undertaken a comprehensive review of tax abatement and the committee's chairman has predicted major changes in the enterprise zone law. *See* Dick Kimmins, Gannett News Service, Mar. 9, 1993, *available in* LEXIS, Nexis Library, Omni File.

[47] A tax abatement or tax exemption is a subsidy just as certainly as is a direct cash grant. *See* Stanley S. Surrey, *Tax Incentives as a Device for Implementing Government Policy: A Comparison with Direct Government Expenditures*, 83 HARV. L. REV. 705 (1970).

inefficient[48] and at worst regressive.[49] Both criticisms weaken support for abatement programs.

A second factor underlying political decisions about where and how public resources are expended is that cities, counties, and other forms of general purpose units of local government that make tax abatement decisions must satisfy not only their own constituencies but those represented by other taxing districts with very different priorities. School, library, and sewer districts located within the boundaries of the political decision-maker are dependent upon the same property tax base for their revenue. As cities shift from a property tax to an income tax base,[50] they can more easily trade any loss in property tax revenue for an increase in sales taxes, income taxes, and utility taxes. On the other hand, other taxing districts do not have that luxury; they are ordinarily wholly dependent on the property tax. Because they can no longer rely on a unity of fiscal interests to protect their revenue source, these other taxing authorities are becoming more vocal in their opposition to municipal tax abatement decisions.

It is difficult to determine empirically how much these districts lose as a result of city and county real estate tax abatement decisions. The study of the 353 program in St. Louis tried to calculate those losses under varying assumptions about project completion with and without the tax abatement. Not surprisingly, no clear results were produced, particularly when assumed decreased expenditures for education by school districts were partially offset by assumed increased expenditures by the city for sanitation and health made possible by the increased economic activity generated by the tax abated projects.[51] But however the numbers are manipulated, the perception remains

---

[48] *See* Alpert, *supra* note 1, at 11–12; Janet Stearns, *The Low Income Tax Credit: A Poor Solution to the Housing Crisis*, 6 YALE L. & POL'Y REV. 203, 225 (1988) (arguing that only about 50% of the federal income tax credit for low- and moderate-housing investment, I.R.C. § 42 (1992), is passed through to the housing consumer).

[49] The argument that tax abatements granted to investors are regressive is based on the alleged disparity between the income group which receives the benefit and the income group which bears the ultimate burden. If one assumes (a) that the services financed by the property tax are used largely by those in the lower-income brackets, not an unreasonable assumption given the demographics of many large "rust belt" cities, and (b) that some benefit to upper-income investors is inevitable (and it is—otherwise the incentive would not be an incentive), the tax abatement is arguably regressive.

[50] *See* DANIEL R. MANDELKER ET AL., STATE AND LOCAL GOVERNMENT IN A FEDERAL SYSTEM 199 (3d ed. 1990).

[51] MANDELKER ET AL., *supra* note 16, at 59–63.

that business interests are enriched at the expense of public school children.[52]

The political dissatisfaction from inside and outside the political unit granting the abatements creates enormous pressure to demonstrate that any revenue lost as a result of the abatement is made up in other ways. The political unit can point to revenue increases from two areas: (1) an increase in employment, which to a city represents both an increase in political capital and an increase in tax revenues through increased payroll taxes and sales tax revenues generated by increased spending in city stores (the economic multiplier effect); and (2) increased contribution to the real estate tax base as the abatement gradually declines in the future. Granting a tax abatement subsidy on the value of improvements not only expends a portion of a city's fiscal resources but a portion of the political capital of those responsible for governing the community as well. Because political capital, like investment capital, is a limited resource, one who expends her political capital hopes to profit from that expenditure.

In terms of the multiplier effect of construction dollars it would not seem to matter whether the public invests in constructing housing for low- and moderate-income families or a major office building in the central business district, assuming those who are employed by both projects are drawn from the same labor pool. A dollar invested in one activity produces the same economic activity as a dollar invested in the other. On the other hand, once the construction period is over, there remains a significant period of time during which real estate taxes are abated on either project. It is during this period that the government and the public perceive that a tax loss is occurring and that the revenue must be recovered in other ways. Low-income residential projects are at a distinct disadvantage here.

Two factors conjoin to encourage the use of tax abatement "expenditures" for commercial and industrial development in the central business district rather than for low- and moderate-income housing in the neighborhoods. First, low-income persons, by definition, have relatively little disposable income.

[52] *See* Krumholz, *supra* note 31, at 297 (noting that tax abatement granted by Cleveland for two bank headquarters cost the Cleveland School District some $136 million over the 20-year period of the abatement). These perceptions, correct or not, are exacerbated by the fact that most abatement decisions are privately negotiated between a city administrator and a developer. Even the requirement of local legislative approval does not totally dispel the idea that the benefits of government inure primarily to the "rich and powerful."

Consequently, compared to the individuals likely to inhabit the brand new corporate offices in the tax-abated local bank building, those who occupy the rehabilitated low- and moderate-income housing are not going to contribute as significantly to the city's payroll tax revenues, the city's share of the sales tax revenues, or the revenues that the city derives from whatever utility taxes it levies.

Second, and quite apart from economic multiplier effects, a community which spends an equal amount of tax abatement dollars on commercial and industrial development as it does on residential development, thus adding the same dollar amount of improvements to the real property tax base, is likely to derive more real estate tax revenues after the abatement period from the former than from the latter because commercial and industrial property usually is assessed at a higher value ratio than is residential property.[53]

Therefore, it should come as no surprise that both studies of Missouri's 353 program indicated not only that the program induced very little investment in low- and moderate-income housing, but that the program induced relatively little private investment in any housing market, low, moderate, or high. While it may be impossible to demonstrate, my suspicion is that these figures are at least as much a political phenomenon as they are a market phenomenon. If that intuition is correct, or even partially correct, the hope that real estate tax abatement can be an effective tool to encourage private investment in housing for lower-income persons is little more than a pipe dream even in the unlikely event that the abatement can be deep enough to overcome market disincentives.

IV. Conclusion

There is one major issue which Alpert discussed that, up to this point, I have chosen to largely ignore—the equity concern.

---

[53] *See, e.g.*, Barr, *supra* note 39, at 352–53. Absent permissive language in the state constitution's tax uniformity clause, some courts have been reluctant to approve legislative or administrative classifications of real property according to use. *See, e.g.*, Bettigole v. Assessors of Springfield, 178 N.E.2d 10 (Mass. 1961) (administrative classification); Gottlieb v. City of Milwaukee, 147 N.W.2d 633 (Wis. 1967) (legislative classification). *See generally* Wade J. Newhouse, Constitutional Uniformity and Equality in State Taxation (2d ed. 1984) (analyzing each state's land tax policy and history).

I have done so not because I think that the issue is unimportant. Indeed it is the equity issue which is at the core not only of the debate about real estate tax abatements to those who invest in lower-income housing, but, more fundamentally, the national debate about indirect incentives through tax expenditures versus direct subsidies to those in need. What I have attempted to demonstrate is that even if all that Alpert proposed to do with tax abatements ensures that their benefits inure only to low-income housing consumers, there is little political reward for expending capital in that way.[54] Even if there were, the economics of the real estate tax system makes success unlikely.

Even if those conclusions prove to be wrong, there are serious questions about whether Alpert's proposals for ensuring that property owners use their abatements only for lower-income persons stand any realistic chance of success. Alpert proposed to limit the locations in which abatements would be available to those areas where low-income housing is concentrated and to establish a process by which the municipality can monitor the occupants of any abated structure, reducing the abatement a certain amount for each ineligible tenant.[55] While a detailed analysis of the benefits received from such a monitoring system versus the administrative costs of securing those benefits is beyond the scope of this commentary, it is overwhelming to contemplate establishing and maintaining a bureaucracy to monitor the tenancy of even as few as 1000 units of lower-income housing located on various sites scattered among many neighborhoods, sometimes, maybe often, separated by long distances and owned by different persons. Indeed, if it is true that cities have been unable to police major tax-abated commercial and industrial projects for compliance with contractual employment promises,[56] it is hard to believe they will be up to the far more difficult task of monitoring the leasing practices of the many and widely dispersed private lessors who have agreed to limit their tenants to those of lower income.

My disagreement with using tax abatements to encourage housing for low-and moderate-income persons is, however,

---

[54] I recognize that "doing good" can have a powerful political appeal. But a cynic might point to the history of the federal housing programs in the last two decades and conclude that "doing good" loses its appeal when the financial sacrifices become apparent. The demise of federal housing subsidies is traced in Nolan, *supra* note 18, at 249–50, 253–57, and Simons, *supra* note 5, at 262–68.

[55] *See* Alpert, *supra* note 1, at 28–30.

[56] *See* Krumholz, *supra* note 31, at 293.

more fundamental. It has now been fifty-five years since the Housing Act of 1937[57] introduced public housing for low-income persons. Since then we have tried a variety of supply-side incentives to encourage private investment in low-income housing, from extending the time over which mortgages are outstanding, thus reducing the monthly amortization costs, to federally insured mortgages at fixed, below-market interest rates, to direct subsidy payments to mortgagees providing financing to private sponsors of low- and moderate-income housing. In 1974 we shifted gears, turning to demand-side subsidies in the form of Section 8 payments to landlords to supplement low-income tenants' rent payments. We then abandoned that approach in favor of several forms of federal grants directly to local governments that those governments could use to promote public/private partnerships to house lower-income families. All of these subsidies, with the exception of the first, were designed to encourage private investment in the lower-income housing market. While housing units were made available with the assistance of these programs, they were not made available in the numbers required.

It may now be time to abandon our attempts to attract private capital into a market which, given the alternatives, will not sustain it, and return to publicly supplied housing for lower-income households. Tax abatements would then be unnecessary; the property would be exempt because it would be publicly owned. Reimbursement for the costs of municipal services could either be required as payments in lieu of taxes, ultimately paid by the federal government with its progressive national income tax, or be borne by local taxpayers in the same way that they bear the cost of local general assistance or the required local share of federal categorical assistance and medicaid, by paying for services enjoyed by others.

I do not suggest that real estate tax abatements have been a direct impediment to moving in this direction; they have not. They have, however, indirectly impeded movement to greater public responsibility by helping to perpetuate the myth that a small amount of public investment can induce the private, profit-motivated sector to produce significant benefits to the poor.

---

[57] 42 U.S.C. §§ 1404a–1440h (1988).

ARTICLE

ASBESTOS: A MULTI-BILLION-DOLLAR CRISIS

CHRISTOPHER F. EDLEY, JR.*
PAUL C. WEILER**

In past issues of the Harvard Journal on Legislation, *authors have delved into the persistent problem of asbestos litigation. In Volume 20, Louis Treiger explored legislative proposals before the Ninety-Seventh Congress in his article,* Relief for Asbestos Victims: A Legislative Analysis. *Bruce H. Nielson's note in Volume 25 explored the potential of class actions to address the needs of victims of asbestos and other mass torts.*

In this Article, Harvard Law Professors Edley and Weiler revisit the asbestos litigation problem and suggest new solutions. Rejecting traditional legal strategies as unworkable, the authors propose a legislative approach that combines the scope of a class action with the consistency of an administrative system. Their proposal envisions a central fund to which claimants can apply without the need to show fault. Clear medical guidelines are applied, and damage awards and legal fees are strictly controlled. Finally, the authors suggest administrative and judicial alternatives that closely resemble the legislative ideal, should legislation be politically impossible.

Federal and state courts are clogged with 100,000 asbestos suits, and that number is rising every month. These suits pose a series of seemingly intractable policy problems:

---

* Professor, Harvard Law School. B.A., Swarthmore College, 1973; M.P.P., Kennedy School of Government, 1978; J.D., Harvard Law School, 1978. Professor Edley has recently become Program Associate Director of the Office of Management and Budget in the Executive Office of the President, where some of his responsibilities include legislative and budget matters affecting the Judiciary and the Departments of Justice and Commerce.

** Professor, Harvard Law School. B.A., University of Toronto, 1960; LL.B., Osgoode Hall Law School, 1964; LL.M., Harvard Law School, 1965. Professor Weiler was Chief Reporter for the American Law Institute's tort reform study, whose report, ENTERPRISE RESPONSIBILITY FOR PERSONAL INJURY (1991), addressed many of the issues discussed in this Article.

We benefitted from the very able research assistance of Takemi Ueno and Timothy Wilkins, and from discussions with our Harvard colleague David Rosenberg and Professor Peter Schuck of Yale Law School. In addition, the research reflected in this Article was funded in part by the Center for Claims Resolution (CCR), an organization established to pursue alternatives to asbestos litigation. The views contained herein are, however, entirely the authors' own and do not necessarily reflect those of anyone else, including the CCR or any of its constituent members.

• Many victims of asbestos exposure are mortally ill from cancer or severe respiratory diseases but must wait years before their claims are resolved—often after the victims have died.

• Tens of thousands of these tort claims have been made, many successfully, by individuals who are understandably worried about their exposure to asbestos but who are not now and *never will be afflicted* by disease.

• The primary early targets of asbestos litigation—the major suppliers of asbestos and asbestos-related products sixty years ago—paid billions of dollars in tort damages. Then, facing many more billions of dollars in prospective tort liability, more than a dozen major American corporations went bankrupt. Others had to cut back severely on their growth and development programs.

• Lawyers then cast the litigation net further to find corporate pockets deep enough to satisfy the vast numbers of pending and future tort claims. Judicial legerdemain helped fill that gap with doctrinal innovations that imposed liability on firms (or their insurers) whose "misdeed," for example, was buying asbestos-related companies in the 1960s and early 1970s—after the human tragedy but before the litigation disaster.

• One key constituency has not been a victim of the asbestos tragedy: lawyers. Of the $7 billion already spent on claims, sixty percent—more than $4 billion—has been spent on fees and expenses of the plaintiff and defense bars. This figure does not account for the large amounts expended by our already overburdened civil justice system.

Throughout the 1980s, many state legislatures passed statutory reforms to their common law standards of product liability and medical malpractice. In the 1990s, the battle over tort reform has shifted to Washington, where the American Tort Reform Association and the American Medical Association are locked in a standoff with the American Trial Lawyers Association. During the 1992 presidential election campaign, former President George Bush and former Vice President Dan Quayle—in an unsuccessful effort to divert the public's attention from our declining economic productivity and spiraling health care costs—continually harped on a "litigation explosion" they attributed to personal injury lawyers wearing "tasseled loafers."[1]

[1] Colman McCarthy, *Joseph Rau and the Public Interest*, WASH. POST, Sept. 15, 1992, at D15.

We believe that the explanation for the lack of action on Capitol Hill is not simply political gridlock: there are serious intellectual disagreements regarding the nature of the tort problem, and what, if anything, should be done about it.[2] We are law professors, not "lawyer-bashers." We acknowledge that tort litigation imposes substantial and often unwarranted economic harm on defendants. However, tort litigation also provides indispensable redress for victims and protection for the public.

Despite conflicts about tort reform in general, there is a consensus among scholars and judges on one point: asbestos litigation presents a tort problem with a unique history, present state, and future course. It is a problem that cries out for major reform.

As a result of corporate actions dating back to the 1930s and 1940s, the American legal system spent approximately $7 billion on asbestos litigation in the 1980s and early 1990s. Many billions of dollars more will be spent well into the next century. The early stages of this litigation spiral displayed the tort system's virtues—hard-working, imaginative lawyers discovering the evidence, and sympathetic, creative judges overcoming a variety of legal obstacles posed by this innovative toxic tort. In the 1970s, tort lawyers identified corporate (and others') responsibility for the asbestos tragedy, and, in the 1980s, escalated their filings until responsible elements of the business community were ready to endorse and fund a more accessible, equitable, and generous system of relief for asbestos victims.

Currently, however, we know that sixty percent of liability funds continue to be spent not on asbestos victims, but on lawyers, expert witnesses, and others involved in disputes over who will pay and who will be compensated. Some workers who were severely injured by asbestos exposure are unable to document their claims or get their cases resolved before they die. At the same time, much of the money that actually reaches claimants goes to people who do not and will not suffer any physical impairment. Furthermore, the ultimate social cost of asbestos litigation is considerably greater even than direct expenditures on tort claims. In the intensely competitive international economy in which we now live and work, there are no

---

[2] *See generally* AMERICAN LAW INST., ENTERPRISE RESPONSIBILITY FOR PERSONAL INJURY (1991) [hereinafter ENTERPRISE RESPONSIBILITY] (Professor Weiler served as Chief Reporter for this report).

true "deep pockets." Economic dislocation from spiraling litigation drives firms to bankruptcy or its brink, and imposes substantial financial burdens on the present shareholders, employees, pensioners, and communities of asbestos defendants.

A drastic overhaul of the current legal approach to asbestos is long overdue. Members of Congress should turn their attention away from the present debate about products liability's impact on small plane manufacturers and the impact of medical malpractice on obstetricians. These are debates involving two distinct, intellectually plausible, and emotionally powerful viewpoints. Far more worthy of congressional attention and action is an asbestos tragedy with human, financial, and legal costs that dwarf those created by other examples of medical and product liability. There is no serious disagreement within the scholarly or judicial arenas that the current asbestos regime is inadequate and in need of reform. Under the current system, remedying the effects of asbestos could cost many billions of dollars. But it need not.

How might Congress reform asbestos litigation? We favor an administrative mechanism, under which compensation of victims would reflect benefit ranges applied through informal adjudication, with access to court preserved only for appellate review of the legality of administrative action. Awards would be financed primarily by asbestos defendants (including their successors and insurers), expenditures on attorneys would be sharply limited, and unimpaired claimants would be the last to be compensated. All pending and future state tort litigation would be preempted.

If, however, our new President and Congress cannot be galvanized, our courts must continue to fill the void. Judges must remake traditional doctrines to deal with the costs of second-generation asbestos litigation just as boldly as they did in the 1970s and 1980s, when they first fashioned tort remedies for asbestos exposure.

The judicial approach we favor involves the use of class action techniques to emulate the key features of a sound administrative solution. A group of plaintiff and defendant attorneys recently negotiated an initiative along these lines. It involves a settlement class action of all future asbestos claims against a group of defendant companies operating under the umbrella of the Center for Claims Resolution, a non-profit entity formed specifically to pursue negotiated alternatives to protracted litigation. The pro-

posed settlement, now pending in the Eastern District of Pennsylvania, is an effort to resolve most claims through a streamlined, non-tort procedure. Impaired claimants would be paid promptly, and the claims of the unimpaired would be deferred unless and until an impairment develops. "Evergreen" funding by the participating companies would ensure eventual payment of all claims by impaired claimants, but total payments for a given year are limited (based on historical claims resolution rates) to make the resulting economic burden manageable. Attorneys' fees would be capped. A right of eventual recourse to litigation is preserved for at least some claimants, but only after full effort to resolve the case through less contentious and costly alternatives.

I. ASBESTOS: THE HUMAN TRAGEDY

A. *Asbestos Use*

When asbestos first appeared on the market in the late nineteenth century, it was touted as a miracle substance—able to withstand punishing forces of fire, corrosion, and acid, while also versatile enough to weave into textiles, line automobile brakes, retard shipboard fires, and bind rockets together. As United States District Court Judge Jack B. Weinstein remarked, "[t]heatrical audiences were once comforted by the thought that huge asbestos curtains between the audience and stage protected against the spread of fire."[3] In its most important use, asbestos was the designated substance incorporated in Navy warships to protect American seamen from fires caused by enemy bombers and submarines in World War II.

Early in the twentieth century, however, evidence began to emerge that these benefits from asbestos use were secured at the price of serious dangers to workers. Those who mined asbestos and manufactured asbestos products for use as insulation in ships, factories, and the like were at risk.[4] Because asbestos-produced disease has a lengthy latency period—anywhere from

[3] *In re* Joint E. & S. Dist. Asbestos Litig. (Findley v. Blinken), 129 B.R. 710, 736 (E. & S.D.N.Y. 1991), *vacated*, 982 F.2d 721 (2d Cir. 1992).

[4] Although certain hazards were becoming known, the full extent of the hazards and the type of response necessary were less clear.

ten to forty years[5]—it took decades to generate epidemiological evidence of the risks this substance poses to human beings.[6] Gradually, medical scientists documented the fact that asbestos exposure can produce a number of respiratory disorders: asbestosis, mesothelioma (a rare malignant cancer), and lung cancer (particularly in combination with smoking).

Evidence of at least some of these risks—in particular, of asbestosis, suffered by asbestos workers exposed to high concentrations of this substance—was beginning to emerge in the scientific literature by the 1930s.[7] From the point of view of the plaintiffs' bar, the true disgrace of the asbestos story is their belief that senior executives of some of the country's leading producers—particularly Johns-Manville—were not only aware of these risks, but took active steps to suppress knowledge of the danger in order to protect the sales of their product.

It is by no means clear that a substantial drop in sales was an inevitable outcome of disclosure. There were employers who took some steps to reduce exposure risks. Asbestos was a life-saving as well as a life-threatening substance, and there was no substitute readily available, especially for ships in the nation's war effort.[8] But if all those who were aware of the risks (including United States Government officials) had heeded the warnings in the medical literature, many protective devices and techniques could have been adopted to reduce occupational hazard levels drastically.

Instead, from the 1940s through the 1960s, millions of American workers were exposed to asbestos with little or no precautions.[9] As a consequence, several hundred thousand Americans

---

[5] PAUL BRODEUR, OUTRAGEOUS MISCONDUCT: THE ASBESTOS INDUSTRY ON TRIAL 185 (1985); Deborah R. Hensler, *Fashioning a National Resolution of Asbestos Personal Injury Litigation: A Reply to Professor Brickman*, 13 CARDOZO L. REV. 1967, 1973 (1992).

[6] BRODEUR, *supra* note 5, at 12–13.

[7] Scientific documentation of the risk of asbestosis among insulators and other workers handling finished asbestos products, and of the risk of cancer among different types of asbestos workers, occurred some time later. *See* Hans Weill & Janet N. Hughes, *Asbestos as a Public Health Risk: Disease and Policy*, 1986 ANN. REV. PUB. HEALTH 171, 173. For a more detailed account, see David E. Lilienfeld, *The Silence: The Asbestos Industry and Early Occupational Cancer Research: A Case Study*, 81 AM. J. PUB. HEALTH 791 (1991).

[8] BRODEUR, *supra* note 5, at 66; Lester Brickman, *The Asbestos Litigation Crisis: Is There a Need for an Administrative Alternative?*, 13 CARDOZO L. REV. 1819, 1884 (1992); Lilienfeld, *supra* note 7.

[9] Although the government caused shipyard workers to be exposed to asbestos during the World War II era, recently the government has regulated workplace and environmental exposure. These efforts, together with the development of alternative materials

were fated to suffer asbestos-related diseases, and many thousands to die, beginning in the 1960s and continuing well into the next century.[10]

B. *An Out-of-Date Morality Play*

It is important to remember this oft-told tale of the "outrageous misconduct" on the part of many companies and the government.[11] Some plaintiffs' lawyers continue to pound the

---

for use in the building trades and other applications, have reduced the proportion of the work force at risk from the effects of excess current exposure to asbestos. For a brief description of the variety of such federal regulatory efforts, see Corrosion Proof Fittings v. EPA, 947 F.2d 1201 (5th Cir. 1991) (reversing an EPA rule that would have eliminated all domestic uses of asbestos).

Researchers at the Rand Corporation have documented the use of asbestos in the United States. During the World War II period, 1935–1950, annual asbestos consumption grew from less than 150,000 metric tons in 1935 to more than 700,000 metric tons. However, though increased asbestos use is often attributed to the wartime effort, United States consumption never dipped below 600,000 metric tons until after its peak in 1974, when it reached 800,000 metric tons. Immediately thereafter, greater awareness of health risks combined with increased government regulation led to a sharp drop in consumption. By 1985 the annual figure was only 200,000 metric tons. *See* DEBORAH HENSLER, ASBESTOS LITIGATION IN THE UNITED STATES: A BRIEF OVERVIEW 5–6 (1992).

Still, asbestos needs continue in a few key industrial areas. *See Corrosion Proof Fittings*, 947 F.2d at 1220–28 (listing current uses, including rocket engines, battery separators, automobile drum and disc brakes, asbestos-cement pipe products, gaskets, roofing, shingles, and paper products).

[10] A decade ago, expert projections were considerably higher than those noted in the text. In 1982, Doctors William J. Nicholson, George Perkel, and Irving J. Selikoff predicted more than 80,000 excess deaths from asbestos-related cancers alone during the period 1967–2027. *See* William J. Nicholson et al., *Occupational Exposure to Asbestos: Population at Risk and Projected Mortality—1980–2030*, 3 AM. J. INDUS. MED. 259 (1982). Recent projections by many of the same scholars, using better data about actual levels of worker exposure, are more conservative. *See* Herbert Seidman & Irving J. Selikoff, *Decline in Death Rates Among Asbestos Workers 1967–1987 Associated with Dimunution of Work Exposure to Asbestos*, 609 ANN. N.Y. ACAD. SCI. 300 (1990); David E. Lilienfeld et al., *Projection of Asbestos-Related Disease in the United States, 1985–2009*, 45 BRIT. J. INDUS. MED. 283 (1988). On the other hand, the great bulk of present claims are by unimpaired individuals who will never become ill, and it is to these individuals—and their lawyers—that the bulk of resources currently flow. *See* HENSLER, *supra* note 9, at 6. As we will observe later in the Article, the question of whether and when a physiological condition caused by asbestos exposure amounts to a "disease" is a matter of intense legal debate. Some physiological changes, such as diffuse "plaque," or fibrous spotting of the pleural membrane, cannot be detected without clinical testing and produce no impairment of the person's ability to function in daily life. At the other end of the scale are the asbestos-related cancers which cause death several decades after the fatal exposure.

[11] *See, e.g.*, BRODEUR, *supra* note 5. This was also the title of Brodeur's journalistic account first published in *The New Yorker*, and then in expanded book form. *See* David Rosenberg, *The Dusting of America: A Story of Asbestos—Carnage, Coverup, and Litigation*, 99 HARV. L. REV. 1693 (1986) (reviewing Brodeur's book). A recent example of plaintiffs' lawyers retelling the same story, supposedly to guide asbestos liability policy for a half century, is Ronald L. Motley & Susan Nial, *A Critical Analysis of the Brickman Administrative Proposal: Who Declared War on Asbestos Victims' Rights?*, 13 CARDOZO L. REV. 1919 (1992).

table in courtrooms and legislative hearing rooms. They tell their listeners that too much blood has been spilled by corporate misconduct for us ever to deprive plaintiffs of the full force of tort retribution by substituting an administrative compensation program, however socially "efficient" the latter might seem.

Whatever those lawyers say about the events that took place in the 1930s and 1940s, this country and its economy need a more analytical perspective about what reforms are appropriate in the 1990s. It was not tort litigation that discovered the asbestos tragedy, it was medical science. The true hero of that story is not the plaintiffs' bar, but Dr. Irving Selikoff. And the first legal response came not via tort litigation, but through workers' compensation and occupational safety and health regulation: two administrative programs that largely removed asbestos from the workplace by the early 1970s—well before the first major jury verdicts were announced.

The importance of tort litigation was its role, beginning in the mid-1970s, in revealing the failure of a number of major corporations to warn workers of the severe risks to which they were exposed. Also crucial was tort's role in meting out the kinds of legal sanctions—approximately $7 billion in claims expenditures[12] and sixteen corporate bankruptcies so far[13]—that should make business executives think twice before following the same path as their asbestos manufacturing predecessors. The tort system, surely, has succeeded in delivering society's punitive response to the events of the 1930s and 1940s.

In any case, the original responsible executives have left their firms—in fact, most are dead. Also, many of the firms most involved—Johns-Manville for example—have themselves gone through bankruptcy proceedings (at severe financial loss to their shareholders) and are thus protected from further tort litigation. The firms now paying the price of the 100,000 suits pending,

---

[12] *See* Suzanne L. Oliver & Leslie Spencer, *Who Will the Monster Devour Next?*, FORBES, Feb. 18, 1991, at 75, 79.

[13] The companies that so far have filed for bankruptcy because of asbestos litigation are Amatex, Carey-Canada, Celotex, Eagle-Picher Industries, Forty-Eight Insulations, Johns-Manville, National Gypsum, Raymark Industries, Standard Insulation, UNARCO, and UNR Industries (the parent of UNARCO), *see* Judicial Conf. Ad Hoc Comm. on Asbestos Litig., Report to the Chief Justice of the United States and Members of the Judicial Conference of the United States 51 n.33 (Mar. 1991) (unpublished manuscript, on file with the *Harvard Journal on Legislation*) [hereinafter Judicial Conf.], Pacor and North American Asbestos, *see* Brickman, *supra* note 8, at 1819 n.2, and H. K. Porter, Nicolet, Brunswick Fabricators, and Hillsborough, *see* Hensler, *supra* note 5, at 1972 n.23.

which will pay for the further 100,000 suits predicted for the future, have only peripheral connections to the original disaster—often just as insurers of or successors to the firms initially involved.

Moreover, it is not corporations, but rather real people, who pay the price of tort litigation—specifically, the present stakeholders in defendant firms such as employees, shareholders, pensioners, and surrounding communities. And the people with the most compelling stake in reform are the victims of asbestos-related disease. Many of these victims will not be able to recover (at least while they are alive) the kind of compensation they need and deserve precisely because of tort litigation's preoccupation with retelling, and retaliating for, a tale of corporate "misconduct" that took place almost sixty years ago.

The asbestos crisis is more than a story of decades-old corporate failures and a contemporary public health tragedy. It is a Dickensian tale about the limitations of the traditional legal process and a general portrait of governance in this country. The picture thus far is discouraging. Judge Weinstein, the preeminent mass tort adjudicator, and Eileen Hershenov put it well:

> To many, particularly those in other countries, it may seem strange that in the United States we leave it to individual courts to provide essentially ad hoc solutions to modern day disasters with their national, social and economic repercussions. In this country, however, three factors have, by default, left the state and federal courts to their own devices: (1) the lack to date of an effective national administrative regulatory scheme capable of controlling undesirable conduct by manufacturers; (2) the absence of a comprehensive social welfare-medical scheme for compensating victims of mass torts, and (3) the lack of adequate state or federal legislation controlling these cases.[14]

II. The Inadequacies of Tort Relief

Asbestos presents a distressing picture of the inadequacies of tort litigation as a vehicle for delivering society's resources to needy and deserving claimants. The following features of the asbestos litigation system make these flaws inescapable.

---

[14] Jack B. Weinstein & Eileen B. Hershenov, *The Effect of Equity on Mass Tort Law*, 1991 U. Ill. L. Rev. 269, 270.

A. *Number of Claims*[15]

More than 100,000 asbestos claims are now pending in court, and the number grows steadily. Approximately two-thirds of these suits are filed in state courts, the other third in federal courts. In the last half of the 1970s there were 1000 asbestos suits filed in federal courts; in the first half of the 1980s, 10,000 claims; and in the last half of the 1980s, 37,000 claims. In 1990 alone there were nearly 14,000 new federal asbestos suits, three times the fewer than 5000 federal suits filed that year for every other type of product liability case. Yet in recent years, both Congress and the states have concentrated their reform efforts on the product liability "tail," rather than the asbestos liability "dog."

B. *Size of Expenditures*

Billions of dollars have already been spent on asbestos litigation: the most recent estimate puts the figure at $7 billion.[16] Though inevitably speculative, estimates have also been made of the total costs of the asbestos litigation crisis. For example, Judge Weinstein estimated that the total cost for current and future personal injury claims would be $26 to $28 billion.[17] Given that there has been no abatement in the flow of and payment for asbestos-related tort claims, it is safe to say that many billions of dollars more will be needed to resolve this human and legal tragedy.

Bankruptcies among asbestos defendants, together with the doctrine of joint and several liability, mean mounting and cumulative financial pressure on the remaining defendants, whose resources are limited. The money spent on unimpaired asbestos claimants and lawyers creates genuine economic risks to workers and pensioners with stakes in the surviving companies. While it is impossible to quantify the magnitude of those risks, the sixteen asbestos-related bankruptcies to date constitute more than one-half of the original twenty-five major asbestos defendants.[18]

---

[15] *See generally* Hensler, *supra* note 5, at 1970–72.

[16] Oliver & Spencer, *supra* note 12, at 79.

[17] *In re* Joint E. & S. Dist. Asbestos Litig. (Findley v. Blinken), 129 B.R. 710, 907 (E. & S.D.N.Y. 1991), *vacated*, 982 F.2d 721 (2d Cir. 1992).

[18] *See* Judicial Conf., *supra* note 13, at 29–30.

C. *Unimpaired Claimants*[19]

Unfortunately, though, it appears that up to one-half of asbestos claims are now being filed by people who have little or no physical impairment. Many of these claims produce substantial payments (and substantial costs) even though the individual litigants will never become impaired. The cases involve pleural plaques—"freckles on the lungs"—or pleural thickening—forms of physical change that are attributable to asbestos exposure. These plaques may or may not be followed by asbestosis or cancers that would disable or kill the asbestos claimants.

While there are understandable legal and emotional reasons why these premature suits are being filed in huge numbers, their presence on court dockets and in settlement negotiations inevitably diverts legal attention and economic resources away from the claimants with severe asbestos disabilities who need help right now.

D. *Legal Costs*

For both severely and potentially impaired litigants, of every dollar paid by defendants, over sixty cents goes to the lawyers. Adding the overhead costs of both the judicial and insurance systems, asbestos litigation consumes two dollars of society's resources in order to deliver a single dollar to people who were exposed.[20]

E. *Contingent Fees*

The usual target of popular concern about legal costs is the contingent percentage fee paid to the plaintiff's lawyer, which usually ranges from thirty to forty percent of the total settlement or award, sometimes reaching fifty percent. We believe that the contingent fee is generally an attractive feature of our tort system because it is an effective mechanism for making justice available to victims of personal injury.[21] In the early stages of

---

[19] *See generally* Peter Schuck, *The Worst Should Go First: Deferral Registries in Asbestos Litigation*, 15 HARV. J.L. & PUB. POL'Y 541 (1992).

[20] *See* HENSLER, *supra* note 9, at 21 (citing JAMES S. KAKALIK & NICHOLAS M. PACE, COSTS & COMPENSATION PAID IN TORT LITIGATION 74 (1986)).

[21] *See* ENTERPRISE RESPONSIBILITY, *supra* note 2, at 274.

asbestos litigation, lawyers who took their chances with these scientifically uncertain and legally difficult cases fully deserved the sizable rewards they won. However, now that the basic legal and scientific issues that figure in asbestos litigation are long since resolved, and huge numbers of new asbestos cases can be processed by paralegals on a largely assembly-line basis, the key plaintiffs' lawyers are earning several thousand dollars per hour of work done on these files.[22] That money would be far better spent on asbestos victims themselves, as well as on investment in better jobs for present-day workers.

F. *Delay*

It is unfair, though, to paint plaintiffs' attorneys as the villains in the picture. While legal fees do make up sixty percent of claims expenditures, half of that amount goes to defense attorneys. The underlying problem escalating legal costs is that the current system still creates too many opportunities and incentives for both sides to litigate aggressively in individual cases, even though the aggregate pattern of asbestos claims disposition has largely stabilized. The unhappy consequence is that asbestos litigation uses up a lot of time as well as money. The average time taken to dispose of an asbestos claim is thirty-one months, about twice the length of the typical civil claim.[23] Indeed, a side effect of the asbestos litigation explosion is that its huge caseload causes delay not only for asbestos claimants, but also for people bringing other types of legal claims to our already overburdened court system. Delay causes special tragedy for asbestos victims: because their exposures occurred so long ago, justice delayed can truly be justice denied. Too often the claimant is dead before payments are finally made to his or her estate.

G. *Searching for Deep Pockets*

The individual actors—the corporate and governmental employees who made those fateful decisions about asbestos use decades ago—have long since departed the scene. Their corporate and government successors, conglomerate affiliates, and

[22] *See* Brickman, *supra* note 8, at 1834–40 (reviewing attorneys' fees).
[23] Judicial Conf., *supra* note 13, at 10–11.

insurers are now left with the legal bills. While an abstract economic argument can be made about undoing the unjust enrichment of these entities, the situation actually involves a search for deep pockets to finance needed (and unneeded) compensation for asbestos claims. This is a worthy social policy objective, but tort is not the ideal mechanism with which to achieve it. Nor is it the historical function of judges and juries to act as roving revenue commissioners to fund social insurance programs.

H. *Erratic Damage Awards*

The most important source of litigation, controversy and delay is legal uncertainty.[24] The most significant source of uncertainty in all forms of tort litigation, including asbestos, is the assessment of damages, especially for the inherently subjective, non-financial damage categories of pain and suffering and punitive awards. The absence of any meaningful guidelines for jury assessment of the value of pain and suffering has produced huge inequities. Similarly-situated asbestos victims may receive payments that vary by a factor of ten or more. The fact that juries are still invited to levy huge punitive awards constitutes equally unfair treatment of defendants faced with the arbitrary prospect of multiple punishment for misconduct that took place long ago.[25]

III. Reforming the System

The previous pages synopsize the pathological nature of present-day asbestos litigation. To reform the system, we must lay to rest several myths about tort law.

---

[24] *See, e.g.*, George Priest, *Measuring Legal Change*, 3 J.L. Econ. & Org. 193, 207 (1987); George Priest, *The Common Law Process and the Selection of Efficient Rules*, 6 J. Legal Stud. 65 (1977).

[25] Indeed, in another context the Supreme Court has written that cumulative punitive damages awards raise important questions of constitutional due process. *See* Pacific Mut. Life Ins. Co. v. Haslip, 111 S. Ct. 1032, 1038–40 (1991). As the Court stated, "[o]ne must concede that unlimited jury discretion—or unlimited judicial discretion for that matter—in the fixing of punitive damages may invite extreme results that jar one's constitutional sensibilities." *Id.* at 1043.

Myth #1. Tort law must continue to exact society's revenge on corporate malefactors. This is the most important and distorting myth about asbestos litigation. Seven billion dollars of tort punishment meted out in the past decade is more than enough to express society's outrage about any misdeeds that may have taken place a half century ago and to create a powerful deterrent against similar misdeeds in the future. In the future, we must divert the remaining billions of liability dollars to redress the present-day needs of asbestos victims and their families.

If we can put tort retribution behind us, there is reason to hope for a consensus about the ingredients of a fair compensation program. Most leaders of the major defendants have clearly stated their willingness to pay for a fair compensation system,[26] among other reasons to protect their businesses from the devastating effects of litigation. Such a system would provide timely and generous benefits to those with significant asbestos-related impairments, even using the existing pattern of tort payments as the basis for calculating benefit amounts. The defendants seek to remove the specter of runaway punitive and pain and suffering verdicts in individual cases and to fund guaranteed and accessible compensation with the revenues now being spent on unimpaired claimants and unnecessary legal services for both sides. This program benefits those who need financial relief for tangible asbestos impairments. It also will preserve the economic resources of defendants for the benefit of future victims of asbestos disease, as well as for the workers, pensioners, and communities that rely on the economic viability of the defendant businesses. History's tragedies cannot be undone. It is wrong to let the memory of them stand in the way of fair redress today.

Myth #2. Asbestos-related injuries are torts, which can only be resolved fairly through individualized, full-blown court proceedings. In fact, legal business-as-usual has wasted billions of dollars, delaying and short-changing deserving claims while attorneys and experts prosper.

---

[26] *See* Andrew Blum, *Playing Asbestos Hardball*, NAT'L L.J., May 18, 1992, at 1 (noting that most companies are willing to settle, but Keene Corp. is an exception); *Keene Corporation's Glenn Bailey Calls Judge Weinstein's Asbestos Proposal "A Step in the Right Direction"*, PR Newswire, May 22, 1992, *available in* LEXIS, Nexis Library, Currnt File; Todd Woody, *As Stemple Sued, Defendants Retreated*, RECORDER, Sept. 12, 1991, at 1 (reporting that defendants were ready to settle).

Myth #3. Courts can handle the problem using massive trials and gigantic batch settlements of hundreds, occasionally thousands, of claims. Actually, this approach inevitably wastes resources on less deserving, unimpaired claimants, threatens further bankruptcies, and does nothing to provide equity among deserving claimants. Experience proves that muddling through simply cannot keep pace with the volume of claims.

Myth #4. Judges lack legal authority to address the crisis now—only Congress and state legislatures can save the day. This, too, is incorrect. Rather than wait for a political solution that may never come, judges can and must act now.

What is to be done? The answer requires consideration of the elements of an ideal solution and appraisal of the institutional capacity of legislatures and courts to approximate that ideal. We briefly describe an administrative compensation mechanism that Congress could enact that would treat claimants fairly and provide economic stability to defendants and their stakeholders. Because political gridlock makes enactment of such legislation highly unlikely, we detail the way courts can and should adopt most of these substantive measures by imaginative use of the variety of judicial powers that have been deployed for other mass tort crises.

IV. A Legislative Approach

Congress should stop relying primarily on court-centered resolution of claims for asbestos compensation. The ideal solution is an administrative alternative to the current litigation morass.

By combining the better features of workers' compensation, the Black Lung program, the national Childhood Vaccine Injury Act, and Social Security Disability Insurance, Congress could replace costly and inequitable tort adjudication in federal and state courts with a simpler, fairer administrative mechanism incorporating the following elements:

1. *Deferral registries for the unimpaired.* Claimants without any present disability (for example, just a clinical diagnosis of pleural plaque) would have their claims put on hold until some tangible physical impairment manifests itself. These claimants would not have to fear that the claim would be barred by the statute of limitations or other procedural obstacles.

2. *No proof of fault.* Claimants would not need to prove the legal fault of any defendant or exposure to any particular company's product, as is required by tort law. Past asbestos exposure combined with present impairments of the type associated with asbestos would establish eligibility for compensation.

3. *Clear, objective medical criteria.* Instead of repeated individual trials involving complex medical evidence, an administrative agency would, through rulemaking, establish clear guidelines to define degrees of impairment based on objective clinical tests. As with the "grid" used in social security disability insurance, there could be a safety valve procedure for truly exceptional cases.[27]

4. *A schedule of compensatory payments.* Instead of repetitive, individualized proceedings to establish the amount of each damage award, a schedule of benefit payments would be created by legislation or public rulemaking. This schedule would build upon the typical experience in asbestos settlements and awards, but would specify a range of appropriate compensation that took into account the financial and non-financial losses of victims of different ages, impairments, and family circumstances.

5. *No punitive damages.* By removing the threat of cumulative and arbitrary punitive damages awards, an administrative system would equalize treatment of similarly situated plaintiffs and conserve the resources of defendants for compensating impaired claimants.

6. *Arbitration of disputes, with meaningful opportunity for appeal.* Claimants dissatisfied with administrative disposition of their eligibility or the amount of their benefit award would have prompt arbitration of their objection, in which they could argue that some exceptional circumstance in their case warrants a departure from the medical eligibility criteria or the benefit schedule. The arbitrator's decision would be subject to review by a neutral administrative tribunal, with limited further appellate review in court for clear legal error.

7. *Rationalized legal fees.* This administrative system would virtually eliminate the contingency element in asbestos claims and greatly simplify calculation of benefits payable. Payment for legal services could thus be set far lower than the present

---

[27] *See* Heckler v. Campbell, 461 U.S. 458, 460–62 (1983) (describing the 1978 Social Security Disability Insurance "grid" regulations adopted by the Department of Health and Human Services, which deal with claims objectively but still may be tailored for individual circumstances).

thirty-three to fifty percent of the unconstrained damage verdicts won under tort law. In view of the compensatory focus of a benefit system, the program should either reimburse legal fees reasonably incurred by successful claimants (as is now done in a number of workers' compensation regimes) or provide professional and semi-professional representation of claimants (as under the National Labor Relations Act).

8. *Financing the program.* The bulk of funding would come from defendant companies and their insurers. Assessments would be paid periodically, in accordance with projected and actual claims experience, rather than in up-front lump sum payments that would unduly strain defendants' present resources. Distribution of the financial burden among different firms would be based on projections of future liability derived from litigation experience and other factors that reflect differences in asbestos risk creation (thereby recognizing firm differences in responsibility for asbestos risk). Because of the documented interaction of smoking and asbestos exposure, a truly fair compensation program would require contribution from tobacco companies (perhaps through an increase in cigarette manufacturer taxes).[28] For the same reason, the significant role of the federal government in promoting the use of asbestos (especially in wartime shipbuilding) would be recognized by a governmental contribution—for example, by absorbing all administrative costs.[29]

There are many advantages to such a legislated administrative solution to the absestos crisis. It would be comprehensive and national in scope, rather than piecemeal and subject to the uncertainties and inconsistencies of our federal court system. Most importantly, it would be *democratic*, reflecting public deliberation about fair resource allocation.

Needless to say, this outline omits a host of subsidiary details about legislative design and administrative implementation.[30]

---

[28] *See* Louis Trieger, *Relief for Asbestos Victims: A Legislative Analysis*, 20 HARV. J. ON LEGIS. 179, 196 (1983).

[29] *Id.* at 192–96.

[30] Legislative treatment of Johns-Manville and other bankrupts poses a difficult question. Considerations of legal repose suggest that already-litigated arrangements for future payments—including not only classic tort actions but also those tort claims resolved through a bankruptcy remedy—should not be disturbed. On the other hand, it can be argued that where a legislative remedy promises procedurally or financially better results for all other future claimants, Congress might well want those future claims against bankruptcy funds folded into the global administrative solution. In such a design, the statute would claim whatever assets and income had been set aside for payment of the

For example, it is not clear if asbestos defendants should be coercively taxed to fund the scheme, or if they should pay a quasi-voluntary "fee" in exchange for legislatively eliminating their potential tort liabilities. Furthermore, it may be necessary to draw upon insurer resources that would be available to satisfy court settlements and tort awards to finance an administrative alternative. We must decide whether almost all claimants could be treated fairly in a matrix or grid system, or whether it is necessary to risk the complexity and disputes of a broad residual discretion to handle "exceptional" cases. Reformers must choose among benefits paid in a lump sum, in installments, or through an actuarially-equivalent annuity. We believe that the experience in other administrative programs can be drawn upon to supply these and many other details. They can be resolved, and the resulting program would be vastly preferable to the current litigation crisis.[31]

The significant problem with a legislative solution has less to do with policy design or legal conundrums than with politics. It is sobering to contemplate the array of interests clamoring to promote, influence, and derail a legislative proposal. Some of these interests, notably the trial bar, are among the most politically powerful lobbies in Washington. The enactment of any piece of controversial legislation, however meritorious, is uncertain at best. In the face of a massive budget deficit and congressional misgivings about the Black Lung and Social Security Disability administrative schemes, enactment of a sound asbestos compensation mechanism is unlikely. The complex legislative agenda announced by President Bill Clinton,[32] much of

---

bankrupt's asbestos liabilities and add those to the pool of resources for all asbestos claims payments under the new, legislated scheme.

[31] *See generally* COMMITTEE FOR EQUITABLE COMPENSATION, OUTLINE OF BILL TO ESTABLISH AN ASBESTOS COMPENSATION PROGRAM (1990) (detailing proposal by five asbestos defendants); Brickman, *supra* note 8 (detailing one administrative solution). Some of the congressional solutions already proposed include: H.R. 3090, 99th Cong., 1st Sess. (1985); H.R. 1626, 99th Cong., 1st Sess. (1985); S. 1265, 99th Cong., 1st Sess. (1985); H.R. 5966, 98th Cong., 2d Sess. (1984); S. 2708, 98th Cong., 2d Sess. (1984); H.R. 5735, 97th Cong., 2d Sess. (1982); H.R. 5224, 97th Cong., 1st Sess. (1981).

As to the feasibility of constructing a benefit schedule and medical criteria, the extensive experience with litigation and settlements over the past two decades provides hope. For example, the fact that defendants' and plaintiffs' lawyers can often negotiate large batch settlements suggests that it is possible to agree on the value of asbestos tort claims, at least in the aggregate. For a more scientific reading, legislative or rulemaking processes could develop detailed analysis of jury awards and settlement amounts. This might clarify the distortions in the data created by the mystery surrounding most settlements, the details of which are never made public.

[32] WILLIAM J. CLINTON, A VISION OF CHANGE FOR AMERICA (1993).

which will involve the same congressional committees as any asbestos proposal, likely puts the legislative ideal beyond the realm of the politically possible.

V. A Judicial Approach

The fact that the legislative ideal is probably unattainable only reinforces the responsibility of courts to address the crisis with imagination and urgency. Courts must work creatively with current substantive and procedural doctrine to devise a mass resolution of the asbestos crisis. An effective judicial solution should include the key elements of the legislative ideal we have described:

- Determination of entitlement based not on company fault, but on victim exposure and impairment.
- Development of objective medical criteria for determining degree of impairment.
- Scheduling of compensatory ranges to allow some flexibility for unusually serious cases.
- Elimination of future punitive damages.
- Arbitration of disputes instead of jury resolution, with meaningful opportunity for appeal.
- Rationalization and reduction of legal fees; simplification of procedures to assure adequate but low-cost representation.
- Financing the scheme through contributions from a large number of asbestos companies and their insurers.

Judges have shown such creative vision in a number of mass tort cases, especially the tort-driven bankruptcy proceedings involving A.H. Robins (the Dalkon Shield)[33] and Johns-Manville (asbestos), the latter led masterfully by Judge Weinstein. These proceedings have adopted several elements of the solution sketched above, including simplified administrative procedures for assessing impairment and calculating benefits, and mechanisms for channeling scarce resources toward current and future claims of a more serious nature. The extraordinary circumstances of bankruptcy forced the judges' hands, but desperate damage control in such a financial crisis is hardly the ideal

[33] *In re* A.H. Robins Co., 88 B.R. 742 (Bankr. E.D. Va. 1988), *aff'd*, 880 F.2d 694 (4th Cir.), *cert. denied*, 493 U.S. 959 (1989).

setting for a comprehensive solution. We must protect the economic viability of solvent enterprises, enabling them to meet their obligations to both claimants and other stakeholders. Solving the problem for individual bankrupts such as Manville by circumscribing their liabilities only adds to the financial pressure on surviving defendants.[34]

Instead, courts should certify broad settlement class actions for purposes of resolving all pending and future asbestos claims against broad groupings of solvent defendants.[35] The recently approved Pfizer heart valve settlement[36] demonstrates how this procedure can produce the various components of a rational compensation scheme that we have emphasized: providing as much as $200 million to settle potential claims from tens of thousands of heart valve recipients; simplifying eligibility; scheduling different categories of benefits; giving priority in payments to those with severe injuries and needs, as opposed to those with only potential health problems; simplifying claims procedures with corresponding reductions in legal costs; and eliminating multiple punitive damages awards.

Because such class action certification would be used only for purposes of settlement, it would avoid the casino-like "bet the company" risks of certification for litigation and trial, with billions of dollars riding on a single jury verdict. Defendants can choose to accept or reject the terms of a negotiated settlement. Rule 23 of the Federal Rules of Civil Procedure requires court approval of the "fairness, adequacy, and reasonableness" of the settlement.[37] Judges can thereby block "sweetheart deals" negotiated between defendants, who want to achieve relief from potentially massive long-term liability with a modest up-front payment, and plaintiffs' lawyers, who want to ensure not only their clients' recovery but their own lavish fees. Judges should

---

[34] This specific feature of Judge Weinstein's restructuring of the Manville Trust has just produced a partial reversal of his ruling by a divided Second Circuit panel. *See In re* Joint E. & S. Dist. Asbestos Litig. (Findley v. Blinken), 982 F.2d 721 (2d Cir. 1992).

[35] For development of the arguments in favor of mass tort settlements (as well as mass tort actions), see 2 ENTERPRISE RESPONSIBILITY, *supra* note 2, at 383–439; David Rosenberg, *Class Actions for Mass Torts: Doing Individual Justice by Collective Means*, 62 IND. L.J. 561 (1987).

[36] The details of the heart valve litigation and class action settlement are set out in the judicial decision approving the terms of the settlement, Bowling v. Pfizer, Inc., 143 F.R.D. 141 (S.D. Ohio 1992).

[37] FED. R. CIV. P. 23(e); *see, e.g.*, Stoetzner v. United States Steel Corp., 897 F.2d 115, 117 (3d Cir. 1990); Piambino v. Bailey, 757 F.2d 1112, 1139 (11th Cir. 1985), *cert. denied*, 476 U.S. 1169 (1986).

not, however, transform the fairness hearing on a proposed settlement into a full-scale trial on the merits. Instead, judges should decide only whether the proposed settlement falls within the parameters of a reasonably negotiated outcome. A complex fairness inquiry is unnecessary if claimants have a meaningful opportunity to opt out of the settlement. Such an opportunity enables plaintiffs to compare the value of the settlement with the value of the tort alternative and to choose accordingly.[38] The opportunity for plaintiffs to opt out, combined with the right typically reserved by defendants to nullify the settlement if too many claimants opt out, provides the appropriate incentives for plaintiffs and defendants alike to fashion a mutually beneficial and final resolution.

Several months after we began this Article, certain asbestos defendants represented by the Center for Claims Resolution ("CCR") negotiated with counsel representing asbestos claimants and reached a class action settlement along the lines suggested above. It was motivated by the Multi-District Litigation Panel's pretrial consolidation of the 30,000 pending federal asbestos claims. The settlement, which governs only *future* federal and state claims (all claims not filed as of January 15, 1993, the filing date of the class action), contains the following key components:

- Instead of litigating all the arguably relevant features of defendant liability, claimants will be entitled to compensation if they satisfy specified criteria regarding exposure to the defendant's asbestos, latency periods, and medical conditions that are typically associated with asbestos.
- Expert medical panels will resolve disputes about whether particular claimants satisfy the criteria. These panels will also determine whether "exceptional" cases that claim to be "substantially comparable" to the standard asbestos diagnoses satisfy the criteria.
- Claims of current medical impairment will be paid promptly. Claims of present indicia of future impairment are not allowed, but when such claims are resubmitted for actual impairment, they will not be time-barred. In addition, claimants who develop asbestosis can collect compensation immediately for that im-

[38] This "market test" is not perfect, however. *See infra* note 41.

pairment, but if a malignancy later develops, can refile for the award established for this more severe disease.

• A schedule consisting of minimum and maximum figures of compensation and an average value range are established, all derived from historical settlement and award payments made by CCR members in tort litigation. For example, for mesothelioma the compensation range runs from a minimum of $20,000 to a maximum of $200,000, with the negotiated average value in the range of $37,000 to $60,000.[39]

• Individual victims may claim "extraordinary" damages, depending on a combination of age, number of dependents, economic factors, and the degree of exposure to the products of CCR members.[40]

• In light of the reduced difficulties and contingencies in representing claimants under this settlement, the maximum attorney fee is held to twenty-five and twenty percent of the compensation received for ordinary and extraordinary compensation claims, respectively.

• On the basis of prior CCR experience, the parties developed annual case-flow caps for "ordinary" claims, "exceptional" diagnostic claims, and "extraordinary" compensation claims. If in any one year the number of claims in a category exceeds that category's cap, the excess claims will be processed in the next year and count against that next year's cap. Because the CCR members are obligated by the settlement to pay *all* qualifying claims, these case flow ceilings affect only the timing of, not entitlement to, payment. Moreover, if the aggregate settlement amount paid to claimants in any one year falls below the average value range, CCR will make up the difference by paying more claims in subsequent years.

• In addition to the front-end opt-out right under the class action settlement procedure, any claimant who meets the exposure and impairment criteria and is dissatisfied with the amount of compensation offered has the option of suing in court for tort damages in lieu of accepting the CCR offer. In such a suit, the defendant waives any issues relating to its legal fault,

---

[39] *See* Stipulation of Settlement Between the Class of Claimants and Defendants Represented by the Center for Claims Resolution at ex. B., Carlough v. Amchem Prods. Inc., CA No. 93-CV 0215 (E.D. Pa. Jan. 15, 1993) (compensation schedule) (on file with the *Harvard Journal on Legislation*).

[40] The negotiated average value for extraordinary mesothelioma claims is set at $300,000, but payments in individual cases can range well beyond that figure. *Id.*

with the only predicates to recovery being claimant exposure and disease. In return, successful plaintiffs are entitled to collect only compensatory, not punitive, damages. A maximum of one percent of claimants can assert this back-end opt-out right in any one year, and any damages award greater than 150 percent of CCR's last offer is paid over a period of five years.

The negotiators appear to have addressed the major pathologies of the present system and adopted a significantly improved method for delivering compensation to future victims of asbestos exposure. Their success, however, will depend on the court's assessment of the proposed settlement's fairness.

There are several reasons to believe that this settlement secures important gains for both sides. For the benefit of plaintiffs, the settlement guarantees compensation for all impaired claimants through "evergreen funding." It also overcomes several obstacles traditionally imposed by tort doctrine by using medically-defined entitlement criteria. In doing so, it both simplifies adjudication and reduces uncertainty and costs for claimants. Prompt administrative processing also sigificantly ameliorates current litigation delays. In addition, the savings resulting from the cap on legal fees, together with any further savings from competitive pressure on fees, will benefit claimants rather than defendants. The compensation schedule specifies a range of payments within each medical category, permitting some individualization. By also specifying an annual average for the range—based on the current pattern of tort awards—the schedule removes the defendant's incentive to offer an unreasonably low or high award to any particular claimant. The front-end opt-out right allows potential claimants who are dissatisfied with the proposed compensation scheme to avoid the settlement. It thus provides an immediate substantive test of the fairness of the proposal. Finally, the back-end opt-out right for claimants who are dissatisfied with the compensation offered through the administrative mechanism provides a gradual market test as to whether settlement benefits remain preferable to the tort litigation alternative.[41]

[41] The front-end opt-out is not a perfect market test of the settlement proposal because some proportion of future claimants, many of whom are currently unimpaired, may be unlikely to exercise the opt-out choice at this time. The back-end opt-out is also flawed to some extent, because the limited number of such tort claims permitted in any one year may make this option less attractive to many claimants than the present unrestricted right to sue if and when they want.

For defendants, the settlement eliminates unimpaired pleural plaque claims, saving both the compensation these claims have usually garnered and the associated defense costs. The use of tangible medical criteria not only simplifies eligibility determinations but also helps screen out some lung cancer and respiratory cases that are not asbestos-related. The combination of objective medical criteria, exclusion of unimpaired claims, and the elimination of statute of limitations problems reduces the incentives for plaintiffs' counsel to search for marginal claims. Finally, the settlement will provide a very important measure of financial certainty for defendants, by specifying a compensation range and average for each impairment category,[42] by eliminating punitive damages awards, and by capping both the annual volume of cases that will receive exceptional compensation and the number of cases that will be handled through the back-end opt- out each year.

There are some important differences between the legislative ideal we sketched earlier and the approach negotiated by the parties to this class action settlement:

• The settlement permits more generous attorneys' fees, although the percentage figures are consistent with those used in the Manville reorganization. But these freely-negotiated caps on attorneys' fees do not preclude competitive pressures generating lower fees if members of the private bar are prepared to bid with smaller percentage rates for the right to represent asbestos victims with less contingent claims under the new system.

• The settlement extinguishes, rather than defers, unimpaired pleural plaque claims until more serious asbestos disease cases are resolved. In effect, the settlement adopts the substantive position that "cancerphobia" actions should be barred. Those claimants, however, will no longer be forced by the statute of limitations in "single-claim" states to bring actions when they are not actually sick and thereby forfeit any recourse to larger compensation when and if serious impairment develops.

• More importantly, a legislated solution would cover current as well as future claims, subject only to minor due process limits. It would also cover more than the subset of companies

[42] With a fixed compensation schedule, defendants are further protected from unexpected real increases in awards. Conversely, future claimants are protected from reductions in tort awards that may result from legislated tort reform.

and insurers participating in the settlement, avoiding the complex problem of rationalizing the indemnification from, and joint and several liability of, the companies and insurers not participating in the settlement.

• A legislated solution would create an administrative rulemaking process that would permit adjustments to the medical and compensation parameters if future circumstances demand it. Such adjustments could be regulated to provide reasonable certainty to claimants and defendants, and they would be much easier to accomplish as administrative matters rather than as modifications of an already approved settlement.[43]

• Finally, a class action settlement involves the prospect of undesirable delay. Although a trial court deems a proposed settlement reasonable, an overly cautious appeals panel with limited time may not see the wisdom of swift and bold action. This danger was demonstrated in the recent reversal of the negotiated reorganization of the Manville Trust.[44] Even an eminently reasonable settlement may take years to work its way through the courts and win final appellate approval. Fortunately, this settlement does provide that it becomes operational—in the sense that CCR will begin processing claims under it—at the conclusion of the front-end opt-out. And CCR defendants have undertaken to resolve more quickly all *pending* tort claims once they receive assurance about their future liability through approval of the class action settlement.

Notwithstanding these considerations, we firmly endorse the fairness and adequacy of this settlement of future asbestos claims. We also hope that this will be the first step towards comparable relief for every one of the more than 100,000 asbestos claims now pending in federal and state courts.

VI. Conclusion

The asbestos crisis can be solved in a way that assures substantial recovery for the sick without bankrupting the businesses that must pay the bill. Whether the solution is administered by

[43] The stipulation includes a procedure for agreed-to adjustment in compensation amounts and ranges. Stipulation, *Carlough*, CA No. 93-CV 0215, at 51–52.

[44] *In re* Joint E. & S. Dist. Asbestos Litig. (Findley v. Blinken), 982 F.2d 721, 750 (2d Cir. 1992).

a court or an administrative agency, it must involve certain key elements: reduction in legal costs, limitation of compensation to claimants who are actually impaired, different but fair levels of benefits for people with different losses, and elimination of excessive and redundant financial punishment for defendants.

There are many good reasons to favor an administrative solution formulated by Congress rather than some scheme created by federal or state judges. Most importantly, the administrative state has demonstrated that it can serve as an excellent alternative to traditional forms of legislation and adjudication. Its strength lies in harnessing the ability of bureaucratic organizations to integrate scientific expertise, individual fairness, and public purpose.[45]

At the same time, however, even those most enthusiastic about a solution that would take the asbestos crisis out of the judicial system must feel reluctant to ask Congress to enact a new administrative scheme. Passage of any piece of controversial legislation is uncertain at best. With the budget deficit and a host of other social and economic priorities, enactment of a sound administrative scheme for asbestos compensation is unlikely.

The judicial solutions we have discussed are a feasible method of resolving the crisis. This is especially true if, as we believe, resolution by settlement class action of pending and future claims can be fashioned to integrate the compensation priorities and case-handling features of a pure administrative approach.

We have had several goals in mind in analyzing the asbestos problem. Most importantly, we have sought to contribute a legal solution to a social and economic problem that has involved enormous human suffering. As students of governance generally, and of administrative systems, tort litigation, and the law of the workplace specifically, we realize that the solution to the asbestos crisis demands the integration of several conceptual frameworks. We hope that judges, legislators, and the involved parties will find our analysis constructive. Without reform, the present system will only lead to deepening crisis in the courts and further injustice to asbestos victims.

---

[45] *See generally* CHRISTOPHER F. EDLEY, JR., ADMINISTRATIVE LAW: RETHINKING JUDICIAL CONTROL OF BUREAUCRACY 13–95, 125–30, 187–99, 213–64 (1990).

ARTICLE

CONTROL OF AIR POLLUTION FROM MOBILE SOURCES THROUGH INSPECTION AND MAINTENANCE PROGRAMS

ARNOLD W. REITZE, JR.*
BARRY NEEDLEMAN**

Twenty years ago, in Volume 10 of the Harvard Journal on Legislation, *Jack Appleman analyzed the difficulties of using the regulatory powers delegated by the Clean Air Act for reducing motor vehicle emissions. This Article revisits the means by which Inspection and Maintenance programs currently instituted under the Clear Air Act work to improve air quality and the problems that prevent I/M programs from realizing this goal. The authors discuss the impact of the 1992 Amendments to the Clean Air Act on these problems, expressing concern regarding the amendments' limited effectiveness and the new problems that could arise during implementation.*

Despite some improvement in recent years, urban air pollution remains a serious problem in the United States.[1] Mobile sources continue to be the primary cause of most urban air pollution, producing half the ozone and almost all of the carbon monoxide pollution in U.S. cities.[2] Therefore, reduction of mobile source emissions is essential in any scheme to improve urban air quality.[3]

Motor Vehicle Inspection and Maintenance ("I/M") programs are the focal point of efforts to reduce mobile source emissions from in-use vehicles. I/M is intended to help reduce urban air pollution by ensuring that light-duty motor vehicles are properly maintained once they are in the consumer's hands.[4] Under the 1970 Clean Air Act ("1970 CAA") states were permitted, but not required, to employ I/M programs to meet the National

* Elyce Zenoff Research Professor of Law and Director of the LL.M. program in environmental law, The George Washington University. B.A., Fairleigh Dickinson University, 1960; J.D., Rutgers University, 1962; M.P.H. (Environmental Health Science), The Johns Hopkins University, 1985.

** Randolph C. Shaw Fellow and LL.M. Candidate (Environmental Law), The George Washington University. B.A., Clark University, 1987; J.D., Lewis and Clark Northwestern School of Law, 1992.

This Article was produced with the assistance of Ms. Winnie Hercules, legal secretary.

[1] NATIONAL RESEARCH COUNCIL, RETHINKING THE OZONE PROBLEM IN URBAN AND REGIONAL AIR POLLUTION 1–2 (1991).

[2] 57 Fed. Reg. 52,950 (1992).

[3] *Id.*

[4] *Id.*

Ambient Air Quality Standards ("NAAQS").[5] The 1977 Clean Air Act Amendments ("1977 CAAA")[6] required I/M programs in carbon monoxide and ozone non-attainment areas that received extensions on deadlines for meeting the NAAQS.[7] The 1990 Clean Air Act Amendments ("1990 CAAA")[8] divide I/M programs into two categories—basic and enhanced—and require all ozone and carbon monoxide non-attainment areas to have one type of I/M, depending on the severity of their non-attainment status.[9]

Parts I and II of this Article provide a general explanation of the Clean Air Act and a brief history of I/M programs prior to the 1990 Amendments. Parts III and IV explain how I/M programs will function after the 1990 Amendments. Part V then discusses the relationship between I/M and Clean Air Act performance warranties. Part VI outlines the deadlines for implementation of the new enhanced I/M regulations while Part VII explains the new regulations in some depth. Parts VIII and IX discuss onboard emissions diagnostic systems and examine the cost of I/M programs. Part X concludes that the I/M program, despite its many problems, is still a useful tool for controlling urban pollution.

I. Regulation of State Pollution Sources Under the Clean Air Act

The Clean Air Act[10] regulates pollution in the ambient air by controlling the atmospheric concentration of six criteria pollutants[11] and emissions of approximately 189 hazardous pol-

[5] *See generally* Arnold W. Reitze, Jr., *Controlling Automotive Air Pollution Through Inspection and Maintenance Programs*, 47 Geo. Wash L. Rev. 705 (1979) (discussing early efforts to use I/M).

[6] Clean Air Act Amendments of 1977, Pub. L. No. 95-95, 91 Stat. 685-796 (codified in scattered sections of 42 U.S.C. (1977)).

[7] *Id.* § 172(b)(11)(B), 42 U.S.C. § 7502(b)(11)(B) (1977).

[8] Pub. L. No. 101-549, 104 Stat. 2399 (codified at 42 U.S.C. §§ 7401–7671 (Supp. II 1990)).

[9] *See infra* notes 74–87 and accompanying text.

[10] 42 U.S.C. §§ 7401–7642 (1988), *amended by* Clean Air Act Amendments of 1990, Pub. L. No. 101-549, §§ 101–1101, 104 Stat. 2399–2712 (codified as amended at 42 U.S.C. §§ 7401–7671(q) (Supp. II 1990)).

[11] Particulate matter (PM_{10}), sulfur dioxide, nitrogen dioxide, carbon monoxide, ozone, and lead. National Primary and Secondary Ambient Air Quality Standards, 40 C.F.R. § 50 (1992). Reactive hydrocarbon emissions are also controlled to prevent the formation of ozone in the atmosphere. Reitze, *supra* note 5, at 706.

lutants.[12] The 1990 Clean Air Act Amendments expanded the CAA to six subchapters: subchapter I primarily regulates stationary sources; subchapter II addresses mobile sources; subchapter III deals with administration and implementation; subchapter IV focuses on acid rain (primarily from electric power plants) and includes the new and innovative market-based emission trading system; subchapter V contains provisions concerning operating permits; and subchapter VI protects stratospheric ozone and implements international agreements concerning this subject.[13]

Motor vehicles are the major sources of three air pollutants regulated by the CAA. In 1985, motor vehicles were responsible for seventy percent of the nation's carbon monoxide ("CO"), forty-five percent of the nitrogen oxides ("NO_x") and thirty-four percent of the volatile organic compounds ("VOCs").[14] The CAA attempts to reduce these emissions through controls on new vehicles and fuels under subchapter II and with programs aimed at in-use vehicles under subchapter I.

The states are the focal point for most of the air pollution control efforts. Each state is divided into Air Quality Control Regions ("AQCRs"), and the AQCRs are further subdivided into "Areas" based on air quality. The Areas are evaluated to determine whether they are above or below the NAAQS for each of the six criteria pollutants.

Once the ambient air quality is determined, the states, with guidance from the Environmental Protection Agency ("EPA"), develop State Implementation Plans ("SIPs").[15] SIPs are detailed strategies committing the states to the control of air pollution. The goals of SIPs are to reduce Area pollution to the NAAQS

---

[12] *See* 42 U.S.C. § 7412(b)(1) (Supp. II 1990), *amended by* Pub. L. No. 102-187, 105 Stat. 1285 (1991) (codified at 42 U.S.C. § 7412(b)(1) (Supp.III 1991)). This section in particular, and other sections generally, control toxics. However, toxic air pollution control was very limited prior to the 1990 Clean Air Act Amendments ("1990 CAAA"); the program under the 1990 CAAA is only beginning to be implemented. Automobile toxics are subject to a new program under the Clean Air Act Amendments of 1990 § 202(l), 42 U.S.C. § 7521(a)(3) (Supp. II 1990). A recent Environmental Protection Agency ("EPA") draft study identified 1,3 butadiene as the motor vehicle air toxic of most concern. Present control programs do not focus on this colorless, flammable, highly-reactive gas formed in vehicle exhaust by incomplete combustion. Although critics and the EPA argue over the effects of 1,3 butadiene, the EPA expects to propose control regulations in 1995. *EPA Study on Motor Vehicle Air Toxics Emissions Singles Out Major Toxin*, INSIDE EPA, Jan. 22, 1993, at 1, 6.

[13] Clean Air Act §§ 101–618, 42 U.S.C. §§ 7401–7671(q) (1988 & Supp. II 1990).

[14] Standards for Emissions, 57 Fed. Reg. 52,912, 52,981 (1992) (to be codified at 40 C.F.R. pts. 85–86, 600) (proposed Nov. 5, 1992).

[15] Clean Air Act § 110(a), 42 U.S.C. § 7410(a) (1988 & Supp. II 1990).

for each of the six criteria pollutants and to prevent significant deterioration of the air quality in those Areas that already meet the national standards.

The CAA provisions serve only as minimum requirements for state control of air pollution. States may select more stringent control measures and may employ any combination of pollution control strategies to clean up or maintain the air in a given area. This is necessary because different areas have unique pollution problems that require different control strategies. For example, excessive automobile pollution in urban locations frequently leads to a non-attainment classification for transportation-related pollutants, thus necessitating particular attention to transportation control measures in the SIPs for those areas.

CAA subchapters I and II work together to control transportation-related pollution. Subchapter II establishes automobile emission standards for new cars and regulates those vehicles until they reach consumers.[16] The federal government has primary responsibility for implementing subchapter II and preempts most state regulation in this area.[17] The states, through subchapter I, primarily regulate automobiles in the hands of consumers through the I/M program.[18] The states also use other measures listed in section 108 that are the basis for the transportation control plan, which is a part of the SIP for any state with air pollution problems related to motor vehicle emissions.

The purpose of the I/M program is to identify and ensure the repair of in-use automobiles that are emitting excessive pollutants. Before the 1977 CAAA,[19] states could choose whether or not to have an I/M program.[20] After the 1977 CAAA, states that requested an extension for attaining the NAAQS were required to implement and comply with all appropriate statutory requirements for I/M programs in locations that had non-attainment areas for transportation-related pollutants.[21] After the 1990 CAAA, states that still had non-attainment areas for transportation-related pollutants were required to implement I/M pro-

[16] *Id.* § 202, 42 U.S.C. § 7521.

[17] *Id.* § 182(b)(4), 42 U.S.C. § 7511a(b)(4) (Supp. II 1990).

[18] *Id.* §§ 174, 176, 182(b)(4), 42 U.S.C. §§ 7504, 7506, 7511a(b)(4).

[19] Pub. L. No. 95-95, 91 Stat. 685–796 (codified in scattered sections of 42 U.S.C. (1977)).

[20] *See generally* Jack M. Appleman, *The Clean Air Act: Analyzing the Automobile Inspection, Warranty, and Recall Provisions*, 10 HARV. J. ON LEGIS. 537 (1973); Reitze, *supra* note 5.

[21] Clean Air Act §§ 110(a)(2)(G), 172, 42 U.S.C. §§ 7410(a)(2)(G), 7502(a)(2)(C) (1988 & Supp. II 1990).

grams. If those areas were not too heavily polluted, they could use a basic I/M program. After 1990, however, the most heavily polluted areas were required to implement a more stringent inspection program called enhanced I/M.[22]

Under the CAA, a state may not adopt any new motor vehicle emission standard applicable to any vehicle prior to its initial titling or registration.[23] There is a waiver of federal preemption, however, for California.[24] Thus, two categories of automobile emission controls have existed since the 1960s. In 1977, the CAA was amended to allow any state that had non-attainment areas for automotive-related pollutants to adopt the California standards.[25] In 1990, section 177 of the CAA was amended to make it clear that no state could limit automobile emissions in any way that would force manufacturers to create a "third" vehicle or engine that was different from a vehicle certified by either the EPA or California.[26] After the passage of the 1990 CAAA, states with ozone and CO non-attainment areas began moving to adopt the California standards.

The federal government had also preempted the control of fuel and fuel additives in section 211. However, if states adopt the California motor vehicle standards, they may also regulate fuels and fuel additives.[27] In addition, states may regulate fuel vapor pressure.[28] Furthermore, states with serious areas for CO non-attainment must, after November 15, 1992, require the sale oxygenated gasoline during winter months in urban areas.[29]

Once motor vehicles are in the hands of the consumer, the states may regulate either vehicle emissions, vehicle use, or both. States usually regulate in-use vehicle emissions through I/M programs. Section 172(b)(11)(B) of the 1977 CAAA required such programs for ozone or CO non-attainment areas. This subsection was eliminated by the 1990 CAAA, but a new section 182(a) requires the continuation and revision of I/M programs.[30]

---

[22] *See id.* § 181(b)(4), 42 U.S.C. § 7511(b)(4) (Supp. II 1990).
[23] *Id.* § 209(a), 42 U.S.C. § 7543(a) (1988).
[24] *Id.* § 209(b), 42 U.S.C. § 7541(b).
[25] *Id.* § 177, Pub. L. No. 95-96, § 129(b), 91 Stat. 750 (1977) (codified at 42 U.S.C. § 7507 (1988)).
[26] Henry A. Waxman et al., *Cars, Fuels and Clean Air: A Review of Title II of the Clean Air Act Amendments of 1990*, 21 ENVTL. L. 1947, 1996 (1991).
[27] Clean Air Act § 211(c)(4)(B), 42 U.S.C. § 7545(c)(4)(B) (1988).
[28] *Id.* § 211(h), 42 U.S.C. § 7545(h) (Supp. II 1990).
[29] *Id.* § 187(b)(3), 42 U.S.C. § 7512a(b)(3).
[30] *Id.* § 182(a), 42 U.S.C. § 7511a(a).

The CAA required the EPA Administrator to promulgate new regulations by November 15, 1991, to guide the states in revising or creating I/M programs.[31] The EPA may grant up to two-thirds of the cost of developing an I/M program.[32] In addition, by May 15, 1992, the EPA was supposed to promulgate regulations requiring manufacturers of all new light-duty vehicles and trucks to install diagnostic systems. Such systems will detect malfunctioning emission controls and, at a minimum, monitor both catalytic converters and oxygen sensors.[33] The states are then required to amend their SIPs to provide for inspection of onboard diagnostic systems.[34]

The 1990 CAAA created five categories of ozone non-attainment areas. Marginal and moderate ozone non-attainment areas must have I/M programs.[35] Serious areas must have enhanced I/M programs with improved monitoring of NO_x emissions and VOCs.[36] There is also a requirement for a clean fuels program as prescribed under part C of subchapter II.[37] Severe and extreme areas must have enhanced I/M programs and meet additional requirements.[38]

The CAA also requires all ozone non-attainment areas to revise their SIPs by November 15, 1992, to require a system for gasoline vapor recovery of emissions from the fueling of motor vehicles.[39] This program is commonly called Stage II Vapor Recovery. Stage I Vapor Recovery involves controls aimed at limiting VOC emissions when the fuel is delivered to the retail gasoline outlet from the refinery or bulk distributor. Stage I has been implemented without controversy, but the petroleum industry has strenuously opposed the Stage II program.

II. Inspection and Maintenance Prior to the 1990 CAAA

The 1970 CAA required each state to develop a detailed program to meet the NAAQS. The SIPs could have used I/M as

[31] *Id.* § 182(a)(2)(B)(ii), 42 U.S.C. § 7511a(a)(2)(B)(ii).
[32] *Id.* § 210, 42 U.S.C. § 7544 (1988).
[33] *Id.* § 202(m)(1), 42 U.S.C. § 7521(m)(1) (Supp. II 1990); *see also* Control of Air Pollution from New Motor Vehicles and New Motor Vehicle Engines, 56 Fed. Reg. 48,272 (1991) (to be codified at 40 C.F.R. pt. 86) (proposed Sept. 24, 1991).
[34] Clean Air Act § 202(m)(3), 42 U.S.C. § 7521(m)(3) (Supp. II 1990).
[35] *Id.* § 182(a)(2)(B), (b)(4), 42 U.S.C. § 7511a(a)(2)(B)(i), (b)(4).
[36] *Id.* § 182(c)(1), (3), 42 U.S.C. § 7511a(c)(1), (3).
[37] *Id.* § 182(c)(4), 42 U.S.C. § 7511a(c)(4).
[38] *Id.* § 182(d)–(e), 42 U.S.C. § 7511a(d)–(e).
[39] *Id.* § 182(b)(3), 42 U.S.C. 7511a(b)(3).

part of their program to reduce motor vehicle pollution.[40] Fewer than half the states, however, developed plans with any transportation control measures, and only a handful had I/M programs.[41]

In the 1977 CAAA, Congress established December 31, 1982, as the new NAAQS state deadline.[42] The deadline could be extended until December 31, 1987, however, for carbon monoxide and ozone violators if specific SIP revisions were made.[43] Non-attainment areas for CO or photochemical oxidants had to submit revised SIPs that were coordinated with the continuing, cooperative, and comprehensive transportation planning process,[44] as well as the air quality maintenance planning process.[45] This required an evaluation of the various measures set forth in CAA section 108(f), primarily involving transportation measures designed to reduce vehicle-miles traveled ("VMT"), even though the I/M provision was aimed at reducing emissions from in-use vehicles. Another section 108(f) provision involved controls for fuel transfer vapor emissions, and was primarily a stationary source measure related to motor vehicle emissions.[46] The state could select from the list of transportation-related measures those necessary to meet the reductions in VOC, NO_x, and CO required for EPA approval of SIPs. Establishing a "specific schedule" for development of an I/M program was not optional, however, but was instead mandatory for any CO or photochemical oxidant non-attainment area receiving a compliance extension to 1987.[47]

---

[40] *See* Reitze, *supra* note 5, at 714–15 (discussing early efforts to use I/M).

[41] Prior to 1977, I/M programs were in operation in New York, New York (for taxi cabs); Chicago, Illinois; Riverside, California; Cincinnati, Norwood, and Hamilton County, Ohio; Portland, Oregon; and Phoenix and Tucson, Arizona. *See* NATIONAL ACADEMY OF SCIENCES, REPORT BY THE COMMITTEE ON MOTOR VEHICLE EMISSIONS 132 (1974).

[42] Clean Air Act Amendments of 1977 § 129, 42 U.S.C. § 7502 (1977).

[43] *Id.* §§ 110(a)(2)(G), 172, 42 U.S.C. §§ 7410(a)(2)(G), 7502.

[44] The CAA requires transportation plans, Transportation Improvement Plans ("TIPs"), and individual projects to conform to the SIP. *Id.* § 176, 42 U.S.C. § 7506 (Supp. II 1990). However, the more significant requirements are imposed by the highway planning requirements of 23 U.S.C. The Intermodal Surface Transportation Efficiency Act, Pub. L. No. 102-240, 105 Stat. 1914 (1991) ("ISTEA"), significantly expanded these requirements in 1991. Of particular importance are §§ 134, 135, and 149 of 23 U.S.C., amended by §§ 1024(a), 1025(a), and 1008(a) of ISTEA, respectively.

[45] Clean Air Act § 175A, 42 U.S.C. § 7505a (Supp. II 1990).

[46] *Id.* § 108(f)(1)(A)(ii) (current version at 42 U.S.C. § 7408(f)(1)(A)(ii) (Supp. II 1990)). Section 182(b)(3) of the CAA now requires moderate and worse ozone non-attainment areas to have vapor recovery systems. *Id.* § 182(b)(3), 42 U.S.C. § 7511a(b)(3).

[47] *Id.* § 172(b)(11)(B), 42 U.S.C. § 7502(b)(11)(B) (1988).

New Jersey's I/M program, the nation's first, became effective February 1, 1974.[48] Oregon and Arizona also had programs that predated the 1977 CAAA.[49] Rhode Island began a program in 1979, but abandoned it when the state successfully met the NAAQS before the December 31, 1982, deadline.[50] After the 1977 CAAA, the I/M program became applicable to thirty states and the District of Columbia, as a direct result of the states' requests for compliance extensions.[51] The EPA, however, generally required I/M only in areas with populations of 200,000 or more.[52] An additional seven of the thirty affected states had acceptable programs by the deadline. By September, 1984, fifteen more states and the District of Columbia had implemented programs, and the four remaining affected states projected program establishment by 1986.[53] In March, 1984, however, New Mexico, one of the first states to have an I/M program, terminated its plan when the state supreme court ruled that the metropolitan area overseeing the program's operation lacked authority to charge an inspection fee.[54] A few other states became subject to I/M requirements when they failed to meet the NAAQS by December 31, 1982, although they had not requested a compliance extension until 1987.[55]

The I/M program was unpopular in many states.[56] Many states opposed I/M because they believed it was unnecessary for achieving NAAQS, was not cost-effective, treated motorists inequitably, and drained scarce state financial resources.[57] State opposition arose despite EPA data showing that ozone levels generally increased between 1975 and 1983, effectively main-

---

[48] *See* U.S. ENVTL. PROTECTION AGENCY, INFORMATION DOCUMENT ON AUTOMOBILE EMISSIONS INSPECTION AND MAINTENANCE PROGRAM 53 (1978) [hereinafter INFORMATION DOCUMENT].

[49] *Id.* at 57.

[50] U.S. GEN. ACCOUNTING OFFICE, GAO/RCED-85-22, VEHICLE EMISSIONS INSPECTION AND MAINTENANCE PROGRAM IS BEHIND SCHEDULE 10 (Jan. 16, 1985) [hereinafter I/M PROGRAM BEHIND SCHEDULE].

[51] *Id.* at 1, 9.

[52] U.S. ENVTL. PROTECTION AGENCY, I/M FACT SHEET (June 4, 1984).

[53] I/M PROGRAM BEHIND SCHEDULE, *supra* note 50, at 10.

[54] Chapman v. Luna, 678 P.2d 687 (N.M. 1984), *cert. denied*, 474 U.S. 947 (1985). The EPA subsequently withheld money from New Mexico as punishment. *See* New Mexico Envtl. Improvement Div. v. Thomas, 789 F.2d 825 (10th Cir. 1986); U.S. GEN. ACCOUNTING OFFICE, GAO/RCED-86-129BR, VEHICLE EMISSIONS—EPA RESPONSE TO QUESTIONS ON ITS INSPECTION AND MAINTENANCE PROGRAM 56–57, 61 (May 1986) [hereinafter EPA QUESTIONS].

[55] I/M PROGRAM BEHIND SCHEDULE, *supra* note 50, at 10.

[56] *See* Reitze, *supra* note 5, at 720.

[57] I/M PROGRAM BEHIND SCHEDULE, *supra* note 50, at 9, 11.

taining ozone as a significant air pollution problem.[58] This difference of opinion led to numerous court cases in which states sought to avoid I/M requirements.[59]

As punishment for not complying with the CAA, in December, 1980, the EPA restricted federal funding in six areas.[60] Other states where the legislature opposed the I/M program included California, Pennsylvania, Arizona, Connecticut, Idaho, Maryland, Michigan, New York, and Oregon.[61] Between May and July, 1982, officials of twelve states wrote to the EPA opposing I/M, in the belief that the program was either not cost-effective or not necessary to meet the NAAQS.[62]

The EPA was generally slow to sanction states not meeting I/M requirements[63] that were set forth in the EPA's final policy on the *Criteria for Approval of the 1982 Plan Revisions*.[64] On December 30, 1982, the Comptroller General issued a legal opinion saying the I/M program was absolutely required by the 1982 non-attainment SIP revisions. On August 3, 1983, the EPA relaxed the I/M requirement, saying SIP provisions could be avoided if the states were making reasonable efforts and some progress was evident.[65] This approach was codified in a new sanctions provision in 1983.[66] At least eleven states thus avoided

---

[58] *See* EPA QUESTIONS, *supra* note 54, at 8.

[59] *See* New Mexico Envtl. Improvement Div. v. Thomas, 789 F.2d 825 (10th Cir. 1986); McCarthy v. Thomas, 17 Envtl. L. Rep. 21214 (D. Ariz.), *petition denied sub nom.* Arizona v. Thomas, 829 F.2d 834 (9th Cir. 1987); Delaware Valley Citizens' Council for Clean Air v. Pennsylvania, 533 F. Supp. 869 (E.D. Pa.), *aff'd*, 678 F.2d 470 (3d Cir.), *cert. denied*, 459 U.S. 969 (1982); *see also* Ora Fred Harris, Jr., *The Automobile Emissions Control Inspection and Maintenance Program: Making It More Palatable to "Coerced" Participants*, 49 LA. L. REV. 1315 (1989); Jerome Ostrov, *Inspection and Maintenance of Automotive Pollution Controls: A Decade-Long Struggle Among Congress, EPA and the States*, 8 HARV. ENVTL. L. REV. 139 (1984).

[60] Federal Assistance Limitations, State of California, 45 Fed. Reg. 81,746 (1980) (to be codified at 40 C.F.R. ch. 1).

[61] *See* I/M PROGRAM BEHIND SCHEDULE, *supra* note 50, at 13. *But see* U.S. ENVTL. PROTECTION AGENCY, I/M UPDATE 1 (Nov. 30, 1979) (reporting surveys showing some popular support for the I/M programs: 58% favoring retention of the program in Arizona; 67% favoring annual emissions inspections in California; 83% favoring continuation of the program in New Jersey; and 83% considering I/M important in Rhode Island); Ostrov, *supra* note 59, at 190 (referring to surveys showing public support for I/M programs generally, as well as very strong support for public funding of cleaner air programs).

[62] *See* I/M PROGRAM BEHIND SCHEDULE, *supra* note 50, at 15.

[63] *Id.* at 17.

[64] State Implementation Plans, 46 Fed. Reg. 7182 (1981) (to be codified at 40 C.F.R. pt. 51).

[65] Federal Assistance Limitation and Construction Moratorium, 48 Fed. Reg. 35,312 (1983) (to be codified at 40 C.F.R. pt. 86) (proposed Aug. 3, 1983).

[66] Compliance with Statutory Provisions of Part D of the Clean Air Act, 48 Fed. Reg. 50,686 (1983) (to be codified at 40 C.F.R. pts. 51–52).

a construction ban. Further delay occurred because states believed that Congress would amend the CAA to reduce the use of I/M programs.[67] Congress, of course, did not amend the CAA until November 15, 1990, and then expanded the I/M program to include the need for enhanced I/M programs in serious to worse ozone non-attainment areas.[68]

The 1990 CAAA required the EPA to issue new guidance for state I/M programs.[69] In developing these guidelines, the EPA must provide some explanation for the results of its internal Office of Mobile Sources ("OMS") audits, which show that even vehicles deliberately modified by the EPA to fail I/M inspections nevertheless passed. The EPA has claimed that test equipment is often unreliable, testing personnel often inadequately trained, and that state enforcement efforts against substandard inspection stations were slow and overly lenient.[70]

Before the 1990 CAAA, the states, partially because of their initial resistance to I/M, were given considerable latitude in developing their I/M programs. The EPA was interested in getting the programs started, even if that meant less stringent oversight that might in turn compromise program effectiveness.[71] In return for adopting I/M programs, states were given credit in their SIPs toward mandated CO and ozone reductions.

The EPA pressed for I/M programs to produce light-duty vehicle hydrocarbon ("HC") and CO exhaust emissions reductions of at least twenty-five percent by December 31, 1987. These reductions were to be made relative to pre-I/M emissions levels, based upon the EPA's own motor vehicle emission factors.[72] Each program is expected to (1) provide for periodic inspection of all vehicles for which emission reductions are claimed (random roadside checks are not acceptable); (2) provide for retesting of failed vehicles to assure compliance; (3) prohibit registration or provide some equally effective mechanism to prevent noncomplying vehicles from operating;

---

[67] *See* I/M PROGRAM BEHIND SCHEDULE, *supra* note 50, at 18.

[68] Clean Air Act § 182(c)(1)–(3), (d)–(e), 42 U.S.C. § 7511a(c)(1)–(3), (d)–(e) (Supp. II 1990).

[69] *Id.* § 182(a)(2)(B)(ii), 42 U.S.C. § 7511a(a)(2)(B)(ii).

[70] U.S. ENVTL. PROTECTION AGENCY, EIKAFO-03-0269-1100359, FINAL REPORT OF AUDIT ON THE VEHICLE INSPECTION/MAINTENANCE PROGRAM 4 (Aug. 1991) [hereinafter EPA AUDIT].

[71] *See* EPA QUESTIONS, *supra* note 54, at 35.

[72] *See* U.S. ENVTL. PROTECTION AGENCY, EPA-460/3-78-013, MOTOR VEHICLE EMISSION INSPECTION MAINTENANCE INFORMATION KIT 1.2-6 (1978) [hereinafter INFORMATION KIT].

(4) provide for quality control over test equipment and procedures; (5) provide either a training program for mechanics or an information service to inform the public of service stations with approved emission analyzers; and (6) inform the public of the reason for the I/M program and the locations and hours of inspection stations.[73]

III. I/M UNDER THE 1990 CAAA

Because the EPA never issued binding regulations for I/M programs, considerable variation developed among state programs. The EPA issued its first policy for I/M programs in 1978 as guidance. This policy addressed the elements to be included in the SIP, minimum emission reduction requirements, administrative requirements, and schedules for implementation.[74] However, the lack of minimum federal requirements led to a less than fully effective I/M program.[75]

The benefit of I/M is that it focuses on the major problem concerning automobile emissions. Fewer than 10% of all cars emit more than 50% of the automobile pollution. The cleanest 50% of the automobiles emit only 3% of the automobile pollution.[76] According to one study, the dirtiest 20% of cars emit 480 times more HC than the cleanest 20%.[77] Old, pre-1975 vehicles do not necessarily produce more pollution than new cars.[78]

However, the current I/M system has deficiencies. The idle mode test is not very effective at identifying either the emissions of today's technically more sophisticated vehicles or the vehicles likely to be produced in the future. These vehicles use onboard computers and sensors that continuously adjust engine performance and emissions. Such vehicles could be more effectively tested by monitoring cycles of acceleration and deacceleration when the vehicle is under load. Moreover, a system

---

[73] *Id.* at 1.2-6 to -7.

[74] 57 Fed. Reg. 52,950, 52,952 (1992) (to be codified at 40 C.F.R. pt. 51); *see also supra* note 72 and accompanying text.

[75] *See* 57 Fed. Reg. 52,953 (1992) (to be codified at 40 C.F.R. pt. 51).

[76] Rick Henderson, *Dirty Driving: Donald Stedman and the EPA's Sins of Emission*, POL'Y REV., Spring 1992, at 56, 57 (discussing a study by the National Academy of Sciences, RETHINKING THE OZONE PROBLEM IN URBAN AND REGIONAL AIR POLLUTION (1991)).

[77] *New State Study Reveals Major Flaw in EPA I/M Approach, Critics Say*, INSIDE EPA, Jan. 24, 1992, at 1, 2.

[78] *See* Henderson, *supra* note 76, at 57.

called the Fuel Efficiency Automobile Test ("FEAT") uses an infrared light beam across a lane of traffic to measure emissions of CO accurately in less than one second with vehicle speeds as high as sixty miles per hour.[79] Such technology, which is already being field tested,[80] may prove to be more cost-effective than other methods of reducing mobile emissions and could make existing I/M tests obsolete.

The increasingly complicated nature of motor vehicle emissions systems makes visual inspections, a component of today's I/M program, less important. Still another weakness of existing I/M programs is that no I/M programs today test for evaporative emissions, an important source of HC. The EPA claims that quality control at decentralized inspection facilities is poor,[81] and some state officials fear I/M programs do not place enough emphasis on training mechanics to fix and maintain vehicles whose problems have been identified.[82] In addition, repairs performed to vehicles that fail inspections are often not completed because of low ceilings on repair costs. These presumed deficiencies prompted Congress to require the development of enhanced I/M programs for certain ozone non-attainment areas.[83] Enhanced I/M improves on existing tests by requiring the use of new technologies and methods.

The 1990 CAAA requires 181 areas in thirty-eight states to have basic or enhanced I/M programs.[84] Ozone and CO non-attainment areas must have at least a basic I/M program. The EPA recommends extending such requirements to areas that significantly contribute to non-attainment from mobile sources. Marginal or worse ozone areas or moderate CO non-attainment areas with a design value less than 12.7 ppm must update exist-

---

[79] *See* Douglas R. Lawson et al., *Emissions from In-use Motor Vehicles in Los Angeles: A Pilot Study of Remote Sensing and the Inspection and Maintenance Program*, 40 J. AIR & WASTE MGMT. ASS'N 1096 (1990).

[80] *See Automakers Embark on Joint Study of Devices for Detecting Car Emissions*, INSIDE EPA, Oct. 16, 1992, at 13, 13; *Remote-Sensing Measurement of Auto Emissions Tested in Michigan by Automakers, Regulators*, Env't Rep. (BNA) 1472 (Sept. 25, 1992).

[81] *See* 57 Fed. Reg. 31,059 (1992) (to be codified at 40 C.F.R. pt. 51) (proposed July 13, 1992); U.S. ENVTL. PROTECTION AGENCY, FUTURE TRENDS IN THE EMISSIONS OF MOTOR VEHICLE CONTROLS 4 (1980) (noting that a 1980 survey showed only 33% of 1900 cars were adequately maintained and had all their emission controls present and functioning).

[82] *See States Fear Lack of Emphasis on Mechanic Training May Doom CAA I/M Program*, INSIDE EPA, Sept. 25, 1982, at 14, 14.

[83] *See infra* notes 125–152 and accompanying text.

[84] U.S. ENVTL. PROTECTION AGENCY, INSPECTION/MAINTENANCE FACTS AND FIGURES (1992) [hereinafter FACTS AND FIGURES].

ing I/M programs to meet the 1990 requirements for basic I/M.[85] Moderate ozone non-attainment areas outside an ozone transport region must also have basic I/M programs.[86] Moderate ozone non-attainment areas that do not have I/M must develop programs unless they are rural areas without any urbanized areas. Basic I/M is required in any 1990 census-defined urbanized area.[87]

The EPA was to upgrade the basic I/M requirements to provide minimum standards concerning inspection frequency, test methods, components covered, quality control, and enforcement.[88] Guidance was to be published by November 15, 1991, and no later than two years after regulations are promulgated the states must submit SIP revisions to meet the requirements.[89] The EPA released a notice of proposed rulemaking on July 13, 1992.[90]

Serious or worse ozone non-attainment areas, or moderate or serious CO non-attainment areas with a design value greater than 12.7 ppm[91] and with a 1980 census population of 200,000 or more, must implement enhanced I/M.[92] Areas needing a program are determined from the 1980 census data, but the boundaries are to be determined from the 1990 census data.[93]

In addition, ozone transport regions are subject to enhanced I/M requirements if they are Metropolitan Statistical Areas ("MSAs") with a 1990 population of 100,000 or more, regardless of attainment status, and if their emissions of ozone precursors contribute to a violation of a state or federal air quality standard for ozone.[94] Also subject to enhanced I/M are commuter corridors, areas not within an enhanced or basic I/M program area

---

[85] Clean Air Act §§ 182(a)(2)(B), 187(a)(4), 42 U.S.C. §§ 7511(a)(2)(B), 7512a(a)(4) (Supp. II 1990). The design value is the ambient level of ozone or CO used by the EPA for classification purposes and is expressed in parts per million. It is calculated using methodology issued by the Administrator. *Id.* §§ 181(a)(1), 186(a)(1), 42 U.S.C. §§ 7511(a)(1), 7512(a)(1).

[86] *Id.* § 182(b)(4), 42 U.S.C. § 7511a(b)(4); 57 Fed. Reg. 52,965 (1992) (to be codified at 40 C.F.R. pt. 51).

[87] 57 Fed. Reg. 56,965–66 (1992) (to be codified at 40 C.F.R. pt. 51).

[88] Clean Air Act § 182(a)(2)(B), 42 U.S.C. § 7511a(a)(2)(B) (Supp. II 1990).

[89] *Id.*

[90] 57 Fed. Reg. 31,058 (1992) (to be codified at 40 C.F.R. pt. 51) (proposed July 13, 1992).

[91] Clean Air Act § 187(a)(6), 42 U.S.C. § 7512a(a)(6) (Supp. II 1990).

[92] *Id.* § 182(c)(3)(A), 42 U.S.C. § 7511a(c)(3)(A); 57 Fed. Reg. 52,966 (1992) (to be codified at 40 C.F.R. pt. 51).

[93] Clean Air Act § 182(c)(3)(A), 42 U.S.C. § 7511a(c)(3)(A) (Supp. II 1990); 57 Fed. Reg. 52,966 (1992) (to be codified at 40 C.F.R. pt. 51).

[94] Clean Air Act § 184(b)(1)(A), 42 U.S.C. § 7511c(b)(1)(A) (Supp. II 1990).

that are origins or destinations of vehicular movements that are significant and regular causes of or contributors to ambient air quality violations within program areas.[95] In the Northeast Ozone Transport Region,[96] however, largely rural counties with less than 200 persons per square mile are exempt if at least fifty percent of any given MSA is included in the enhanced I/M program.[97] Islands off the Northeast United States coast unconnected to the mainland are also exempt.[98] In multi-state areas, urbanized areas of 50,000 persons within a state, as determined by the 1990 census, are subject to I/M requirements.[99]

The federally-defined boundaries may not always be consistent with logical administration, but the enhanced I/M boundaries may be adjusted. States may add vehicles registered in commuter corridors serving urbanized areas or ozone transport areas or other areas with less severe pollution where such expansion is cost-effective and contributes to achieving air quality standards.[100] Enhanced programs may choose to base coverage on ZIP code boundaries, rather than on census-based boundaries. ZIP code boundaries make it easier to implement and enforce programs based on vehicle registration. So long as the EPA's geographical coverage requirements are met, the states can expand coverage to meet their needs.[101]

If a state expands I/M coverage beyond what the EPA requires, the extra emission credits cannot be applied to the minimum performance standard. They can be used only to meet "reasonable further progress" requirements or as an offset.[102] In some circumstances, urban population may be excluded if an equal number of contiguous non-urban residents who live in the same MSA are included to make the I/M program jurisdiction correspond to county boundaries.[103] Ozone or CO non-attainment areas that are serious, severe, or extreme must include the entire MSA in the non-attainment area.[104]

---

[95] *See* CALIFORNIA I/M REVIEW COMM., PROPOSED LEGISLATIVE CHANGES TO IMPLEMENT ENHANCED VEHICLE INSPECTION AND MAINTENANCE 9 (1992).
[96] Clean Air Act § 184(a), 42 U.S.C. § 7511c(a) (Supp. II 1990).
[97] 57 Fed. Reg. 52,966 (1992) (to be codified at 40 C.F.R. pt. 51).
[98] *Id.*
[99] *Id.*
[100] *See* CALIFORNIA I/M REVIEW COMM., *supra* note 95, at 3.
[101] *See id.* at 9.
[102] 57 Fed. Reg. 52,967 (1992) (to be codified at 40 C.F.R. pt. 51).
[103] *Id.*
[104] Clean Air Act § 107(d)(4)(A)(iv), 42 U.S.C. § 7407(d)(4)(A)(iv) (Supp. II 1990).

For serious or worse ozone non-attainment areas with a 1980 population of 200,000 or more, the states were required to submit enhanced I/M programs by November 15, 1992.[105] The EPA is to publish guidance that will include a performance standard based on emission testing, including on-road emission testing and inspection to detect tampering with emission controls. The programs are applicable to all light-duty vehicles and trucks.[106] The statute goes on to specify seven elements required for an enhanced I/M program:[107] (1) computerized emission analyzers, including on-road testing devices; (2) no waivers for vehicles or parts covered by emission control performance warranties; (3) a minimum expenditure by the consumer of $450 for repairs, adjusted annually by reference to the Consumer Price Index, before the state can waive emission requirements; (4) enforcement through denial of vehicle registration unless the state can demonstrate a more effective enforcement program; (5) annual inspections unless biennial inspections are as effective; (6) a centralized program unless the state can demonstrate a decentralized program is as effective; and (7) a program for inspection and repair of emission control diagnostic systems.

Prior to the 1990 CAAA, the EPA had only issued guidance documents for I/M programs. Guidance continues to be required for the basic I/M program under section 182(a)(2)(B)(ii), but the

---

[105] *Id.* § 182(c)(3)(A), 42 U.S.C. § 7511a(c)(3)(A).

[106] *Id.* § 182(c)(3)(B), 42 U.S.C. § 7511a(c)(3)(B).

Light-duty vehicles are divided into categories based on loaded vehicle weight ("LVW"), which is curb weight plus 300 pounds. Vehicles under 3750 pounds LVW are held to almost the same standards as passenger cars. Vehicles between 3751 and 5750 pounds LVW have slightly less stringent standards.

Light-duty vehicles can exceed 6000 pounds, and are divided into subcategories of vehicles with test weight of 3751 to 5750 pounds and those over 5750 pounds test weight. *Id.* § 202(h), 42 U.S.C. § 7521(h). Test weight is the curb weight added to the gross vehicle weight rating ("GVWR") and divided by two. *Id.* § 216(8), 42 U.S.C. § 7550(8). These vehicles have less stringent requirements than lighter vehicles, but the requirements tighten the standards beginning with model year ("MY") 1994. What constitutes a heavy-duty truck is ultimately determined by regulations. It should be noted that on its face the provisions of § 202(b)(3)(C) and (h) do not mesh. Heavy-duty vehicles must meet standards based on "the greatest degree of emission reduction achievable through the application of technology which the Administrator determines will be available for the model year to which such standards apply, giving appropriate consideration to cost, energy, and safety factors associated with the application of such technology." *Id.* § 202(a)(3)(A), 42 U.S.C. § 7521(a)(3)(A). The explanation is that light-duty trucks can be heavy-duty vehicles. If GVWR is over 6000 pounds the heavy-duty vehicle regulations under § 202(a)(3)(A) apply. A light-duty truck over 6000 pounds GVWR must also meet the requirements of § 202(h). *See* Natural Resources Defense Council v. Thomas, 805 F.2d 410, 438 (D.C. Cir. 1986) (declining to apply different standards for light-duty trucks that are heavy-duty vehicles than for other heavy-duty vehicles).

[107] Clean Air Act § 182(c)(3)(C), 42 U.S.C. § 7511a(c)(3)(C) (Supp. II 1990).

EPA interprets section 182(c)(3)(B) to require binding standards for enhanced I/M. In promulgating standards, rather than providing guidance, the EPA must satisfy all required rulemaking procedures, including compliance with the required notice and comment procedures.[108] The EPA decided to issue binding regulations for enhanced I/M where required and for basic I/M where rulemaking is optional.[109] Under this approach the EPA has promulgated a performance standard that provides states with the flexibility to design their programs as long as these programs meet the requirements of the CAA and the overall effectiveness of the performance required by the regulation.[110] The most important new provision is the requirement of centralized testing unless the state can demonstrate that decentralized testing is equally effective.[111]

Section 182(c)(3)(C)(iv) requires denial of motor vehicle registration unless the owner complies with enhanced I/M requirements. A registration-based enforcement program could prevent the processing of registration renewals for registrants that have not submitted proof of a successful emission inspection within the previous two years. States may choose a different method of enforcement in areas subject to existing I/M programs if the alternative is more effective. In other basic I/M areas, equally effective alternatives are allowed. For newly implementing areas there is no alternative to registration denial.

IV. The Structure of the I/M Program

The EPA has allowed each state to choose a centralized or decentralized structure for its I/M program.[112] The state or municipal government could own and operate the inspection facility or a contractor or other private entity could run the program. New Jersey, for example, used a government-ownership approach.[113] This approach required more capital outlay and greater start-up expenses than other methods, but it entailed less administrative, auditing, and surveillance costs than a contractor-operated or private garage system. When a state, such

[108] *See* PPG Indus. v. Costle, 659 F.2d 1239 (D.C. Cir. 1981).
[109] *See* 57 Fed. Reg. 52,953 (1992) (to be codified at 40 C.F.R. pt. 51).
[110] *Id.* at 52,951, 52,953.
[111] *Id.* at 52,953.
[112] *See* Information Document, *supra* note 48, at 19, 22, 38.
[113] *See* Information Kit, *supra* note 72, at 53.

as New Jersey, already has a centralized safety inspection program, adding emission inspection capability may not be difficult.[114]

Most I/M programs require annual inspection.[115] The in-use vehicle emissions tests use a short test, applicable to warmed-up vehicles, that could identify high-emitting vehicles.[116] Two distinct emission testing procedures were developed: the idle mode and loaded mode tests.[117]

The idle mode test measures exhaust emissions with the vehicle in a neutral gear and the engine at idle. To pass the test, the vehicle's HC and CO levels must meet EPA standards at both normal and high idle speeds. The idle mode test is easy to perform and requires little technical training.[118] The test can easily be duplicated with equipment that most service stations can afford. If this test is used for I/M, the service station repairing a failed vehicle can confirm that emission-related maintenance has been successfully performed.[119]

The loaded mode test measures exhaust emissions with the vehicle in a forward drive gear and operating under simulated driving conditions. Because it partially simulates actual driving conditions, the loaded mode test provides a better indication of actual emissions than does the idle mode test. The simulation is also capable of diagnosing engine maladjustments and malfunctions. In addition, loaded tests can measure NO_x emissions, which the idle mode tests cannot measure, because NO_x emissions are negligible when a vehicle idles. Loaded mode tests are more expensive, however, because they require a chassis dynamometer, greater technical skills, and more time.[120] Because these tests are not easily duplicated at a repair facility, owners

---

[114] *Id.* at 3.1-7.
[115] *Id.* at 4.2-4.
[116] *See* INFORMATION DOCUMENT, *supra* note 48, at 19.
[117] *Id.*
[118] *Id.* at 21–22.
[119] *Id.*
[120] Two kinds of loaded mode test procedures exist: steady state and transient. The transient test will yield better correlation than the steady state test with respect to the federal test procedure. However, the transient test is more expensive to perform and requires more time. Because the steady state test uses a volumetric procedure, a standard exhaust emission analyzer can be used. The transient loaded mode test, by contrast, collects a composite emission sample from a specified driving schedule. The composite sample is collected into a constant volume sample unit for further analysis to determine pollutant concentration. A chassis dynamometer loads the vehicle to simulate the desired driving schedule. However, for this test, the dynamometer must be capable of performing at variable inertia weight and road load settings. *Id.*

of vehicles that fail I/M tests may have to make repeated trips between I/M and repair facilities before passing the test.

The SIP requires a certain percentage of automobiles with the highest emissions to be rejected for failure to meet emission standards. The cut point is the level of emissions that distinguishes between those vehicles requiring emissions-related maintenance and those that do not. The cut points define a stringency factor (flunk rate) that is a measure of the program's rigor. The more stringent the program, the greater the pollution reductions the state may claim in the SIP. Because the public may react negatively to an excessive number of vehicles failing the first inspection, the cut points were originally set high enough to reduce emissions to meet atmospheric goals but low enough to be politically acceptable.[121] For example, in the late 1970s, Arizona and New Jersey used a stringency factor of sixteen percent, Chicago, Illinois, used twenty-eight percent, and Cincinnati, Ohio, used eighteen percent.[122] However, despite a targeted failure rate, some states passed too many vehicles.[123] For example, as of May, 1984, Virginia was failing about 8000 vehicles annually rather than the 80,000 it was expected to fail.[124] Thus the benefits given to the states' SIPs by the EPA's computer simulation models may be unduly generous.[125] For basic I/M programs pre-1981 vehicles are subject to a twenty percent failure rate.[126] For 1981 and later vehicles the failure rate is based on the performance standard used for SIP approval with emission standards no weaker than specified in the regulations[127] and subject to a further requirement that NO_x emissions not increase because of the I/M program.[128]

Under the 1990 CAA's enhanced I/M program there must be a twenty percent failure rate among pre-1981 vehicles.[129] For model year ("MY") 1981 and later vehicles the regulation limits HC, CO, and NO_x emissions by model year, with the limitations

---

[121] *See generally* Stephen J. Lynton, *N. Va. Rate Lowest for Car Exhaust Test Failures*, WASH. POST, Mar. 18, 1983, at A-1 (describing political difficulties of differential failure rates across states).

[122] *See* INFORMATION DOCUMENT, *supra* note 48, at 44–53.

[123] *See, e.g.*, Lynton, *supra* note 121, at A-1.

[124] *See* I/M PROGRAM BEHIND SCHEDULE, *supra* note 50, at iii.

[125] *Benefits of State's Vehicle I/M Program Overestimated, Auditor General Says in Report*, Env't Rep. (BNA) 1809 (Jan. 6, 1989).

[126] 57 Fed. Reg. 52,989 (1992) (to be codified at 40 C.F.R. pt. 51).

[127] 40 C.F.R. § 85.2201–.2217 (1992).

[128] 57 Fed. Reg. 52,989 (1992) (to be codified at 40 C.F.R. pt. 51).

[129] *Id.*

becoming more stringent for MY 1986 vehicles and still more stringent for MY 1994 vehicles. There are less stringent requirements for light-duty trucks. Beginning with MY 1994, HC emissions limits will be changed to a measurement based on nonmethane hydrocarbons ("NMHC").[130] Three percent of the vehicles that fail may receive waivers, and it is expected that ninety-six percent of the vehicles subject to inspection will comply.[131]

The states must demonstrate that the emission levels achieved by the I/M programs are equivalent to the reductions projected in their SIPs using the most current version of the EPA's mobile source emission model or an EPA-approved alternative.[132] This will require testing random samples of vehicles subject to the I/M program. The states will monitor IM240 transient exhaust tests, purge tests, and pressure tests on these samples. The results will be compared to the SIP emission factor projections. States that do not achieve their emission factor targets must take corrective actions.[133]

Under the EPA regulations, the enhanced I/M targets are based on the average emission level for on-road vehicles tested under a centralized, test-only annual inspection applicable to 1968 and later MY light-duty vehicles using an enhanced I/M test regime. Based on the EPA's current model, this requirement is equivalent to a reduction of twenty-eight percent for VOC, thirty percent for CO, and nine percent for NO_x. The HC reduction standard applies to both exhaust and evaporative emissions, which differs from the traditional approach of reporting emission reductions based only on exhaust emissions. Therefore, a twenty-eight percent reduction in tailpipe HC emissions will not meet the new requirements.[134]

The trend is to perform I/M testing using contractors, with testing at centralized or decentralized stations.[135] The central-

---

[130] *Id.* at 52,988.

[131] *Id.* at 52,988–89.

[132] *Id.* The EPA is in the process of moving from the Mobile 4.1 emission factor model to Mobile 5.0. *See infra* note 244.

[133] *See* STATE OF NEW YORK DEP'T OF ENVTL. CONSERVATION & DEP'T OF MOTOR VEHICLES, PROPOSED NEW YORK STATE INSPECTION AND MAINTENANCE COMMITMENTS TO BE USED IN A STATE IMPLEMENTATION PLAN 3 (1992) [hereinafter NEW YORK I/M COMMITMENTS].

[134] *See* CALIFORNIA I/M REVIEW COMM., *supra* note 95, at 6. California's rules base the reduction on the use of the EPA's latest Mobile 5 model, but the regulations project the identical reduction using the predecessor Mobile 4.1 model. 57 Fed. Reg. 52,954 (1992).

[135] Systems Control, the leading I/M company in the United States, ran programs in

ized network uses a high-volume, multi-lane station that may be run by the government or a contractor. These facilities are usually highly automated and only perform testing.[136] The decentralized network uses gasoline stations or repair facilities as test centers. These facilities, by contrast, both test and repair vehicles. The 1990 CAAA does not require basic I/M to be test-only, and a reasonably comprehensive, conventional test and repair system can meet EPA requirements.[137] However, the marginal and moderate areas that are allowed to use basic I/M must meet the NAAQS by 1993 and 1996, which means basic programs may not exist after 1996. The enhanced programs mandated by the 1990 CAAA must operate centrally, unless the State demonstrates that a decentralized program will be equally effective.[138] The EPA believes that the standard can be met with a private or government-run centralized system. Alternatively, the standard can also be met with either a test-only, high-volume, decentralized, multi-participant system or with a test-only, multi-contractor system such as the one used in Florida.[139] The EPA believes, however, that it is not possible for a decentralized test and repair facility to meet the performance standard for an enhanced I/M program.[140]

Under the test-only approach being aggressively pushed by the EPA, a vehicle is initially tested at a test-only station. If the vehicle passes, a certificate of compliance is issued, but if the vehicle fails it must be repaired by the owner or at a licensed repair facility. After repair the vehicle must be returned and retested at the test-only facility. If it passes the emissions test, or if the amount spent on repairs exceeds the waiver amount (at least $450 for enhanced I/M programs), a certificate is issued. If the vehicle fails again and the amount of money spent on repairs is less than the waiver amount, the process is repeated.

1990 in Maryland, Florida, Illinois, Minnesota, Washington, California, and Alaska. These programs include roller dynamometer testing, which tests for NO_x as well as HC and CO.

[136] *See* 57 Fed. Reg. 52,958 (1992).

[137] *Id.*

[138] Clean Air Act § 182(c)(3)(C), 42 U.S.C. § 7511a(c)(3)(C)(iv) (Supp. III 1991).

[139] 57 Fed. Reg. 52,959 (1992). California, while providing for test-only stations, plans to allow such stations to make repairs, to fix parts damaged by station personnel, to make vehicles safe for operation while at a station, and to do other minor repairs requiring less than five minutes. All such repairs must be free of charge. *See* CALIFORNIA I/M REVIEW COMM., *supra* note 95, at 11.

[140] CALIFORNIA I/M REVIEW COMM., *supra* note 95, at 11.

This process, known as "ping-ponging," can lead to public opposition to test-only programs.[141]

California has proposed a limited program of "Gold Shield" smog check stations that may perform repairs and issue certificates of compliance if safeguards are met. These stations must meet more rigorous licensing requirements than ordinary repair facilities, and any testing at Gold Shield stations must precede or follow a test at a test-only station. Certificates may not be issued to noncomplying vehicles based on waivers or to vehicles that were either tampered with or were classified as gross emitters. The number of Gold Shield stations and the number of vehicles that may be tested will be limited, however, until California can document the emission reductions in order to overcome the heavy discount factors that the EPA will apply. In addition, there is a more stringent training and certification program for Gold Shield technicians.[142] California hopes to use this program to demonstrate that test and repair facilities can work effectively, thereby forcing the EPA to reduce or eliminate the fifty-percent discount for exhaust emission reductions and the seventy-five percent discount for evaporative emission reductions that it applies to test and repair stations.[143] Since Gold Shield stations must run loaded mode tests on dynamometers but need not use the expensive IM240 instrument, the California program, if successful, could reduce the EPA's ability to force states to use IM240 technology.[144] Although the EPA is convinced that test and repair stations cannot meet performance standards,[145] the statute authorizes states to demonstrate that these stations are equivalent to the EPA's test-only approach.[146]

While the EPA admits centralized and decentralized test centers can provide convenient service, it also believes that contractor-operated centralized systems provide the added benefit of more accurate testing.[147] Such systems, however, need to be carefully designed to avoid the adverse ping-pong effect and the repair cost limitation/waiver. The EPA encourages states to deal with ping-ponging through mechanic training programs, technician outreach programs, and technician performance monitor-

---

[141] *Id.* at 13.
[142] *Id.* at 14, 34.
[143] *Id.*
[144] *Id.* at 16.
[145] *See* 57 Fed. Reg. 52,959 (1992).
[146] Clean Air Act § 182(c)(3)(C), 42 U.S.C. § 7511a (c)(3)(C) (Supp. III 1991).
[147] 57 Fed. Reg. 52,959 (1992).

ing.[148] Towards this end, the EPA has awarded a three-year, $700,000 grant to the National Automotive Technicians Education Foundation to develop mechanic curricula and training materials.[149]

There were 127 areas in 1992 that had I/M programs and this number is expected to increase to 181 areas under the 1990 CAAA.[150] The EPA estimates that over a billion dollars per year is spent nationwide on I/M programs and that this amount may increase under the 1990 requirements.[151] In addition, the widespread use of decentralized inspection facilities has resulted in less than adequate I/M programs. Thus, it is expected that the EPA's new regulations will lead to much greater controls over decentralized programs and an effort to force the states to adopt centralized testing facilities.

In April, 1991, the State and Territorial Air Pollution Program Administrators and the Association of Local Air Pollution Control Officials asked the EPA to require emerging technologies, if feasible, to be incorporated into I/M programs. If decentralized programs are to be converted to centralized programs the affected agencies should have the maximum flexibility to phase in such programs.[152] The Northeast States for Coordinated Air Use Management has also called on the EPA to centralize vehicle testing; require transient loaded emission testing for NO_x, CO, and HC; require evaporative emission purge and pressure testing; plan to link testing to onboard diagnostic systems; improve enforcement; require anti-tampering checks; develop effective program evaluation measures; and develop measures to ensure effective repairs, including advanced mechanics training and certifications.[153]

The EPA's preference for centralized programs has found support in Congress. For example, Chairman Henry Waxman (D-Cal.) of the House Energy & Commerce, Health and the Environment subcommittee of the House Committee on Energy has supported the EPA's view.[154] However, the Bush adminis-

---

[148] *Id.* at 52,961.
[149] *Id.* at 52,963.
[150] FACTS & FIGURES, *supra* note 84.
[151] *See infra* note 322 and accompanying text.
[152] *State Regulators Urge EPA to Mandate Strong Auto Inspection, Maintenance*, INSIDE EPA, Apr. 5, 1991, at 12.
[153] *Id.*
[154] *Lieberman Joins Chorus of Supporters for Tough Centralized Auto Inspection Plan*, CLEAN AIR REP., Sept. 12, 1991, at 14.

tration lobbied for a decentralized program because of the benefits to small businesses that perform I/M testing.[155] Similarly, some states with large investments in decentralized programs, such as Texas and California, have opposed changing to a centralized system.[156] California has argued that the new BAR 90 analyzer system greatly improves the monitoring and enforcement of the I/M testing process and allows decentralized programs to be as effective as centralized programs.[157] The automobile industry, on the other hand, has endorsed a strengthened I/M program without taking a position on whether a centralized program is superior to a decentralized one. The industry hopes that a strict I/M program will keep the states from adopting California's low emission standards.[158] The industry has also voiced support for I/M programs over higher tailpipe emission standards, in part because "cranking down on tailpipes . . . isn't giving us as much [emissions reductions] as we thought."[159]

The 1990 CAAA approach aims to test motor vehicles in the most polluted areas at a level of sophistication greater than both the common idle/2500 rpm test and the loaded test. Enhanced I/M will require testing under load during cycles of acceleration and deceleration, and this test is called the transient loaded, high-tech, or IM240 exhaust test.[160] It will also require an evaporative system integrity test (pressure test) and an evaporative performance test (purge test).[161] Moreover, imposition of enhanced I/M programs will adversely affect the many decentralized programs. For instance, decentralized stations may have test equipment or other capital investment that may not be fully amortized. Furthermore, a more centralized system will force many existing test stations to discontinue testing and thus lose

---

[155] *Proposal to Tighten Auto I/M Regulations Would Threaten Repair Industry, Officials Say*, Env't Rep. (BNA) 1247 (Aug. 21, 1992) [hereinafter *Proposal to Tighten I/M Regulations*]; *White House May Fight EPA Plan on Centralized Auto Inspection, Some Say*, CLEAN AIR REP., Sept. 12, 1991, at 13.

[156] *Proposal to Tighten I/M Regulations*, *supra* note 155.

[157] *Proponents of Decentralized I/M Programs Float New Study to Sway EPA*, INSIDE EPA, Sept. 20, 1991, at 13.

[158] *Auto Industry in Significant Move Expected to Endorse Strengthened I/M*, INSIDE EPA, Nov. 15, 1991, at 16; *Chrysler Backs 'High Tech' Auto Inspection as Alternative to Other Reduction Programs*, Env't Rep. (BNA) 1793 (Nov. 22, 1991); *EPA Staff Pleased with Industry Support for Stiffer Inspection Program*, CLEAN AIR REP., Dec. 5, 1991, at 9.

[159] *New Study Shows Car Emissions Hugely Underestimated, Questioning Science*, INSIDE EPA, Apr. 3, 1992, at 1, 12.

[160] 57 Fed. Reg. 52,951, 52,953–54 (1992).

[161] *Id.* at 52,951.

testing revenue and profits from related repair-work. Possible remedies proposed by the EPA include financial assistance to test facilities adversely affected by the transition to a centralized, high-tech system, delays to reduce the adverse impact of the new inspection systems, and state programs to retrain inspectors.[162] However, there is no assurance that money will be made available to assist decentralized service stations, and compensation for the loss of future income seems even more unlikely. Despite the optimistic language of the regulations, the new approach will have losers—the test and repair facilities or some of the decentralized facilities.

While the EPA is proposing centralized testing only, some state programs—such as the management contractor/franchise system being proposed in Texas—are considered equivalent to the proposed single contractor, test-only program.[163] The EPA has disagreed that decentralized systems are more convenient and has cited a Missouri study that demonstrated that testing at decentralized stations still took an average of forty-eight minutes even with an appointment. This study reinforced earlier findings that test results in decentralized systems take longer to obtain than in more efficient centralized systems. A California study estimated that the entire time required for a test, including repairs and retesting, averaged seventy-six minutes in centralized programs and 83.25 minutes in decentralized programs. The EPA maintains that the states can reduce inconvenience further by testing biennially and providing other services at the testing station, such as driver license renewal, tax payments, etc.[164]

The EPA has dropped the provisional equivalency option from the final rule, effectively requiring either a centralized or decentralized, test-only program. However, states can attempt to show on a case-by-case basis that decentralized test and repair programs will be as effective as centralized systems. States will have to make this demonstration together with their SIP submittal. Despite this option, the EPA has not yet found a test and repair program equal in effectiveness to a test-only program. The EPA has made reference to the experiences of California and New York in order to demonstrate that decentralized test and repair programs achieve at best approximately fifty percent

---

[162] *Id.* at 52,964.
[163] *Id.* at 52,972.
[164] *Id.* at 52,973.

of the potential emission reduction. Moreover, even that level of effectiveness can only be obtained by expending a vast amount of resources in a largely ineffective effort to prevent fraud.[165]

The details of the inspection are set forth in state statutes and regulations. In Virginia, for example, vehicles up to 8500 pounds and between one and twenty-one years of age must be tested.[166] Motorcycles and diesel vehicles are excluded. Inspections costing $12.50 must be done every two years or upon the transfer of the vehicle's title. A basic inspection includes a visual check of the fuel neck restrictor, the catalytic converter, positive crankcase ventilation ("PCV") valve, air pump, and evaporative emissions control system. A probe is placed in the vehicle's tailpipe while the engine is idling to measure CO and HC content of the exhaust. Vehicles pass or fail based on the allowable CO and HC levels for the model year of the vehicle.[167] States that use loaded mode tests can also test NO_x. If a vehicle fails it must be repaired and retested until it passes, and a vehicle cannot be registered or reregistered until it passes the I/M test or the owner obtains a waiver. A waiver can be obtained for a failed vehicle if all pollution control equipment is installed and operating and a minimum amount—which depends on the model year—has been spent on emission-related repairs. For pre-1972 model years the minimum amount is $60, $125 for MY 1972–1974, $175 for MY 1975–1979, and $200 for 1980 and newer models.[168] The minimum amount increases to $450 under the 1990 CAAA for enhanced I/M programs and is adjusted anually.[169] In 1993 the adjusted waiver amount is about $500.[170]

The EPA also allows a three percent waiver rate in enhanced I/M programs[171] but is moving towards a zero waiver for basic I/M programs.[172] States may, however, have a more stringent waiver requirement to qualify for increased emission reduction credits in their SIP. New York, for example, projects a waiver

---

[165] *Id.* at 52,974.

[166] Enhanced I/M is required for MY 1968 and later vehicles by federal regulation. *See id.* at 52,988 (to be codified at 40 C.F.R. § 51.351(a)(4)).

[167] EPA AUDIT, *supra* note 70, at 6.

[168] 57 Fed. Reg. 52,964 (1992).

[169] Clean Air Act § 182(c)(3)(C)(iii), 42 U.S.C. § 7511a(c)(3)(C)(iii) (Supp. III 1991).

[170] CALIFORNIA I/M REVIEW COMM., *supra* note 95, at 26.

[171] 57 Fed. Reg. 52,989 (1992) (to be codified at 40 C.F.R. § 51.351(a)(11)).

[172] *Id.* (to be codified at 40 C.F.R. § 51.352(a)(10)).

rate of one percent of the initially failed vehicles with corrective action required if that rate was exceeded.[173]

The EPA interprets the waiver provision[174] stringently.[175] As a condition for a waiver, there must be a thorough diagnosis and inspection of the vehicle that certifies that all reasonable cost-effective repairs have been properly performed and that additional repairs costing less than $450 will not further reduce emissions. For vehicles subject to the basic I/M program, the minimum waiver expenditure is $75 for pre-1981 vehicles and $200 for 1981 and later MY vehicles before the waiver of further expenditures is applicable.[176] The EPA estimates the average cost of repairs to vehicles that fail transient tests will be $120. Evaporative pressure test failures will cost $38 to fix, and purge test failures will cost $70 to repair.[177] The EPA also encourages states to purchase and scrap vehicles that cannot be repaired cost effectively.[178]

The 1990 CAAA will require the states to make some adjustment to their basic I/M programs. For those states required to implement an enhanced I/M program, substantial changes will be necessary to meet the statutory requirements.[179] The 1990 requirements are discussed below.[180]

V. I/M AND PERFORMANCE WARRANTIES

The two Clean Air Act warranty provisions are intimately related to the I/M program. The CAA requires manufacturers to warrant that properly maintained vehicles will meet emission standards over their useful lives. The more completely manufacturers meet this requirement, the less need for I/M programs.

The first of the two CAA warranty provisions is section 207(a), which has been in effect since MY 1972.[181] Section 207(a) requires the vehicle or engine to meet applicable emission standards at the time of sale and to be free from defects that may cause the vehicle or engine to fail to comply with the emission

---

[173] NEW YORK I/M COMMITMENTS, *supra* note 133, at 7.
[174] Clean Air Act § 182(a)(2)(B), 42 U.S.C. § 7511a(a)(2)(B) (Supp. III 1991).
[175] 57 Fed. Reg. 52,963–64 (1992).
[176] *Id.* at 52,964.
[177] *Id.* at 52,963.
[178] *Id.*
[179] *See supra* text accompanying note 107.
[180] *See infra* text accompanying notes 231–279.
[181] Clean Air Act § 207(a), 42 U.S.C. § 7541(a) (Supp. III 1991).

standards during the useful life of the car.[182] This warranty provision has been of little benefit to consumers because of the difficulty in determining which components and failures the warranty covers. When consumers actually do make claims, they are unable to prove that the component failure caused the emissions to exceed federal standards. The 1977 CAAA strengthened this warranty by subjecting manufacturers that improperly dishonored a claim to civil penalties of up to $10,000.[183] The EPA expected this provision would enable automobile owners to bring warranty claims for defective components that were related to vehicle emissions.[184]

The second provision, section 207(b), is known as the "performance warranty." It requires a manufacturer to bring any car that fails an I/M test into compliance with emission standards. Under the pre-1990 CAA, the warranty applied for twenty-four months or until 24,000 miles were driven.[185] A manufacturer could escape the warranty only by showing that the owner did not perform the required maintenance or repairs as set forth in the owner's manual or abused the vehicle during its operation. After the warranty expired, coverage continued only for an "emission control device" or component designed for emission control, including a catalytic converter or thermal reactor.[186] This protection lasted sixty months. The limitations in section 207(b) did not apply to section 207(a) warranties or to section 207(c) recall actions.

The performance warranty was implemented slowly because of the technical difficulty of developing quick and inexpensive tests that could be reasonably correlated with the sophisticated test used in prototype certification.[187] However, the EPA has

---

[182] "Useful life" of a car was defined as five years of use or until 50,000 miles are driven. *Id.* § 202(d)(1), 42 U.S.C. § 7521(d)(1). The 1990 CAAA increased the useful life to 10 years or 100,000 miles.

[183] *Id.* §§ 203(a)(4)(D), 205, 42 U.S.C. §§ 7522(a)(4)(D), 7524.

[184] Reitze, *supra* note 5, at 711.

[185] Clean Air Act § 207(b), 42 U.S.C. § 7541(b) (Supp. III 1991).

[186] *Id.*

[187] Reitze, *supra* note 5, at 711. The federal test procedure measures CO, HC, and NO_x in terms of grams per mile ("gpm"). Short tests measure CO as a percentage of total emissions and HC by parts per million. There was no test for NO_x. The federal test procedure tests at varying speeds with hot and cold starts, and the short tests have usually been performed at both low and high idle speeds; *See* INFORMATION DOCUMENT, supra note 48, at 12; Jerome Ostrov, *Controlling Automobile Pollution Through Manufacturer's Performance Warranties*, 61 DET. J. URBAN L. 65 (1983).

developed and proposed test procedures and performance warranty regulations to satisfy the requirements of section 207(b).[188]

After the passage of the 1977 CAAA, the EPA began redrafting the warranty regulations. The EPA believed that a short test could implement the section 207(b) warranty provisions.[189] The automobile industry disputed the correlation of these tests with the federal test procedure used for certification,[190] but the performance warranty requirements were upheld in three cases decided by the U.S. Court of Appeals for the D.C. Circuit.[191] After these decisions, the EPA proposed new regulations.[192] In 1989, the EPA proposed a short test for aftermarket part certification known as the cold 505 test procedure that could be used to avoid the more costly testing under the FTP.[193]

On June 12, 1984, the EPA published a final rule creating three new warranty short tests and made minor changes to the existing short tests.[194] The EPA found that the three new tests, which were slight variations of the existing tests, satisfied the provisions of section 207(b): the tests were available, they were in accordance with good engineering practices, and they could be correlated with the certification test.[195] The EPA also found that the amendments to the existing short tests were so minor as to have no effect on their prior finding of compliance with section 207(b).[196]

The new rule allowed any state I/M authority to petition the agency for approval of alternative quality control procedures.

---

[188] INFORMATION DOCUMENT, *supra* note 48, at 13.

[189] *Id.* at 14.

[190] *See* Reitze, *supra* note 5, at 712.

[191] *See* Specialty Equipment Market Ass'n v. Ruckelshaus, 720 F.2d 124 (D.C. Cir. 1983) (holding that certification under § 207(a)(2) could be extended to all parts affecting emissions, including specialty parts used to alter original equipment); Automotive Parts Rebuilders Ass'n v. EPA, 720 F.2d 142 (D.C. Cir. 1983) (upholding the EPA's interpretation that the warranty extends to primary emissions-related parts and all other parts necessary to the proper functioning of the primary parts); Motor Vehicle Manufacturers Ass'n v. Ruckelshaus, 720 F.2d 142 (D.C. Cir. 1983) (holding that state and local I/M "short test" procedures could be the basis for warranty violations because the tests were reasonably able to identify vehicles likely to fail the Federal Test Procedure ("FTP")).

[192] *See* 52 Fed. Reg. 924 (1987).

[193] 54 Fed. Reg. 32,598 (1989).

[194] 49 Fed. Reg. 24,320 (1984). The three new tests were all forms of idle testing. The EPA called them the "Engine Restart Idle Test," the "2500 rpm/Idle Test," and the "Engine Restart 2500 rpm/Idle Test." The tests were developed in response to complaints from Ford Motor Company, which claimed that some of their 1981 and later models were incorrectly failing the existing short tests. *Id.*

[195] *Id.*

[196] *Id.*

A state could request to use alternative quality control procedures equivalent to those required by the regulations.[197] However, the state must provide data and technical support to the EPA to substantiate its claim of equivalency.[198] The provision also required that, following a preliminary determination that the alternative procedures are equivalent, the EPA would publish a notice announcing the request and explaining the EPA's preliminary determination of equivalence.[199] If the EPA did not receive any adverse comments, a final notice would confirm the alternative procedures as valid. Vehicles failing I/M tests conducted in accordance with the alternative procedures would then be eligible for the performance warranty remedy, provided that other warranty conditions were met.

In 1987, Maryland and New York requested that the EPA approve variations of the federal short test used as part of their I/M programs. On December 7, 1987, the EPA approved these tests, finding them equivalent to the federal short test.[200]

A decade of experience with the present short test procedures had revealed problems. One of the most significant problems was "pattern failure." Pattern failures occur when specific groups of vehicles fail I/M procedures and neither owner action nor performance warranty coverage adequately addresses the failures.

Pattern failures arise under a variety of circumstances. They may be caused by malfunctions of a particular component in an engine family or by the vehicle design's incompatibility with the short test procedures or testing conditions.[201] Pattern failures of the short test do not necessarily mean that the vehicles will fail the federal test procedure.[202] The EPA has identified five vehicle design characteristics that account for the majority of pattern failures resulting from incompatibility with the short test procedures: (1) fuel control systems that react slowly to abrupt rpm changes; (2) air injection or fuel metering timers (used in some vehicles to prevent excess catalyst temperatures during idle); (3) catalysts and oxygen sensors that cool down while waiting in line to take the I/M test; (4) variations in evaporative system

---

[197] 40 C.F.R. § 85.2217 (1992).
[198] *Id.* § 85.2208(b).
[199] *Id.*
[200] 52 Fed. Reg. 46,354 (1987).
[201] 58 Fed. Reg. 3380, 3381 (1993) (to be codified at 40 C.F.R. pts. 85–86) (proposed Jan. 8, 1993).
[202] *Id.*

purge technology; and (5) unconventional ignition systems that prevent vehicles from completing the I/M tests.[203]

On January 8, 1993, in response to these problems with the short test, the EPA proposed changes to the existing section 207(b) scheme.[204] The first change would create a "Certified Short Test" ("CST") as part of the vehicle certification process to protect owners of light-duty vehicles and trucks from incurring I/M-related repair expenses in vehicles that have a history of pattern failures. The CST, which would be performed at the time of vehicle certification, would act as an early indicator of vehicle performance on the short test. The second change would replace the current I/M test procedures with six alternative tests.[205]

The EPA decided to amend the current short test procedures for several reasons. Changes in analyzer technology since the creation of the original short tests now permit improved and more rigorous testing.[206] Also, the proposed modifications to the current section 207(b) tests would significantly reduce errors committed during administration of the tests.[207]

The 1990 CAAA modified section 207(a)(1) to make the warranty period for light-duty vehicles and trucks either two years or 24,000 miles, whichever occurs first. The new warranty period applies to 1995 and later model years. For major emission control components, the warranty period is either eight years or 80,000 miles, whichever occurs first. Such controls are limited to the catalytic converter, electronic emissions control unit, onboard emissions diagnostic device, and other components (to be designated by the Administrator) that have a retail cost over $200 (in 1989 dollars) and were not in general use prior to MY 1990. The Administrator may, by regulation, establish a warranty period for other types of vehicles. Until new regulations are promulgated, the pre-1990 regulations will continue to apply to those vehicles.[208]

---

[203] *Id.* at 3381–82.

[204] *Id.* at 3380.

[205] *Id.* at 3382; *see also* EUGENE J. TIERNEY ET AL., RECOMMENDED I/M SHORT TEST PROCEDURES FOR THE 1990'S: SIX ALTERNATIVES (1990).

[206] 58 Fed. Reg. 3380, 3382 (1993) (to be codified at 40 C.F.R. pts. 85–86) (proposed Jan. 8, 1993).

[207] *Id.*

[208] Clean Air Act Amendments of 1990, Pub. L. No. 101-549, § 209, 104 Stat. 2399, 2484.

The warranty provisions of the 1990 CAAA[209] were the only environmental provisions in the bill passed by the House of Representatives that required a contested floor vote. The new provisions, which strengthened the reported bill, were a compromise between environmentalists and service station representatives.[210] Under a section 207(b) performance warranty, a vehicle's manufacturer must pay for the replacement of emissions control components when their replacement is called for by the vehicle's maintenance instructions.[211] The EPA considered certification of aftermarket parts to be a solution to potential anticompetitive effects of the emission warranty provisions. The more than 1700 parts manufacturers and 22,000 wholesalers and distributors must do business with the small number of manufacturers that dominate the field.[212] The parts industry feared that the added expense caused by government regulation of parts would force small companies out of business and prevent new entries. In addition, the parts industry feared that the warranty provisions would enable the automobile manufacturers to force car owners to have their vehicles serviced exclusively by franchised dealers who use only manufacturer-approved parts.[213]

During the late 1970s, the EPA and the aftermarket parts industry developed technical criteria and standards for parts certification. The certification program enabled aftermarket manufacturers to qualify their parts for use in maintenance required by the emission warranties. It thus eliminated the car owner's fear that use of aftermarket products would void emission warranty coverage. The program required aftermarket manufacturers to test and perhaps warrant their products, much like vehicle manufacturers are required to do.[214] The testing requirement mandated by the regulations can be onerous for a small business.[215]

---

[209] *Id.*

[210] *See* Waxman et al., *supra* note 26, at 1967.

[211] Clean Air Act § 207(a)(3), 42 U.S.C. § 7541(a)(3) (1988). The instructions must state that parts that have been certified pursuant to § 207(a)(2) may be used in the repairs. *Id.* § 207(c)(3)(A), 42 U.S.C. § 7541(c)(3)(A).

[212] *Monopolistic Tendencies of Auto Emission Warranty Provisions: Hearings Before the House Subcomm. on Environmental Problems Affecting Small Business of the Permanent Select Comm. on Small Business*, 93d Cong., 2d Sess. 1–2 (1974) (statement of Chairman William L. Hungate (D-Mo.)).

[213] *Id.* at 9.

[214] Reitze, *supra* note 5, at 712.

[215] *See* 40 C.F.R. § 85.2114–.2115 (1992). The EPA tried to write regulations that it

The 1990 CAAA, on balance, appears to benefit the repair industry. The change in section 207 extends the warranty period to eight years/80,000 miles for catalytic converters, electronic control units, and onboard diagnostic systems but reduces the warranty period to two years/24,000 miles for all other equipment. The EPA can impose a longer warranty only on parts that cost over $200 and were not used on vehicles in 1989.[216] This warranty rollback should encourage consumers to use independent repair shops. In addition, the more stringent I/M requirements[217] and the increase to $450 of the amount consumers must pay to repair a vehicle that fails an I/M test should lead to more use of automobile repair shops.[218] I/M tests will not play much of a role in the section 207(b) performance warranty under the 1990 CAAA because many vehicles will be past the two-year/24,000-mile period for general coverage before they are first tested. The I/M program will, however, cover the major specified emission control components that are covered for eight years or 80,000 miles.[219]

VI. Implementation Deadlines

For marginal or worse ozone non-attainment areas that were required to have an I/M program before the 1990 CAAA, the state had to submit a SIP revision immediately if the I/M program did not meet requirements imposed under the 1977 CAAA.[220] By November 15, 1991, the EPA was to update its I/M guidance.[221] Subsections (a)(2)(B)(i) and (b)(4) of section 182 of the CAA do not specify a date for the implementation of a

believed would not be anticompetitive; it developed the regulations in cooperation with the Automotive Liaison Council and the Automotive Products Emissions Committee. *See generally* Automotive Liaison Council, Voluntary Self-Certification Program for Certain Aftermarket Parts Parent Emissions—Related Standard for Aftermarket Parts and Systems (1978). The parts industry was opposed to any mandatory service-life interval for certified products. If a warranty is required the impact would be greatest on small manufacturers of inexpensive parts, who would be less able to investigate the validity of the warranty claims. Letter from Marc Fleischaker, Automotive Liaison Council, to David Feldman, U.S. Envtl. Protection Agency (Aug. 18, 1978) (on file with the *Harvard Journal on Legislation*).

[216] Clean Air Act § 207(b), 42 U.S.C. § 7541(b) (Supp. III 1991).

[217] *Id.* § 182(a)(2)(B), 42 U.S.C. § 7511a(a)(2)(B).

[218] *Auto Repair Industry Expected To Be Big Beneficiary of New CAA*, Clean Air Rep., Nov. 8, 1990, at 8.

[219] *Id.*

[220] Clean Air Act § 182(a)(2)(B), 42 U.S.C. § 7511a(a)(2)(B) (Supp. II 1990).

[221] *Id.* § 182(a)(2)(B)(ii), 42 U.S.C. § 7511a(a)(2)(B)(ii).

basic I/M program. There are only three states that will be required to develop I/M programs for the first time.[222] By November 15, 1992, all states must submit plans that include formal commitments to implement I/M programs.[223] By November 15, 1993, complete SIP revisions must be submitted. The basic I/M program is to be implemented for decentralized programs by January 1, 1994, and for centralized programs by July 1, 1994. If an area that is only required to have a basic program opts for an enhanced I/M program, it may receive additional time to comply.[224] For serious or worse ozone non-attainment areas, section 182(c)(3)(B) requires an enhanced I/M program to take effect by November 15, 1992.[225] The EPA recognizes that this date is impossible to meet.[226] The EPA plans to give the states one year from the initial SIP commitment submission to adopt the necessary statutory and regulatory authority for implementation. When this is done, the EPA will consider satisfied the statutory requirement that the program "take effect."[227]

Enhanced I/M programs must be fully implemented by January 1, 1995, but may be phased in by initially covering only thirty percent of the fleet that is required to be covered in 1999. There may also be a phase-in of cutpoints with full stringency to be reached by January 1, 1998.[228] The test and repair stations are to be phased out, with fifty percent of the fleet to be subject to test-only operations by January 1, 1995, and all vehicles subject to test-only operations by January 1, 1996. The EPA takes the position that under the authority of section 110(k)(4) it may grant conditional approval of I/M SIP submittals to states that commit to adopt and submit I/M regulations by November 15, 1993.[229] This conditional approval should prevent sanctions from applying.[230]

VII. Testing Under the 1992 I/M Regulations

The 1990 CAAA required the EPA to issue regulations for enhanced I/M programs in urban ozone non-attainment areas by

---

[222] 57 Fed. Reg. 52,950, 52,976 (1992) (to be codified at 40 C.F.R. pt. 51).
[223] *Id.* at 52,970.
[224] *Id.* at 52,971.
[225] Clean Air Act § 182(c)(3)(B), 42 U.S.C. 7511a(c)(3)(B) (Supp. II 1990).
[226] 57 Fed. Reg. 52,950, 52,971 (1992) (to be codified at 40 C.F.R. pt. 51).
[227] *Id.*
[228] *Id.*
[229] *Id.* at 52,976.
[230] *Id.* at 52,977.

November 15, 1992.[231] The EPA failed to meet the rulemaking deadline. The rule was finished in February, 1992, but the EPA claimed that the Office of Management and Budget review delayed its release.[232] After the Natural Resources Defense Council brought suit in federal district court, Judge Dennis R. Hurley ruled that the EPA must publish regulations by November 6, 1992. This deadline would have given the states nine days to develop a program and comply with regulations that were expected to run several hundred pages.[233]

On July 13, 1992, the EPA issued the proposed I/M rule.[234] Under this rule, the fifty-five urban areas that now have no I/M program will be required to implement one by July, 1993. For more seriously polluted areas, including much of the northeastern United States, enhanced I/M will be necessary.[235] The final rule was issued on November 5, 1992.[236] Under the CAA, enhanced I/M was to be implemented by November 15, 1992. Because the rule was late, the proposed regulations will not begin the program until July 1, 1994, and will not be fully operational until 1998.[237]

The enhanced program has several components. The EPA has developed an improved emissions test called IM240.[238] IM240 is based on the Federal Test Procedure, the method by which new vehicles are certified.[239] Prior to 1992, no I/M program tested evaporative losses, although these losses are a significant source of pollutants. The EPA developed two new tests that measure evaporative emissions. The Evaporative Systems Integrity Test uses a simple pressure test to determine if there are evaporative leaks in the fuel system. The second test is a test of the "purge" system, which uses a canister that stores evaporated fuel and routes it to the engine when it is operating. The EPA believes

---

[231] Clean Air Act § 182(c)(3), 42 U.S.C. § 7511a(c)(3) (Supp. II 1990).

[232] Natural Resources Defense Council v. EPA, 797 F. Supp. 194, 198 (E.D.N.Y. 1992).

[233] *Id.* at 194; *see also* Matthew L. Wald, *EPA Told to Publish New Air Pollution Rules Before Compliance Deadline*, N.Y. TIMES, July 3, 1992, at A10.

[234] 57 Fed. Reg. 31,058 (1992) (to be codified at 40 C.F.R. pt. 51) (proposed July 13, 1992).

[235] Keith Schneider, *EPA Is Proposing Strict New Testing of Auto Emissions*, N.Y. TIMES, July 14, 1992, at A1.

[236] 57 Fed. Reg. 52,950 (1992) (to be codified at 40 C.F.R. pt. 51).

[237] Schneider, *supra* note 235, at A1, A22.

[238] *See supra* note 160 and accompanying text.

[239] 57 Fed. Reg. 31,058, 31,061 (1992) (to be codified at 40 C.F.R. pt. 51) (proposed July 13, 1992).

that the purge system is best tested under operating conditions.[240]

The EPA has defined enhanced I/M to include annual testing of MY 1986 and later light-duty vehicles and trucks, and to require transient mass-emission testing using an IM240 driving cycle for MY 1986 and later vehicles. In addition, for MY 1983 and later vehicles, an evaporative system integrity test is required, and for MY 1986 and later vehicles, a transient evaporative system purge test is required. For MY 1984 and later vehicles, a visual inspection of the catalytic converter and the fuel inlet restrictor is also necessary.[241]

The proposed standard was based on a pre-1981 failure rate of 20%, a waiver rate of 1%, and a compliance rate of 98%.[242] In the final rule, the EPA relaxed the stringency requirements and required only a 96% compliance rate for the model program and a waiver rate of 3% of failed vehicles.[243] When a state submits its SIP, it must use the most current version of the EPA's mobile source emissions model to demonstrate that its program will achieve VOC, NO_x, and/or CO emission levels that are equal to or less than those that would be reached using the model program.[244] NO_x reductions may not be required if such reductions will not result in lower ozone levels. This requires an appropriate finding under section 182(f).[245]

The enhanced I/M program will require the use of $140,000 dynamometers, as opposed to the $15,000–$40,000 idle mode testing equipment currently in use.[246] This new equipment will test vehicles under operating conditions resulting in better tailpipe emissions measurements. The new equipment can also measure NO_x emissions and better identify vehicles requiring

---

[240] *Id.* at 31,062.

[241] *Id.* at 52,978.

[242] *Id.*

[243] *Id.* at 52,980.

[244] *Id.* at 52,978. The EPA is scheduled to release the final version of Mobile 5, its mobile source emissions computer model, in 1993. Mobile 5 was released for comment in August, 1992, and the EPA has since made changes in the final version. The EPA has also warned that those who used the old Mobile 4.1 model for their California clean car program may have to rerun their data on Mobile 5. *EPA: New Emissions Model Should Be Run Before Making 'California Car' Claims*, INSIDE EPA, Feb. 5, 1993, at 14.

[245] 57 Fed. Reg. 52,950, 52,979 (1992) (to be codified at 40 C.F.R. pt. 51).

[246] *Draft Final Inspection-And-Maintenance Rules Would Require Separate Testing, Repair System*, 23 Env't Rep. (BNA) 1586 (1992) [hereinafter *Draft Final I/M Rules*].

adjustments. Vehicles that fail will be required to have repairs up to the waiver amount.[247]

The EPA claims its "high-tech" I/M test is at least three times as effective as the best-designed and well-run existing programs. The EPA also claims it is so effective that tests need to be done only biennially, cutting costs and consumer inconvenience in half while only losing about three percent of the potential emission reduction.[248] The emission reductions from high-tech I/M are expected to be a major factor in meeting the overall twenty-four percent VOC reduction required for most polluted cities by the year 2000.[249]

The new rule does not mandate the use of centralized testing stations, but it encourages their use by prohibiting those who test vehicles from participating in repairs. This prohibition was adopted mainly in response to concern about quality control problems.[250] The EPA also considered whether lube shops, tire stores, and other automotive businesses that did not perform engine repairs should be allowed to provide test service. The agency concluded that objectionable conflicts of interest would still exist and, therefore, that these businesses should not be allowed to perform I/M tests.[251] Small inspection and repair shops vigorously fought this scheme, appealing to the Bush administration and claiming the program would have devastating economic effects.[252]

According to the EPA, the three major keys to an effective I/M program are: (1) use of technology able to fail or pass cars accurately; (2) quality control and enforcement to assure that testing is properly performed; and (3) skillful diagnosis and repair of failed cars to fix them properly.[253] The EPA claims most existing programs fail these requirements. Tests at idle or at low and high idle (2500 rpm) worked adequately for pre-1981 carbureted vehicles whose emission problems usually involved rich air/fuel mixtures. If these vehicles were brought into specifica-

---

[247] Clean Air Act § 182(c)(3)(C)(iii), 42 U.S.C. § 7511a(c)(3)(C)(iii) (Supp. II 1990); *see also supra* notes 168–176 and accompanying text.

[248] *See* 57 Fed. Reg. 31,058, 31,060 (1992) (to be codified at 40 C.F.R. pt. 51) (proposed July 13, 1992); *id.* at 52,952; *States, Environmentalists, Industry Clash over EPA's Proposed Car Tests*, CLEAN AIR REP., Aug. 27, 1992, at 12.

[249] 57 Fed. Reg. 52,950, 52,954 (1992) (to be codified at 40 C.F.R. pt. 51).

[250] *Draft Final I/M Rules*, *supra* note 246, at 1586.

[251] 57 Fed. Reg. 52,950, 52,975 (1992) (to be codified at 40 C.F.R. pt. 51).

[252] *Auto Shops Urge Penalties on Test, Repair Stations Be Dropped from CAA Rule*, INSIDE EPA, Oct. 23, 1992, at 11.

[253] 57 Fed. Reg. 52,950, 52,951 (1992) (to be codified at 40 C.F.R. pt. 51).

tions at idle they would usually perform properly at cruising speed.[254] The visual inspection of emission control devices would prevent tampering with the earlier emissions-related components. Visual inspections are less relevant now because of the difficulty of tampering with current high-tech emissions control systems.[255] Today vehicles with sensors and computers that continuously adjust engine operations need to be tested under load conditions during acceleration and deceleration to effectively test the vehicle's emissions control.[256] Tampering, which was a significant problem in the past, has become more difficult because the emission controls are an integral part of the engine and also because leaded gasoline is generally unavailable.[257]

Another weakness of existing tests is that they do not detect excessive evaporative emissions, which in today's vehicles generally exceed the HC emissions from the tailpipe.[258] Most existing tests also do not test NO_x emissions, which cannot be effectively measured at idle.[259] Other tests that are performed under simulated operating conditions can include NO_x measurement. Some I/M tests use a loaded, steady-state test that can test for NO_x. The new "high-tech" transient test that is part of the enhanced I/M requirements measures pollutants during acceleration and deceleration rather than at steady-state conditions and so is able to provide a very accurate picture of vehicle emissions.[260]

The program for basic I/M areas, after the 1990 CAAA, remains about the same as the program required by the 1977 amendments. The performance standard is based on the I/M program established by New Jersey in the early 1970s.[261] The basic I/M standard in the 1990 CAAA required the use of a computerized BAR 90 analyzer or a similar quality analyzer.[262] The 1990 basic I/M standard also required only a basic idle test, but the more sophisticated steady-state tests could be used.

---

[254] *Id.* at 52,953.
[255] *Id.* at 52,951.
[256] *Id.* at 52,953.
[257] Leaded gasoline destroys the effectiveness of the catalytic converter, allowing emissions to increase. It was a common form of tampering because leaded gasoline was less costly. NATIONAL ACADEMY OF SCIENCES, *supra* note 41, at 38; *see also Environmental Regulation of the Automobile*, 13 Env't Rep. (BNA) 9–11 (1982).
[258] 57 Fed. Reg. 52,950, 52,954 (1992) (to be codified at 40 C.F.R. pt. 51).
[259] *Id.* at 52,958.
[260] *Id.*
[261] *Id.* at 52,954.
[262] *Id.* at 52,968.

Basic I/M results in about a five percent reduction in highway mobile source emissions of VOCs, while enhanced I/M is claimed to produce a thirty percent reduction. The basic I/M program does not regulate NO_x, but there is a 1990 requirement that the overall NO_x will not increase unless it will not affect attaining the NAAQS.[263] For basic I/M programs, the final rule was similar to the proposed rule.

The enhanced I/M program is expected to yield a 28% reduction in VOCs, a 31% reduction in CO, and a 9% in NO_x from highway mobile sources.[264] This estimate is based on the use of the EPA's mobile source emission factor model, Mobile 4.1, although a new model is expected to be approved in 1993.[265] It should be noted that the 25% VOC reduction previously projected from basic I/M was a reduction in exhaust emissions (not total emissions) from light-duty vehicles.[266]

The CAA requires the EPA to establish a performance standard for enhanced I/M programs, but specifies only minimum requirements.[267] The regulations adopted by the EPA require testing of MY 1968 and later light-duty vehicles and trucks. The stringency of the I/M test depends on the model year of the vehicle. The transient exhaust emission IM240 test and a transient purge test are required on MY 1986 and later vehicles, and pressure testing is required for MY 1983 and later vehicles. Pre-1981 MY vehicles need only an idle exhaust test and MY 1981–1985 vehicles must have a two-speed idle exhaust test.[268] The cutoff points were selected to fail only vehicles with emissions over twice the design standards applicable to new motor vehicles.[269] Biennial tests are recommended.[270] The transient test will make visual inspection generally unnecessary except for older vehicles not subject to transient testing.[271]

Diesel engines have been exempt from I/M requirements under the EPA's regulations, although states could choose to regulate them. This exemption continues under the enhanced I/M requirements. New York, for example, might only require in-

---

[263] *Id.* at 52,954, 52,980.
[264] *Id.* at 52,954.
[265] *See supra* note 244.
[266] 57 Fed. Reg. 52,950, 52,955 (1992) (to be codified at 40 C.F.R. pt. 51).
[267] Clean Air Act § 182(c)(3)(B)–(C), 42 U.S.C. § 7511a(c)(3)(B)–(C) (Supp. II 1990).
[268] 57 Fed. Reg. 52,950, 52,956 (1992) (to be codified at 40 C.F.R. pt. 51).
[269] *Id.*
[270] *Id.* at 52,957.
[271] *Id.*

spections of gasoline-powered vehicles.[272] In California, however, by January 1, 1996, test procedures for diesel-powered vehicles over 8500 pounds gross vehicle weight rating ("GVWR") must be established as required by EPA regulations.[273] For basic I/M programs the control of such heavy-duty diesel vehicles is optional in California. The Bureau of Automotive Repair (the "BAR") is given authority to adopt such procedures. The BAR is not required to use a loaded-mode test procedure for diesels, but may use a procedure that focuses on excessive smoke. The I/M tests for diesel must be consistent with certification and in-use enforcement as well as the roadside smoke testing and underhood inspection programs applied in California to heavy-duty diesel vehicles.[274]

Remote sensing devices must be used in enhanced I/M areas to evaluate the in-use performance of at least 0.5% of the fleet that is subject to inspection each year. The EPA believes this technology is a useful supplement to enhanced I/M, but not enough information is available to allow a general credit for its use in a model for on-road testing.[275] The remote sensing technology is accurate for CO but is only crudely accurate for HC and has no NO_x testing capability. The EPA is prepared to work with states that are interested in establishing appropriate credit for its use. EPA regulations limit the mandatory use of remote sensing to 0.5% of the fleet or 20,000 vehicles a year, whichever is less.[276]

If there are EPA-ordered or voluntary emissions-related recalls, then the repairs involved must be completed either as part of the I/M inspection process or as part of the registration process.[277] Manufacturers will have to give the EPA the relevant information concerning recalls.[278]

---

[272] *See* NEW YORK I/M COMMITMENTS, *supra* note 133, at 4.

[273] *See* CALIFORNIA I/M REVIEW COMM., *supra* note 95, at 21.

[274] *Id.*

[275] *New Study Touts Cost-Effectiveness of On-Road Vehicle Emission Testing*, INSIDE EPA, Feb. 5, 1993, at 3. A new study also found that remote sensing technology is a cost-effective method of reducing mobile source emissions and is particularly effective at targeting high-emitting vehicles, as opposed to all vehicles under traditonal I/M. While the EPA recognizes the effectiveness of this technology for detecting HC and CO emissions, it points out that remote sensing still cannot test effectively for NO_x or evaporative emissions. *Id.* at 3–5.

[276] 57 Fed. Reg. 52,950, 52,957, 52,978 (1992) (to be codified at 40 C.F.R. pt. 51).

[277] *Id.* at 52,957–58.

[278] *Id.* at 52,958.

These requirements are minimum requirements. States are free to establish more stringent standards.[279]

VIII. I/M AND ONBOARD EMISSION DIAGNOSTIC SYSTEMS

Manufacturers of automobiles began to use onboard computer and memory systems in 1981 to monitor and control engine systems. These mechanisms can also be used to diagnose performance-related components and to monitor emission-control components as part of an I/M program.

California air pollution regulations required ECD systems and warning lights for emission-related functions beginning with 1988 MY automobiles. These regulations were significantly strengthened beginning with the 1994 MY vehicles. In addition to emission monitoring, these regulations also required monitoring of air conditioning systems for CFC coolant leaks as well as leakage of any other coolant identified as presenting a significant risk of damage to stratospheric ozone.

The 1990 CAAA requires further monitoring of the deterioration or malfunction of catalytic converters and oxygen sensors.[280] The proposed legislation also requires monitoring exhaust gas recirculation systems, evaporative emission control systems, auxiliary air systems, and fuel metering and ignition systems, but the monitoring requirement for these systems depends on regulations that were to be promulgated by May 15, 1992.[281] The law also authorizes regulations to standardize connectors to the ECD system, fault code identification, and computer access protocols. Such standardization would allow the use of a single, standardized off-vehicle diagnostic device. Since regulations adopted under section 202 can trigger EPA-initiated recalls under section 207(c), the Administrator will be able to recall classes or categories of vehicles with malfunctioning ECD systems. Even vehicles that meet the emissions standards can be recalled if the ECD system is not working properly.[282]

On September 24, 1991, the EPA published a notice of proposed rulemaking ("NPRM").[283] The proposal requires light-

---

[279] *Id.* at 52,978.

[280] Clean Air Act § 202(m)(1)(A), 42 U.S.C. § 7521 (Supp. III 1991).

[281] 56 Fed. Reg. 48,272 (1991) (to be codified at 40 C.F.R. § 86).

[282] *See* SENATE COMM. ON ENV'T & PUBLIC WORKS, REPORT ON CLEAN AIR ACT AMENDMENTS OF 1989, S. REP. NOS. 323–24, 101st Cong., 1st Sess. 98 (1989).

[283] 56 Fed. Reg. 48,272 (1991) (to be codified at 40 C.F.R. § 86).

duty vehicles and trucks to have systems that monitor the functioning of emission-control components and inform the vehicle operator of needed repairs. When a malfunction occurs, the diagnostic information must be stored in the vehicle's computer to assist the mechanic. The proposed rule also requires that the information necessary to repair the on-board diagnostic ("OBD") equipment and other emission-related vehicle components be available to the service and repair industry.

The NPRM would impose most of its requirements on MY 1994 and later light-duty vehicles and trucks. The OBD system would be required to track the performance of the catalyst and oxygen sensor, detect engine misfire, and monitor for electrical disconnect of any emission-related component that sends or receives information from the vehicle's on-board computer. The computer would have to be protected from tampering to ensure the availability of the information to independent repair facilities and other interested persons as well as to new vehicle facilities. Manufacturers would have to make available, at a reasonable cost, the information necessary to properly use the OBD system. The OBD is also expected to be checked as part of the I/M procedure so that failure codes will be discovered and repair required.

Unfortunately, the EPA was slow to issue final regulations dealing with diagnostic systems. The delay led to lawsuits by Representative Waxman and the Sierra Club to force the EPA to carry out its CAA responsibilities. On November 20, 1992, a consent order was entered mandating a schedule for the EPA to issue a number of rules.[284] This order contained an agreement to issue regulations by January 29, 1993, which would require automobile manufacturers to equip vehicles with diagnostic systems.[285]

IX. The Cost of I/M Programs

Representative Waxman is one of the major supporters of a more stringent I/M program. Waxman argues that stepped-up testing will remove VOC emissions at less than $3,500 per ton,

[284] Waxman v. EPA, No. 92-1320 (D. Cal. Nov. 20, 1992), *available in* WL 23 ER 1938; Sierra Club v. EPA, No. 92-1749 (D.D.C. Nov. 20, 1992), *available in* WL 23 ER 1938.

[285] *Waxman, Sierra Club Settle Lawsuits; Schedule Set for Issuing 19 Air Regulations*, 23 Env't Rep. (BNA) 1938 (Dec. 4, 1992).

which is well below the $5,000 per ton average cost of removing VOCs from stationary sources.[286] The EPA, for its part, estimates a cost of only $500 per ton of pollutant removed per vehicle in enhanced inspections, versus ten times that cost for comparable reductions from stationary sources.[287] The EPA also claims I/M is cost-effective even if no value is given to the CO and NO_x reductions obtained.[288]

The projected cost of enhanced I/M testing is one of the most important forces driving the program toward centralization. Current testing costs are about eighteen dollars per car in decentralized programs and about eight dollars per car in centralized programs. A high-tech test will cost about seventeen dollars per vehicle. The price increases because new state-of-the-art test equipment costs about $140,000, while existing equipment used for idle testing costs $15,000 to $40,000. Testing time for new tests averages ten to fifteen minutes versus about five minutes to perform present tests. The new tests also result in increased costs because fewer vehicles can be tested in a work day. The EPA claims these new tests save from six to thirteen percent of the fuel used due to the increased fuel economy of a properly tuned car.[289] The EPA also claims that with testing every other year the costs of the I/M program would be about nine dollars per vehicle per year.[290] The EPA predicts such a program would reduce emissions by twenty-eight percent.[291]

Yet the cost-effectiveness of high-tech versus current tests is still in question. The cost to repair a transient test failure that would also fail a 2500 rpm/idle test is estimated at $75; the repair of a vehicle that would fail only the transient test is estimated at $150. The overall average repair cost for transient failures is estimated to be $120. Pressure and purge test repairs are estimated to be $38 and $70, respectively. NO_x failures are esti-

---

[286] *Waxman Urges Reilly to Adopt Most Stringent I/M Test, Backing EPA Staff*, INSIDE EPA, Sept. 13, 1991, at 12.

[287] *Tough Vehicle Emissions Inspection Gets Nod from EPA Advisors but Not Without an Argument*, 22 Env't Rep. (BNA) 1646 (Nov. 1, 1991); *see also* 57 Fed. Reg. 52,950, 52,952 (1992). This $500 per ton estimate seems to be based on biennial testing.

[288] *Inspection & Maintenance—EPA Urges States to Adopt 'High Tech' Program*, ENVTL. POL. ALERT, Nov. 13, 1991, at 6.

[289] Inspection/Maintenance Program Requirements, 57 Fed. Reg. 52,950, 52,981 (1992) (to be codified at 40 C.F.R. § 51).

[290] *Id.*

[291] *Id.* at 52,950. Fuel economy improvements are projected to be 6.1% for repair of pressure test failures and 5.7% for repair of purge test failures. *Id.* at 52,981.

mated to cost about $100 to repair.[292] While warranties may shift some of these costs to the vehicle manufacturer,[293] the EPA's cost estimates appear unduly optimistic.[294] Other scenarios with different assumptions produce significantly higher costs.

Lowest cost scenario. Assume eight gpm of HC[295] is emitted from twenty percent of the vehicles, resulting in an eighty percent pass rate. The average U.S. VMT per vehicle per year is 9880 miles. According to the EPA's average cost estimate, the emissions from the failed vehicles could be corrected to zero emissions at a cost of $120 per vehicle. This scenario uses a very high emissions level and an unrealistic level of reduction to produce the most optimistic, cost-effective picture of the program. Working through the cost analysis, ten tests at $17 per vehicle cost $170. Repair costs for the two failed vehicles are $120 multiplied by two, or $240. Total costs are $240 plus $170, or $410. The two repaired vehicles result in a total sixteen gpm improvement, which at 9880 miles per year over two years (16 x 9880 x 2) comes to 316,160 grams of HC reduction. This amount equals 696 lbs or 0.35 tons. The cost per ton of HC reduction then comes to $410/.35 ton or $1,171 per ton. This cost projection is below the $1,600 per ton for HC reduction that the EPA estimates.[296]

More realistic scenario. Assume two out of every ten cars tested fail the HC test at four gpm (approximately the 1970 standard) and cost $150 each to repair to 0.5 gpm. Miles driven per year in urban non-attainment areas average fewer miles per year than the national average, so assume 8000 miles as a rough estimate of urban VMT per year. Benefits of repair extend for two years. Seven gpm multiplied by 8000 miles equals 56,000 grams multiplied by two years, which equals 112,000 grams or 246 lbs. This 246 lbs. or 0.123 tons of HC reduction, at a cost of $170 plus $300 equals $470, works out to a cost of $3,821/ton of HC reduction, which is well above the EPA estimate.

---

[292] *Id.*

[293] *Id.*

[294] In discussing issues that affect the future of I/M programs, the General Accounting Office in 1986 recognized that cost effectiveness is a potential weakness: "The cost of I/M programs is expected to be substantial, but the benefits of such programs are at best inconclusive and at worst may not justify the cost of the program implementation." I/M PROGRAM BEHIND SCHEDULE, *supra* note 50, at 33.

[295] HC is usually considered VOC for mobile sources by regulatory agencies, although VOCs only constitute 85% of HCs.

[296] Inspection/Maintenance Program Requirements, 57 Fed. Reg. 52,950, 52,952 (1992) (to be codified at 40 C.F.R. § 51).

Still more realistic scenario. Yet costs may be even higher. If the EPA's estimated repair costs are too optimistic and if the value of citizens' time is included, then the scenario that emerges is closer in cost to traditional stationary source control. For example, two cars out of every ten tested fail at three gpm, are corrected to 0.5 gpm, cost $250 each to repair and require 0.75 hours of consumers' time (valued at twenty dollars per hour for each ten vehicles).[297]

At five grams per mile multiplied by 8000 miles per year there is a 40,000-gram reduction. After two years the reduction is 80,000 grams. Adding $170 to $500 and $150 (for lost time) comes to $820 to reduce 0.088 tons of HC, or a cost of $9,318/ton of HC reduction.[298]

As for VOCs, the EPA estimates that within a high-tech I/M program VOC reduction will cost $880 per ton. This figure is based on the assumption that all program costs are allocated to VOC reduction, that the test is performed biennially, and that no value is placed on the time lost by the vehicle owner. The EPA therefore sees the high-tech result on VOCs as much more cost-effective than the present $5,400 per ton of VOC reduction from a basic I/M program.[299] If the time of a consumer is valued at $20 per hour for 0.75 hours, then the $880 per ton for VOC removal increases to $2,000 per ton.[300]

The EPA's figure for repair costs is further reduced by the EPA's claim that owners get the money back in gas mileage improvement. A repair of a twenty-mpg car to obtain a ten percent fuel improvement saves about sixty dollars per year if the car is driven the national average number of miles, using gasoline priced at $1.20 per gallon. Thus a hypothetical $120 repair would require about two years of fuel savings to pay for itself.

---

[297] These time and dollar values are consistent with assumptions by the EPA that are considered appropriate, and that leads to an EPA estimate of $1,600 per ton of HC reduced. *Id.* However, a California study estimated the time required at 76 minutes. CALIFORNIA I/M REVIEW COMM., *supra* note 95, at 26.

[298] This is consistent with a Resources for the Future study that estimated the cost of removing VOCs through I/M to be $9,000 per ton per year. *Vehicle Inspection-and-Maintenance Programs Cost More, Help Less Than Estimated, Group Says*, 22 Env't Rep. (BNA) No. 45, at 2472 (Mar. 6, 1992).

[299] Inspection/Maintenance Program Requirements, 57 Fed. Reg. 52,950, 52,984 (1992) (to be codified at 40 C.F.R. § 51).

[300] The EPA says this is still significantly lower than the costs of other strategies that could be used to control air pollution. *Id.*

If the three scenarios on HC reduction are recalculated with the EPA's assumption that the repair costs are offset by $120 in gas savings per repaired vehicle over two years, then the cost per ton of HC reduction becomes even more favorable. Under the lowest cost scenario, the cost per ton of HC reduction drops to $486 per ton. The more realistic scenario, with higher repair costs, would result in a cost per ton of $1,870. The still more realistic scenario, with its assumed repair costs of $250, would have a cost of $6,591 per ton of HC reduction.

Whether repairs would really average $120 per vehicle failing a transient high-tech test, as the EPA claims, or whether repair costs would be higher will be unknown until the program is implemented and credible data is collected. Whatever the case may be, under all these scenarios an additional benefit will be CO and NO_x reductions. Fuel economy benefits are projected at $825 million with $617 million in benefits attributable to tailpipe emissions tests and $208 million in benefits from the evaporation tests.

An estimated 64 million vehicles are subject to I/M inspections in the United States. The EPA estimates the cost of these inspections to run an estimated $747 million a year and the costs of the required repairs to be an estimated $392 million. The costs, however, are reduced by fuel economy benefits estimated at $245 million. By the year 2000, the inspection costs for enhanced I/M programs are projected to increase to $451 million for a biennial program, with repair costs rising to $710 million. For basic I/M, program costs for that year are projected at $162 million and repair costs at $113 million. Overall, costs are expected to decline by the year 2000 because of the enhanced I/M program, the change to biennial testing, and the projected increase in fuel economy benefits.[301]

In 1992 there were approximately 11,000 jobs associated with the I/M program. About 9100 jobs were in the inspection field, and about 2300 jobs were to perform necessary repairs. Under the new 1990 CAAA enhanced I/M program, with centralized inspections run by contractors, the EPA estimates about 2700 inspectors will be needed; at least the same number of inspectors will be needed for the basic I/M program, along with about 6200 repair technicians. Thus the CAAA should create an additional 3900 repair technician jobs, with possible gains in inspector jobs

[301] *Id.* at 52,982.

as well.[302] Multiple independent test facilities might produce a more significant increase in the number of new jobs.

X. Conclusion

The problems plaguing present I/M programs include excessive waivers, motorist noncompliance, inadequate quality control, outdated test procedures, insufficient enforcement against inspectors who violate the I/M regulations, inadequate data collection and analysis, inadequate resources, and improper testing.[303] The EPA claims that decentralized programs, with rates of fuel switching and tampering from twenty to fifty percent higher, are not as effective as centralized programs.[304] When errors are found in I/M programs, it is much easier to correct the problem in a centralized system. Problems also accompany the oversight and management of test procedures in some I/M programs. Some inspectors fail to perform inspections properly even when they know they are being audited. Covert audits have discovered improper testing fifty percent of the time in test and repair stations.[305] Even in California, which has the most intensive management of any decentralized program in the country, management problems persist.

The lack of funds for management and oversight has been a problem for I/M programs, especially if decentralized. Centralized programs spend about one to two dollars per vehicle on oversight costs while decentralized systems spend up to six dollars per vehicle. California is spending six to seven dollars per vehicle to address problems in operating their program.[306] Decentralized systems thus have a high hidden cost that reduces their ability to control pollution.

Furthermore, in decentralized systems analyzers frequently malfunction due to leaks or calibration failures, problems that rarely occur in centralized systems.[307] The EPA believes that separating testing from repair operations can improve the operation of decentralized facilities. However, technology such as the BAR 90 system, more advanced than idle tests yet less

---

[302] *Id.* at 52,984.
[303] *Id.* at 52,967.
[304] *Id.*
[305] *Id.* at 52,951.
[306] *Id.* at 52,969.
[307] *Id.*

advanced than the high-tech test, is insufficient to ensure acceptable performance of I/M tests by inspection facilities.

The 1992 regulations set out new requirements for both basic and enhanced I/M programs.[308] To reduce cheating, the enhanced I/M programs require real-time data transfer to those who monitor the I/M program. The regulations also require that computerized BAR 90 (or comparable) analyzers, designed to reduce the opportunity for cheating by the technician performing the test, be used in basic I/M programs.[309]

The changes in test instruments and procedures as well as the changes in motor vehicle on-board diagnostics can be expected to reduce cheating and thus improve the overall effectiveness of the I/M program. The move to centralized, test-only facilities should also help. However, even if the efficacy of I/M programs increases, failed vehicles must still be properly repaired if emissions are to be reduced. Improvements in this aspect of the program will be more difficult to achieve.

Problems also exist with the new enhanced I/M program. The EPA asserts that the key to an effective I/M program is accurate identification of non-complying vehicles. To accomplish this goal, the EPA has developed and placed great reliance upon the IM240 test. However, the effectiveness of this test has been questioned.[310] Furthermore, since enhanced I/M requires centralized testing, motorists may be forced to have their vehicles tested in two separate locations: once at the centralized facility and once at a separate repair facility. Obtaining consistent test results will be a critical challenge to the program's effectiveness.

The EPA conducted sample IM240 tests at two separate facilities in Indiana and obtained inconsistent results. Sixty-four vehicles, MY 1986 or later, failed the test at the first site and were retested at the second site, where eighteen of the sixty-four vehicles (twenty-eight percent) passed the IM240 test, though none had been repaired.[311]

During the first test, one of these eighteen vehicles was five times over the 0.8 gpm HC standard. During the same test at the second facility, the vehicle passed with a 0.4 gpm reading.

---

[308] *Id.* at 52,968.

[309] *Id.*

[310] U.S. GEN. ACCOUNTING OFFICE, UNRESOLVED ISSUES MAY HAMPER SUCCESS OF EPA'S PROPOSED EMISSIONS PROGRAM (1992); *see also supra* notes 286–302 and accompanying text.

[311] U.S. GEN. ACCOUNTING OFFICE, *supra* note 310, at 5.

Another vehicle failed the fifteen-gpm CO standard by emitting 22.2 gpm at the first test. At the second facility the same vehicle passed the CO test with a reading of 2.9 gpm.[312] The differences in the other sixteen vehicles were less dramatic but still affected their compliance status.[313]

The EPA argued that the variations were due to changes in the vehicles and the ambient air during the time between the two test procedures.[314] Preconditioning the vehicles at the second testing facility to simulate conditions at the first facility did not solve the problem.[315]

The EPA has also considered correcting the variability problem by providing two ways for vehicles to pass the IM240 test: meeting the emission standard during the entire 240-second test procedure or during the last 150 seconds. However, data available on eleven vehicles indicated that even with this revised test, six of them would still have failed, then passed, the IM240 test. Consequently, this alternative testing method does not assure consistency.[316]

Since IM240 is the centerpiece of enhanced I/M, the program's success depends on the test's accuracy and consistency. Although the problems associated with the test's accuracy raise concerns, other problems merit attention as well.

The IM240 test is more complex than earlier I/M tests. Its complexity may affect how failed vehicles are repaired in several ways: (1) it may be more difficult to diagnose problems in vehicles that only marginally exceed emission limits; (2) there is a shortage of mechanics trained to diagnose and repair marginal failures of high-tech vehicles; and (3) repair shops probably will not be able to afford the expensive IM240 equipment and therefore may be unable to confirm the effectiveness of their repairs.[317] This fault will lead to the ping-pong effect, discussed earlier, that the EPA has tried to minimize.

The EPA has also been criticized for promoting the IM240 test without completing studies on possible alternatives.[318] The EPA will give states until November, 1993, to commit to a specific test procedure. However, if the EPA's analysis of the

---

[312] *Id.* at 5–6.
[313] *Id.* at 6.
[314] *Id.*
[315] *Id.* at 6–7.
[316] *Id.* at 6.
[317] *Id.* at 7.
[318] *Id.* at 11.

alternatives is not complete by then, states may be forced to choose a test procedure before all the information is available.[319] The General Accounting Office believes there may be alternative procedures that cost less yet provide similar emissions control benefits.[320]

If these issues are not addressed and resolved, the result could be an erosion of public confidence and acceptance of the I/M program. Such a result could hamper the program and block the emissions reductions envisioned by the 1990 CAAA.[321]

The EPA also considers it important to improve the effectiveness of repairs in order to realize the goals of the I/M program. States are therefore required to provide technical assistance to repair and monitor facilities. The past fifteen years of I/M experience have shown that market mechanisms cannot ensure good training or an adequate number of skilled technicians. The EPA expects to establish national examples and guidelines but considers the states responsible for ensuring that adequate training is provided.[322] However, given the broad problems of obtaining competent and honest automotive repairs at a fair price, it is unlikely that the EPA will achieve major improvements in vehicle repairs with their limited efforts at training mechanics.

The 1992 CAAA may improve the I/M program, but at a price. The EPA has attempted to show that the I/M program can be made more effective at little if any additional cost. As has been previously discussed, the EPA's cost/benefit projections may be overly optimistic. Many of the EPA's assertions are suspect. For example, it is difficult to believe that vehicles failing enhanced I/M tests will be repaired for $120 on average or that the public will be able to meet the I/M requirements, including the repairs and retesting for failed vehicles, in forty-five minutes on average. It is also unlikely that the existing decentralized systems can be phased out without injuring many small business owners or that a testing program of significantly increased stringency can be developed without a corresponding increase in inconvenience to the public. Despite the EPA's claims to the contrary, the ping-pong effect between different testing centers is likely to be significant.

---

[319] *Id.* at 14.
[320] *Id.*
[321] *Id.*
[322] *Id.* at 9.

However, these problems with I/M pale by comparison with the larger issues. After more than twenty years of effort, the federal and state governments have yet to clean up the air in urban areas. In many urban areas where motor vehicle emissions dominate the air pollution problem, there are no stationary sources left to regulate; further air quality improvements must come from emission control. Even where stationary sources can be further controlled, I/M may still be necessary to meet SIP requirements. Thus, whether I/M is significantly lower in cost than other procedures is unimportant. It is one of the few measures left to air quality managers that holds any promise of demonstrable air quality improvements.

Because mobile sources have been regulated more effectively than other sources, the proportion of emissions from transportation sources has been declining.[323] The improvements in emissions from mobile sources occurred during years when VMT were increasing. Passenger car emission standards did not change from MY 1981 until MY 1994, but VMT continued to increase. In 1950, VMT were less than 500 billion. By the time the 1970 CAA was enacted VMT had grown to 1.1 trillion; by 1977 VMT was about 1.5 trillion; by 1990 VMT was about 2 trillion.[324] This increase in motor vehicle use has essentially canceled the reductions in air pollution due to emission controls. In addition, the size of the vehicle fleet has grown, curtailing the overall effectiveness of evaporative controls. In 1970 there were 89.2 million automobiles in use in the United States; in 1987 there were 139 million.[325] Fossil fuels used for transportation went from 8.38 quadrillion Btu ("quads") in 1950 to 16.04 quads in 1970, increased to 19.77 quads in 1977, and are projected to have grown to 22.36 quads in 1990.[326] Even improved

---

[323] Emissions of CO from transportation sources decreased during the 1981–1990 period from 55.4 million tons ("MMTs") to 37.6 MMTs while VMT increased by 37%. Total CO emissions dropped from 77.5 MMTs to 60.1 MMTs. NO_x emissions from transportation decreased from 9.9 MMTs in 1981 to 7.5 MMTs in 1991, but total NO_x in the same period decreased from 20.9 MMTs to 19.6 MMTs. VOCs from transportation decreased from 8.9 MMTs to 6.4 MMTs between 1981 and 1990, while total VOCs dropped from 21.3 MMTs to 18.7 MMTs. U.S. ENVTL. PROTECTION AGENCY, NATIONAL AIR QUALITY AND EMISSIONS TRENDS REPORT, 1990, at 3-18, -22, -27 (1991) [hereinafter 1990 EPA REPORT].

[324] U.S. DEP'T OF TRANSP., FED. HIGHWAY ADMIN., HIGHWAY STATISTICS 1989, at 181, *cited in* OAK RIDGE NAT'L LAB., TRANSPORTATION ENERGY DATA BOOK 3–5 (12th ed. 1992).

[325] COUNCIL ON ENVTL. QUALITY, ENVIRONMENTAL QUALITY, TWENTIETH ANNUAL REPORT 452 (20th ed. 1989).

[326] *Id.* at 451.

automotive fuel efficiency over the past decade has not offset the increase in fuel consumption due to increases in VMT. Today motor vehicles emit from sixty to eighty percent (depending on the pollutant) less pollution per vehicle-mile traveled than the vehicles of the 1960s, but motor vehicles still account for about forty percent of the VOC and NO_x emissions[327] and about two-thirds of the CO emitted in the United States.[328]

Despite the stringent controls on new vehicles, the average in-use emissions and evaporative losses have been higher than the applicable standards, even where vehicles are tested under the specific conditions of the FTP.[329] Controls on some categories of light-duty trucks and on heavy-duty trucks have been more lenient than for automobiles. Heavy-duty diesel trucks have been lightly controlled, with NO_x not being subject to significant regulation until the 1990 CAAA.[330]

VMT can be expected to increase further, but it will be difficult to achieve significant additional reductions in emissions. Thus it is unlikely that the reductions in transportation-related pollutants achieved since the mid-1970s will be repeated while VMT keep increasing. As stationary sources come under more stringent control from the 1990 CAAA, the proportion of emissions from mobile sources will probably reverse its historic pattern and also increase.

To deal with the diminishing benefits of more stringent emission controls on new vehicles, the 1990 CAAA has imposed more stringent requirements on in-use vehicles, primarily through enhanced I/M. It would be more effective to reduce VMT, reduce the total number of vehicles (even parked vehicles emit HC), and increase the fuel efficiency of vehicles while maintaining the effectiveness of pollution controls. Such an approach would require more governmental action to control land use and energy consumption. To date, only limited efforts have been made in this direction. Although the 1990 CAAA include some new programs to develop low-polluting vehicles[331] and to improve motor vehicle fuels,[332] air pollution control efforts continue to center on emission limitations on new vehicles and on

---

[327] NATIONAL RESEARCH COUNCIL, *supra* note 1, at 284.
[328] 1990 EPA REPORT, *supra* note 323, at 3-18, -22, -27.
[329] Michael P. Walsh, *Motor Vehicles and Fuels: The Problem*, 17 EPA J. 12 (1991).
[330] Clean Air Act § 202(a)(3)(B), 42 U.S.C. § 7521(a)(3)(B) (Supp. III 1991).
[331] *Id.* §§ 241–250, 42 U.S.C. §§ 7581–7590.
[332] *Id.* § 211, 42 U.S.C. § 7545.

the I/M requirements for in-use vehicles that are the subject of this Article. These efforts will help protect the environment, but an expanding world population of both people and motor vehicles will eventually require either a new technology for transportation, a rationing of vehicle/fuel use, a reduction in demand for transportation due to better communication or land use technology, or some mix of these approaches. An improved I/M program buys some additional time in which to develop a more effective program. But until the nation is willing to deal with the root causes of air pollution—population and consumption—we can expect to see more complex regulatory programs mandated, programs that are increasingly intrusive and increasingly less effective at the margin. I/M is an example of such intrusive regulation[333] and is only modestly effective; yet it may be one of the only tools available that can be made more effective at a politically and economically acceptable cost.

[333] New York state expects to create a database for each vehicle subject to enhanced I/M that will include inspection date and location, emissions readings, safety items, type of repair, repair cost, repair shop number, reinspection information, etc. The inspection record database will be available to DMV and DEC. *See* NEW YORK I/M COMMITMENTS, *supra* note 133, at 8.

STATUTE

EMERGENCY MEDICAL CARE AND INJURY/ILLNESS PREVENTION SYSTEMS FOR CHILDREN

R. CHRISTOPHER BARDEN
ROBERT KINSCHERFF
WILLIAM GEORGE III
RICHARD FLYER
JAMES S. SEIDEL
DEBRA PARKMAN HENDERSON
HARVARD LAW SCHOOL LAW AND MEDICINE SOCIETY*

Each year, thousands of children suffer accidental deaths because they do not receive appropriate emergency care. The authors of this Article present evidence that a concerted effort by state and federal legislators to implement emergency medical care systems and injury prevention programs would greatly reduce these needless deaths.

---

* R. Christopher Barden, Member, Lindquist & Vennum, Minneapolis, Minn; B.A., University of Minnesota, 1976; Ph.D., University of Minnesota, 1982; J.D., Harvard Law School, 1992. Dr. Barden is also a licensed psychologist and serves as an Assistant Research Professor of Medicine at the University of Utah School of Medicine as well as an Adjunct Professor of Law at the Hamline University Law School. Robert Kinscherff, Forensic Psychologist, Law and Psychiatry Service of Massachusetts General Hospital, Harvard Medical School; B.A., University of California, Berkeley, 1977; M.A., University of Chicago, 1980; M.A., City University of New York, 1983; Ph.D., City University of New York, 1988; J.D., Harvard Law School, 1992. William George, Member, Ropes & Gray, Boston, Mass.; B.A., Brigham Young University, 1988; J.D., Harvard Law School, 1992. Richard Flyer is a practicing pediatrician and advocate for children's health care in New Jersey; B.A., Yale University, 1972; M.D., Harvard Medical School, 1976. James S. Seidel, Professor of Pediatrics, University of California, Los Angeles School of Medicine; Chief, General and Emergency Pediatrics at Harbor-U.C.L.A. Medical Center; B.S., Michigan State University, 1964; M.S.P.H., University of North Carolina, 1967; M.D., University of California, Los Angeles, 1973; Ph.D., University of California, Los Angeles, 1976. Debra Parkman Henderson, Administrator, EMSC Project of the Harbor-UCLA Medical Center; A.A., Pasadena City College, 1978; B.A., University of San Francisco, 1983; M.A., Pacific Oaks College, 1987. Ms. Henderson also serves as Co-Director of the National Emergency Medical Services for Children Resource Alliance.

This work was supported in part by a Faculty Scholar Award (1987–1992) from the W.T. Grant Foundation to Dr. Barden. Thanks are due to Judge Stephen Breyer (Chief Judge, United States First Circuit Court of Appeals), Richard Zeckhauser (Frank Plumpton Ramsey Professor of Political Economy, JFK School of Government, Harvard University), Dr. Troyen Brennan (Professor, Harvard Law School and Harvard Medical School), Jerome Kagan (Professor and Chairman, Department of Psychology, Harvard University), Robert Mnookin (Professor, Harvard Law School), Jim Ramstad (United States Congressman, 3d District-Minnesota), the Legislature and Governor of the State of New Jersey, and especially Dr. Alan A. Stone (Touroff-Glueck Professor of Law and Psychiatry, Harvard Law School and Harvard Medical School). Many thanks are also due to Karen Campbell, Christina Slawson, Samir Jain, Alan Romero, and Tara Isa Koslov.

The authors argue that while government regulation of health and safety risks often produces little real benefit, their proposals for improvement have been demonstrated to save children's lives at relatively little cost. They also offer a model act to implement their proposals.

Most researchers agree that between 6000 and 10,000 of the 22,000 American children who die from accidents each year could be saved by inexpensive injury prevention programs and emergency medical systems for children ("EMSC"). Although the leaders of both major political parties agree that children are our most precious resource, little attention has been given to their emergency health care needs.

Saving children's lives is an objective that is both economically and morally compelling. Experts have documented the "hidden taxation" of unnecessary childhood injuries and deaths, estimating direct and indirect lifetime costs of injuries to children fourteen years old and under at over $13.8 billion. If ages fifteen through twenty-four are included, this estimate balloons to $53 billion.[1] Billions of dollars in medical and disability expenses, enormous losses in future wages and taxes, and much pain and suffering could be prevented with low-cost changes to the present EMSC and injury prevention programs. The empirically demonstrable need for improved EMSC contrasts sharply with many governmental health and safety regulations that consume vast sums of public resources with little verified benefit.

The vast majority of Americans view children as innocent, needy, and deserving of protection. Their problems are not the result of antisocial or negligent behavior, like taking illegal drugs, or deliberate refusal to behave safely, such as not using seat belts. While American values emphasize individual choice, character, and free will in determining an adult's quality of life, injured children, within the innocence of their tender years, are almost always viewed as casualties of the actions of others.

Emergency medical services in the United States are largely the result of systems developed for adults during wartime. Tragically, many responsible emergency department personnel are simply unaware of and untrained in basic pediatric emergency medicine. Inadequacies in emergency medical care and injury/illness prevention programs have put American children at far higher risk of death following traumatic accidents than children in other countries with similar cultures and histories. Standard

[1] *See infra* note 64 and accompanying text.

be used to create economic incentives for the medical profession to improve current emergency care. The need for ongoing post-implementation assessment of the proposals is discussed in Part VII. Finally, we provide model legislation that addresses the problems identified by the Article.

I. THE NEED FOR EMSC

A. *Childhood Deaths from Injury*

> *"If some infectious disease came along that affected one out of every four children in the United States, there would be a huge public outcry and we would be told to spare no expense to find the cure—and to be quick about it."*
>
> —Former Surgeon General C. Everett Koop[3]

Injury is the single most important public health problem for American children, "however it is measured—number of deaths, dollar costs for treatment, or relative rankings with other health problems."[4] Injuries to children occur in epidemic numbers throughout this country. The number of injury deaths per capita for children and adolescents under nineteen years of age is strikingly higher in the United States than in Canada, Norway, France, England, or the Netherlands (Figure 1).[5] In 1986, more than 22,000 children and adolescents, ages nineteen and under, died from injuries in the United States.[6] Injury deaths among children continue to surpass all major disease groups as the leading cause of years of life lost, and injuries are also the leading cause of child disability (Figure 2).[7] It is estimated that

---

[3] *Childhood Accidents and Injuries: Hearing Before the Subcomm. on Children, Family, Drugs, and Alcoholism of the Senate Comm. on Labor and Human Resources*, 101st Cong., 1st Sess. 6 (1989) (statement of C. Everett Koop, Surgeon General).

[4] CHILDREN'S SAFETY NETWORK, NATIONAL CTR. FOR EDUC. IN MATERNAL & CHILD HEALTH, A DATA BOOK OF CHILD AND ADOLESCENT INJURY vii (1991).

[5] Bret C. Williams & Jonathan B. Kotch, *Excess Injury Mortality Among Children in the United States: Comparison of Recent International Statistics*, 86 PEDIATRICS 1067, 1068 (1990).

[6] *See* Juan G. Rodriguez & Stuart T. Brown, *Childhood Injuries in the United States*, 144 AM. J. DISEASES CHILDREN 627 (1990).

[7] Injuries account for approximately 40% of deaths among children ages 1 to 4 and 70% of deaths of children and adolescents ages 5 through 19. *See* NATIONAL CTR. FOR EDUC. IN MATERNAL & CHILD HEALTH, EMERGENCY MEDICAL SERVICES FOR CHILDREN, A REPORT TO THE NATION 84 (James S. Seidel & Debra P. Henderson eds., 1991) [hereinafter EMSC REPORT]; *see also* CENTERS FOR DISEASE CONTROL, U.S. DEP'T OF HEALTH & HUMAN SERVS., CHILDHOOD INJURIES IN THE UNITED STATES: A REPORT TO CONGRESS (1989) (providing information about the causes, effects, and preventability of childhood injury).

emergency medical systems have reduced mortality and morbidity in adults, but not in children. Similarly, trauma centers have significantly reduced mortality and morbidity for adults, but have not improved the chances for children. For example, detailed studies of trauma victims in Los Angeles indicate that the death rate for child trauma victims is almost twice that for adults with injuries of similar severity.[2]

Correcting the inadequacies of emergency medical care for children will require cooperative efforts between medical providers, the medical marketplace, and government. A number of specific and effective societal responses to this problem are possible, including the setting of standards by business, professional, and governmental organizations, mandating disclosure of critical information to the public, increasing social awareness via education, and clearly delineating costs and benefits via ongoing empirical research programs. In sum, EMSC and injury prevention programs are the most cost-effective methods currently available for improving the health and future productivity of American children while reducing overall health care expenditures.

Despite the pressing need to synthesize divergent analytic methods, few sources provide integrated analyses of current social problems. As a result, legislative and administrative efforts to ameliorate intricate social problems too frequently rely upon opinion, political theory, anecdotal evidence, and rhetoric instead of methodologically sophisticated scientific data and analyses. This discussion seeks to integrate legal, economic, philosophical, epidemiological, and medical analyses of a pressing health and safety problem—inadequate EMSC and injury prevention programs in our society. The analysis will highlight philosophical and methodological assumptions and identify strengths and weaknesses of a number of analytic approaches.

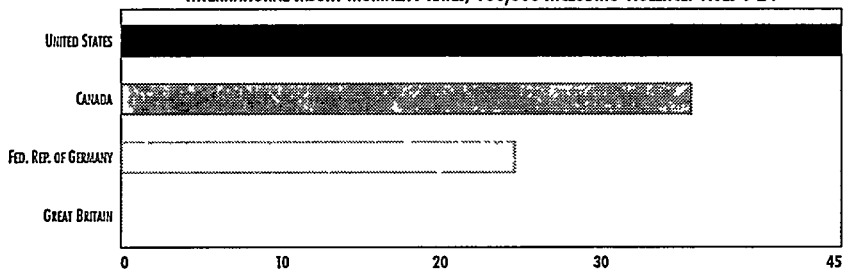
Parts I and II of this Article demonstrate the importance of establishing emergency medical care systems and injury prevention programs that specifically address the unique needs of children. In order to realize these goals, Part III offers specific proposals for reform. An illustration of the cost-effectiveness of the proposals and potential sources of funding are then outlined in Parts IV and V. Part VI suggests ways the legal system might

---

[2] *See, e.g.*, James S. Seidel et al., *Emergency Medical Services and the Pediatric Patient: Are the Needs Being Met?*, 73 PEDIATRICS 769 (1984).

FIGURE 1

INTERNATIONAL INJURY MORTALITY RATES/100,000 INCLUDING VIOLENCE: AGES 0-24

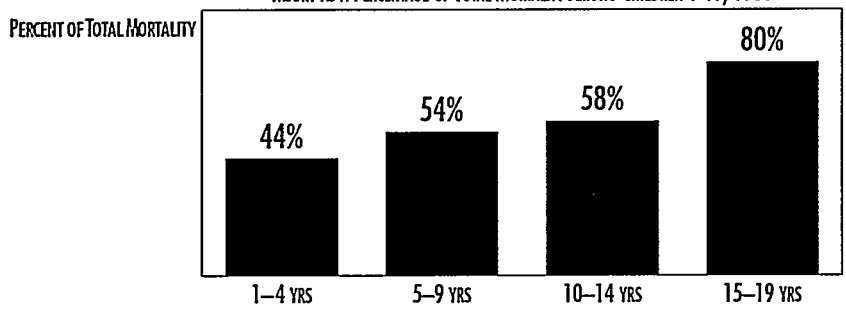


MORTALITY RATE/100,000

SOURCE: WORLD HEALTH ORGANIZATION, 1990 WORLD HEALTH STATISTICS ANNUAL (1991).

FIGURE 2

INJURY AS A PERCENTAGE OF TOTAL MORTALITY AMONG CHILDREN 1-19, 1988



SOURCE: LOIS A. FINGERHUT & JOEL C. KLEINMAN, U.S. DEP'T OF HEALTH & HUMAN SERVS., TRENDS AND CURRENT STATUS IN CHILDHOOD MORTALITY: UNITED STATES, 1980-85, 20 TBL. M (1989).

each year more than 80,000 American children suffer permanent disabilities from injuries, 600,000 children are hospitalized because of injuries, and 16 million American children are treated in emergency departments. Injuries result in more days of hospital care than any childhood disease, cause the highest proportion of discharges to rehabilitation and long-term care facilities, and result in the greatest need for home health care services after being discharged from a hospital.[8] Injured American chil-

[8] Mark L. Rosenberg et al., *Childhood Injuries: Where We Are*, 86 PEDIATRICS 1084, 1084 (1990).

dren spend more than 10 million days in bed annually, often as the result of preventable injuries.[9] To reduce the number of child and adolescent deaths from injury, EMSC and injury prevention programs are essential.

Historically, there has been a consistent and dramatic reduction in child mortality from natural causes such as diseases (including congenital anomalies), while mortality due to injury has steadily increased. For children ages one to nineteen, injuries far exceed cancer and congenital anomalies as the leading cause of death.[10]

Perhaps the most serious long-term effects of childhood injuries result from traumatic brain injuries which lead to disabling conditions that are often preventable and frequently require rehabilitation services. Brain injuries are the most common childhood injuries, resulting from falls, motor vehicle accidents, sports, and assaults.[11] Almost thirty percent of all childhood injury deaths result from head injuries, and over 150,000 children suffered traumatic brain injuries in 1986. Nearly 7000 of these children died from such injuries. The long-term emotional and financial costs to the survivors are great. Children occupy hospital beds for more than 550,000 days per year due to traumatic brain injuries,[12] with hospital costs alone exceeding $1 billion per year. Children with brain injuries often suffer from long-term residual effects,[13] especially because intellectual impairment frequently persists despite a child's recovery of basic motor and sensory abilities.[14] Common sequelae include decrements in I.Q. test scores, school failure, behavioral problems,

---

[9] EMSC REPORT, *supra* note 7, at 64.

[10] LOIS A. FINGERHUT & JOEL C. KLEINMAN, U.S. DEP'T OF HEALTH & HUMAN SERVS., TRENDS AND CURRENT STATUS IN CHILDHOOD MORTALITY: UNITED STATES, 1900–85 (1989).

[11] *See generally id.* at 65.

[12] Jess F. Kraus et al., *The Causes, Impact, and Preventability of Childhood Injuries in the United States: Brain Injuries Among Infants, Children, Adolescents, and Young Adults in the United States*, 144 AM. J. DISEASES CHILDREN 684, 689 (1990).

[13] *See* CYNTHIA J. WRIGHT, CHILDREN'S HOSP. NAT'L MEDICAL CTR., MILD HEAD INJURY: CARE OF THE CHILD AT HOME (1990) (describing symptoms of head injury that may persist after children are discharged from the hospital); John F. Doronzo, *Mild Head Injury*, *in* PSYCHOLOGICAL MANAGEMENT OF TRAUMATIC BRAIN INJURIES IN CHILDREN AND ADOLESCENTS 207, 211 (Ellen Lehr ed., 1990).

[14] Harvey S. Levin et al., *Neuropsychologic Findings in Head Injured Children*, *in* PEDIATRIC HEAD TRAUMA 223, 226–27 (Kenneth Shapiro ed., 1983); D. Shaffer et al., *Psychiatric Outcome of Localized Head Injury in Children*, 34 CIBA FOUND. SYMP. 191 (1975).

and the need for special education programs.[15] These impairments, often subtle, may persist for years after the injury.[16]

B. *Increased Mortality Rates in the Absence of EMSC*

The U.S. infant mortality rate is the highest among virtually all major industrialized countries.[17] This unusually high child mortality rate is largely attributable to the greater number of deaths from traumatic injuries.[18] In the United States, mortality rates for critically ill and injured children are significantly higher in communities without access to EMSC.[19] For example, in Los Angeles, the death rate for pre-hospital care was almost twice as high for children as adults and most child deaths occurred along a corridor of the county in which pediatric services were limited.[20] In Illinois the introduction of trauma centers significantly reduced mortality and morbidity for adults, but not for children.[21] In a particularly detailed investigation, one hundred consecutive pediatric trauma deaths in Mobile, Alabama, were reviewed. The authors' analysis revealed that with optimum care, fifty-three percent of these patients could have been saved. Among the potentially fatal deficiencies identified in this study were: (1) failure to use the existing EMSC; (2) failure to provide appropriate pre-hospital treatment in the field; (3) failure to send the child to the appropriately staffed and equipped hospital; (4) delay in arrival of the treating physician at the hospital; and (5) treatment in the hospital that was inappropriate for children.[22]

A study of pediatric deaths reported to coroners showed that many child deaths occurred on adult wards and in adult intensive

---

[15] EMSC REPORT, *supra* note 7, at 65.

[16] *See* Jytte Flash & Richard Malmros, *A Long-Term Follow-Up Study of Children with Severe Head Injury*, 4 SCANDINAVIAN J. REHABILITATIVE MED. 9 (1972).

[17] *Surgery Needed: Some ABCs of Mediconomics*, ECONOMIST, July 6, 1991, at 4, 5 [hereinafter *Surgery Needed*].

[18] *See* Rosenberg et al., *supra* note 8, at 1084.

[19] Seidel et al., *supra* note 2; Albert Tsai & Gene Kallsen, *Epidemiology of Pediatric Prehospital Care*, 16 ANNALS EMERGENCY MED. 284, 289–91 (1987).

[20] Seidel et al., *supra* note 2.

[21] *See* Ross Mullner & Jack Goldberg, *The Illinois Trauma System: Changes in Patient Survival Patterns Following Vehicular Injuries*, 6 J. AM. C. EMERGENCY PHYSICIANS 393 (1977).

[22] Max L. Ramenofsky et al., *Maximum Survival in Pediatric Trauma: The Ideal System*, 24 J. TRAUMA 818, 818–21 (1984); *see also* Mullner & Goldberg, *supra* note 21 (reporting the effects of trauma centers on patient mortality, based on a four-year study of vehicular injuries).

care units.[23] Despite the well-documented need for specialized care for children, there are still many areas in this country, particularly in rural and remote communities, that are not linked by a pediatric critical care system or EMSC project to tertiary intensive care services for children.

Not surprisingly, research indicates that the time from the initial incident to the provision of critical emergency care is extremely important. In Chicago, a study by Holmes and Reyes analyzed two groups of pediatric patients with serious traumatic injuries. The first group, which was transported directly to a tertiary care center with a pediatric intensive care unit ("PICU"), had no deaths and only one case of permanent disability. Thirty-two percent of the patients initially transported to community hospitals and subsequently transferred to the tertiary care facility had poor outcomes—either death or permanent disability. The authors offered a compelling explanation for these differences: the group of children transported to community hospitals had, on average, a four-hour delay before the initiation of optimal care.[24]

A state-wide study in Oregon compared children who received care in PICUs to those kept in community hospitals without specialized pediatric intensive care services. Outcomes were compared using a scoring system developed to classify critically ill and injured children (Pediatric Risk of Mortality Scale, or "PRISM"). The study showed that the death rate for children who had moderate illness or injury and who were kept in community hospitals was higher than would have been predicted by their PRISM score. Although the overall death rate was actually higher in the PICUs, their mortality was consistent with their predicted outcomes based on PRISM scores because these patients tended to be much more critically injured than those kept in community hospitals. Overall, the PICUs were able to offer significantly superior care.[25]

The clearest empirical support for the implementation of EMSC technology comes from San Diego and Sweden, where

---

[23] Marianne Gausche et al., *Pediatric Deaths and Emergency Medical Services in Urban and Rural Areas*, 5 PEDIATRIC EMERGENCY CARE 158, 160–61 (1984).

[24] Michael J. Holmes & Hernan M. Reyes, *A Critical Review of Urban Pediatric Trauma*, 24 J. TRAUMA 253, 253–55 (1984).

[25] *See* Murray M. Pollack et al., *Comparison of Tertiary and Nontertiary Intensive Care: A Statewide Comparison*, 23 PEDIATRIC RES. 234 (1988); *see also* Gausche et al., *supra* note 23, at 161 (stating that high mortality rate in rural region may be due to lack of PICUs and lack of formal transfer policy).

the kind of system suggested by this analysis has been implemented. For example, in San Diego County, pediatric trauma centers must be approved for critically ill children. Trauma scores are used to determine the need for direct transport. The use of ground or air transport (after initial stabilization at the local emergency department) is determined by transport time. Sweden's implementation of similar systems, together with an extensive injury prevention program, has demonstrated both the cost-effectiveness of such programs and their success in significantly decreasing trauma-related morbidity and mortality in children.[26]

C. *The Importance of Providing Specialized Equipment and Care for Children*

Emergency medical systems are a relatively new component of our national health care system. These systems were developed in the United States following recognition that an organized approach to the care of critically ill and injured patients could significantly reduce death and disability. Prior to 1966, emergency transport to a hospital was often provided by undertakers in hearses or by well-meaning citizens in private vehicles.

It was not until experienced Vietnam medical corps veterans returned home that pre-hospital care changed significantly. Upon the return of these war veterans, the evolution of a formal emergency medical system progressed more rapidly. As the medical community recognized the impact of pre-hospital care on the overall outcome of patients, the American College of Emergency Physicians ("ACEP") began to take a more active role in providing direction for care in the pre-hospital arena.[27]

A systems approach to emergency care was thus stimulated by wartime experiences with adult patients. The systems con-

---

[26] *See* Abraham B. Bergman & Frederick P. Rivara, *Sweden's Experience in Reducing Childhood Injuries*, 88 PEDIATRICS 69, 69–74 (1991). In the 1950s, Sweden suffered from one of the highest child death rates from injuries of any country in Western society. An inexpensive program of injury prevention and EMSC has reduced the child death rate from injuries by approximately 70%. *See also* COUNTY OF SAN DIEGO, DEP'T OF HEALTH SERVS., TRAUMA SYSTEM ANNUAL REPORT (1989) [hereinafter SAN DIEGO TRAUMA REPORT] (showing that the survival rate for severely injured trauma victims in San Diego County surpassed the expected national average, since 80% of victims who reached a trauma center survived).

[27] William R. Roush, *Emergency Medical Service Systems*, *in* MANAGEMENT OF EMERGENCY SERVICES (John H. van de Leuv ed., 1987).

cept was relatively slow to develop in the civilian sector because of the traditional emphasis on medical specialties and the economic and administrative independence of hospitals and private health care providers. Because the historical models for emergency medical systems were military operations and the first major focus of civilian emergency care was the treatment of cardiac patients, the special needs of children have generally been overlooked in emergency medical systems.

Injured children present many unique problems to health care providers. These include such diverse considerations as psychological management, problems of vascular access, and management of the airway. The treatment of seriously injured children requires the services of physicians and nurses with specialized training in pediatric care, both surgical and medical. It also requires intensive care facilities, laboratory, and support services specifically designed to care for children.

In addition, medical, laboratory, and epidemiological research increasingly demonstrates that children suffering from trauma differ from adults in important physical and emotional ways. Injured children require care from professionals with specialized training and expertise in the assessment and treatment of children.[28] However, education and training in emergency care for children has lagged behind other programs designed to improve emergency medical systems.[29]

The most glaring deficiency in present emergency department care for children is simple ignorance of children's special medical needs. Many hard-working, responsible emergency department physicians, nurses, and other providers are simply unaware of, and untrained in, basic pediatric emergency medicine.[30]

---

[28] Stephen Ludwig et al., *Pediatric Training in Emergency Medicine Residency Programs*, 11 ANNALS EMERGENCY MED. 170, 170–73 (1982); James S. Seidel, *A Needs Assessment of Pediatric Advanced Life Support and Emergency Medical Services for the Pediatric Patient: State of the Art*, 74 CIRCULATION 129, 129–33 (Supp. IV 1986).

[29] Pediatric training was not included in the comprehensive Emergency Medical Services Systems Act of 1973, Pub. L. No. 93-154, 87 Stat. 594, which targeted personnel training as one of the 15 essential components of emergency medical systems. Before the creation of EMSC demonstration grants within the U.S. Public Health Service's Maternal and Child Health Bureau, little training was offered in the management of pediatric emergencies. In addition, prior to the development of the Pediatric Advanced Life Support course in 1988 by the American Heart Association ("AHA") and American Academy of Pediatrics ("AAP"), there were no national training programs for physicians in pediatric emergency care. *See* EMSC REPORT, *supra* note 7, at 51.

[30] "Although continuing education in pediatric emergency nursing is offered periodically by regional organizations and some proprietary concerns, the Emergency Nurses Association did not sanction the development of a standard curriculum for emergency nurses until 1990." EMSC REPORT, *supra* note 7, at 52.

Improved education in pediatric emergency care is essential to insure that injured children have access to care from an organized team of health care providers with the knowledge, skills, and attitudes necessary to perform their lifesaving tasks effectively.

For example, a study of sixty-three emergency medical technician ("EMT") and paramedic training programs in the United States demonstrated the extent of current deficiencies in the education of pre-hospital care providers regarding pediatric emergencies. Paramedic/EMT education takes place primarily at colleges and universities, with only half of all students being taught at hospitals and emergency medical systems agencies. Five percent of the trainees receive no training in pediatrics and forty-one percent receive less than ten hours of training. The majority of programs offer less than ten hours in clinical pediatric training and twenty-one percent have none. Fifty-five percent of the students have no preceptorship in pediatric emergencies at an emergency department or other clinical facility. Pediatric education was particularly deficient in the areas of pediatric field simulations, arrhythmias, envenomation, hypotension, coma, and drowning. Incredibly, advanced pediatric life support was not taught in twenty-two percent of the programs surveyed.[31]

In addition to deficiencies in training, many pre-hospital care providers and emergency departments lack equipment that is basic to the treatment of children's injuries. Appropriate equipment is often not carried in pediatric sizes. The most common items not carried include such simple and inexpensive items as backboards, hard-neck collars, endotracheal tubes, appropriately sized blood pressure cuffs, pediatric syringes, and small-gauge intravenous catheters. Although most units had a bag-

---

[31] *Id.* at 52 (noting that basic EMT training devotes only 2 of 110 course hours to pre-hospital emergency care of children); James S. Seidel, *Emergency Medical Services and the Pediatric Patient: Are the Needs Being Met? II. Training and Equipping Emergency Medical Services Providers for Pediatric Emergencies*, 78 PEDIATRICS 808, 810 tbl. 3 (1986); Jerry Buckley, *The Shame of Emergency Care for Kids*, U.S. NEWS & WORLD REP., Jan. 27, 1992, at 34, 37. In response to concerns about the inadequacy of present emergency medical systems for children, the American Board of Medical Specialties approved a new subspecialty in pediatric emergency medicine in April, 1991. This new subspecialty will hopefully reduce the longstanding rivalry between pediatricians and emergency care physicians and make the field more attractive to future physicians.

Table 1
Deficiencies in Pediatric Equipment and Supplies Carried by Emergency Medical Systems Providers

Equipment and Supplies	*% of Deficient Providers*
Doppler blood pressure device	98%
Feeding tubes	
5F	87%
8F	81%
Naloxone for neonatal use	74%
Pediatric backboards	63%
24-gauge intravenous catheters	62%
Blood pressure cuffs	
Thigh size	52%
Infant size	51%
Dextrostix	51%
Pediatric sodium bicarbonate syringe	49%
Pediatric atropine syringe	48%
Pediatric defibrillator paddles	43%
22-gauge intravenous catheter	30%

Source: James S. Seidel, *Emergency Medical Services and the Pediatric Patient: Are the Needs Being Met? II. Training and Equipping Emergency Medical Services Providers for Pediatric Emergencies*, 78 PEDIATRICS 808, 810 tbl. 5 (1986).

Table 2
Important Equipment and Supplies Frequently Not Found in Ambulatory Care Centers

Equipment and Supplies	*% of Centers NOT Having the Item*
Intraosseous needles	93.0%
Infusion pump	82.3%
Venous cutdown tray	77.4%
Pediatric femur splint	73.6%
Premature infant bag-valve mask	72.9%
Pediatric McGill forceps	69.2%
Pediatric spine board	64.3%
Infant blood pressure cuff	40.3%
Infant bag-valve mask	30.4%
Infant-sized laryngoscope blade	29.6%
Child-sized laryngoscope blade	24.9%

Source: James S. Seidel et al., *Emergency Medical Services and the Pediatric Patient III: Resources of Ambulatory Care Centers*, 88 PEDIATRICS 230, 233 tbl. 2 (1991).

valve mask resuscitator, seventy-nine percent did not have a complete set of masks to fit infants and children.[32]

[32] Seidel, *supra* note 31, at 810–11.

These problems are distressingly widespread. For example, only forty-three percent of the pediatricians surveyed last year by Johns Hopkins Medical School said their offices contained all of the equipment and drugs on a list of commonly used pediatric emergency equipment.[33] This lack of attention and substandard emergency care for children has appalled many practitioners. Dr. Frank Costello, Director of Pediatric Intensive Care at Children's Hospital of New Jersey in Newark, has stated, "If this same situation were happening to old people, there would be a march on Washington."[34]

II. The Need for Community Injury/Illness Prevention Programs

Although the primary role of EMSC has been improving acute care, more attention and resources must be focused on injury and illness prevention if we hope to achieve a meaningful reduction in unnecessary death and disability among children. An injury prevention program should be established at the community level as an integral component of any EMSC because: (1) primary prevention saves lives; (2) primary prevention reduces costs; (3) EMSC providers, as credible spokespersons, could play a key role in education to prevent injuries; and (4) EMSC providers could make a unique contribution to research on injury/illness prevention by providing reliable information about trauma-producing situations.

The importance of preventive measures in the overall effort to reduce mortality and morbidity caused by injury and illness is receiving increasing attention. A significant proportion of injury and illness deaths will occur regardless of improvements in emergency response and trauma care. Only preventive measures to avoid injuries and illnesses can lower the number of deaths further. A growing body of knowledge is now available about the effectiveness of various prevention strategies.[35]

Michael Matlak has estimated that among children, immediate deaths and deaths that occur within three hours of the injury together account for close to ninety percent of all pediatric

---

[33] Buckley, *supra* note 31, at 37.

[34] *Id.* at 36.

[35] *See* National Comm. for Injury Prevention and Control, Injury Prevention: Meeting the Challenge (1989).

trauma-related fatalities. He and others suggest that more aggressive injury prevention programs might avert 6000 child deaths annually. In comparison, strategies such as improved transport and initial aggressive care may be capable of saving approximately 2500 lives, and new clinical research, directed at late deaths, may save an additional 750 to 1000 children annually.[36] These figures indicate that prevention is the method of maximum utility for saving lives.

The best-documented approaches to injury prevention are sometimes referred to as the "three E's": education, enforcement, and engineering. For example, most analysts attribute the decline in the United States in motor vehicle fatalities per vehicle miles driven to these strategies, including improvements in the interstate highway system and vehicle design, enactment of adult and child seat belt legislation, public safety education, and increased enforcement of laws against drunk driving.[37]

In the area of childhood injury, several prevention measures have been quite successful. A good example is the dramatic increase in the use of child safety seats since 1985, when their use was mandated by legislation in all fifty states. Surveys by the National Highway Traffic Safety Administration ("NHTSA") indicate a continuing increase in child restraint usage rates, from seventeen percent in 1979 to seventy-eight percent in 1986.[38] An econometric model for estimating the lifesaving benefit of legislation on child restraint use suggests that restraint legislation has reduced car occupant fatalities by thirty-nine percent for infants and by thirty percent for toddlers.[39] These data indicate that mandatory child restraint laws are some of the most cost-effective health care measures in existence.

According to one estimate, the enactment of twelve currently available prevention strategies could reduce deaths of children under age fifteen by twenty-nine percent.[40] These methods for reducing fatalities include infant car seat restraints and air bags

---

[36] Michael E. Matlak, *Current Problems in the Management of Pediatric Trauma*, *in* EMERGENCY MEDICAL SERVICES FOR CHILDREN: REPORT OF THE 97TH ROSS CONFERENCE ON PEDIATRIC RESEARCH 2, 6–7 (J. Alex Haller, Jr. ed., 1989) [hereinafter ROSS CONFERENCE].

[37] EMSC REPORT, *supra* note 7, at 86.

[38] *Id.*

[39] William N. Evans & John D. Graham, *An Estimate of the Lifesaving Benefit of Child Restraint Use Legislation*, 9 J. HEALTH ECON. 121, 134 (1990).

[40] Frederick P. Rivara, *Traumatic Deaths of Children in the United States: Currently Available Preventive Strategies*, 75 PEDIATRICS 456 (1985).

for older children, motorcycle and bicycle helmets, expansion and enforcement of the Poison Prevention Packaging Act, isolation fencing around swimming pools, self-extinguishing cigarettes, smoke detectors, window bars in high-rise buildings, and regulations on toys and other children's products to reduce choking and suffocation. Other effective measures include safer playground equipment, more resilient playground surfaces, and installation of mixing valves and water heater controls to prevent tap water scalds. These technologies for reducing the toll of childhood injury are currently available, but require widespread political and legislative support to insure full implementation.[41]

III. The Need for Government Action

Federal regulation of health and safety risks has expanded rapidly in recent decades.[42] Unfortunately, the limitations and inertia of regulatory bureaucracies have often made government regulation an ineffective tool of social engineering,[43] and well-intentioned efforts have often proven counterproductive or wasteful. Such errors demonstrate that federal regulation should be used only to correct formidable defects in the free market.[44]

Such defects are evident in the realm of health care. Health risks can never be completely eliminated, and since reductions in risk must be purchased with limited societal resources, se-

[41] Injury prevention is increasingly viewed as an integral component of practice by a number of professional organizations. For example, both the American College of Surgeons and the ACEP have published guidelines for trauma care systems that include injury prevention as a system component. *See* Erwin R. Thal, American College of Surgeons Comm. on Trauma, *Resources for Optimal Care of the Injured Patient: An Update*, Am. C. Surgeons Bull., Sept. 1990, at 20; American College of Emergency Physicians Trauma Comm., *Guidelines for Trauma Care Systems*, 16 Annals Emergency Med. 459 (1987). The ACEP has also developed a position statement on motor vehicle safety, stating the organization's commitment to legislation, public and professional education, and research. The statement encourages members to take the lead in these activities at the local, state, and national levels. American College of Emergency Physicians, *Motor Vehicle Safety*, 14 Annals Emergency Med. 822 (1985).

[42] The federal government's total regulatory budget grew from just over $1 billion in 1971 to over $6 billion in 1980. *See* Stephen Breyer, Regulation and Its Reform 376 fig. 6 (1982). Similarly, the number of regulations pages in the *Federal Register* grew from 2599 in 1936 to over 1 million in 1981. *Id.* at 377 tbl. 16.

[43] *See* Nathan Glazer, The Limits of Social Policy (1988) (analyzing the limitations of social engineering by government).

[44] Even when market defects are chronic and serious, private sector approaches to correction, such as tax incentives, standard setting, disclosure of important information to consumers, bargaining, marketable rights, and other less restrictive forms of intervention, should arguably be attempted first. *See* Breyer, *supra* note 42, at 185.

lecting rational goals and appropriate targets for health regulation can be difficult. Due to methodological difficulties, errors in data collection and analysis, and various influences upon policy formation, the cost-effectiveness of government regulation for the reduction of health risks varies wildly, from $54,000 per life saved by breast cancer screening programs, to $200,000 per life saved by airplane cabin fire protection, to over $130 million per life saved by the 1979 regulation of diethylstilbestrol ("DES") in cattlefeed.[45] Extraordinarily expensive crusades against minuscule but well-publicized risks lead to the perpetuation of widespread health risks in situations where cost-effective intervention would be possible.[46]

As a result, while billions of taxpayers' dollars have been spent on the relatively trivial risks posed by DES in cattlefeed or Atrazine/Alachor in drinking water,[47] the prevention of significant health risks to children has never been a high priority issue within the U.S. political system. In 1989, more than 9 of every 1000 American infants perished before their first birthday.[48] Moreover, the U.S. infant mortality rate has worsened comparatively during the twentieth century. In 1918, the U.S. Children's Bureau found that the U.S. infant mortality rate was sixth among twenty nations. In 1986, it was thirteenth among these same twenty nations.[49] At the present infant mortality rate, we will lose more infants between 1988 and 2000 than the total number of Americans killed in combat in World War I, World War II, Korea, and Vietnam combined.[50]

---

[45] STEPHEN BREYER, BREAKING THE VICIOUS CIRCLE: TOWARD EFFECTIVE RISK REGULATION (forthcoming 1993); *see also* Richard J. Zeckhauser & W. Kip Viscusi, *Risk Within Reason*, SCIENCE, May 4, 1990, at 562 (noting that methodologies vary widely across government agencies; for example, the FAA values lives based only on lost earnings, while regulations of food additives are set without consideration of cost).

[46] *See* Zeckhauser & Viscusi, *supra* note 45, at 560 (calling for systematic strategies for assessing and responding to risks, and explaining how misperceptions of risk have led to distortions of economic valuation and misapplication of public funds).

[47] The cost of regulations for these risks is over $100 million per life saved and over $92 billion per life saved, respectively. John Morral III, *A Review of the Record*, 10 REGULATION 25, 30 (1986); *see also* BREYER, *supra* note 45; *cf.* Zeckhauser & Viscusi, *supra* note 45, at 564 n.36 (demonstrating by polling data that many Americans want legislators to seek safety regardless of cost).

[48] *Surgery Needed*, *supra* note 17, at 5; *see also* NATIONAL COMM'N TO PREVENT INFANT MORTALITY, TROUBLING TRENDS: THE HEALTH OF AMERICA'S NEXT GENERATION 2 (1990).

[49] NATIONAL COMM'N TO PREVENT INFANT MORTALITY, *supra* note 48; Julie Johnson, *Congress Shows Signs of Spending to Fight Infant Deaths*, N.Y. TIMES, May 21, 1989, § 4, at 4.

[50] NATIONAL COMM'N TO PREVENT INFANT MORTALITY, HOME VISITING: OPENING DOORS FOR AMERICA'S PREGNANT WOMEN AND CHILDREN 8 (1989).

Clearly, many causal factors interact to create this complex phenomenon, including births to teenage mothers,[51] lack of parental education, maternal drug use during pregnancy,[52] lack of stable family support systems, domestic and street violence, irresponsible and negligent parenting,[53] and cultural tolerance of these risk factors, as well as inadequate prenatal, maternity, and infant medical care. The current U.S. health care system for infants and children often fails to provide even the simplest of preventive measures such as immunization. In 1985, more than one-fifth of all two-year-olds in America were not fully immunized against polio, rubella, mumps, or measles.[54] Similarly, U.S. immunization rates for DPT (diphtheria, pertussis, and tetanus) among children under the age of one are half those of Western Europe and Israel even though the United States spends a larger proportion of its GNP on health care than does any other country.[55]

Despite the inherent limitations of governmental intervention in complex social problems, it is an empirically demonstrable fact that improved emergency health services and injury prevention programs for rationally targeted populations could save both lives and resources.[56] Researchers agree that between 6000 and 10,000 of the 22,000 children who die from accidents each

---

[51] Births to women under the age of 16 present higher medical risks for both mother and child. Births to teenagers between the ages of 15 and 19 were three times more frequent in the United States than in Europe during the 1980s. C. Arden Miller, *Prenatal Care Outreach: An International Perspective*, *in* PRENATAL CARE: REACHING MOTHERS, REACHING INFANTS 210 (Sarah Brown ed., 1988).

[52] Fetal drug exposure has become an increasingly major health hazard for children in the United States. Affected infants may face a long list of potentially devastating developmental, psychological, physical, educational, and economic effects. *See* Barry Zuckerman et al., *Effects of Marijuana and Cocaine Use on Fetal Growth*, 320 NEW ENG. J. MED. 762 (1989); R. Christopher Barden et al., The Prevention of Drug Exposure in Infancy: A Legal, Biomedical, Economic, and Psychological Analysis (May 27, 1992) (unpublished manuscript, on file in the Harvard Law School Library).

[53] Recent studies showing that only one in four children injured in automobile accidents was wearing a seat belt offer a striking example of how parental negligence can lead to infant death. *See* Buckley, *supra* note 31, at 34, 38.

[54] *See* U.S. CONGRESS, OFFICE OF TECH. ASSESSMENT, HEALTHY CHILDREN: INVESTING IN THE FUTURE 142–43 (1988).

[55] In 1987, health care accounted for 11.1 % of the U.S. GNP. *Id.*

[56] *See* NATIONAL COMM'N TO PREVENT INFANT MORTALITY, 1985 INDIRECT COSTS OF INFANT MORTALITY AND LOW BIRTHWEIGHT 1 (1988). This commission estimates that if the 40,030 babies who died in 1985 had lived to become productive members of society, the current value of their future earnings would have been between $10.2 and $18.9 billion. Similarly, if the United States could reduce its number of disabled low-birthweight babies by one-half, the present value of wages that children spared these disabilities could earn would be between $0.9 and $1.9 billion.

year could be saved by injury prevention and better emergency treatment by medical personnel.[57]

State and federal governments should create multi-disciplinary regulatory councils that will mandate safety and quality standards and the disclosure of critical information to the public. Immediate action to improve EMSC at the state and/or federal level is possible with minimal funding. For example, by delegating standard-setting and information disclosure responsibility to an office and director of EMSC programs within existing state and federal Departments of Public Health, America can begin the evaluation, research, disclosure, and standard-setting that will create incentives for hospital systems to improve emergency services without further government action or the creation of unwanted bureaucracy.

As in all areas, government action should be limited and should maximize the impact of private efforts. Medical professionals and the hospital industry should be given opportunities, as well as incentives, to implement high-quality injury/illness prevention programs and EMSC. On the other hand, the epidemic of injury-related deaths among American children demands immediate action on both private and governmental levels. Given the current number of deaths and injuries among American children and the inadequacy of current EMSC, immediate governmental action, at least in setting standards and facilitating the disclosure of information, is clearly necessary.

A classic method by which government may regulate imperfect markets without creating more bureaucracy is by setting standards. Minimal quality standards for EMSC may be enforced through liability, removal of licenses, civil fines, or negative publicity.[58] Investigations of existing research and theoretical discussions on appropriate standards for pediatric emergency care should be conducted by government and professional agencies as the first step towards public notice of standards. Discussions on the implementation of standards should include the target of the standards, degree of specificity, performance or design standard issues, and enforceability. Policy an-

---

[57] Although this is the clear consensus among experts in this field, these figures must be documented by further empirical research. However, given the extraordinary seriousness of the problem, it would be ethically impermissible to wait for more rigorous verification before acting. *See* EMSC REPORT, *supra* note 7, at 85; Buckley, *supra* note 31, at 38.

[58] *See* BREYER, *supra* note 42, at 96–119 (giving an overview of auto safety standards).

alysts should note the inherent tension between allowing maximum flexibility to the industry so they may implement the standards as efficiently as possible, versus specific, detailed standards that are more easily enforced.

Providing optimal emergency care for children will require a range of standards for the implementation of a fully integrated EMSC including pre-hospital services, emergency department and inpatient care, interfacility transport systems, and specialized pediatric centers offering pediatric surgical, intensive, and subspecialty care (Figure 3). The essential components of effective systems of trauma care include: (1) identification of the injured patient; (2) interventions which sustain life and prevent further injury until definitive care is available; (3) triage of patients to an appropriate facility; (4) transport of patients; (5) definitive care at a level that meets the patient's needs; and (6) rehabilitation.[59] Each component of the EMSC will require legislative and/or regulatory attention. The implementation of such standards should be conducted via an open, public, and empirically directed process.

Perhaps the most important component of EMSC should be ongoing education programs for health professionals, including primary care professionals. Programs should include: (1) current concepts; (2) techniques and procedures in pediatric critical care; (3) an introduction to EMSC services and resources and how to use them; (4) an introduction to necessary research protocols; and (5) discussions of the best systems for routing critically ill and injured children to facilities where optimal care is available.[60]

IV. Cost-Benefit Analysis

EMSC and injury prevention programs could save both lives and tax revenue. Given limits on societal resources, difficult decisions must be made as to how they will be used. However, simply focusing on expenditures may fail to assess some of the more intangible benefits of health care programs (e.g., reduction

---

[59] Max L. Ramenofsky, *How Can We Address the Differences in Trauma Versus Illness Systems?*, *in* Ross Conference, *supra* note 36, at 51.

[60] *See* EMSC Report, *supra* note 7 (detailing minimum standards for effective EMSC education and training).

FIGURE 3
COMPONENTS OF EMERGENCY MEDICAL SYSTEMS FOR CHILDREN

of pain, suffering, and long-term disability).[61] Regardless of the methodological difficulties involved, EMSC and injury prevention programs, as with all public policy initiatives, must be guided by considerations of cost-effectiveness.

In addition to extraordinary human suffering, the economic costs of children's injuries are enormous. The direct economic cost of more than 22,000 annual child deaths from injuries has been estimated at nearly $8.3 billion a year.[62] A 1986 study found that the cost of future lost productivity resulting from the 22,411 child fatalities that year amounted to another $8 billion.[63]

In addition to mortality costs, 16 million children are seen in emergency departments for non-fatal injuries each year. Estimates of direct and indirect costs of such injuries are even higher. One study of the "hidden taxation" of childhood injuries and deaths estimated the lifetime costs (both direct and indirect) of injuries to children ages fourteen and under in 1985 to total over $13.8 billion. If ages fifteen to twenty-four are included, their estimate balloons to $53 billion.[64] The annual direct cost of non-fatal childhood injuries is also in the billions of dollars. (See Figure 4). Although additional data would always be helpful, judging from the reductions in mortality rates demonstrated by EMSC programs in Sweden, San Diego, and parts of Los Angeles, at least a third of these death and disability costs could be eliminated by adequate EMSC care and injury prevention programs.[65] Few governmental health regulatory programs enjoy such demonstrable cost-effectiveness prior to implementation.

Additionally, EMSC and injury/illness prevention programs can be viewed as an investment in the future American workforce. If our infant/child mortality rate were reduced to that of Japan's (which has the lowest rate in the world), the additional 20,000 children who would survive each year could contribute

---

[61] *See* MICHAEL F. DRUMMOND, ECONOMIC APPRAISAL OF HEALTH TECHNOLOGY IN THE EUROPEAN COMMUNITY (1986).

[62] CENTERS FOR DISEASE CONTROL, U.S. DEP'T OF HEALTH & HUMAN SERVS., CHILDHOOD INJURIES IN THE UNITED STATES (1990); *see also* EMSC REPORT, *supra* note 7, at 84.

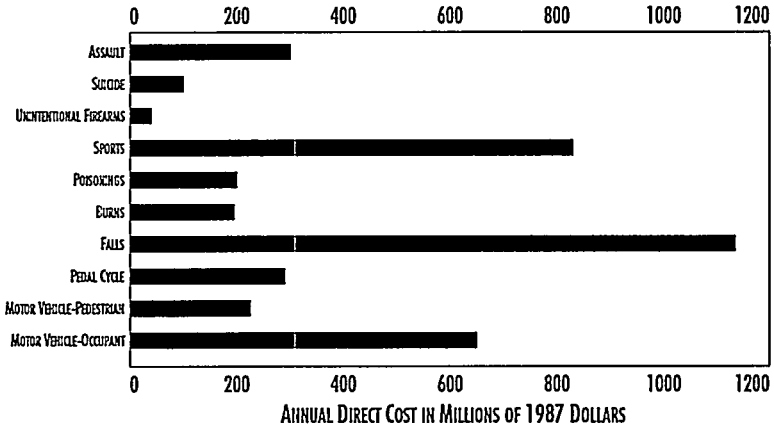
[63] Rodriguez & Brown, *supra* note 6, at 627–35.

[64] DOROTHY P. RICE & ELLEN J. MACKENZIE, COST OF INJURY IN THE UNITED STATES (1989).

[65] *See* Bergman & Rivara, *supra* note 26, at 69–74; *see also* SAN DIEGO TRAUMA REPORT, *supra* note 26 (reporting that the direct cost to San Diego County from trauma patients in 1987–88 was 31% lower than the year before, and 40% lower than 1985–86, and that many of the trauma injuries treated were preventable).

FIGURE 4

ANNUAL DIRECT MEDICAL COSTS OF NON-FATAL CHILDHOOD INJURIES



SOURCE: MARVIN MALEK ET AL., *THE COST OF MEDICAL CARE FOR INJURIES TO CHILDREN*, 20 ANNALS EMERGENCY MED. 997, 1003 TBL. 9 (1991).

$10 billion to the economy as employees.[66] "We ought to invest in human capital with the same entrepreneurial spirit and concern for long-range payoffs that venture capitalists bring to investments in [other] enterprises."[67] An approach that provides children with high-quality health care and protection has the best chance of winning liberal and conservative support and becoming a reality for all U.S. children.[68]

Weighing unnecessary injuries and resulting disabilities of innocent children in the equations of economic decision theory is a challenging task. Unfulfilled development and misery cannot

[66] NATIONAL COMM'N TO PREVENT INFANT MORTALITY, DEATH BEFORE LIFE: THE TRAGEDY OF INFANT MORTALITY 10 (1988).

[67] FORD FOUND. PROJECT ON SOCIAL WELFARE & THE AM. FUTURE, THE COMMON GOOD: SOCIAL WELFARE AND THE AMERICAN FUTURE 46 (1989).

[68] Efforts to improve the health of American children will only succeed if public planners appreciate the limitations of governmental action. Social control systems beyond government (individual action, families, church groups, volunteer organizations, community organizations, and the media) must be involved if these efforts are to be successful. In the 1960s and early 1970s, many political theoreticians believed that government policies alone could transform and improve human behavior. In hindsight, this simplistic view overestimated the powers of academic knowledge and underestimated the limitations of social/governmental policy unaided by the private sector. *See* GLAZER, *supra* note 43.

be adequately measured, or compared, but must be included in our philosophical analysis of this problem. Behind the quantitative hypotheses regarding the number of children injured and the benefits of well-designed EMSC networks is the fact that each unnecessarily injured child may face minutes, days, months, and years of suffering, disability, and loss.

In the context of childhood injuries, the existence of large numbers of unnecessarily injured and disabled children implies fewer unimpaired children. Functioning members of society support families, produce the resources that are taxed, and ultimately allow economic growth. Thousands of serious unnecessary injuries and long-term disabilities impair the functioning of individuals in society and shrink economic resources while simultaneously increasing needs. Even a moderate number of these injuries and subsequent disabilities create a cumulative effect that will decrease growth and lessen overall resources.

While this analysis catalogs the costs and potential benefits of improved EMSC and injury prevention programs, the essential issue is the priority that this problem should receive. While the enormous costs of injured and disabled children need immediate attention, there are many competing social needs. Rational decisions must be made among a number of important options. For example, spending more on emergency medical care may mean less for education or other needs. Although the assignment of priority is ultimately the prerogative of individuals and government, the relevant costs and expected benefits must be honestly assessed. Compared to the billions of dollars spent on the questionable benefits of many other government regulatory programs,[69] EMSC and injury prevention programs appear to be a far wiser investment of our finite resources.

Moreover, although cost is always an issue, surprisingly little would be required to upgrade the pediatric emergency care system to acceptable levels. The price of education and equipment needed to properly train providers to carry out essential tasks is only a few thousand dollars.[70] The time required to ensure adequate training could be credited toward the annual continuing education requirements already mandated for recertification.

---

[69] *See supra* notes 45–46 and accompanying text.

[70] Buckley, *supra* note 31, at 38.

In sum, EMSC and injury prevention programs are clearly among the least expensive and most cost-effective policies that have been introduced in recent decades for improving the health and welfare of American children. Federal regulatory safety and health programs have all too frequently failed even the most generous assessments of cost-effectiveness. In contrast, EMSC and injury prevention programs will save lives every day with little added expense. Furthermore, when considering costs and funding for these improvements, we should remember that we are deciding whether children are entitled to the same quality of emergency health care adults already receive.

V. Sources of Funding

Integrating EMSC and injury prevention programs into existing emergency medical system programs will require the modest expenditure of federal and state funds. In the absence of direct legislative appropriations, several programs in the U.S. Department of Health and Human Services disburse grant funds that should be considered by states, cities, and other communities attempting to improve EMSC and injury prevention programs. Examples of these programs include the U.S. Public Health Service Centers for Disease Control; the Maternal and Child Health Bureau; the Alcohol, Drug and Mental Health Administration; and the National Institute for Child Health and Human Development. Funds for EMSC and injury prevention programs are also available from the U.S. Department of Transportation and the NHTSA. Finally, the Trauma Care Systems Planning and Development Act recently passed by Congress may include EMSC and injury prevention funds.

State general funds are an additional source of funding. A strategy for introducing or supporting EMSC-oriented state legislation with attached funding should also be considered. Enlightened state lawmakers should quickly realize that EMSC and injury prevention programs are an effective way for states to reduce expenditures for chronic care by reducing the number of unnecessary childhood injuries and resulting disabilities. A dedicated tax scheme could be pursued in states where it is constitutional. For example, taxes on cigarettes and alcohol are used in some states to support various health and social services programs. Similar tax schemes could support improvements in

EMSC and injury prevention programs. Two logical choices would be surtaxes on homeowners insurance polices and automobiles. The reductions in death and disability by EMSC and injury prevention programs would produce direct financial incentives for insurance companies to support these improvements. Vehicle surcharges are already in place in Idaho and Virginia to support EMSC programs. Additional surcharges on drivers licenses and license plates could be added as well. Finally, all or a portion of traffic fines, or an added assessment on traffic fines for speeding, transporting children without appropriate restraints, or drunk driving, could help subsidize EMSC efforts. Arizona, Mississippi, and Rhode Island are already using this funding mechanism.[71]

In addition to grants and legislative appropriations, EMSC and injury prevention programs could rapidly become self-supporting via ambulance fees, certification/licensing fees for professional providers, or requiring additional instruction for licensing. Increasing such fees might require strong public support that could result from wide publication of the tremendous need and cost-effectiveness of EMSC and injury prevention programs.

Local newspapers, radio stations, and television stations can be tapped for public service announcements, advertising space, and sponsorship of other EMSC and injury prevention programs. Local media can be particularly helpful in providing public information on injury/illness prevention activities and when and how to call 911. Local merchants, manufacturers, and distributors can also sponsor various EMSC and injury prevention activities. Injury/illness prevention activities are especially appealing to the public and to merchants when sponsors' names are attached to the activities as "free advertising." For example, bicycle shops or bicycle helmet manufacturers or distributors can sponsor bicycle safety workshops and promote helmet usage. Sponsors may even contribute free helmets to be sold at cost by project coordinators (the beginnings of a self-sustaining helmet sales program). Sponsors can also be asked to provide grants for equipment and training for local ambulance squads.

Other sources of private sector funding include foundations. There are currently more than 5000 foundations in the United States with assets in excess of $1 million, and annual donations

[71] EMSC REPORT, *supra* note 7, at 151.

exceeding $100,000. Certain private sector entities such as pharmaceutical and insurance companies also may provide financial support for EMSC and injury prevention activities.

VI. Legal Incentives

In addition to the adoption of state and federal legislation, improvements in EMSC could be implemented by creating economic incentives for medical systems to raise standards for EMSC. Constitutional arguments against discriminatory treatment of children in emergency health care systems receiving public funding may be effective. In addition, the use of the tort system to highlight liability rules for the allocation of the burden of paying for the harm of improper care could also be a powerful tool for creating incentives for change. As a mechanism for monitoring the quality of care, civil law has the advantage of providing patients (and their lawyers) with a financial incentive to bring information regarding inferior care into the public domain. As EMSC standards become more widely known and accepted, avoiding malpractice litigation will become an increasingly powerful incentive for the implementation of adequate EMSC.

Informed consent is another avenue of tort litigation for children receiving inadequate emergency care.[72] Although the duty of disclosure is usually limited to "material risks," it seems reasonable to assume that inadequately staffed, trained, or equipped emergency medical systems may need to disclose such risks to the families of injured children who may be treated there.[73] In practice, of course, parents of seriously injured children will lack genuine options in emergency situations, raising the question of whether parents can give true informed consent to inadequate care when a delay of seconds may mean death.

The major difficulty with incentives produced by litigation is the inefficiency involved. Litigation is extremely costly. In addition, given the lack of consumer information, many patients

[72] *See, e.g.*, Canterbury v. Spence, 464 F.2d 772, 780–81 (D.C. Cir. 1972) (holding that every person has the right to determine what shall be done to his body); Truman v. Thomas, 611 P.2d 902, 905 (Cal. 1980) (holding that physicians have a duty of reasonable disclosure to patients regarding the available choices of therapy and the potential dangers).

[73] *See, e.g.*, Crain v. Allison, 443 A.2d 558, 562 (D.C. 1982) (acknowledging that only material risks need be disclosed).

will not know they have been injured by inadequate care. Others will opt not to sue for personal reasons. The legal process can thus be extraordinarily costly and unpredictable, greatly weakening any potential deterrent effect. The unpredictability of the system is exacerbated by the backlog of cases in the courts that results in years of delay. Further, verdicts in such cases may bear little rational relationship to the damage suffered. In sum, yet another rationale for government action is that judicial remedies are too costly, time consuming, or inaccurate to adequately encourage EMSC.[74]

VII. Measuring the Benefits and Allowing Consumer Choice

Responsible allocation of societal resources requires measurement of results. Assessment methodologies have already been developed to measure the efficacy of EMSC. For example, the Major Trauma Outcome study involved over 100 institutions and analyzed data on over 100,000 injured patients.[75] From this evaluation has come the first attempts at objective assessment of system function through a research process known as TRISS methodology.[76] TRISS methodology uses two proven scales for assessing and charting the extent of injury in children: the Injury Severity Score and the Trauma Score. The child's actual prognosis is then compared to typical expectations given the severity of the child's injuries. The methodology is designed to identify patients of all ages who die of injuries which should not result in fatality, and thus provides a measure of how well the system is functioning. This is useful not only as a means of assessing the emergency care system, but also for examining the care delivered by hospitals. This monitoring may be used to evaluate the effectiveness of hospitals and EMSC.

The complex structure of EMSC makes the quality-assurance process complicated, as EMSC combines multiple organizations, facilities, and authorities. Ongoing empirical investigations will allow constant monitoring of cost-effectiveness data

---

[74] *See* Breyer, *supra* note 42, at 27.

[75] Howard R. Champion et al., *The Major Trauma Outcome Study: Establishing National Norms for Trauma Care*, 30 J. Trauma 1356 (1990).

[76] *See* Carl R. Boyd et al., *Evaluating Trauma Care: The TRISS Method*, 27 J. Trauma 370 (1987); E.M. Guirguis et al., *Trauma Outcome Analysis of Two Canadian Centres Using the TRISS Method*, 30 J. Trauma 426 (1990).

in these complex, rapidly moving systems. It is important for a clinical data management program to be comprehensive, reliable, and valid, so that both similarities and differences between the patients and groups can be explored. The collection of EMSC data should have the following characteristics: (1) it must identify disease processes which require urgent or emergent care; (2) it should provide data for continuous evaluation of these disease processes; (3) it should provide sufficient information for comparison of different areas or regions of the country at various time intervals in order to gain an understanding of regional differences in EMSC; (4) it should monitor seasonal variations along with other factors that affect human behavior to assure that EMSC is effective above normal variations in mortality and morbidity data; and (5) it should assure that relevant data for EMSC surveillance are available in a timely manner.

Data collection should occur at regular time intervals, and data should be readily available and easy to enter into national and international databases. In addition, general output measures should be developed to assess the efficacy of various emergency health care systems. Adequate measures of outcomes must acknowledge the multi-dimensional nature of health and encompass assessments of the patient's physical mobility, ability to perform activities of daily living, social functioning, and psycho-social status.

The empirical analysis of EMSC data should lead to understandable rating scales, the results of which should be disclosed to the public. Although minimal standards for EMSC should be mandated by law, variations in programs will continually be created by market forces. Thus, disclosure would allow consumers to make more rational decisions in a complex market. In the context of emergency medical care, published ratings of EMSC at local hospitals would be an enormous incentive for hospitals to either maintain compliance with EMSC standards or create adequate transfer policies and systems.

Disclosure standards offer a less restrictive means of obtaining a regulatory end. While regulations and standards may limit consumer freedom, disclosure of information does not impede consumer flexibility. Disclosure regulation does not require officials to fine-tune standards, thus allowing more freedom for technological improvements. In addition, disclosure rules create fewer legal dilemmas than formal regulation, and are thus more likely to survive judicial scrutiny.

The most important limitation on policies of mandated disclosure is that the information offered must be of practical use to the consumer. If the information is too complex, it is not likely to have much of an effect on consumer behavior. In the current emergency medical systems market, the incentives to produce and disseminate information are skewed by the agency relationship between doctor and patient. Labelling regulations in the drug industry address this problem, by informing consumers of similarities and differences between complex products that would otherwise be indistinguishable to the non-expert. This type of labelling requirement lowers the cost of information-seeking by requiring the best source of the necessary information to disseminate it.[77] Government action may be required initially to fund the research necessary to create empirically derived standards, describe the kind of information that must be provided, and help buyers evaluate the information that is being supplied.

Other justifications for disclosure regulation include: (1) paternalism—medical treatments and technology are simply too complex for the average citizen to understand under any circumstances; (2) spillover—inadequate care produces increased mortality and disability in thousands of children creating a loss of societal productivity; (3) inadequate coverage—under current medical insurance systems, neither patients' families nor insurers always pay the price of additional health care that follows inadequate initial care, leaving the state to pick up these often enormous expenses.

Arguments against government mandates for disseminating information include beliefs that: (1) the information is not needed; (2) the information will mislead; (3) providing the information is too expensive; or (4) the information will interfere with a well-functioning marketplace. These arguments are quite baseless in the case of lifesaving emergency medical care for infants and children. Given the uncertainty, changeability, and complexity of emergency medical systems, however, even nearly perfect information would not enable people to accurately evaluate the risks their children face from inadequate emergency care.[78] The limitations of time and realistic options in emergency situations argue for adequate pre-emergency information, ra-

---

[77] *See* Abbott Labs. v. Gardner, 387 U.S. 136 (1967).

[78] *See* Thomas Schelling, *The Life You Save May Be Your Own*, *in* PROBLEMS IN PUBLIC EXPENDITURE ANALYSIS 136 (Samuel B. Chase ed., 1968).

tional planning, and incentives to maintain acceptable and uniform minimum standards for emergency services.

Disclosure provisions should insure that all people have a choice among competing alternatives, that they have accurate information on which to base their choice, and that competition emphasizes the quality of benefits and total costs. Competition between emergency health care providers and rational decision-making by both patients and referring physicians could be improved by public education and advertising of standardized ratings of emergency care facilities.[79] Beginning in 1976, following a supportive Supreme Court decision,[80] the Federal Trade Commission has issued less stringent regulations governing hospital and physician advertising of services directly to the public. Public dissemination of objective ratings of hospital EMSC quality would create powerful economic incentives for hospitals to improve standards or improve transfer policies. Many economic analysts have suggested that public disclosure of hospital mortality rates would improve the efficiency of the health care market.[81]

VIII. Model Statute

The following model state legislation is offered to encourage states to adopt EMSC standards and advisory panels for both EMSC and injury prevention programs.

CHILDREN'S EMERGENCY MEDICAL AND INJURY PREVENTION SYSTEMS ACT OF _____[82]

An ACT concerning emergency medical systems for children and supplementing Chapter ____ of Title ____ of the Revised Statutes of the State of _____.

---

[79] Unfortunately, the medical profession has long opposed advertising to inform consumers of differences in quality of care. Although ostensibly considered "unprofessional conduct," the primary effect of this reduction in public information regarding medical care may well have been to limit competition and maintain high prices. *See* JOHN C. GOODMAN, THE REGULATION OF MEDICAL CARE: IS THE PRICE TOO HIGH? (Cato Inst. Pub. Policy Monograph No.L203, 1980).

[80] *See* Bates v. Bar of Arizona, 433 U.S. 350 (1977).

[81] *See* Clark Havighurst, *Regulations of Health Facilities and Services by Certificate of Need*, 59 VA. L. REV. 1143, 1163 n.76 (1973).

[82] This Model Act is a collaborative and ongoing process with assistance from the authors, the Legislature of the State of New Jersey, Children's Hospital of Washington, D.C., and the Association for Safe Kids of New Jersey. Since the initial drafting of this manuscript, virtually all of the provisions of this model act were passed unanimously

BE IT ENACTED by the Senate and General Assembly of the State of _____.

1. The Legislature finds and declares that:

a. Traumatic injuries, such as automobile accidents, bicycle accidents, drownings, and poisonings, are the most common cause of death in children over the age of one; and children have a high death rate in these emergency situations.

b. Children react differently than adults to stress, metabolize drugs differently, and suffer different illnesses and injuries. Because of these differences, children's emergency medical needs should be recognized.

c. Emergency medical systems training programs focus almost exclusively on adults and therefore currently offer inadequate hours of pediatric training. In addition, many emergency medical systems personnel have insufficient clinical experience with children, indicating a public health need to improve training of these personnel in pediatric emergencies.

d. It is the public policy of this State that children are entitled to comprehensive emergency medical and injury prevention systems, including pre-hospital, hospital, and rehabilitative care, and that the public is entitled to accurate information regarding the availability of such systems.

2. As used in this Act:

"Advanced life support" means an advanced level of pre-hospital, interhospital, and emergency service care which includes basic life support functions, cardiac monitoring, cardiac defibrillation, telemetered electrocardiography, administration of antiarrhythmic agents, intravenous therapy, administration of specific medications, drugs and solutions, use of adjunctive ventilation devices, trauma care, and other techniques and procedures authorized in writing by the commissioner pursuant to department regulations and P.L. _____.

"Advisory council" means the Emergency Medical and Injury Prevention Systems for Children Advisory Council established pursuant to section 5 of this Act.

"Basic life support" means a basic level of pre-hospital care which includes patient stabilization, airway clearance, cardiopulmonary resuscitation, hemorrhage control, initial wound care and

---

by the New Jersey House and Senate and signed into law by Governor Florio on September 10, 1992. The passage of this bill capped a decade of effort by Dr. Richard Flyer to improve EMSC services in the state of New Jersey. Similar laws are currently under consideration by the state legislatures of Texas and Minnesota.

fracture stabilization, and other techniques and procedures authorized by the commissioner.

"Commissioner" means the Commissioner of Health.

"Coordinator" means the person coordinating the EMSC program within the Office of Emergency Medical Systems in the Department of Health.

"Department" means the Department of Health.

"EMIPSC program" means the Emergency Medical and Injury Prevention Systems for Children program established pursuant to section 3 of this act, and other relevant activities conducted by the Office of Emergency Medical Systems in the Department of Health in support of prevention programs and the appropriate treatment, transport, and triage of ill or injured children.

"Emergency medical systems personnel" means persons trained to provide emergency medical care, and certified or licensed to do so, whether on a paid or volunteer basis, as part of a basic life support or advanced life support pre-hospital emergency care service or in an emergency department or pediatric critical care or specialty unit in a licensed hospital.

"Pre-hospital care" means emergency medical care or transportation by persons trained to provide emergency medical care, and certified or licensed to do so at the scene of an emergency and while transporting sick or injured persons to a medical care facility or provider.

3. Office of Emergency Medical Systems, Coordinator

a. There is established within the Office of Emergency Medical Systems in the Department of Health, the Emergency Medical and Injury Prevention Systems for Children (EMIPSC) program.

b. The commissioner shall hire a full-time coordinator for the EMIPSC program in consultation with and by the recommendation of the advisory council.

c. The coordinator shall implement the statewide EMIPSC following consultation with, and at the recommendation of, the advisory council. The coordinator shall serve as a liaison to the advisory council.

d. The coordinator may employ professional, technical, research, and clerical staff as necessary within the limits of available appropriations. The provisions of Title ____ of the _____ Statutes shall apply to all personnel so employed.

e. The coordinator may solicit and accept grants of funds from the federal government and from other public and private sources.

f. The coordinator shall annually file a public report with the Legislature on the state of emergency medical services for children in the state of _____.

4. The EMIPSC program shall include, but not be limited to, the establishment of the following:

a. Initial and continuing education programs for emergency medical systems personnel that include training in the emergency care of infants and children;

b. Guidelines for referring children to the appropriate emergency treatment facility;

c. Pediatric equipment guidelines for pre-hospital care;

d. Guidelines for hospital-based emergency departments appropriate for pediatric care to assess, stabilize, and treat critically ill infants and children, either to resolve the problem or to prepare the child for transfer to a pediatric intensive care unit or a pediatric trauma center;

e. Guidelines for pediatric intensive care units, pediatric trauma centers, and intermediate care units fully equipped and staffed by appropriately trained critical care pediatric physicians, surgeons, nurses, and therapists;

f. An interhospital transfer system for critically ill or injured children;

g. Pediatric rehabilitation units staffed by rehabilitation specialists and capable of providing any service required to assure maximum recovery from the physical, emotional, and cognitive effects of critical illness and severe trauma;

h. Guidelines for the implementation of injury prevention programs throughout the state in conjunction with local fire, public safety, and school personnel; and

i. Guidelines for the collection, analysis, and public dissemination of quality assurance information regarding ongoing improvements in the EMIPSC program.

5. Emergency Medical and Injury Prevention Systems for Children Advisory Council

a. There is created an Emergency Medical and Injury Prevention Systems for Children Advisory Council to advise the Office of Emergency Medical Systems and the coordinator of the EMIPSC program on all matters concerning emergency medical systems for children. The advisory council shall assist in the formulation of policy and regulations to effectuate the purposes of this act.

b. The advisory council shall consist of a minimum of 12 public members to be appointed by the Governor, with the advice and consent of the Senate, for a term of three years. Membership of the advisory council shall include: one practicing pediatrician, one pediatric critical care physician, and one pediatric psychiatrist or licensed psychologist, to be appointed upon the recommendation of the _____ chapter of the American Academy of Pediatrics; one pediatric surgeon, to be appointed upon the recommendation of the _____ chapter of the American College of Surgeons; one emergency physician, to be appointed upon the recommendation of the _____ chapter of the American College of Emergency Physicians; one emergency medical technician and one paramedic, to be appointed upon the recommendation of the _____ State First Aid Council; one family practice physician, to be appointed upon the recommendation of the _____ State Nurses Association; and three members, each with a non-medical background, two of whom are parents with one or more children under the age of 18, to be appointed upon the joint recommendation of the Association for Children of _____ and the Children's Defense Fund of _____.

c. Vacancies on the advisory council shall be filled for the unexpired term by appointment of the Governor in the same manner as originally filled. The members of the advisory council shall serve without compensation, but shall be reimbursed for necessary expenses incurred in the performance of their duties. The advisory council shall elect a chairperson, who may select from among the members a vice-chairperson and other officers or subcommittees which are deemed necessary or appropriate. The council may further organize itself in any manner it deems appropriate and enact bylaws as deemed necessary to carry out the responsibilities of the council.

6. Additional administrative procedures. The commissioner shall, pursuant to administrative procedures legislation, P.L. _____, adopt rules and regulations necessary to implement the purposes of this Act.

7. Date of effect. This Act shall take effect immediately.

LEGISLATIVE OVERVIEW

This bill establishes the Emergency Medical and Injury Prevention Systems for Children Program in the Office of Emergency

Medical Systems within the Department of Health of the State of _____. A full-time coordinator of the program shall be hired by the Commissioner of Health upon the recommendation of the Emergency Medical and Injury Prevention Systems for Children Advisory Council established pursuant to section 5 of the bill.

The bill requires the coordinator to implement a statewide program of emergency medical systems for children. The coordinator may employ necessary personnel and solicit and accept grants of public and private funds. The EMIPSC program shall include, but not be limited to, establishment of the following:

(1) initial and continuing education programs for emergency medical systems personnel that include training in the emergency care of infants and children;

(2) guidelines for referring children to the appropriate emergency treatment facility;

(3) pediatric equipment guidelines for pre-hospital care;

(4) guidelines for hospital-based emergency departments appropriate for pediatric care to assess, stabilize, and treat critically ill infants and children either to resolve the problem or to prepare the child for transfer to a pediatric intensive care unit or a pediatric trauma center;

(5) guidelines for pediatric intensive care units, pediatric trauma centers, and intermediate care units fully equipped and staffed by appropriately trained critical care pediatric physicians, surgeons, nurses, and therapists;

(6) an interhospital transfer system for critically ill or injured children;

(7) pediatric rehabilitation units staffed by rehabilitation specialists and capable of providing any service required to assure maximum recovery from the physical, emotional, and cognitive effects of critical illness and severe trauma;

(8) guidelines for the implementation of injury prevention programs throughout the state in conjunction with local fire, public safety, and school personnel; and

(9) guidelines for the collection, analysis, and public dissemination of quality assurance information regarding ongoing improvements in the EMIPSC program.

The commissioner is authorized, pursuant to administrative procedures legislation, P.L. _____, to adopt rules and regulations necessary to effectuate the purposes of the bill.

IX. CONCLUSION

A wealth of empirical evidence clearly demonstrates that effective improvements to EMSC and injury prevention programs could significantly improve the health and future productivity of America's children while saving billions of taxpayers' dollars. Such an approach would provide comprehensive preventive, emergency medical, and educational programs. A reliably funded, multi-modal effort is essential if unnecessary child/infant mortality is to be significantly reduced in the United States.

The proposed multi-modal health improvement program for America's infants and children will require at least some funding at both state and federal levels. Such funding will require a committed coalition of legislators and administrators who understand this rare opportunity to pursue an objective that is both morally and economically compelling.

Concerted action by concerned citizen, governmental, religious, community, and professional groups could produce enormous gains. Building a broad-based political coalition to move responsible programs for children's emergency health care and injury prevention through the legislative process during times of fiscal austerity should be a first order of business.

Fiscal responsibility is important in generating this broad political base for children's health programs. Investing in injury/illness prevention programs and EMSC is both compassionate and financially prudent because it can reduce future health care expenditures. As outlined above,[83] cost-benefit analyses indicate that improving children's health and reducing long-term disability is an effective, responsible way of reducing long-term health care expenditures.

According to polling research done in 1987, U.S. voters do not resent government policies that help other people's children.[84] Moreover, the proposed assistance consists entirely of health care services and education for children that could not be misused by irresponsible adults. With such compelling goals and with the public's justified fear of misuse of funds laid to rest, we must now create the necessary coalitions and legislation

---

[83] *See supra* part IV.

[84] E.J. Dionne, Jr., *Children Emerge as Issue for Politicians*, N.Y. TIMES, Sept. 27, 1987, at A36.

to improve the health and future economic productivity of American infants and children.

What is needed now is action. Systems for injury prevention, improved emergency medical care, and preventive health services for infants and children need not wait for reform in the adult health care system. An extensive bipartisan political and educational campaign will help to make emergency medical systems and injury prevention programs for children a top political priority in the United States. Such a campaign should involve the health care industry, families, business, labor, the national media, and religious and professional groups, as well as state and federal government. With limited expenditures we can save lives and billions of taxpayers' dollars while providing a model of how multi-disciplinary analysis can improve the nature of governmental decisionmaking. We should do no less for the children who will become the future of our great nation.

COMMENT

DEFINING A "BUSINESS TRUST": PROPOSED AMENDMENT OF SECTION 101(9) OF THE BANKRUPTCY CODE

If a business organization wishes to file for bankruptcy, the first issue that it will confront is whether it is eligible to file. This may seem to be a simple question, but for a business trust wishing to file under chapter 11 the answer is unclear. This lack of clarity is inefficient because it leads to unnecessary lawsuits.

After describing the current treatment of business trusts by the Bankruptcy Code and the courts, this Comment will propose a modification of section 101(9) of the Code.[1] Consistent with the idea of law as a seamless web, the proposed definition of business trust refers to both state law and to federal securities law.

I. WHAT IS A BUSINESS TRUST?

First, a preliminary matter: what is a business trust? It is like a corporation, except that the managers are called trustees and the shareholders are called beneficiaries. Certificates are issued instead of shares, and the firm's capital is called its corpus.[2] A business trust is sometimes called a Massachusetts trust or a common law trust.

Why would an organization form a business trust rather than a corporation or a partnership? A business trust is more flexible than a corporation,[3] and it allows for limited liability, which a

---

[1] 11 U.S.C. § 101(9) (Supp. II 1990).

[2] For definitions, see BLACK'S LAW DICTIONARY 199 (business trust), 277 (common-law trust), 974 (Massachusetts trust), 1515 (trust estates as business companies) (6th ed. 1990). For a general discussion of business trusts, see Herbert B. Chermside, Jr., Annotation, *Modern Status of the Massachusetts or Business Trust*, 88 A.L.R.3D 704 (1978); 13 AM. JUR. 2D *Business Trusts* (1964); 12A C.J.S. *Business Trusts* (1980).

[3] *See* 16A WILLIAM M. FLETCHER, FLETCHER CYCLOPEDIA OF THE LAW OF PRIVATE CORPORATIONS § 8232 (perm. ed. 1988); 13 AM. JUR. 2D *Business Trusts* § 5 (1964); 12A C.J.S. *Business Trusts* § 7 (1980). For example, a business trust's Declaration of Trust might allow it to set its quorum at 30%, whereas a corporation would be obliged by law to set its quorum at 50%. *See, e.g.*, JANUS VENTURE FUND, INC. & JANUS TWENTY FUND, INC., PROXY STATEMENT 3 (May 28, 1992) [hereinafter JANUS PROXY STATEMENT].

partnership does not. Administrative costs are sometimes lower for a business trust than for a corporation.[4] Finally, some types of business trusts, such as Real Estate Investment Trusts ("REITs") and regulated investment companies, escape corporate double taxation.[5]

In the past, some well-known business organizations, such as the Standard Oil Trust,[6] have been business trusts. Today, the giant Fidelity Magellan Fund, which has over $21.6 billion in assets[7] and is the largest general equity mutual fund in the United States,[8] is organized as a Massachusetts business trust,[9] and half of the new investment companies in the mid-1980s were Massachusetts business trusts.[10] Clearly, business trusts are worth some study.[11]

II. CURRENT TREATMENT OF BUSINESS TRUSTS IN BANKRUPTCY

A. *The Bankruptcy Code*

The starting point for an analysis of the status of business trusts under the Bankruptcy Code is the definition of who may be a debtor under chapter 11: "Only a person that may be a debtor under chapter 7 of this title, except a stockbroker or a commodity broker, and a railroad may be a debtor under chapter 11 of this title."[12] One then turns to the definition of who may

---

[4] *See* JANUS PROXY STATEMENT, *supra* note 3, at 2.

[5] *See* 26 U.S.C. §§ 851–859; Chermside, *supra* note 2, at 711; Robert D.M. Flannigan, *Business Trusts—Past and Present*, 6 EST. & TR. Q. 375, 382 (1984); Sheldon A. Jones et al., *The Massachusetts Business Trust and Registered Investment Companies*, 13 DEL. J. CORP. L. 421, 448–49, 453–54 (1988).

[6] Flannigan, *supra* note 5, at 380.

[7] FIDELITY MAGELLAN FUND, SEMIANNUAL REPORT 9 (Sept. 30, 1992).

[8] *See, e.g.*, Carole Gould, *Another New Face at Magellan*, N.Y. TIMES, May 17, 1992, § 3, at 16.

[9] FIDELITY MAGELLAN FUND, PROSPECTUS 9 (May 30, 1991).

[10] Jones et al., *supra* note 5, at 422.

[11] To be fair, business trusts do have limitations. "To date [the business trust] appears to have seen only limited use as an entity through which the actual physical operations of a business (other than an investment business) are carried on." Flannigan, *supra* note 5, at 384.

[12] 11 U.S.C. § 109(d) (1988). In addition, § 109(a) imposes a territorial requirement: "Notwithstanding any other provision of this section, only a person that resides or has a domicile, a place of business, or property in the United States, or a municipality, may be a debtor under this title." *Id.* § 109(a).

be a debtor under chapter 7: basically, any "person," except for a railroad, insurance company, or bank, "may be a debtor under chapter 7."[13]

What, then, is a "person" under the Bankruptcy Code? "'[P]erson' includes individual, partnership, and corporation, but does not include governmental unit . . . ."[14] In turn, "corporation" includes a "business trust."[15] However, estates and trusts are not "persons" (and are therefore ineligible to be debtors) because they are defined elsewhere as "entities."[16]

Thus, a business trust may be a debtor eligible under chapter 11, but a mere estate or trust may not. Unfortunately, the Bankruptcy Code nowhere defines "business trust," and legislative history is not very helpful.

The original 1898 Bankruptcy Act did not mention business trusts.[17] In 1926, the Act was amended to bring within the definition of a corporation "any business conducted by a trustee or trustees wherein beneficial interest or ownership is evidenced by certificate or other written instrument."[18] The Bankruptcy Commission which drafted the current Bankruptcy Code eliminated the reference to certificates or other written instruments because it felt that "[t]he requirement gives undue significance to an evidentiary formality."[19] This implies that business trusts without certificates or written instruments are now eligible to be debtors. On the other hand, the House and Senate reports did not intend "to expand the category of trusts which are included within the term corporation,"[20] and, consequently, some bankruptcy courts act as if the law still required transferable certificates.[21]

---

[13] *Id.* § 109(b).

[14] *Id.* § 101(41) (Supp. II 1990).

[15] *Id.* § 101(9)(A)(v).

[16] *Id.* § 101(15).

[17] FLETCHER, *supra* note 3, § 8267; 1 WILLIAM L. NORTON, JR., NORTON BANKRUPTCY LAW AND PRACTICE § 8.08 (1992).

[18] Bankruptcy Act of 1898, Pub. L. No. 62-57, § 1(8), 30 Stat. 544, 544 (1926).

[19] H.R. Doc. No. 137, 93d Cong., 1st Sess., pt. 2, at 7 (1973).

[20] NORTON, *supra* note 17, § 8.08 (citing H.R. REP. No. 595, 95th Cong., 1st Sess. 309 (1977), *reprinted in* 1978 U.S.C.C.A.N. 5963, 6266, and S. REP. No. 989, 95th Cong., 2d Sess. 22 (1978), *reprinted in* 1978 U.S.C.C.A.N. 5787, 5808).

[21] *See, e.g.*, *In re* Heritage N. Dunlap Trust, 120 B.R. 252 (Bankr. D. Mass. 1990); *In re* Woodsville Realty Trust, 120 B.R. 2 (Bankr. D.N.H. 1990); *In re* Village Green Realty Trust, 113 B.R. 105 (Bankr. D. Mass. 1990); *In re* L & V Realty Trust, 61 B.R. 423 (Bankr. D. Mass. 1986).

B. *The Bankruptcy Courts*

At least one court has referred to the decisions on business trusts' eligibility to be debtors as "if not hopelessly divided, at least certainly divergent."[22] Out of forty-nine cases since 1978 that have dealt with this topic, thirty-six ("negative cases") have held that the entity was not a business trust and, hence, ineligible to be a debtor;[23] the remaining thirteen ("positive cases") have found that the entity was a business trust and was eligible to be a debtor.[24] In analyzing these cases, one can take two

---

[22] *Village Green*, 113 B.R. at 113.

[23] *See In re* Action Roofing & Supply Co., 137 B.R. 217 (Bankr. S.D. Tex. 1991); *In re* Mohan Kutty Trust, 134 B.R. 987 (Bankr. M.D. Fla. 1991); *In re* Parade Realty, Inc., Employees Retirement Pension Trust, 134 B.R. 7 (Bankr. D. Haw. 1991); Findley v. Blinken (*In re* Joint E. & S. Dist. Asbestos Litig.), 129 B.R. 710 (E. & S.D.N.Y. 1991), *vacated*, 982 F.2d 721 (2d Cir. 1992); *In re* Hemex Liquidation Trust, 129 B.R. 91 (Bankr. W.D. La. 1991); *In re* BKC Realty Trust, 125 B.R. 65 (Bankr. D.N.H. 1991); *Heritage N. Dunlap Trust*, 120 B.R. 252; *Woodsville Realty Trust*, 120 B.R. 2; *In re* Constitutional Trust No. 2-562, 114 B.R. 627 (Bankr. D. Minn. 1990); *In re* Margaret E. DeHoff Trust I, 114 B.R. 189 (Bankr. W.D. Mo. 1990); *In re* Ralph Faber Trust, 113 B.R. 599 (Bankr. D.N.D. 1990); *Village Green*, 113 B.R. 105; *In re* The Ophir Trust, 112 B.R. 956 (Bankr. E.D. Wis. 1990); Westchester County Civil Serv. Employees Ass'n Benefit Fund v. Westchester County (*In re* Westchester County Civil Serv. Employees Ass'n Benefit Fund), 111 B.R. 451 (Bankr. S.D.N.Y. 1990); *In re* St. Augustine Trust, 109 B.R. 494 (Bankr. M.D. Fla. 1990); *In re* Medallion Realty Trust, 103 B.R. 8 (Bankr. D. Mass. 1989), *aff'd*, 120 B.R. 245 (D. Mass. 1990); *In re* Vivian A. Skaife Irrevocable Trust Agreement No. 1, 90 B.R. 325 (Bankr. E.D. Tenn. 1988); *In re* Johnson, 82 B.R. 618 (Bankr. S.D. Fla. 1988); *In re* Walker, 79 B.R. 59 (Bankr. M.D. Fla. 1987); *In re* Betty L. Hays Trust, 65 B.R. 665 (Bankr. D. Neb. 1986); *In re* Estate of Whiteside by Whiteside, 64 B.R. 99 (Bankr. E.D. Cal. 1986); *In re* Jay M. Weisman Irrevocable Children's Trust of 1981, 62 B.R. 286 (Bankr. M.D. Fla. 1986); *In re* Milani Family Irrevocable Trust, 62 B.R. 6 (Bankr. S.D. Fla. 1986); *L & V Realty Trust*, 61 B.R. 423; Mosby v. Boatmen's Bank (*In re* Mosby), 46 B.R. 175 (Bankr. E.D. Mo.), *aff'd*, 61 B.R. 636 (E.D. Mo. 1985), *aff'd*, 791 F.2d 628 (8th Cir. 1986); *In re* Don A. Johnson Revocable Trust, Bankr. No. 82C-02471 (Bankr. D. Utah Feb. 15, 1984) (LEXIS, Bkrtcy library, Cases file); *In re* Armstead & Margaret Wayson Trust, 29 B.R. 58 (Bankr. D. Md. 1982); *In re* Dolton Lodge Trust No. 35188, 22 B.R. 918 (Bankr. N.D. Ill. 1982); *In re* SCR Trust, 20 B.R. 17 (Bankr. M.D. La. 1982); *In re* North Shore Nat'l Bank, Land Trust No. 362, 17 B.R. 867 (Bankr. N.D. Ill. 1982); *In re* Cahill, 15 B.R. 639 (Bankr. E.D. Pa. 1981); *In re* Citizens Bank & Trust Co., 8 B.R. 812 (N.D. Ill. 1981); *In re* Old Second Nat'l Bank, 7 B.R. 37 (Bankr. N.D. Ill. 1980); *In re* G-2 Realty Trust, 6 B.R. 549 (Bankr. D. Mass. 1980); *In re* Cohen, 4 B.R. 201 (Bankr. S.D. Fla. 1980); *In re* Treasure Island Land Trust, 2 B.R. 332 (Bankr. M.D. Fla. 1980).

[24] *See In re* Affiliated Food Stores, Inc. Group Benefit Trust, 134 B.R. 215 (Bankr. N.D. Tex. 1991); Loux v. Gabelhart (*In re* Carriage House, Inc.), 120 B.R. 754 (Bankr. D. Vt. 1990), *aff'd*, 146 B.R. 362 (D. Vt. 1992); *In re* 640 Harvard St. Trust, Civ. No. 90-18-D, 1990 WL 176100 (D.N.H. Feb. 14, 1990); *In re* Michigan Real Estate Ins. Trust, 87 B.R. 447 (E.D. Mich. 1988); *In re* Cooper Properties Liquidating Trust, 61 B.R. 531 (Bankr. W.D. Tenn. 1986); *In re* Captran Creditors Trust, 53 B.R. 741 (Bankr. M.D. Fla. 1985); *In re* Arehart, 52 B.R. 308 (Bankr. M.D. Fla. 1985); *In re* Gonic Realty Trust, 50 B.R. 710 (Bankr. D.N.H. 1985); Merrill v. Abbott (*In re* Independent Clearing House Co.), 41 B.R. 985 (Bankr. D. Utah 1984); *In re* Tru Block Concrete Prods., Inc., 27 B.R. 486 (Bankr. S.D. Cal. 1983); *In re* Dreske Greenway Trust, 14 B.R. 618 (E.D. Wis. 1981); *In re* Ponn Realty Trust, 4 B.R. 226 (Bankr. D. Mass. 1980); *In re* Maidman, 2 B.R. 569 (S.D.N.Y. 1980), *aff'd*, 668 F.2d 682 (2d Cir. 1982).

approaches. The first approach is a categorical approximation: whole categories of trusts either qualify or fail to qualify as business trusts. The second approach looks for underlying characteristics of business trusts.

1. Categories of Trusts

a. *Types of trusts that are ineligible to be debtors.* There are three main types of trusts that are ineligible to be debtors. First, a family trust rarely qualifies as a business trust, even when it does business.[25] Nineteen of the thirty-six negative cases involved family trusts.[26]

On the other hand, a few cases have gone the other way. The trust in *In re Carriage House, Inc.*[27] was held to be a business trust, even though it was run by and for a married couple, and the trust in *In Re Gonic Realty Trust*[28] was found to be a business trust even though the beneficiaries were the trustee and his wife. In *In re Dreske Greenway Trust*, the grantors and trustees were Dreske's children, and the beneficiaries were the grantors' descendants, but the court held that the trust was a business trust and eligible as a debtor because it was "created by the Dreskes for the specific purpose of engaging in business."[29]

The exclusion of family trusts makes sense textually because, as stated earlier, estates and trusts are "entities," not "persons,"[30] and family trusts are what come to mind when one thinks of estates and trusts.

A category that is related to the family trust is the trust that is nothing more than the alter ego of its creator. It, too, is ineligible to be a debtor. For example, in *In re Parade Realty,*

---

[25] By "family trust," this Comment means both a traditional trust in which the corpus is preserved for the comfort and maintenance of the beneficiaries—who are usually the settlor's or grantor's relatives—and a trust in which the trustees and beneficiaries are relatives or close friends—the trust equivalent of a close corporation.

[26] *See BKC*, 125 B.R. 65; *Woodsville Realty Trust*, 120 B.R. 2; *Margaret E. DeHoff Trust I*, 114 B.R. 189; *Ralph Faber Trust*, 113 B.R. 599; *Ophir Trust*, 112 B.R. 956; *St. Augustine Trust*, 109 B.R. 494; *Vivian A. Skaife Irrevocable Trust*, 90 B.R. 325; *Johnson*, 82 B.R. 618; *Walker*, 79 B.R. 59; *Betty L. Hays Trust*, 65 B.R. 99; *Whiteside*, 65 B.R. 99 (probate estate); *Jay M. Weisman Irrevocable Children's Trust*, 62 B.R. 286; *Milani Family Irrevocable Trust*, 62 B.R. 6; *L & V Realty Trust*, 61 B.R. 423; *Mosby*, 46 B.R. 175; *Don A. Johnson Revocable Trust*, Bankr. No. 82C-02471 (LEXIS, Bkrtcy library, Cases file); *Wayson Trust*, 29 B.R. 58; *Cohen*, 4 B.R. 201.

[27] 120 B.R. 754.

[28] 50 B.R. 710.

[29] 14 B.R. 618, 623 (E.D. Wis. 1981).

[30] *See supra* notes 14–16 and accompanying text.

Inc., Employees Retirement Pension Trust,[31] the creator of the trust was its trustee, sole beneficiary, and sole employee. In denying bankruptcy protection, the court emphasized the lack of outside investors.

Land trusts constitute a second major category of trusts that are usually ineligible to be debtors. A land trust, also called an Illinois land trust, is an arrangement whereby a recorded deed gives the trustee full power to deal with the property.[32] However, an unrecorded agreement can return a great deal of power to the beneficiary, so the beneficiary of an Illinois land trust is often more active than the beneficiary of a Massachusetts business trust. Moreover, an Illinois trust usually lasts only for a definite term, while a Massachusetts trust can last indefinitely.

Six of the thirty-six negative cases involved land trusts.[33] On the other hand, *In re Arehart* held that a trust that was formed "to acquire, develop, improve, operate, lease,, [sic] and/or sell the trust lands . . . to the economic benefit of the BENEFICIARIES" was a business trust.[34] Similarly, *In re Maidman*[35] held that the trustee of a land trust could qualify as a debtor under chapter XII of the old Bankruptcy Act, which corresponds to chapter 11 of the current Bankruptcy Code.

One commentator suggests a policy reason why land trusts should be ineligible to file for bankruptcy:

> The effect of holding that such a trust qualified for bankruptcy would be to permit an individual to fragment bankruptcy proceedings among various assets which would be separately administered. An individual might, for example, file a reorganization case for "land trust" property and a liquidation case for other property, resulting in differing treatment of creditors.[36]

Norton suggests that land trusts should sometimes be allowed to qualify as business trusts, depending on the purpose of the trust and how business-like the trust appears to be.[37]

---

[31] 134 B.R. 7 (Bankr. D. Haw. 1991).

[32] *See generally* NORTON, *supra* note 17; BLACK'S LAW DICTIONARY 880 (6th ed. 1990).

[33] *See In re* Dolton Lodge Trust No. 35188, 22 B.R. 918 (Bankr. N.D. Ill. 1982); *In re* North Shore Nat'l Bank of Chicago, Land Trust No. 362, 17 B.R. 867 (Bankr. N.D. Ill. 1982); *In re* Citizens Bank & Trust Co., 8 B.R. 812 (N.D. Ill. 1981); *In re* Old Second Nat'l Bank, 7 B.R. 37 (Bankr. N.D. Ill. 1980); *In re* Cohen, 4 B.R. 201 (Bankr. S.D. Fla. 1980); *In re* Treasure Island Land Trust, 2 B.R. 332 (Bankr. M.D. Fla. 1980).

[34] 52 B.R. 308, 308 (Bankr. M.D. Fla. 1985).

[35] 2 B.R. 569 (S.D.N.Y. 1980).

[36] NORTON, *supra* note 17, § 8.08.

[37] *Id.*

The third category of trusts that fail to qualify as business trusts is nominee trusts. A nominee trust is "[a]n arrangement for holding title to real property under which one or more persons or corporations, pursuant to a written declaration of trust, declare that they will hold any property that they acquire as trustees for the benefit of one or more undisclosed beneficiaries."[38] Four of the thirty-six negative cases dealt with nominee trusts.[39]

If the trust agreement says that the trust is not to be considered a business trust, the court will usually agree.[40] However, the opposite is not always true: even if a trust calls itself a business trust, the court will not necessarily accord that statement binding effect.[41]

b. *Types of trusts that are eligible to be debtors.* Liquidating trusts, which are created to wind up the affairs of a predecessor corporation which has gone bankrupt, usually qualify as business trusts. Three of the thirteen positive cases involved liquidating trusts.[42] However, two cases have gone the other way: *In re Action Roofing & Supply Co.*[43] held that business trusts do not exist under Texas law, so a liquidating trust is not a proper debtor under chapter 11; and *In re Hemex Liquidation Trust*[44] held that a liquidating trust is not a business trust.

---

[38] BLACK'S LAW DICTIONARY 1050 (6th ed. 1990).

[39] *See In re* Heritage N. Dunlap Realty Trust, 120 B.R. 252 (Bankr. D. Mass. 1990); *In re* Woodsville Realty Trust, 120 B.R. 2 (Bankr. D.N.H. 1990); *In re* Village Green Realty Trust, 113 B.R. 105 (Bankr. D. Mass. 1990); *In re* Medallion Realty Trust, 103 B.R. 8 (Bankr. D. Mass. 1989), *aff'd*, 120 B.R. 245 (D. Mass. 1990).

[40] *See In re* Constitutional Trust No. 2-562, 114 B.R. 627 (D. Minn. 1990); *In re* Cohen, 4 B.R. 201 (Bankr. S.D. Fla. 1980); *In re* Treasure Island Land Trust, 2 B.R. 332 (Bankr. M.D. Fla. 1980). Similarly, in *In re* Roxy Roller Rink Joint Venture, 67 B.R. 474 (Bankr. S.D.N.Y. 1985), *aff'd sub nom.* Twins Roller Corp. v. Roxy Roller Rink Joint Venture, 67 B.R. 479 (S.D.N.Y. 1986), the court decided that the joint venture was a partnership, not a corporation, because the joint venture agreement said that it would be governed by the laws relating to general partnership. *Id.* at 482.

[41] *See Medallion Realty*, 120 B.R. 245 (finding that entity registered as business trust under state law was a partnership); *In re* The Ophir Trust, 112 B.R. 956 (Bankr. E.D. Wis. 1990) (disregarding declaration in trust agreement that it was a business trust). *But see* Loux v. Gabelhart (*In re* Carriage House, Inc.), 120 B.R. 754 (Bankr. D. Vt. 1990), *aff'd*, 146 B.R. 352 (D. Vt. 1992).

[42] *See In re* Cooper Properties Liquidating Trust, 61 B.R. 531 (Bankr. W.D. Tenn. 1986); *In re* Captran Creditors Trust, 53 B.R. 741 (Bankr. M.D. Fla. 1985); *In re* Tru Block Concrete Prods., Inc., 27 B.R. 486 (Bankr. S.D. Cal. 1983).

[43] 137 B.R. 217 (Bankr. S.D. Tex. 1991).

[44] 129 B.R. 91 (Bankr. W.D. La. 1991).

c. *Trusts that may or may not be eligible to be debtors.* The courts are divided about pension plans, benefit plans, and the like. Two cases say that they are business trusts, and two say that they are not. *In re Affiliated Food Stores, Inc. Group Benefit Trust* held that a trust which provided low-cost group medical benefits was a business trust because its "principal purpose is not effectuating a settlor's estate plan through the preservation of a trust *res*," and because the trust's capital was the product of voluntary pooling, not a donor's largesse.[45] Similarly, *In re Michigan Real Estate Insurance Trust* held that "an unincorporated entity funded by voluntary contributions from individuals in the real estate industry in southeastern Michigan for the purpose of reducing the cost of health care for its members" was eligible to be a chapter 7 debtor.[46] On the other hand, *In re Westchester County Civil Service Employees Ass'n Benefit Fund* held that a trust which provided health and welfare benefits for civil service employees was not eligible for chapter 11 relief because it "is not engaged in any business activities and does not generate any business profits,"[47] and *In re Cahill*[48] held that a pension plan was not a business trust.

2. Characteristics of a Business Trust

Courts rarely conclude that a trust is eligible to be a debtor simply because of its category; rather, they find that a trust is a business trust on a case-by-case basis. Thus, it is important to examine the characteristics of a business trust. The most important characteristics are a profit-making purpose, voluntary pooling of capital into transferable certificates or shares, and creditors. The outward trappings of a business are sometimes cited. Simply doing business may be sufficient, provided that the trust falls outside the "ineligible" categories.

Almost seventy years ago, in *Hecht v. Malley*,[49] the Supreme Court defined a Massachusetts trust as

> a form of business organization . . . consisting essentially of an arrangement whereby property is conveyed to trustees, in accordance with the terms of an instrument of trust, to

[45] 134 B.R. 215, 218 & n.2 (Bankr. N.D. Tex. 1991).
[46] 87 B.R. 447, 449 (E.D. Mich 1988).
[47] 111 B.R. 451, 456 (Bankr. S.D.N.Y. 1990).
[48] 15 B.R. 639 (Bankr. E.D. Pa. 1981).
[49] 265 U.S. 144 (1924).

> be held and managed for the benefit of such persons as may from time to time be the holders of transferable certificates issued by the trustees showing the shares into which the beneficial interest in the property is divided. These certificates, which resemble certificates for shares of stock in a corporation and are issued and transferred in like manner, entitle the holders to share ratably in the income of the property, and, upon termination of the trust, in the proceeds.[50]

Eleven years later the Court said that a business trust was characterized by business purpose, title to property held by trustees, centralized management, continuity of existence, transferability of interests, and limited liability.[51]

a. *Profit-making purpose*. Profit-making purpose is similar to the "business purpose" mentioned in *Morrissey*.[52] An early and influential Bankruptcy Code case on business trusts, *In re Treasure Island Land Trust*,[53] distinguished a business trust from a nonbusiness trust as follows:

> [B]usiness trusts are created for the purpose of carrying on some kind of business or commercial activity for profit; the object of a nonbusiness trust is to protect and preserve the

---

[50] *Id.* at 146–47. Many state statutes and cases echo the *Hecht* definition. *See, e.g.*, ALA. CODE § 19-3-60 (1990); ARIZ. REV. STAT. ANN. § 10-501 (1990); IND. CODE § 23-5-1-2(a) (1988); KAN. STAT. ANN. § 17-2028(a) (1988); KY. REV. STAT. ANN. § 386.370 (Michie/Bobbs-Merrill 1984); MD. CORPS. & ASS'NS CODE ANN. § 3-101(d) (1993); MISS. CODE ANN. § 79-15-3(1) (1989); MONT. CODE ANN. § 35-5-101 (1991); N.Y. GEN. ASS'NS LAW § 2(2) (McKinney 1942); N.C. GEN. STAT. § 39-44 (1984); OHIO REV. CODE ANN. § 1746.01(A) (Anderson 1992); OR. REV. STAT. § 128.560 (1991); S.C. CODE ANN. § 33-53-10 (Law. Co-op. 1990); S.D. CODIFIED LAWS ANN. § 47-14-1 (1991); TENN. CODE ANN. § 48-3-202(a) (Supp. 1992); TEX. REV. CIV. STAT. ANN. art. 6138A, § 23.1(A)(1) (West Supp. 1993); UTAH CODE ANN. § 7-1-103(5) (Supp. 1992); WASH. REV. CODE § 23.90.020 (1985); Mosby v. Boatmen's Bank (*In re* Mosby), 46 B.R. 175, 177 (Bankr. E.D. Mo.), *aff'd*, 61 B.R. 636 (E.D. Mo. 1985), *aff'd*, 791 F.2d 628 (8th Cir. 1986); *In re* Armstead & Margaret Wayson Trust, 29 B.R. 58, 59 (Bankr. D. Md. 1982).

[51] Morrissey v. Commissioner, 296 U.S. 344, 359 (1935).

[52] At least nine bankruptcy cases mention the six *Morrissey* characteristics. *See* Loux v. Gabelhart (*In re* Carriage House, Inc.), 120 B.R. 754 (Bankr. D. Vt. 1990), *aff'd*, 146 B.R. 352 (D. Vt. 1992); *In re* Constitutional Trust No. 2-562, 114 B.R. 627 (Bankr. D. Minn. 1990); *In re* Margaret E. DeHoff Trust I, 114 B.R. 189 (Bankr. W.D. Mo. 1990); *In re* Vivian A. Skaife Irrevocable Trust Agreement No. 1, 90 B.R. 325 (Bankr. E.D. Tenn. 1988); *In re* Betty L. Hays Trust, 65 B.R. 665 (Bankr. D. Neb. 1986); *Mosby*, 61 B.R. 636; *In re* Don A. Johnson Revocable Trust, Bankr. No. 82C-02471 (Bankr. D. Utah Feb. 15, 1984) (LEXIS, Bkrtcy library, Cases file); *In re* Tru Block Concrete Prods., Inc., 27 B.R. 486 (Bankr. S.D. Cal. 1983); *In re* Cohen, 4 B.R. 201 (Bankr. S.D. Fla. 1980).

[53] 2 B.R. 332 (Bankr. M.D. Fla. 1980).

> trust res . . . . It is the business trust's similarity to a corporation that permits it to be a debtor in bankruptcy.[54]

In re Universal Clearing House Co.[55] emphasized the importance of this characteristic: "The primary consideration in most cases has been the overt purpose of the trust. If its purpose is to protect the trust res, the trust is found to be ineligible for bankruptcy protection. If the purpose is profit oriented, the trust is found to be an eligible business trust."[56]

The importance of the trust's profit-making motive is evident in *In re Arehart*.[57] Even though the trust had no "income from any source, and never conducted any business in the conventional sense,"[58] the court held that it was a business trust because it was "created for the purpose of carrying on a business or commercial activity for profit."[59]

Sometimes a trust that does not start with a business purpose is held to be a business trust if it subsequently acts like a business. For example, the court found that the trust in *In re Captran Creditors Trust* was eligible to be a debtor, even though it originally had "no objective to continue or engage in the conduct of a trade or business," because it had constructed two additional buildings at one of its projects.[60]

b. *Voluntary pooling of capital.* Various courts distinguish a business trust from a nonbusiness trust by saying that the beneficiaries of a business trust contribute something of value (either money or services) to the trust. In contrast, the beneficiaries of a nonbusiness trust merely benefit from the grantor's or settlor's largesse. For example, *In re Treasure Island Land Trust*[61] noted:

---

[54] *Id.* at 334; *see also In re* Ralph Faber Trust, 113 B.R. 599, 601 (Bankr. D.N.D. 1990) (looking at "whether the express purpose of the trust was principally profit or protection of the trust property").

[55] 60 B.R. 985 (D. Utah 1986), *aff'd in relevant part sub nom. In re* Independent Clearing House Co., 77 B.R. 843 (D. Utah 1987).

[56] *Id.* at 991; *see also In re* BKC Realty Trust, 125 B.R. 65, 68–69 (D.N.H. 1991) ("To qualify as a business trust the trust must not only be doing business, and have some of the significant attributes of a corporation, but also must have been formed primarily for a business purpose."); *Betty L. Hays*, 65 B.R. at 668 (holding that trust's purpose must be profit-oriented).

[57] 52 B.R. 308 (Bankr. M.D. Fla. 1985).

[58] *Id.* at 309.

[59] *Id.* at 311.

[60] 53 B.R. 741, 744 (Bankr. M.D. Fla. 1985).

[61] 2 B.R. 332 (Bankr. M.D. Fla. 1980).

> The Trust was not created by a grant of a settlor, but was formed through the voluntary association of unrelated persons and subscriptions sold through a prospectus. The "beneficial interests" are very much like shares in that they are equal in value, held by a large number of people in varying amounts, and are transferrable [sic].[62]

Similarly, *In re Medallion Realty Trust* noted: "Congress intended to permit bankruptcy relief for all trusts which are created for the purpose of transacting business and whose beneficiaries make a contribution in money or money's worth to the enterprise, without regard to whether the trust has characteristics of a corporation such as separate certificate of ownership."[63]

Other cases lend further support to the importance of contributions by the beneficiaries. At least two cases in which entities were found not to be business trusts mentioned that the beneficiaries had made no contribution to the trust.[64]

c. *Creditors.* No case flatly says that a business trust must have creditors, perhaps because this characteristic seems too obvious in the bankruptcy context. However, four cases in which an entity was judged not to be a business trust mentioned that the trust had no trade creditors,[65] and one case said that it had only a few.[66] Conversely, in five cases, the trust did have unsecured creditors, and it was found eligible to be a debtor.[67]

d. *The outward trappings of a business.* Many cases say that a business trust should have its own name, office, officers, employees, and bank account; keep proper books and records;

---

[62] *Id.* at 334; *see also In re* Margaret E. DeHoff Trust I, 114 B.R. 189, 191 (Bankr. W.D. Mo. 1990); Mosby v. Boatmen's Bank (*In re* Mosby), 46 B.R. 175, 177 (Bankr. E.D. Mo.), *aff'd*, 61 B.R. 636 (E.D. Mo. 1985), *aff'd*, 791 F.2d 628 (8th Cir. 1986); *In re* Armstead & Margaret Wayson Trust, 29 B.R. 58, 59 (Bankr. D. Md. 1982).

[63] 103 B.R. 8, 11–12 (Bankr. D. Mass. 1989), *aff'd*, 120 B.R. 245 (D. Mass. 1990); *see also In re* 640 Harvard St. Trust, Civ. No. 90-18-D, 1990 WL 176100 (D.N.H. Feb. 14, 1990).

[64] *See In re* BKC Realty Trust, 125 B.R. 65 (Bankr. D.N.H. 1991); *In re* Walker, 79 B.R. 59 (Bankr. M.D. Fla. 1987).

[65] *See In re* Jay M. Weisman Irrevocable Children's Trust of 1981, 62 B.R. 286 (Bankr. M.D. Fla. 1986); *In re* Dolton Lodge Trust No. 35188, 22 B.R. 918 (Bankr. N.D. Ill. 1982); *In re* Cohen, 4 B.R. 201 (Bankr. S.D. Fla. 1980); *Treasure Island*, 2 B.R. 332.

[66] *See In re* Village Green Realty Trust, 113 B.R. 105 (Bankr. D. Mass. 1990).

[67] *See In re* Affiliated Food Stores, Inc. Group Benefit Trust, 134 B.R. 215 (Bankr. N.D. Tex. 1991); *Medallion Realty Trust*, 103 B.R. at 9 ($4 million in secured debts and $300,000 in unsecured debts owed to 100 creditors); *In re* Cooper Properties Liquidating Trust, 61 B.R. 531 (Bankr. W.D. Tenn. 1986); *In re* Arehart, 52 B.R. 308, 309 (Bankr. M.D. Fla. 1985) ($2 million of secured debt and $18,856 of unsecured debt); *In re* Gonic Realty Trust, 50 B.R. 710 (Bankr. D.N.H. 1985).

and file tax returns.[68] However, such cases usually involve family trusts or land trusts, so it is unclear whether a non-family or non-land trust would have to meet these standards. It is unlikely that the absence of just one or two of these factors would make a trust ineligible to be a debtor, especially given cases where trusts had almost no characteristics of a business but were still found to be business trusts. For example, the trust in *In re Carriage House, Inc.*[69] never issued certificates or filed tax returns, but it was still found to be a business trust. In *In re Captran Creditors Trust*, the trust "ha[d] no offices of its own, no employees, no business license, ha[d] not filed income tax returns, and ha[d] not obtained loans or been extended credit."[70] Nevertheless, this liquidating trust was deemed eligible to be a debtor under chapter 11.[71] In *In re Arehart*, "the Trustee never maintained a place of business for the Trust, never had any employees, nor has he filed any tax returns [or] obtained an occupational license."[72] Nonetheless, the court held that it was a business trust.[73] Lastly, the trust in *In re Gonic Realty Trust* had no office and no employees, but it was held to be a business trust because it was conducting business and had unsecured debt.[74]

e. *Does any trust that does business qualify as a "business trust"?* A pre-1978 case, *Pope & Cottle Co. v. Fairbanks Realty Trust*, implied that simply doing business was insufficient: "If Congress had intended to include within its definition of 'corporation' all trusts . . . in which the trustees are empowered to conduct a business, it could easily have done so."[75] However, the specific point about which *Pope & Cottle* was concerned (the fact that the shares were not evidenced by certificates) was addressed by the 1978 Code.[76]

---

[68] *See, e.g.*, *In re* Mohan Kutty Trust, 134 B.R. 987, 989 (Bankr. M.D. Fla. 1991); *In re* L & V Realty Trust, 61 B.R. 423 (Bankr. D. Mass. 1986); *Dolton Lodge Trust*, 22 B.R. at 923; *In re* Old Second Nat'l Bank, 7 B.R. 37, 38 (Bankr. N.D. Ill. 1980); *Cohen*, 4 B.R. at 203–04, 207.

[69] Loux v. Gabelhart (*In re* Carriage House, Inc.), 120 B.R. 754 (Bankr. D. Vt. 1990), *aff'd*, 146 B.R. 352 (D. Vt. 1992).

[70] 53 B.R. 741, 742 (Bankr. M.D. Fla. 1985).

[71] *Id.* at 744.

[72] 52 B.R. 308, 309 (Bankr. M.D. Fla. 1985).

[73] *Id.* at 311.

[74] 50 B.R. 710, 712 (Bankr. D.N.H. 1985).

[75] 124 F.2d 132, 135 (1st Cir. 1941).

[76] *See supra* note 19 and accompanying text.

Even after 1978, bankruptcy courts have declared that simply engaging in business is insufficient.[77] However, all of these cases involved family trusts. In addition, one case was about a trust that was both a family trust and a nominee trust,[78] and another involved a probate estate.[79] If a trust is neither a family trust nor a nominee trust nor a probate estate, courts are less likely to say that engaging in business is insufficient to qualify as a business trust.

Various cases since 1978 have said that engaging in business is an important characteristic of a business trust.[80] Conversely, the trust in *In re Westchester County Civil Service Employees Ass'n* was found not to be a business trust because it was "not engaged in any business activities and d[id] not generate any business profits."[81]

If one considers the cases as a whole, it is probably true that any trust that is substantively engaged in business will be considered a business trust as long as it is neither a family trust, nor a land trust, nor a nominee trust.

III. THE PROPOSAL

A. *State and Federal Law*

At present, in the absence of a definition of "business trust" in the Bankruptcy Code, courts are developing their own definitions on a case-by-case basis. One court expressed the hope

---

[77] *See, e.g., In re* St. Augustine Trust, 109 B.R. 494, 496 (Bankr. M.D. Fla. 1990) ("[T]he mere fact that the trust happens to engage in business activities does not therefore make it a 'business trust' within the meaning of the Code."); *In re* Margaret E. Dehoff Trust I, 114 B.R. 189, 191 (Bankr. W.D. Mo. 1990) ("A business trust is something more than a trust that carries on business activities."); Mosby v. Boatmen's Bank (*In re* Mosby), 61 B.R. 636, 638 (E.D. Mo. 1985), *aff'd*, 791 F.2d 628 (8th Cir. 1986) ("[A] business trust is something more than simply a trust that carries on a business.").

[78] *In re* Woodsville Realty Trust, 120 B.R. 2 (Bankr. D.N.H. 1990).

[79] *In re* Estate of Whiteside by Whiteside, 64 B.R. 99 (Bankr. E.D. Cal. 1986) (holding that a probate estate, even if it is doing business, is not a business trust).

[80] *See, e.g., In re* Constitutional Trust No. 2-562, 114 B.R. 627, 633 (Bankr. D. Minn. 1990) (holding that, even if an entity is formally a business trust, it must also engage in business); *In re* Cooper Properties Liquidating Trust, 61 B.R. 531, 536 (Bankr. W.D. Tenn. 1986) (deciding that the trust must be "conducting business of some kind"); *In re* Tru Block Concrete Prods., Inc., 27 B.R. 486, 489 (Bankr. S.D. Cal. 1983) (observing that California cases "focus . . . on whether the trust was authorized to, or was actually doing business").

[81] 111 B.R. 451, 456 (Bankr. S.D.N.Y. 1990).

that case-by-case adjudication would produce a uniform national definition,[82] but this has not turned out to be the case. Some courts hold that state law definitions of "business trust" are irrelevant, while other courts turn to state law.

In the first camp, *In re Arehart* held that the trust's failure to register as a business trust in accordance with Florida law was "without legal significance" because "[w]hether an entity is eligible for relief under title 11 of the United State Code is purely a matter of federal law."[83] The court believed that Congress could not have intended fifty different definitions of "business trust."[84] Similarly, bankruptcy courts in Maryland, Tennessee, and Wisconsin have chosen not to refer to their state statutes on business trusts.[85]

In the other camp, *In re Village Green Realty Trust* required the debtor to meet state statutory requirements, reasoning that, "[s]ince the Bankruptcy Code does not define what constitutes a business trust, and since the decisions are, if not hopelessly divided, at least certainly divergent, this Court must look to state law for guidance."[86] Courts in California, Florida, Louisiana, Massachusetts, Minnesota, Texas, and Vermont have turned to state law or mentioned that the trust had or had not complied with state statutory requirements such as registering with the secretary of state.[87]

It is true that the Constitution grants Congress the power to establish "uniform Laws on the subject of Bankruptcies through-

---

[82] *Woodsville Realty*, 120 B.R. at 4 n.2.

[83] *See In re* Arehart, 52 B.R. 308, 310 (Bankr. M.D. Fla. 1985).

[84] *Id.* at 311.

[85] *See In re* The Ophir Trust, 112 B.R. 956 (Bankr. E.D. Wis. 1990); *In re* Vivian A. Skaife Irrevocable Trust Agreement No. 1, 90 B.R. 325 (Bankr. E.D. Tenn. 1988); *In re* Cooper Properties Liquidating Trust, Inc., 61 B.R. 531 (Bankr. W.D. Tenn. 1986); *In re* Armstead & Margaret Wayson Trust, 29 B.R. 58 (Bankr. D. Md. 1982).

[86] 113 B.R. 105, 113 (Bankr. D. Mass. 1990); *see also In re* Heritage N. Dunlap Trust, 120 B.R. 252, 254 (Bankr. D. Mass. 1990) ("Since the Code does not define what constitutes a business trust, we look to state law.").

[87] *See, e.g., In re* Action Roofing & Supply Co., 137 B.R. 217, 219 (Bankr. S.D. Tex. 1991) ("There is no effective device creating a business trust under Texas law. A business trust is expressly excluded under the Texas Trust Code . . . ."); *In re* Mohan Kutty Trust, 134 B.R. 987 (Bankr. M.D. Fla. 1991); Loux v. Gabelhart (*In re* Carriage House, Inc.), 120 B.R. 754 (Bankr. D. Vt. 1990), *aff'd*, 146 B.R. 352 (Bankr. D. Vt. 1992) (turning to treatises only after finding no applicable Vermont cases or statutes); *In re* Constitutional Trust No. 2-562, 114 B.R. 627 (Bankr. D. Minn. 1990); *In re* Medallion Realty Trust, 103 B.R. 8 (Bankr. D. Mass. 1989), *aff'd*, 120 B.R. 245 (D. Mass. 1990); *In re* Tru Block Concrete Prods., Inc., 27 B.R. 486 (Bankr. S.D. Cal. 1983); *In re* SCR Trust, 20 B.R. 17, 18 (Bankr. M.D. La. 1982) ("[A] business trust is not authorized under the Louisiana Trust Code.").

out the United States."[88] However, this does not necessarily mean that every last detail of the Bankruptcy Code must be the same in all fifty states. For example, section 522(b) gives debtors a choice between federal and state law when deciding what property to exempt from the property of the bankruptcy estate.[89] Scholars note that "Congress has consistently deferred to local law and custom in framing bankruptcy statutes"[90] and that, without question, bankruptcy law "relies on state law for the definition of substantive rights."[91] Lastly, by analogy, *United States v. Kimbell Foods, Inc.*[92] held that even when federal courts have competence to make law, they have discretion to rely on state law instead of creating a federal rule. The Supreme Court rejected the generalized claim that uniformity was necessary; it said that state law could be adopted as long as it was consistent with federal policy. One reason for the Court's decision is judicial economy: creating new federal rules requires more time and expense than using pre-existing state law.[93]

In the interests of judicial economy, "business trust" could be defined by reference to existing state law. Thirty-four states have statutes on business or investment trusts.[94] Most state statutes require business trusts to register with the secretary of state and/or the county clerk,[95] so the Bankruptcy Code could easily declare that a business trust which is registered under state law is automatically a "business trust" (and hence eligible to be a debtor) for the purposes of the bankruptcy law.

However, state statutes are not the only statutes that require business trusts to register with a government body. The federal securities laws include business trusts within their coverage.[96]

---

[88] U.S. CONST., art. I, § 8, cl. 4.

[89] 11 U.S.C. § 522(b).

[90] TERESA A. SULLIVAN ET AL., AS WE FORGIVE OUR DEBTORS 27 (1989).

[91] Elizabeth Warren, *Bankruptcy Policy*, 54 U. CHI. L. REV. 775, 781 (1987).

[92] 440 U.S. 715 (1979).

[93] Other possible reasons include: (a) the federal government is a government of limited powers, so it should not displace state law unless there is a real need for displacement; (b) some areas have traditionally been left to the states; (c) if a federal court displaces one state law, it might have to displace other connected state laws, interfering with the fabric of state law and creating instability; (d) if there are different rules in state and federal court, there will be intrastate disuniformity. *See generally* PAUL M. BATOR ET AL., HART AND WECHSLER'S THE FEDERAL COURTS AND THE FEDERAL SYSTEM 849–905 (3d ed. 1988).

[94] *See infra* appendix.

[95] *See infra* appendix.

[96] *See* 2 LOUIS LOSS & JOEL SELIGMAN, SECURITIES REGULATION 1062 (3d ed. 1989); 13 AM. JUR. 2D *Business Trusts* § 82 (1964). A business trust certificate would qualify as a "security" under the Securities Act of 1933, 15 U.S.C. § 77b(1) (1988); the Securities

Thus, for states that do not require registration of business trusts, or for business trusts that prefer federal to state law, the Bankruptcy Code could provide that a business trust which is registered with the Securities and Exchange Commission is automatically a "business trust" for the purposes of section 101(9) of the Bankruptcy Code.

In sum, this Comment proposes that section 101(9)(A)(v) be amended as follows:

> "corporation"—
> (A) includes—
>
> . . .
>
> (v) business trust, which means
> (a) any trust which is registered as a business trust in accordance with state law, or
> (b) any trust which is registered as a business trust with the Securities and Exchange Commission.

B. *Effect of Proposal*

If a business trust is too small to register with the SEC and is located in a state that has no statute on business trusts, it might not be able to be a debtor under this proposal. Unfortunately, there appear to be no statistics on how many business trusts there are in the United States and how large they are, so it is hard to say how many business trusts would be affected.

Alternatively, an amended section 101(9)(A)(v) might declare that a business trust need only be "recognized," not "registered," as a business trust under state law. "Law" would then include state case law as well as state statutes. However, there are two problems with this idea. First, some states have neither statutes nor case law defining "business trust."[97] Second, relying

---

Exchange Act of 1934, 15 U.S.C. § 78c(a)(10) (1988); the Public Utility Holding Company Act of 1935, 15 U.S.C. § 79(a)(16) (1988); the Trust Indenture Act of 1939, 15 U.S.C. § 77ccc(16) (1988); and the Investment Company Act of 1940, 15 U.S.C. § 80a-2(a)(36) (1988).

[97] Alaska, Arkansas, Nebraska, New Mexico, Vermont, and Wyoming apparently have no cases at all dealing with business trusts. Search of Westlaw, Allstates database (Feb. 22, 1993) (search for cases satisfying search "sy, di ("business trust" "Massachusetts trust" "common law trust")"). In addition, cases in Georgia, Hawaii, Iowa, and Maine merely involve a business trust as a party; they do not define "business trust."

on case law would merely shift the problem of multiple definitions from the bankruptcy courts to the state courts.

Another factor to bear in mind is efficiency. If the sample of forty-nine cases examined in this Comment is representative, almost three-fourths[98] of the entities claiming to be business trusts are not business trusts. If the law were clear enough to prevent such entities from filing suit in the first place, both the parties and the court system would save money. The current system of uncertainty imposes other costs in addition to initial litigation costs. For example, if an entity is not a business trust and therefore cannot be a debtor, the Bankruptcy Code may not protect the trust's property because it is not part of the bankruptcy estate. Second, if an entity wants to ensure that it will qualify as a business trust, it might dissolve, reorganize, and refile. Although reorganization might be achieved without adverse tax consequences,[99] it will take time and money.

There will always be tension between the individualized justice that is available through case law and the certainty and accessibility of statutes.[100] However, the twentieth century is the age of statutes,[101] so it is time to change the definition of "business trust" in the Bankruptcy Code from a case law definition to a statutory definition.

—*Takemi Ueno**

---

[98] Thirty-six cases out of 49 is 73.5%.

[99] *See* 1A MYRON M. SHEINFELD ET AL., COLLIER ON BANKRUPTCY ¶ 17 (15th ed. 1993).

[100] *See, e.g.*, James C. Thomas, *Statutory Construction when Legislation Is Viewed as a Legal Institution*, 3 HARV. J. ON LEGIS. 191, 205–06 (1966) (describing the debate between David Dudley Field and James Carter in the late 19th century).

[101] *See, e.g.*, Erwin N. Griswold, Preface, *The Explosive Growth of Law Through Legislation and the Need for Legislative Scholarship*, 20 HARV. J. ON LEGIS. 267, 268–69 (1983); Willard Hurst, Foreword, *Legislation as a Field of Legal Research*, 2 HARV. J. ON LEGIS. 3, 3 (1965).

* I wish to thank Robert W. Gray, Esq., and Glenn E. Siegel, Esq., of Winthrop, Stimson, Putnam & Roberts for patient instruction in bankruptcy and Professor Elizabeth Warren for her helpful suggestions and kind encouragement. Any errors are mine, not theirs.

APPENDIX: SELECTED STATE STATUTES ON BUSINESS TRUSTS

1. *Alabama*: ALA. CODE § 19-3-60 (1990) (definition), § 19-3-61 (purposes for which business trust may be established), § 19-3-64 (fee; recordation of declaration of trust with county official).
2. *Arizona*: ARIZ. REV. STAT. ANN. § 10-501 (1990) (definition), § 10-504 (filing of declaration of trust with state official).
3. *California*: CAL. REV. & TAX. CODE § 23038(b) (West 1992) (definition).
4. *Connecticut*: CONN. GEN. STAT. § 47-6a (1991) (powers of foreign business trust; filing with state official).
5. *Delaware*: DEL. CODE ANN. tit. 12, § 3801(a) (Supp. 1992) (definition), § 3807 (residence requirement for trustees), § 3810 (filing of certificate of trust with state official), § 3813(a)(2) (fee).
6. *Florida*: FLA. STAT. § 609.01 (1991) (prohibited transactions), § 609.02 (fee; filing with state official), § 609.05 (permit from Department of Banking and Finance before selling securities).
7. *Idaho*: IDAHO CODE § 26-501(3) (1990) (definition).
8. *Indiana*: IND. CODE § 23-5-1-2(a) (1988) (definition), § 23-5-1-8 (powers of business trust; prohibited business), § 23-5-1-4 (fee; filing with state official), § 23-5-1-5 (additional fee), § 23-5-1-10 (annual reports).
9. *Kansas*: KAN. STAT. ANN. § 17-2028(a) (1988) (definition), § 17-2030 (filing with state official), § 17-2031 (fee), § 17-2036(c) (annual reports).
10. *Kentucky*: KY. REV. STAT. ANN. § 386.370 (Michie/Bobbs-Merrill 1984) (definition), § 386.380 (purposes for which business trust may be established), § 386.420(2) (fee; filing with both state and county officials).
11. *Louisiana*: LA. REV. STAT. ANN. § 12-491(B) (West 1993) (definition of real estate investment trust), § 12-492(A) (West 1969) (filing with both state and local officials), § 12-492(B) (West 1969) (taxes and fees), § 12-492(C) (West 1993) (rights and duties).
12. *Maryland*: MD. CORPS. & ASS'NS CODE ANN. § 3-101(d) (1993) (definition).
13. *Massachusetts*: MASS. GEN. L. ch. 182, § 1 (1991) (definition), § 2 (filing with state and local officials), § 12 (annual reports). Sections 3 and 4 lay down special requirements for trusts owning or controlling public utilities.
14. *Minnesota*: MINN. STAT. § 318.01 (1969) (prohibited business), § 318.02, subdiv. 1 (Supp. 1993) (fee; filing with state

official), § 318.03 (Supp. 1993) (registration of securities), § 318.04 (Supp. 1993) (business trusts engaged in insurance).
15. *Mississippi*: MISS. CODE ANN. § 79-15-3(1) (1989) (definition of "investment trust"), § 79-15-5(1) (purposes for which investment trust may be established), § 79-15-5(2) (requirements for trustees), § 79-15-7 (contents of declaration of trust), § 79-15-19 (filing with state and local officials).
16. *Montana*: MONT. CODE ANN. § 35-5-101 (1991) (definition), § 35-5-201 (filing with state official).
17. *Nevada*: NEV. REV. STAT. § 666.065(3) (1992) (definition).
18. *New Hampshire*: N.H. REV. STAT. ANN. § 293-A:107 (Supp. I 1992) and § 293-A:113 (1987) (foreign business trust; registration). Effective January 1, 1993, ch. 293-B (Supp. 1992) creates a new kind of organization, the New Hampshire Investment Trust.
19. *New York*: N.Y. GEN. ASS'NS LAW § 2(2) (McKinney 1942) (definition). Section 18(2), read in conjunction with § 2(4), requires filing with the secretary of state.
20. *North Carolina*: N.C. GEN. STAT. § 39-44 (1984) (definition).
21. *Ohio*: OHIO REV. CODE ANN. § 1746.01(A) (Anderson 1992) (definition), § 1746.04 (fee; filing with state official), § 1746.09 (powers of business trust).
22. *Oklahoma*: OKLA. STAT. tit. 68, § 202(4) (1991) (definition).
23. *Oregon*: OR. REV. STAT. § 128.560 (1991) (definition), § 128.575 (filing with state official), § 128.595 (annual report), § 128.600 (fees).
24. *Pennsylvania*: 15 PA. CONS. STAT. ANN. § 9501(a) (Supp. 1992) (definition), § 9502 (powers of business trust), § 9503(a) (filing with state official).
25. *Rhode Island*: R.I. GEN. LAWS § 7-1.1-99.1 (1992) (foreign business trust).
26. *South Carolina*: S.C. CODE ANN. § 33-53-10 (Law. Co-op. 1990) (filing with state and local officials).
27. *South Dakota*: S.D. CODIFIED LAWS ANN. § 47-14-1 (1991) (definition), § 47-14-2 (prohibited business), § 47-14-3 (filing with state official), § 47-14-4 (fee), § 47-14-8 (annual reports), § 47-14-7 (registration of securities).
28. *Tennessee*: TENN. CODE ANN. § 48-3-202(a) (Supp. 1992) (definition), § 48-3-204 (1988) (filing with state and local officials).
29. *Texas*: TEX. REV. CIV. STAT. ANN. art. 6138A, § 23.1(A)(1) (West Supp. 1993) (definition).

30. *Utah*: UTAH CODE ANN. § 7-1-103(5) (Supp. 1992) (definition).
31. *Virginia*: VA. CODE ANN. § 6.1-344(1)(b) (Michie 1988) (definition of real estate investment trust), § 6.1-345(2) (filing with local official).
32. *Washington*: WASH. REV. CODE § 23.90.020 (1985) (definition), § 23.90.040(1) (filing with state official).
33. *West Virginia*: W. VA. CODE § 47-9A-1 (1992) (powers of business trust), § 47-9A-3 (filing with local official), § 47-9A-5 (filing with state official).
34. *Wisconsin*: WIS. STAT. § 226.14(1) (1989–90) (filing with state and local officials), § 226.14(3) (fee), § 226.14(4) (annual report). Subsection (8) imposes further requirements on business trusts that sell beneficial certificates.

RECENT DEVELOPMENT

THE POTENTIAL EFFECT OF MANAGED COMPETITION IN HEALTH CARE ON PROVIDER LIABILITY AND PATIENT AUTONOMY

As early as 1983, one commentator noted that "[g]reater competition in health care is trendy."[1] In 1978, Alain Enthoven introduced a competitive health system model to solve the elusive problem of soaring health care costs.[2] Enthoven's proposal came to be known as "managed competition,"[3] and rapidly attracted the attention of several members of Congress.[4] At least four bills are pending in Congress with provisions for increasing competition in health care.[5] Indeed, the increasing support among legislators for using market reform to solve the nation's ailing health care system has led one commentator to preface his address to the House by stating that "managed competition" has "bec[o]me a buzzword here on Capitol Hill and in the Presidential election campaign. Now, it is almost all we are talking about with regard to health care reform."[6]

Although President Clinton's administration has yet to enunciate a clearly defined health care reform package, he repeatedly alluded to a "strategy of competition within a budget" during his campaign.[7] The President has indicated that he favors mar-

---

[1] Eli Ginzberg, *The Grand Illusion of Competition in Health Care*, 249 JAMA 1857, 1857 (1983).

[2] *See* Alain C. Enthoven, *Consumer-Choice Health Plan* (pts. 1 & 2), 298 NEW ENG. J. MED. 650, 709 (1978).

[3] An informal group of health industry leaders, public officials, and health services researchers formed the "Jackson Hole Group," which has been at the forefront of the managed competition movement. Alain Enthoven, a Stanford professor and economist, participated in the group discussions from the outset. *See* Paul M. Ellwood et al., *The Jackson Hole Initiatives for a Twenty-First Century American Health Care System*, 1 HEALTH ECONOMICS 149, 149 (1992).

[4] *See, e.g.*, Paul M. Ellwood, *Competition: Medicine's Creeping Revolution*, *in* FINANCING HEALTH CARE: COMPETITION VERSUS REGULATION 69, 73–74 (Duncan Yaggy & William G. Anlyan eds., 1982) (listing at least five "procompetition" bills introduced in 1981).

[5] H.R. 30, 103d Cong., 1st Sess. § 2(c)2(C) (1993) (Universal Health Benefits Empowerment and Partnership Act of 1993); H.R. 191, 103d Cong., 1st Sess. § 213(2) (1993) (American Consumers Health Care Reform Act of 1993); H.R. 200, 103d Cong., 1st Sess. §§ 221–252 (1993) (Health Care Cost Containment and Reform Act of 1993); S. 223, 103d Cong., 1st Sess. §§ 100–139 (1993) (Access to Affordable Health Care Act).

[6] 139 CONG. REC. H534 (daily ed. Feb. 4, 1993) (statement of Rep. Dan Glickman (D-Kan.)).

[7] *Clinton/Gore Health Care Plan*, National Desk, Political Writer, Sept. 24, 1992, *available in* LEXIS, Legis Library, USNWR File (statement of Clinton/Gore Campaign).

ket-based reform and managed competition.[8] On the other hand, Clinton seems reluctant to abandon the government's role in regulation.[9] De-emphasizing the difference between the two approaches, Clinton stated that although it was "traditional and politically convenient rhetoric" to advocate choosing between either extreme of government regulation or market-based competition, he "reject[ed] such polarizing rhetoric . . . in health care."[10] It appears, then, that the President's reform package, scheduled to be unveiled on May 1, 1993, will be an amalgam of different strategies for reform, including top-down government regulation[11] and increased competition within the market.[12] This Recent Development examines managed competition and its possible implications for patient autonomy and for liability arising from cost-containment measures. Part I describes the conflicting concerns of cost, quality, and access in health care. Part II presents the managed competition view of the problem

[8] *See, e.g., Hillary Clinton Pushes Health Reform, Promises Results this Year in Pennsylvania*, Daily Report for Executives (BNA), Feb. 12, 1993, *available in* LEXIS, Exec. Library, DREXEC File (noting that the First Lady is seeking reform through managed care and competition); *Less Health Care Seen as Possible*, N.Y. TIMES, Mar. 3, 1993, at A20 (stating that President Clinton intends to submit to Congress a health plan centered on the concept of managed competition by May, 1993); Dana Priest, *First Lady Holds Health Teach-In: Hillary Clinton Begins Effort to Humanize Issue with Citizens' Input*, WASH. POST, Feb. 12, 1993, at A4 (reporting that Hillary Clinton presided over courses in "managed competition"—a strategy favored by the Administration).

[9] Perhaps this ambivalence reflects the deep disagreement among policy-makers over whether to follow a course of "command and control regulation" or "deregulation and competition." *See* Alan A. Stone, *Law's Influence on Medicine and Medical Ethics*, 312 NEW ENG. J. MED. 309, 310 (1985) (deeming contradictory cost-containing legislation a strange compromise between "command and control regulation" on the one hand and "deregulation and competition" on the other).

[10] Bill Clinton, *The Clinton Health Care Plan*, 327 NEW ENG. J. MED. 804, 805 (1992).

[11] *See, e.g., id.* (proposing a National Health Board "to establish annual budget targets and [to] define a core benefit package").

[12] *See id.* (encouraging insurers, physicians, and health care institutions to collaborate in local health networks that will compete for patients on the basis of cost and quality); *Hearing of the Senate Labor and Human Resources Committee: Confirmation Hearing for Health and Human Services Secretary-Designate Donna Shalala*, Federal News Service, Jan. 15, 1993, *available in* LEXIS, Nexis Library, FEDNEW File ("[T]he President-elect has outlined a managed competition plan which has a global budget on top of it as a check, so . . . we're trying to marry two ideas in some sense."); Janice Castro, *Paging Dr. Clinton*, TIME, Jan. 18, 1993, at 24. The feasibility of merging the two approaches—regulation and competition—has been a subject of great concern. *See generally Achieving Effective Cost Control in Comprehensive Health Care Reform, 1992: Hearings on Examining Methods to Control Excessive Health Care Costs, Focusing on the Relationship Between Managed Competition and Enforceable National Spending Goals Before the Senate Comm. on Labor and Human Resources*, 102d Cong., 2d Sess. (1992) [hereinafter *Hearings*]; Uwe E. Reinhardt, *Commentary: Politics and the Health Care System*, 327 NEW ENG. J. MED. 809, 811 (1992) (pointing out the lack of clarity in President Clinton's plan for transmitting the global budgets to individual health care providers and the method of triggering a backup rate-setting system).

and solution. Part III explores some of the potential effects of managed competition on physician incentives and on liability for injuries resulting from improperly motivated clinical decisions. Part IV argues that informed consent fails to protect against the negative effects of managed competition on patient autonomy. Finally, Part V brings the development of "informed refusal" into the debate and explains why that doctrine may be a less-than-perfect remedy for managed competition's encroachment on patient autonomy.

I. THE PROBLEMS OF COST, QUALITY, AND ACCESS

The problem of soaring health care costs is not a new one. From 1970 to 1990, health care expenditures in the United States increased at a yearly rate of 12%.[13] Americans spent $42 billion on health care in 1965—6% of the gross national product in that year.[14] That figure rose to $287 billion, or 9.8% of the GNP, in 1981.[15] During this time, local, state, and federal government expenditures were responsible for an increased share of total health care financing—from 26% in 1965 to almost 43% ($123 billion) in 1981.[16] The Commerce Department estimated that spending on health care totaled $838.5 billion for 1992—more than 14% of the GNP for that year.[17] According to current projections, annual national health expenditures may reach almost $1.7 trillion by the year 2000.[18]

Although health care costs have been rising, so has the number of people without health insurance coverage. In the 1970s, when national health expenditures were about eight percent of the GNP, 25 million Americans were uninsured.[19] The gap widened in the 1980s, and by 1991, the number of uninsured Amer-

[13] John K. Iglehart, *The American Health Care System: Introduction*, 326 NEW ENG. J. MED. 962, 964 (1992).

[14] 1 PRESIDENT'S COMM. FOR THE STUDY OF ETHICAL PROBLEMS IN MEDICINE & BIOMEDICAL & BEHAVIORAL RESEARCH, SECURING ACCESS TO HEALTH CARE: THE ETHICAL IMPLICATIONS OF DIFFERENCES IN THE AVAILABILITY OF HEALTH SERVICES 184 (1983) (citing statistics from Robert M. Gibson & Daniel R. Waldo, *National Health Expenditures, 1981*, HEALTH CARE FINANCING REV., Sept. 1982, at 2).

[15] *Id.*

[16] *Id.*

[17] Robert Pear, *Health-Care Costs Up Sharply Again, Posing New Threat*, N.Y. TIMES, Jan. 5, 1993, at A1.

[18] *Hearings*, *supra* note 12, at 4.

[19] Uwe E. Reinhardt, *Health Insurance For All—Now*, N.Y. TIMES, Dec. 14, 1992, at A17.

icans was estimated at 36.6 million.[20] Of the 35.7 million uninsured Americans in 1990, three-fourths of them were full-time workers and their dependents.[21] The vast majority of firms not providing insurance had twenty-five or fewer employees.[22] Addressing this problem, one senator remarked, "we are no longer really arguing whether there should be some kind of universal health care coverage, but really what kind."[23] And yet, at least one commentator has remarked that "[a] country that is determined to provide its uninsured populations with access to medical care does not talk about lowering the taxes of the middle class, particularly one as debt-ridden as the United States."[24] Indeed, a guiding principle of American health policy is that universal insurance coverage must be considered only after costs are contained—a principle "as politically imprudent as it is morally bankrupt."[25] Policy-makers have been seeking a solution that would contain costs without compromising access or quality of care. This presents a difficult challenge for President Clinton because almost every move designed to solve one conflicting problem may have unforeseen effects on others.[26]

II. THE MANAGED COMPETITION SCHEME

A. *The Problem of Market Failure*

Advocates of managed competition claim that it provides one solution that would contain costs while providing universal coverage without compromising quality. Although there are many

[20] *Hearings*, *supra* note 12, at 4.

[21] Iglehart, *supra* note 13, at 962 (footnote omitted). To acquire medical care, the uninsured rely on visits to hospital emergency rooms, physicians' offices, and clinics, and are treated on a payment out of pocket basis or as charity. *Id.* at 967.

[22] *See* Mark A. Hall, *Reforming the Health Insurance Market for Small Businesses*, 326 NEW ENG. J. MED. 565, 565 (1992) (citation omitted).

[23] *Hearings*, *supra* note 12, at 7 (statement of Sen. Paul D. Wellstone (D-Minn.)).

[24] Iglehart, *supra* note 13, at 963 (quoting economist Victor Fuchs).

[25] Reinhardt, *supra* note 19.

[26] *See* David M. Kinzer, *Universal Entitlement to Health Care: Can We Get There from Here?*, 322 NEW ENG. J. MED. 467, 467 (1990); Sylvia Law, *Introduction to Health Care Delivery in the United States*, *in* AMERICAN HEALTH LAW 2 (George Annas et al. eds., 1990).

possible explanations for steadily rising health care costs,[27] advocates of managed competition view the primary evils as cost-unconscious demand and perverse incentives created by the fee-for-service, third-party insurance system.

The American form of health service financing is unique in that it relies predominantly on private, employment-based, insurance-type financing schemes that divorce financing from the delivery of services.[28] The problem with our current system of employer-based health insurance is its promotion of "cost-unconscious demand."[29] Whereas medical care is often thought of as an insurable event, the casualty insurance model does not fit health care because of uncertainty which, in the third-party-payer system, is usually resolved in favor of excessive and costly care.[30] Under the traditional insurance scheme, third parties are predominantly accountable for the medical bills that arise from the doctor-patient encounter. In 1987, insured patients paid directly, on average, only ten cents for every dollar of hospital care and approximately twenty-six cents for every dollar paid to physicians.[31] Both patients and physicians are less sensitive to rising health care costs than if patients had to pay for services directly.[32] Providers, consequently, have little incentive to seek and utilize medical practices that yield the same health outcomes at less cost, and little competition exists among providers to produce services efficiently and pass savings on to the consumers.[33]

---

[27] *See, e.g.*, ALAIN C. ENTHOVEN, HEALTH PLAN: THE ONLY PRACTICAL SOLUTION TO THE SOARING COST OF MEDICAL CARE 6 (1980) (attributing rising costs, in part, to the general misconception that more medical care is better than less); David U. Himmelstein & Steffie Woolhandler, *Cost Without Benefit: Administrative Waste in U.S. Health Care*, 314 NEW ENG. J. MED. 441, 442–43 (1986) (citing runaway administrative costs as source of cost increases); Edward L. Schneider & Jacob A. Brody, *Aging, Natural Death, and the Compression of Morbidity: Another View*, 309 NEW ENG. J. MED. 854, 855 (1983) (noting that costs are rising because the fastest growing segment of the population is over 85); Enthoven, *supra* note 2, at 651 (blaming expensive technology, government-sponsored proliferation of health professionals, "defensive medicine" induced by fears of malpractice, and patient demands and expectations); Arnold S. Relman, *The New Medical-Industrial Complex*, 303 NEW ENG. J. MED. 963 (1980) (blaming exploding costs on growth of the "medical-industrial complex").

[28] *See* Law, *supra* note 26, at 20. In 1988, 73% of Americans with insurance obtained coverage through employers. Iglehart, *supra* note 13, at 965 (citing HEALTH INS. ASSOCS. OF AM., SOURCE BOOK OF HEALTH INSURANCE DATA (1990)).

[29] Ellwood et al., *supra* note 3, at 155.

[30] ENTHOVEN, *supra* note 27, at 9–11.

[31] Henry Aaron & William B. Schwartz, *Rationing Health Care: The Choice Before Us*, 247 SCIENCE 418, 418 (1990).

[32] *See* Iglehart, *supra* note 13, at 963.

[33] Enthoven, *supra* note 2, at 652.

The resulting health care delivery system is what Enthoven characterizes as the "paradox of excess and deprivation."[34] Health plans in an inherently uncompetitive market use numerous strategies that destroy efficiency and fairness: risk selection, market segmentation, and refusals of insurance for some groups of people.[35] The fee-for-service payment system for physician services contributes to the inability of the current private health insurance market to allocate resources properly. Under the traditional fee-for-service system, a physician is paid for each service rendered to a patient and thus rewarded for providing more and more costly services.[36] The incentives for a physician are always to do more regardless of the benefits to the patient;[37] too many inappropriate and unnecessary procedures are carried out in the practice of "flat-of-the-curve medicine."[38] The diminishing returns from additional input suggests health care's "virtually unlimited capacity to consume additional resources."[39]

From the patient's perspective, the fee-for-service, third-party-reimbursement system provides a choice among doctors and hospitals, but leaves no opportunity to keep the savings that the patient could receive by selecting effective, less costly care.[40] Not only do most consumers generally have poor information regarding health care alternatives, but they also have little incentive to compare the costs of care, which will be borne by employers or the government.[41]

According to Enthoven, the Internal Revenue Code is another culprit in encouraging excessive spending for health insurance and in eliminating incentives for consumers to shop among competing plans. Federal tax provisions allow employees to characterize their health insurance premium contributions—no mat-

---

[34] Alain C. Enthoven & Richard Kronick, *A Consumer-Choice Health Plan for the 1990s: Universal Health Insurance in a System Designed to Promote Quality Economy* (pt. 1), 320 NEW ENG. J. MED. 29, 29 (1989).

[35] *See id.* at 34; Ellwood et al., *supra* note 3, at 155.

[36] *See* ENTHOVEN, *supra* note 27, at 16; Enthoven, *supra* note 2, at 651.

[37] *Hearings*, *supra* note 12, at 33 (statement of Alain C. Enthoven).

[38] *Id.* In the health care context, the "flattening end of the curve" refers to the point at which input of an additional unit of care yields minimal or no benefit to the patient. *See* ENTHOVEN, *supra* note 27, at 45.

[39] John H. Morrow & Arch B. Edwards, *U.S. Health Manpower Policy: Will the Benefits Justify the Costs?*, 51 J. MED. EDUC. 791, 795 (1976); *cf.* ENTHOVEN, *supra* note 27, at 3 ("The medical care system can legitimately absorb every dollar society will give it.") (quoting WALTER MCCLURE, MEMORANDUM I: ESSENTIAL POINTS TO UNDERSTANDING THE MAJOR OPTIONS AVAILABLE IN NATIONAL HEALTH INSURANCE, INTERSTUDY (1974)).

[40] Enthoven, *supra* note 2, at 652.

[41] *Id.* at 652–53.

ter how large—as nontaxable employer contributions. If an employee chooses a health plan that is more expensive, rather than less, the government, in effect, subsidizes one third of the difference in cost in the form of tax relief.[42] Consequently, the employee is less cost-conscious, and health providers have less incentive to reduce prices to attract subscribers.[43]

B. *The Solution*

Managed competition proposes to solve the problem of health care costs by restructuring the market to promote "cost-conscious consumer choice among health plans in the pursuit of equity and efficiency in health care financing and delivery."[44] As Enthoven has presented the plan, the market involves three key actors: consumers, health plans, and sponsors.[45]

1. Consumers

For consumers, managed competition emphasizes informed, cost-conscious choice from among competing health plans. Employers would be required to offer employees choices of health plans.[46] In order to encourage employees to choose the most economical plan, the plan limits tax-free employer contributions to the cost of the lowest-priced plan.[47] Individuals joining more costly health plans would not receive extra subsidies, as they currently do, but would have to pay the extra cost out of their own pocket.[48] Finally, the needy would receive vouchers to be used only as a premium contribution to the qualified plan they choose.[49]

---

[42] Treating employer contributions to health insurance premiums as part of employees' taxable income could yield an estimated $40 billion in additional income taxes and $27 billion in Social Security taxes. STUART M. BUTLER, A POLICY MAKER'S GUIDE TO THE HEALTH CARE CRISIS. PART II: THE HERITAGE CONSUMER CHOICE HEALTH PLAN (1992), *cited in* Reinhardt, *supra* note 12, at 810.

[43] *See* Ellwood et al., *supra* note 3, at 155 (lamenting "perverse public subsidies"); Enthoven, *supra* note 2, at 651–52; Enthoven & Kronick, *supra* note 34, at 30.

[44] Alain C. Enthoven, *Managed Competition of Alternative Delivery Systems*, *in* COMPETITION IN THE HEALTH CARE SECTOR: TEN YEARS LATER 83, 85 (Warren Greenberg ed., 1988).

[45] *Id.*

[46] Alain C. Enthoven, *Commentary: Measuring the Candidates on Health Care*, 327 NEW ENG. J. MED. 807, 807 (1992).

[47] *Id.*

[48] *See id.*; Enthoven, *supra* note 2, at 710.

[49] Enthoven, *supra* note 2, at 712.

2. Health Plans

"Accountable health partnerships" that integrate financing and delivery of health care, as health maintenance organizations ("HMOs") do today, form the foundation of the managed competition reform model.[50] These health plans, which integrate doctors, hospitals, and insurance providers, would compete for patient enrollment and would be forced to "deliver value for money."[51] The more an organization is able to improve quality and cut costs, the more subscribers would choose that organization for their plan.[52] Organizing doctors into competing units would necessarily entail patients' acceptance of "closed panel plans," in which patients can select from only those doctors participating in or contracting with their plan in exchange for lower costs or other perceived benefits.[53] Under the managed competition scheme, every doctor would be fully committed to one organization.[54] Although a few physicians would continue to operate under the traditional fee-for-service system of payment, competition would promote the development of "alternative delivery systems," such as prepaid group practices, because they would foster more efficient medical practice.[55]

To qualify for tax credits, vouchers, or Medicare payments, a health plan would be required to conform to a set of rules designed to foster a "fair and socially desirable competition based on quality and cost effectiveness."[56] Every plan would be required to participate in a "periodic government-run open enrollment," to charge according to community rating, "to set community rates according to market area," "to offer a low option limited to basic benefits," and to state clearly an annual limit on individual out-of-pocket expenses.[57] Standardization of coverage, compensation for enrollment of high-risk patient pop-

---

[50] *Hearings*, *supra* note 12, at 33; Enthoven & Kronick, *supra* note 34, at 95.

[51] *Hearings*, *supra* note 12, at 34.

[52] Enthoven, *supra* note 46, at 807.

[53] Alain C. Enthoven, *Does Anyone Want Competition? The Politics of NHI*, *in* NEW DIRECTIONS IN PUBLIC HEALTH CARE: A PRESCRIPTION FOR THE 1980s 227, 233 (Cotton M. Lindsay ed., 1980).

[54] Richard Kronick et al., *The Marketplace in Health Care Reform: The Demographic Limitations of Managed Competition*, 328 NEW ENG. J. MED. 148, 148 (1993).

[55] Enthoven, *supra* note 53, at 233–34; *see also* Enthoven & Kronick, *supra* note 34, at 96. Most doctors, then, would be affiliated with some sort of managed-care practice group. Enthoven, *supra* note 2, at 710.

[56] Enthoven, *supra* note 2, at 713.

[57] *Id.* at 713–14.

ulations, outcome management, and technology assessment would all facilitate comparison of competing health plans.[58] All of these rules would be fair to cost-effective, organized systems.[59]

3. Sponsors

The third group of actors in a managed competition system is the "public sponsor" agency in each state, which would act as the final guarantor of health coverage.[60] For anyone without employment-based coverage, the state's public sponsor would subsidize enrollment and contract with private-sector health plans.[61]

The public sponsors would act as collective purchasing agents for employment groups of 100 or less and for the self-employed—groups that are currently too small to spread the risks or to achieve economies of scale in administration.[62] Small employers could obtain coverage for their groups, for example, by paying an eight percent payroll tax to sponsors.[63] By pooling hundreds of thousands of workers, Health Insurance Purchasing Cooperatives ("HIPCs") could aggregate the buying power of small employment groups, contract with qualified health plans to offer enrollees a price-competitive choice, process the information needed to facilitate good decisions, manage competition, "monitor the quality of care and compliance, and relieve their participating employers of costly administrative burdens."[64] The sponsor's role as a broker would entail structuring coverages, "contracting with health plans and beneficiaries for the rules of participation, managing the enrollment process, collecting premium contributions from beneficiaries, paying premiums to

---

[58] *See* Enthoven, *supra* note 46, at 807; Enthoven & Kronick, *supra* note 34, at 34–35.

[59] *See* Enthoven, *supra* note 2, at 715.

[60] Enthoven, *supra* note 44, at 85; Enthoven & Kronick, *supra* note 34, at 31.

[61] *See* Alain C. Enthoven & Richard Kronick, *Universal Health Insurance Through Incentive Reform*, 265 JAMA 2532, 2533–34 (1991); Enthoven & Kronick, *supra* note 34, at 31.

[62] *See Hearings*, *supra* note 12, at 34–35; Enthoven & Kronick, *supra* note 61, at 2534.

[63] *See* Enthoven & Kronick, *supra* note 61, at 2533–34; Enthoven & Kronick, *supra* note 34, at 32–33.

[64] *See Hearings*, *supra* note 12, at 35; Enthoven, *supra* note 46, at 807–08.

health plans, and administering both cross-subsidies among beneficiaries and subsidies available to the whole group."[65]

III. The Role of Physicians

A. *Physician Incentives*

A central premise of managed competition is the primacy of physician incentives. Though doctors receive only about twenty percent of each health care dollar, they influence seventy percent of total health care spending.[66] Influencing physician incentives is essential because doctors make the most costly decisions, such as hospital admissions, diagnostic tests, and procedures.[67]

The strategy of managed competition alters the basic framework of financial incentives to promote cost-reducing behavior and to reward doctors for cost-effective care.[68] In contrast to the fee-for-service system of payment, HMOs agree to accept a fixed payment and then provide all necessary care. This motivates HMO management to encourage their physicians to minimize costs.[69]

Although they may do so in different ways, all forms of managed care attempt to contain costs by modifying the behavior of doctors. One primary method managed health plans use to limit costs is utilization review—evaluation of quality and appropriateness of medical care at various stages in delivery based

---

[65] Enthoven & Kronick, *supra* note 34, at 31; *see also* Enthoven, *supra* note 44, at 85.

[66] *See* Alain C. Enthoven, *Supply-Side Economics of Health Care and Consumer Choice Plan*, *in* A New Approach to the Economics of Health Care 467, 470 (Mancur Olson ed., 1981); Enthoven, *supra* note 2, at 652.

[67] Enthoven, *supra* note 27, at 9–10. One commentator has framed the situation more starkly: "To doctors, hospital officials and other providers of health care, someone else's 'health-care spending' translates into their 'health-care revenue'; likewise, 'cost control' spells 'income control' to the medical industry." Reinhardt, *supra* note 19, at A17.

[68] *See Hearings*, *supra* note 12, at 185 (statement by Dr. Philip R. Lee) (noting that desirable changes are unlikely without such rewards); Alain C. Enthoven, *Shattuck Lecture—Cutting Cost Without Cutting the Quality of Care*, 298 New Eng. J. Med. 1229, 1229–30 (1978).

[69] *See* Arnold S. Relman, *Controlling Costs by "Managed Competition"—Would It Work?*, 328 New Eng. J. Med. 133 (1993).

on "established clinical criteria."[70] The financial success of these organizations will probably depend on selecting and recruiting physicians sympathetic to the managed-care philosophy.[71] Furthermore, many managed-care arrangements use financial incentives for physicians to reduce the number of referrals, tests, procedures, and hospital admissions. The primary-care doctors thus act as "gatekeepers."[72] When a physician pledges her services to one health plan exclusively, the financial well-being of that system becomes much more relevant to her clinical decisions.

B. *Appeal to Politicians*

To politicians, one of the greatest appeals of managed competition is that, despite numerous references to a "basic health plan," the proposed model does not define with certainty what the basic level of services would include.[73] For the most part, managed competition relies on incentives to influence physicians to render less costly care, but exactly how those costs will be cut is left to the discretion of physicians. All countries allocate scarce resources and withhold some forms of treatment from some people who might benefit from them. The differences lie in the process by which these allocational decisions are made.[74] This country "has managed to address quality/cost/access tradeoffs in subtle ways, avoiding direct confrontation."[75] Politicians tend to endorse systems that pass difficult cost-balancing deci-

---

[70] David D. Griner, Note, *Paying the Piper: Third-Party Payor Liability for Medical Treatment Decisions*, 25 GA. L. REV. 861, 883 (1991). Griner notes that "[u]tilization review, particularly prospective and concurrent review, strikes at the heart of health care delivery: the physician-patient relationship." *Id.* at 885.

[71] John K. Iglehart, *The American Health Care System: Managed Care*, 327 NEW ENG. J. MED. 742, 746 (1992).

[72] *Id.* at 745.

[73] *See, e.g.*, Enthoven, *supra* note 2, at 714 (proposing that "[q]ualified plans must offer a low option limited to the basic benefits defined in the national-health-insurance law"); Enthoven & Kronick, *supra* note 34, at 31. Enthoven provides a sketchy list of "basic health services"—those in the HMO Act of 1973—as a common yardstick for comparing the covered benefits of various health plans, and he points out the need for some "agreed on, uniform definition of 'all the care you need'" applicable to all health plans. He does not, however, adhere steadfastly to this particular list and does not provide guidance as to what the list should include. *See* ENTHOVEN, *supra* note 27, at 127–28.

[74] *See* Alan M. Garber, *No Price Too High?*, 327 NEW ENG. J. MED. 1676, 1677 (1992).

[75] Marshall B. Kapp, *Legal and Ethical Implications of Health Care Reimbursement by Diagnosis Related Groups*, 12 L. MED. & HEALTH CARE 245, 247 (1984).

sions from government officials to health care providers.[76] Managed competition replaces what one court described as the legislature's responsibility to make the "macro-decision" with the physician's "micro-decision."[77]

C. *From Patient Advocate to Resource Allocator*

During his campaign, Bill Clinton recognized the need to "shelter the provider-patient relationship from some of the intrusive methods of many of today's efforts to control costs," and claimed that collaborative networks of health care providers would serve that function.[78] Unaccustomed to considering the costs of health care, American physicians have traditionally deemed the patient's welfare as their first, if not only, concern.[79] In a managed-care plan, however, the doctor's role as advocate for the patient may be compromised by her increased responsibilities to ensure against excessive use of costly resources.

Enthoven insists that physicians are the best qualified to make these difficult decisions of need and cost-effectiveness.[80] However, the appropriateness of thrusting hospitals and physicians into the role of rationer has been seriously questioned.[81] Some would resolve the allocation issue, instead, in favor of the person who will enjoy the benefits or suffer the consequences of

---

[76] *See* Stone, *supra* note 9, at 310 (pointing to Diagnosis Related Groups, Health Maintenance Organizations, and Professional Standard Review Organizations as systems popular with politicians); *cf.* Kapp, *supra* note 75, at 247 (noting the argument that the "hard and tragic" rationing choices are properly borne by society as a whole); E. Haavi Morreim, *The MD and the DRG*, HASTINGS CENTER REP., June 1985, at 30, 36 (arguing that the decision whether the patient merits the benefit is a matter of social policy and falls outside the professional purview of the doctor).

[77] *See* Jackson v. Stockdale, 264 Cal. Rptr. 525, 529 (Cal. Ct. App. 1990) (describing the two levels of judgment regarding medical necessity in the Medi-Cal scheme).

[78] Clinton, *supra* note 10, at 805. *Contra* Relman, *supra* note 69, at 135 (arguing that increasing price competition among insurers would actually increase third-party intrusion into health care decisions and the doctor-patient relationship).

[79] During their training, physicians identify with conscientious and compassionate role models who insist on certainty of diagnosis and every possibly beneficial treatment, even at the cost of what is now termed the flattening end of the curve. *See* Stone, *supra* note 9, at 312.

[80] Enthoven, *supra* note 68, at 1237; Enthoven, *supra* note 2, at 709.

[81] Kapp, *supra* note 75, at 247. Those critics who worry only about rising health care costs "fail to appreciate . . . the ethical void created when medical practice is viewed through the prism of cost-benefit analysis." Stone, *supra* note 9, at 312 (footnote omitted).

the decision—the patient.[82] They argue that forcing the role of rationer onto doctors not only threatens to undermine the basis of the fiduciary patient-doctor relationship, but also increases doctors' exposure to potential legal liability.[83] Moreover, even if societal obligations necessitate that the expanding role of health providers today include that of scarce resource allocation, there must be explicit uniform criteria to guide physicians in determining which of the patients competing for the same limited health goods will prevail.[84]

Enthoven further insists that, in the managed-care setting, the doctor's interests equal the patient's interests—high-quality *economical* care.[85] Again, this premise may be flawed. In the managed-care setting there is often a divergence of objectives. As Professor Stone has characterized the conflict of interests:

> What is the patient's interest in reducing the economic risk to the doctor or the aggregate cost of health care by foregoing a bed in the coronary care unit or a CAT scan? It is one thing to entrust your life and health at times of crisis to a physician who is committed to the practical ethics that involves a quest for excellence and who may err on the side of doing too much. It is quite another to entrust your life and health at times of crisis to a physician whose diagnostic and therapeutic interventions are limited by new regulatory constraints or incentives of competitive efficiency that "place the provider at economic risk."[86]

The concept of "medical necessity" precludes a doctor from striving exclusively for the best interests of her patients.[87] The role of allocator hinders her loyalty to the best interests of those patients.[88]

---

[82] *See* Marjorie M. Shultz, *From Informed Consent to Patient Choice: A New Protected Interest*, 95 YALE L.J. 219, 271 (1985); *cf.* Matthew R. Gregory, *Hard Choices: Patient Autonomy in an Era of Health Care Cost Containment*, 30 JURIMETRICS J. 483, 494–95 (1990) (arguing that prospective reimbursement need not replace the patient with the doctor as the primary treatment decision-maker).

[83] Kapp, *supra* note 75, at 247; *see also* Stone, *supra* note 9, at 310 (describing the confusing and contradictory legal messages urging physicians to consider aggregate cost of health care in deciding clinical course of treatment on the one hand, and prohibiting cost from serving as an ethically or clinically relevant consideration on the other).

[84] Kapp, *supra* note 75, at 247–48; Morreim, *supra* note 76, at 36 (arguing that ad hoc bedside budget balancing is inappropriate).

[85] Enthoven & Kronick, *supra* note 34, at 96.

[86] Stone, *supra* note 9, at 312 (citation omitted).

[87] FRANK H. MARSH & MARK YARBOROUGH, MEDICINE AND MONEY: A STUDY OF THE ROLE OF BENEFICENCE IN HEALTH CARE COST CONTAINMENT 88–89 (1990).

[88] Shultz, *supra* note 82, at 295; *cf.* Harold S. Luft, *Health Maintenance Organizations and the Rationing of Medical Care*, *in* 3 SECURING ACCESS TO HEALTH CARE 313, 327 (President's Comm'n for the Study of Ethical Problems in Medicine & Biomedical &

D. *Liability for Cost-Containment Measures*

Because managed competition motivates physicians to restrict utilization of resources, it raises the issue of who will be held liable for injuries resulting from cost-constraining measures. Although the general perception is that much of the medical care that is currently rendered is unnecessary, and even harmful,[89] it is simply too optimistic and naive to believe that managed care will be able to trim all the fat out of the current system without sacrificing some quality. The line that doctors draw between "necessary" and "unnecessary" procedures will depend, in large part, on the degree of uncertainty that they will accept. If the distinction between necessary and unnecessary care were easy to make, then physicians would not overutilize resources to the extent that they do now. Inevitably, the impetus to minimize utilization of resources will result in some instances of medically necessary procedures being cut as well.[90] The immediate concern is patient injury from inadequate health care and liability for situations where intrusions into the physician-patient relationship turn sour. Assuming it impossible for anyone to exert control over the medical decisions of physicians, courts have historically held physicians solely responsible for the adverse consequences of their treatment of patients.[91] Until recently courts had not found HMOs or managed-care insurers directly liable for injuries caused by cost-containment requirements.[92] As these cost-containment measures become more widespread, however, plaintiffs have continued to sue on the theory that

---

Behavioral Research ed., 1983) (arguing that a doctor swayed by financial incentives can no longer serve as a perfect agent for the patient).

[89] Griner, *supra* note 70, at 882 (1991) (noting that the general perception that much medical care is unnecessary and harmful results in optimism that external review of physician utilization can reduce costs without compromising quality).

[90] *See id.* at 883 (observing unavoidable conflict between controlling health care costs and maintaining quality of care); *cf.* Stephen M. Shortell & Edward F.X. Hughes, *The Effects of Regulation, Competition, and Ownership on Mortality Rates Among Hospital Inpatients*, 318 NEW ENG. J. MED. 1100, 1103 (1988) (finding significant association between higher inpatient mortality rates and intensity of marketplace competition, as measured by enrollment in HMOs).

[91] *See, e.g.*, Wickline v. State, 239 Cal. Rptr. 810, 819 (Cal. Ct. App. 1986) (refusing to hold public insurer liable where it did not override the medical judgment of the plaintiff's treating physicians at the time of her discharge); Schloendorff v. Society of New York Hosp., 105 N.E. 92, 93–94 (N.Y. 1914) (holding a hospital not liable for physician's negligence because it did not have power to direct the work of its physicians).

[92] *See* Rex O'Neal, Note, *Safe Harbor for Health Care Cost Containment*, 43 STAN. L. REV. 399, 400 (1991).

HMO cost-reduction incentives have influenced physicians' decisions to deny or defer essential diagnostic procedures and treatments.[93] Courts have correspondingly demonstrated a willingness to recognize the growing risks to patients of cost-containment measures and to extend tort liability to insurers where their utilization review process has displaced accepted medical judgment.[94]

Although managed-care plans have not been held directly liable for the adverse consequences of cost-containment measures, they can be held accountable for medical decisions to withhold potentially beneficial treatment or diagnostic procedures under several theories of liability. The first is respondeat superior, the doctrine holding an employer liable for the negligent acts of an employee committed within the scope of employment.[95] Respondeat superior would apply in the case of a typical health plan in the managed competition scheme, such as a staff-model HMO where the physician is an actual employee, and where the third-party payor, as the employer of the physician, exerts direct control over her practice methods.[96]

Managed-care plans may also be found liable for patient injury resulting from cost-containment measures under the theory of apparent authority or ostensible agency. Ostensible agency liability may attach where the defendant represents that another is her servant and where the plaintiff relies on this representation in such a manner that it causes injury.[97] Health plans that offer a closed panel of providers,[98] such as those in the managed competition scheme, can give the appearance of an agency re-

---

[93] *See id.* at 411; Daniel Q. Haney & Fred Bayles, *Paying a Price for Cost-Conscious HMOs*, L.A. TIMES, Jan. 28, 1990, at A3 (describing claims by patients from whom treatment was inappropriately withheld).

[94] *See, e.g.*, *Wickline*, 239 Cal. Rptr. at 819.

[95] *See* RESTATEMENT (SECOND) OF AGENCY § 219 (1957).

[96] *See* Griner, *supra* note 70, at 892 (noting that courts consider physicians' salary in staff-model HMO as proof of master-servant relationship).

[97] RESTATEMENT (SECOND) OF AGENCY §§ 8, 159, 267 (1957); *see, e.g.*, Mehlman v. Powell, 378 A.2d 1121 (Md. 1977) (holding that emergency-room physician is an agent of the hospital where patient had relied on hospital's representation that doctor was its servant); Arthur v. St. Peter's Hosp., 405 A.2d 443 (N.J. Super. Ct. Law Div. 1979) (holding hospital liable for conduct of independent contractor where patient justifiably believed physician to be agent of hospital). The fact that a physician is actually an independent contractor rather than an employee is irrelevant as long as it appeared that she was an employee. *See* David J. Oakley & Eileen M. Kelly, *HMO Liability for Malpractice of Member Physicians: The Case of IPA Model HMOs*, 23 TORT & INS. L.J. 624, 630 (1988).

[98] In a closed-panel health plan, subscribers are allowed to see only the providers that have been approved by the third-party payor.

lationship by virtue of their control over the delivery of medical care. It is reasonable for a patient to imply this relationship where the health plan purports to provide all the medical care for the subscriber and where the subscriber must choose a pre-approved physician from the closed panel.[99] Accordingly, courts in Indiana and Pennsylvania have held HMOs liable for patient injuries under the doctrine of ostensible agency.[100]

Finally, courts have indicated an increasing willingness to hold hospitals and insurers directly liable for patient injuries under a theory of corporate negligence. In the health care context, corporate negligence consists of a hospital's breach of its duty to the patient to "insure that its medical staff is qualified for the privileges granted"[101]—or possibly, an insurer's breach of its duty not to endanger the patient's health by measures such as cost containment.[102] It has been argued that application of corporate negligence doctrine to HMOs and Preferred Provider Organizations is appropriate because they restrict a patient's choice of physician, and should therefore have a duty to investigate the professional competence of member physicians.[103]

Wickline v. State[104] was the first case to accept the notion that cost-containment measures affecting a physician's medical judgment could result in a third-party payor's liability for patient injury.[105] Lois Wickline underwent surgical treatment for arteriosclerosis. Medi-Cal, the state medical-assistance program, au-

---

[99] *See* Griner *supra*, note 70, at 894.

[100] Sloan v. Metropolitan Health Council of Indianapolis, Inc., 516 N.E.2d 1104 (Ind. Ct. App. 1987) (finding that HMO could be held liable for physician's failure to diagnose); Boyd v. Albert Einstein Medical Ctr., 547 A.2d 1229 (Pa. Super. Ct. 1988) (finding that HMO could be held liable for actions of physicians under theory of ostensible agency).

[101] Johnson v. Misericordia Community Hosp., 301 N.W.2d 156, 165 (Wis. 1981). *Darling v. Charleston Memorial Hospital* was the first case to find that a hospital has an independent duty to supervise the medical treatment of staff physicians. 211 N.E.2d 253 (Ill. 1965), *cert. denied*, 383 U.S. 946 (1966). Decisions after *Darling* have limited its scope and imposed on hospitals only a duty to "select and retain competent physicians" and to ensure against grossly negligent care by a physician. *See* Modla v. Parker, 495 P.2d 494 (Ariz. Ct. App.) (refusing to apply *Darling* because the hospital staff's treatment did not actively retard or worsen the plaintiff's treatment and because their conduct was not "outrageous"), *cert. denied*, 409 U.S. 1038 (1972); Insigna v. Labella, 543 So.2d 209 (Fla. 1989) (requiring hospitals to use due care in selection and retention of physicians on medical staff).

[102] O'Neal, *supra* note 92, at 413.

[103] *See* Griner, *supra* note 70, at 896.

[104] 239 Cal. Rptr. 810 (Cal. Ct. App. 1986).

[105] *See* Griner, *supra* note 70, at 863–64.

thorized ten days of hospitalization for the surgical treatment.[106] Although her surgeon requested an eight-day extension for her to remain in the hospital, Medi-Cal granted only four additional days, in accordance with its utilization review determination.[107] She was discharged after four days and later suffered complications that required the amputation of one leg above the knee. The surgeon testified that, had Wickline remained in the hospital for the eight additional days he had requested, she would not have lost her leg.[108]

Wickline sued the state of California and alleged that the state negligently discontinued her Medi-Cal eligibility and caused her premature discharge from the hospital while still in need of continuing in-patient care, which resulted in amputation of her right leg.[109] The court of appeal held that Medi-Cal was not liable for Wickline's injuries.[110] As a factual matter, the court found that Wickline's physicians did not file a further request to discharge her hospital stay, even though they believed it medically necessary for her to remain hospitalized. The court emphasized, nonetheless, that the decision to discharge Wickline met the medical standard of care and that "Medi-Cal did not override the medical judgment of Wickline's treating physicians at the time of her discharge."[111]

Although the *Wickline* court pinned liability for the patient's injuries on the treating physician, it indicated a willingness to hold third-party payors accountable for patient injury arising out of negligence in the utilization review process or out of undue influence on medical decisions:

> Third party payors of health care services can be held legally accountable when medically inappropriate decisions result from defects in the design or implementation of cost-containment mechanisms as, for example, when appeals made on a patient's behalf for medical or hospital care are arbitrarily ignored or unreasonably disregarded or overridden . . . . [I]t is essential that cost limitation programs not be permitted to corrupt medical judgment.[112]

---

[106] *Wickline*, 239 Cal. Rptr. at 812. Medi-Cal administers distribution of federal Medicaid benefits to eligible California residents. *Id.*

[107] *Id.* at 813–14.

[108] *Id.* at 816–17.

[109] *Id.* at 811.

[110] *Id.* at 819.

[111] *Id.*

[112] *Id.* at 819–20.

The *Wickline* court also noted that California law specifically authorized Medi-Cal to conduct prior authorization of hospital admissions and extensions.[113] Thus, *Wickline* did not resolve the potential liability of a private, rather than public, program, such as a health plan in the managed competition scheme, where utilization review personnel conduct similar cost-containment measures.

Four years later, the California Court of Appeal addressed this issue in *Wilson v. Blue Cross*.[114] The plaintiff had subscribed to a conventional fee-for-service health insurance plan without explicit provisions for cost-containment procedures. He was admitted to a hospital for treatment of major depression, drug dependency, and anorexia. Though his treating physician had recommended three to four weeks of hospitalization, the plaintiff's insurance company refused to pay for hospital care beyond the tenth day of treatment because of a determination by an independent utilization review organization. The plaintiff could not afford to pay for his continued stay himself and was discharged. Twenty days later, he committed suicide.

The plaintiff's survivors sued his insurer and claims-review agent and alleged breach of insurance contract and tortious denial of payment for "medically necessary hospitalization."[115] The court of appeal overturned the lower court's summary judgment for defendants, which was based largely on an application of *Wickline*.[116]

The *Wilson* court distinguished *Wickline* on three grounds. First, the discharge decision in *Wickline* met Medi-Cal's standard for determining whether to provide care, which was the same as the standard of care for physicians.[117] Second, the state's duty in *Wickline* arose from statute and provisions of the California Administrative Code, which altered the normal course of tort liability in that case.[118] Third, "*Wickline* was not a case where a cost limitation program . . . was 'permitted to corrupt medical judgment.'"[119]

The *Wilson* court characterized as dicta the language in *Wickline* that suggested that liability for a discharge decision rests

---

[113] *Id.* at 820.
[114] 271 Cal. Rptr. 876 (Cal. Ct. App. 1990).
[115] *Id.* at 880.
[116] *Id.* at 878.
[117] *Id.* at 879.
[118] *Id.*
[119] *Id.* (citation omitted).

solely on the treating physician in all contexts.[120] The court found that

> [t]he legitimate rationale of *Wickline* was that the normal tort responsibility principles . . . were modified by the [statutory] provisions . . . so that a Medi-Cal recipient was entitled to medical care within 'the usual standards of medical practice within the community' . . . and that the discharge decision in that case fell within the standard of medical practice.[121]

The court found that sufficient evidence existed to raise a triable issue of whether the claims-review agent's conduct was a "substantial factor" in causing the plaintiff's death.[122] The court refused to hold that, analagous to *Wickline* where state policy mandated the utilization review process that was used, public policy considerations warranted protection of private insurance companies' utilization review procedures.[123] Finally, the court rejected the claims-review agent's contention that Wilson's treating physician was liable for failure to appeal the denial of insurance benefits for his patient because this argument relied solely on dicta in *Wickline* that had "no application to this case."[124]

---

[120] The defendants had relied on this broad language to argue that Wilson's treating physician was solely responsible for his discharge and subsequent suicide. *See* Wickline v. State, 239 Cal. Rptr. 810, 819 (Cal. Ct. App. 1986) ("The decision to discharge is, therefore, the responsibility of the patient's own treating doctor . . . . [T]he physician who complies without protest with the limitations imposed by a third party payor, when his medical judgment dictates otherwise, cannot avoid his ultimate responsibility for his patient's care.").

[121] *Wilson*, 271 Cal. Rptr. at 880 (quoting *Wickline*, 239 Cal. Rptr. at 819).

[122] *Id.* at 883. Evidence that might prove that Western Medical's conduct was a substantial factor included the fact that, "[o]nce the insurance benefits were terminated, there were no other funds to pay for the decedent's hospitalization. The sole reason for the discharge, based on the evidence adduced in connection with the summary judgment motion, was that the decedent had no insurance or money to pay for any further in-patient benefits." *Id.*

[123] The defendants cited *Wickline* in contending that the court should extend immunity to a health care payor that refuses to provide insurance benefits on the basis of a utilization review determination. *Id. Wickline* had indicated that utilization review involved "issues of profound importance to the health care community and to the general public." *Wickline*, 239 Cal. Rptr. at 811. Unlike *Wickline*, the court could not find any clearly and statutorily expressed public policy that applied to the plaintiff's contract with his insurer. *Wilson*, 271 Cal. Rptr. at 884.

[124] *Wilson*, 271 Cal. Rptr. at 884. One defendant cited the *Wickline* court's general language that the physician "who complies without protest with the limitations imposed by a third party payor, when [her or] his medical judgment dictates otherwise, cannot avoid his [or her] ultimate responsibility for [her or] his patient's care." *Id.* (quoting *Wickline*, 239 Cal. Rptr. at 819). The court found that the defendants failed to prove that, even if Wilson's physician had filed a reconsideration request, it would have been granted. *Id.* at 884–85.

Wilson held that cost-containment measures, such as utilization review, may breach a private insurance contract and narrowed the scope of *Wickline*'s exception from liability.[125] Under *Wilson*, if a patient can prove an injury resulting from an unwarranted discharge or limited treatment, negligence in the insurer's performance of a cost-containment measure can trigger liability.[126] Both *Wilson* and *Wickline* indicate courts' greater willingness to hold private health insurers accountable for the increasingly influential role that cost-containment measures may play in physician decision-making. Under managed competition, physicians would have even greater pressure to consider cost as a factor influencing their clinical decisions. These California cases suggest that a health plan's use of various strategies and incentives to induce a member physician to minimize costs may be sufficient to trigger its liability.[127]

IV. The Limits on Patient Choice

A. *The Doctor's Incentive Not to Disclose*

An intensely competitive health care environment may burden the doctor-patient relationship in another way. The conflict between the physician's interests and the patient's interests may crystallize in the area of information disclosure. Under managed competition, a doctor's allegiance is divided between her patient and the health plan that employs her. The pressure to limit medical resource consumption may lead the doctor not to mention to her patient possible tests and procedures. In her own mind, the doctor will conduct a cost-benefit analysis and conclude that the potential benefits of certain alternative tests or procedures are not worth the cost. Discussing additional diagnostic or therapeutic options may result in a confrontation with

---

[125] *See* O'Neal, *supra* note 92, at 416–17.

[126] *See id.* at 417.

[127] *Wickline* does not specify how cost-containment programs may "corrupt medical judgment." 239 Cal. Rptr. at 820. However, measures such as capitation payments, monthly length of stay reports by physicians, and sanctions for exceeding level of care targets have all been found to affect physician decisions. *See* Alan L. Hillman, *Financial Incentives for Physicians in HMOs: Is There a Conflict of Interests?*, 317 New Eng. J. Med. 1743, 1747–48 (1987); Alan L. Hillman et al., *How Do Financial Incentives Affect Physicians' Clinical Decisions and the Financial Performance of Health Maintenance Organizations?*, 321 New Eng. J. Med. 86, 88–89 (1989); O'Neal, *supra* note 92, at 417 n.89.

the patient, who may demand the further treatment. Thus, the risk that the doctor may simply say nothing is substantial.[128]

B. *Informed Consent*

The doctrine of informed consent offers little protection against this danger of nondisclosure. Although long recognized as an important component of medical decision-making, patient autonomy has not been recognized as a legally protected interest. Rather, it is vindicated as a corollary of two other interests: bodily security as protected by the rules against battery, and bodily well-being as protected by the rules of professional competence.[129] A group of early twentieth century cases protected physical security by applying rules that prohibit unconsented touching to mandate patient consent to specific procedures.[130] In the most famous of these cases, *Schloendorff v. Society of New York Hospitals*,[131] Justice Cardozo wrote the classic statement defending a patient's right to decide the course of treatment: "Every human being of adult years and sound mind has a right to determine what shall be done with his own body . . . ."[132]

Although battery doctrine protects patient autonomy to some extent, many aspects of these legal rules against unconsented touching do not correspond to the doctor-patient relationship. Intentional torts such as battery imply an antisocial motivation that courts are reluctant to associate with doctors. Thus, many courts and legislatures have limited battery actions to the rare

---

[128] *See* Gregory, *supra* note 82, at 493–94 (describing the pressure not to disclose potentially beneficial tests and procedures under the prospective payment system). The danger of nondisclosure is even more likely in the managed cometition setting, which targets *physician* incentives, than in the prospective payment setting, which affects *hospital* reimbursement.

[129] Shultz, *supra* note 82, at 219.

[130] *See, e.g.*, Pratt v. Davis, 79 N.E. 562 (Ill. 1906) (holding physician who performed hysterectomy without patient's consent liable for battery); Mohr v. Williams, 104 N.W. 12 (Minn. 1905) (finding for the plaintiff patient in her battery action against a physician who obtained consent to operate on her right ear but operated on her left ear instead), *overruled in part by* Genzel v. Halverson, 80 N.W.2d 854 (Minn. 1957).

[131] 105 N.E. 92 (N.Y. 1914), *overruled by* Bing v. Thunig, 143 N.E.2d 3 (N.Y. 1957).

[132] *Id.* at 93.

situations where the patient has not given *any* consent, rather than where the consent has merely been uninformed.[133]

The informed consent doctrine places an affirmative duty on physicians to disclose to their patients information sufficient to allow them to give intelligent and informed consent. Liability is predicated on a lack of consent, not negligent treatment.[134] To recover in an action predicated on nondisclosure, the plaintiff patient must first show a violation of the duty to inform.[135] Then the plaintiff must prove that the nondisclosure caused an injury. Most states have adopted an objective standard of causation requiring the plaintiff to show that the undisclosed information would have induced a reasonable patient to withhold consent to the proposed treatment.[136]

C. *A Critical Gap in Informed Consent Doctrine*

The doctrine of informed consent was developed in the fee-for-service regime where doctors had every incentive to do everything remotely beneficial for their patients. This kind of aggressive treatment was not only in the patients' best interests, but was also very lucrative for doctors. The fear that doctors were doing too much without their patients' consent spurred the

---

[133] *See* Cobbs v. Grant, 502 P.2d 1, 8 (Cal. 1972) ("We agree with the majority trend. The battery theory should be reserved for those circumstances when a doctor performs an operation to which the patient has not consented."); Shultz, *supra* note 82, at 224–26.

[134] *See* W. PAGE KEETON ET AL., PROSSER AND KEETON ON THE LAW OF TORTS § 32 (5th ed. 1984).

[135] *See* Shultz, *supra* note 82, at 226–27. As articulated in the landmark case of Nathanson v. Kline, 350 P.2d 1093 (Kan. 1960), the physician's duty to inform is initially measured against the yardstick of professonal custom. *Id.* at 1106 ("The duty of the physician to disclose . . . is limited to those disclosures which a reasonable medical practitioner would make under the same or similar circumstances."). Canterbury v. Spence, 464 F.2d 772 (D.C. Cir.), *cert. denied*, 409 U.S. 1064 (1972), establishes the more modern rule which guides a significant minority of jurisdictions. The new standard of disclosure focuses more on the informational needs of the reasonable patient than on the standards of disclosure established by doctors themselves. "[T]o bind the disclosure obligation to medical usage is to arrogate the decision on revelation to the physician alone. Respect for the patient's right of self-determination on particular therapy demands a standard set by law for physicians rather than one which physicians may or may not impose on themselves." *Id.* at 784 (footnotes omitted). The court went on to hold that "the patient's right of self-decision shapes the boundaries of the duty to reveal . . . . [T]he test for determining whether a particular peril must be divulged is its materiality to the patient's decision: all risks potentially affecting the decision must be unmasked." *Id.* at 786–87 (footnote omitted); *see also* ARNOLD J. ROSOFF, INFORMED CONSENT: A GUIDE FOR HEALTH CARE PROVIDERS 34–41 (1981) (describing the shift from the old "professional community" standard to the new "reasonable patient" standard).

[136] *See, e.g.*, *Canterbury*, 464 F.2d at 791.

development of informed consent. As a result, informed consent doctrine fails to protect patients in an array of situations where the doctor has made a unilateral decision not to act at all. For example, a doctor may be satisfied with a given level of diagnostic clarity or a certain outcome of treatment, or a doctor may decide not to administer any medical treatment at all. Battery doctrine is inapposite in each of these cases where no touching occurs. Informed consent doctrine may not provide any additional protection because the rules governing informed consent are still touch-oriented. Consequently, although each of these situations implicates important autonomy interests, the patient's consent to these decisions may not be required.[137]

Marjorie Maguire Shultz has argued that an informed consent theory of negligence subordinates protection of patient choice to the dominant interest in physical well-being.[138] As a result, the duty of disclosure may not exist where the case is not the standard prototype of an unconsented touching. For example, in cases not involving physical touching, analysis of the duty to disclose may shift to issues about professional care of physical well-being.[139] In *Roark v. Allen*,[140] the defendant doctor noticed indentations on the plaintiff's baby's head after delivery and examined the infant for skull fractures. He concluded that none existed without ordering confirmatory x-rays. After fractures were later detected and corrected, the parents sued the doctor for lack of informed consent on the theory that his failure to inform them of the possibility of fractures was a breach of his duty to protect their interest in choosing further tests. The Texas Supreme Court held that "the doctrine of informed consent applies only to medical procedures which have yet to be performed and . . . is inapplicable . . . where the patient has already undergone the proposed treatment and been injured."[141] The court shifted the issue from the patient's interest in choice of possible courses of action to the physician's professional competence in diagnosing the skull fractures.[142] Although the uncer-

---

[137] Shultz, *supra* note 82, at 230.
[138] *Id.* at 232–41.
[139] *Id.* at 238–40.
[140] 633 S.W.2d 804 (Tex. 1982).
[141] *Id.* at 808.
[142] Shultz, *supra* note 82, at 239. Shultz also cites Sinkey v. Surgical Assocs., 186 N.W.2d 659 (Iowa 1971) (transforming the issue of the duty to inform the parents of a child patient about the radiologist's opinion into an issue of the competence of the reading of the x-ray).

tainty of medical judgment legitimately excuses a failure to diagnose correctly, it should not excuse a failure to disclose information that would have induced a patient to choose a different option, such as further diagnostic tests.[143]

Some courts have held that reasonable alternative procedures fall within the scope of disclosure requirements.[144] In these cases, the patient's consent to a procedure is rendered invalid by the doctor's failure to mention a reasonable alternative procedure.[145] By this reasoning, the doctor's failure to mention any potential treatment options at all may result in liability as well. In practice, however, the requirement to disclose alternative procedures does little to further the patient's interest in learning about her choices when the doctor does not recommend any treatment. Courts have used this theory only to render consent to some affirmative treatment invalid, but have not applied it in instances where there has been no treatment at all.[146]

When the doctor's recommendation is not to act, she can implement that unilateral decision without the patient's awareness. Existing touch-oriented informed consent legal doctrines reinforce the view of these unilateral decisions as issues of professional competence rather than patient choice.[147] The danger in the managed competition setting is that the physician's incentive not to disclose costly diagnostic or therapeutic options

---

[143] Shultz, *supra* note 82, at 240; *cf.* Michael J. Rider, Comment, *Informed Refusal: Physician Liability for Failure to Inform of the Risks Associated with Refusing Diagnostic Tests*, 19 SAN DIEGO L. REV. 823, 824–25 (1982) (arguing that, although diagnostic decisions are traditionally made only by physicians, the role of the patient to participate in diagnostic decision-making has been expanded by the doctrine of "informed refusal").

[144] *E.g.*, Holt v. Nelson, 523 P.2d 211 (Wash. Ct. App. 1974) (holding that failure to advise parents of alternative to proposed treatment for their child was lack of informed consent); *see also* ROSOFF, *supra* note 135, at 47–50 ("Several courts have held that consent is not informed unless the physician discloses to the patient at least those alternatives that would be generally acknowledged within the medical community as feasible in the patient's case."); FAY A. ROZOVSKY, CONSENT TO TREATMENT: A PRACTICAL GUIDE 47 (1984) ("In addition to risk and benefit information, a patient should be told of the availability of *reasonable* alternative procedures.").

[145] *See* John H. Derrick, Annotation, *Medical Malpractice: Liability for Failure of Physician to Inform Patient of Alternative Modes of Diagnosis or Treatment*, 38 A.L.R.4TH 900 (1992).

[146] *See, e.g.*, Saliv v. United States, 522 F. Supp. 989 (M.D. Pa. 1981) (holding that, under Pennsylvania law, the staff's failure to inform patient of conservative therapy as an alternative to an angiogram vitiated patient's consent to the angiogram); Logan v. Greenwich Hosp. Ass'n, 465 A.2d 294 (Conn. 1983) (holding that failure to discuss alternative procedure of open biopsy rendered consent to close needle biopsy invalid); Sard v. Hardy, 379 A.2d 1014 (Md. 1977) (finding surgeon's failure to inform patient of different methods of sterilization as alternatives to tubal ligation created jury question of whether information withheld by surgeon was material to patient's decision).

[147] *See* Shultz, *supra* note 82, at 254.

corresponds exactly to the major gap in informed consent doctrine where patient autonomy is not protected: a physician's recommendation of inaction.[148]

The doctor-patient relationship has often been characterized as one that is fiduciary in nature.[149] As Angela Roddey Holder has explained:

> The relationship between patient and physician is one known to the law as a "fiduciary relationship." Any person such as a physician, attorney, priest or other who enters into a relationship of trust and confidence with another has a positive obligation to disclose all relevant facts . . . [s]ince the essence of a professional relationship is that the professional knows more about his subject than the person who seeks his help . . . .[150]

The key elements of the fiduciary relationship between a doctor and patient are trust and asymmetry in knowledge. The doctor, in the position of superior knowledge, has a special obligation to convey this information to the patient so that she may truly arrive at informed consent. A commentator remarked about the prospective payment system:

> If DRGs implicitly influence physicians to withhold from a patient material information about the patient's own diagnosis, prognosis, and potential alternatives and their risks and benefits, and thus infringe on the patient's decision-making autonomy, then DRGs will have struck at the heart of the physician/patient relationship and have exposed physicians to legal liability for failure to fulfill the fiduciary duty to communicate.[151]

The patient, in turn, must trust and rely on her doctor to act in her best interests and to relay information regarding the clinical

---

[148] *Cf.* Gregory, *supra* note 82, at 489, 493–94. Gregory describes how, in the prospective payment setting, a physician whose allegiance is divided between his patient's best interests and the financial well-being of the medical facility he utilizes might conclude that the cost of a procedure outweighs any potential benefit and unilaterally decide to forego the procedure. Gregory notes that, to avoid confrontation with the patient, who would demand the additional service if informed about the possible benefits, the doctor might be tempted simply not to mention the additional procedure. The outcome Gregory describes in the prospective payment setting would be even more likely under managed competition where the pressure to limit resource competition is more directly targeted at *doctors* rather than at *hospitals*.

[149] *See, e.g.*, Miller v. Kennedy, 522 P.2d 852 (Wash. Ct. App. 1974), *aff'd*, 530 P.2d 334 (Wash. 1975).

[150] ANGELA R. HOLDER, MEDICAL MALPRACTICE LAW 225 (1975).

[151] Kapp, *supra* note 75, at 248.

problem, alternative treatments, and their attendant risks.[152] Where a conflict of interest, such as pressure to remain economically competitive, is possible, the fiduciary's accountability for disclosure should not be confused with her separate and cumulative accountability for professional competence. Thus, the view of the physician as a fiduciary may offer greater protection against nondisclosure than informed consent doctrines.[153]

Because of the patient's reliance on her physician for information, nondisclosure induced by economic incentives threatens to remove the patient from the decision-making process altogether. A patient should have a right to know when cost-benefit analysis affects her doctor's recommendation. She should also have right to know potentially beneficial treatments.[154] Certainly, the decision not to recommend a diagnostic or therapeutic option due to cost constraints is an independent decision. A reasonable person in the patient's circumstance would want to know about potentially beneficial procedures and his doctor's financial incentives. The baseline assumption should not be that patients are incapable of taking into account financial considerations when making their health care decisions.[155] With full information, a patient might in fact forego a costly procedure.

Moreover, a belief in patients' inability to make decisions in the best interests of society is unjustified medical paternalism and provides no excuse to conceal information regarding either potential treatments or the financial considerations that militate

---

[152] *See* Cobbs v. Grant, 502 P.2d 1, 9 (Cal. 1972) ("[T]he patient, being unlearned in medical sciences, has an abject dependence upon and trust in his physician for the information upon which he relies during the decisional process, thus raising an obligation in the physician that transcends arm-length transactions."); Michael Bayles, *Obligations to Clients*, *in* MEDICAL ETHICS 107, 111–12 (Natalie Abrams & Michael D. Buckner eds., 1983); Shultz, *supra* note 82, at 235 ("[R]eceiving information from the doctor is the only way a patient can become aware of a pending choice. Nondisclosure here is tantamount to loss of the choice interest itself.").

[153] *See* Shultz, *supra* note 82, at 260–61. *But cf.* Joan Vogel & Richard Delgado, *To Tell the Truth: Physicians' Duty to Disclose Medical Mistakes*, 28 UCLA L. REV. 52, 67 (1980) (observing that the duties of disclosure imposed on doctors have been less extensive than those imposed on other fiduciaries).

[154] *See* Stone, *supra* note 9, at 312.

[155] To the contrary, consumers are showing more and more willingness to consider the financial impact of treatment decisions. A study conducted by the President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research revealed that two-thirds of the public desire more information about the costs of various treatments. *See* 1 PRESIDENT'S COMM'N FOR THE STUDY OF ETHICAL PROBLEMS IN MEDICINE & BIOMEDICAL & BEHAVIORAL RESEARCH, SECURING ACCESS TO HEALTH CARE 197 (1983).

against them.[156] John Stuart Mill's explanation of autonomy interests remains compelling:

> The only freedom which deserves the name is that of pursuing our own good in our own way, so long as we do not attempt to deprive others of theirs or impede their efforts to obtain it. Each is the proper guardian of his own health, whether bodily *or* mental and spiritual. Mankind are greater gainers by suffering each other to live as seems good to themselves than by compelling each to live as seems good to the rest.[157]

The fact that doctors are faced with medical uncertainty means that there is a high level of election in decision-making, and extra-medical values necessarily define the resulting choices.[158] In the end, it should be the patient herself who makes the difficult decisions on the basis of fully disclosed costs and benefits, and the law should fortify the patient's rights of choice, autonomy, and self-determination.[159]

V. INFORMED REFUSAL

Two decisions in Washington[160] and one in California[161] have forged a basis for vindicating patients' rights in situations where the doctor has arrived at a decision not to use a particular procedure.[162] These cases have developed an analogue to informed consent, "informed refusal,"[163] which refers to the pa-

---

[156] *See* Allen Buchanan, *Medical Paternalism, in* MORAL PROBLEMS IN MEDICINE 49 (Samuel Gorovitz et al. eds., 2d ed. 1983) (arguing that paternalistic withholding of information is unjustified and perpetuated by the distinction between "ordinary" and "extraordinary" therapeutic measures); Shultz, *supra* note 82, at 274–75 (arguing that the problem of medical paternalism transcends boundaries of traditional conflict of interest).

[157] JOHN STUART MILL, ON LIBERTY 12–13 (Alburey Castell ed., 1947).

[158] *See* Shultz, *supra* note 82, at 272, 276.

[159] *See* Gregory, *supra* note 82, at 495–96; Shultz, *supra* note 82, at 223 (arguing that final authority for important decisions should remain with the patient); *cf.* Stone, *supra* note 9, at 312 ("[W]hy should a sick and anxious patient accept the doctor's economic calculation?").

[160] Keogan v. Holy Family Hosp., 622 P.2d 1246 (Wash. 1980); Gates v. Jensen, 595 P.2d 919 (Wash. 1979).

[161] Truman v. Thomas, 611 P.2d 902 (Cal. 1980).

[162] *See* Gregory, *supra* note 82, at 489–92 (analyzing these three cases and their effect on the physician's incentive not to disclose in the prospective payment system).

[163] This term was coined by the California Court of Appeal in Truman v. Thomas, 155 Cal. Rptr. 752, 757 (Cal. Ct. App. 1979), *rev'd*, 611 P.2d 902 (Cal. 1980), to describe a situation where the doctor failed to inform a patient of the possible consequences of refusing a pap smear.

tient's right to be informed of the risks and potential consequences of refusing or omitting[164] a diagnostic test.

The first of these cases, *Gates v. Jensen*,[165] considered whether the doctrine of informed consent requires a physician to inform a patient of diagnostic procedures that may be taken to determine the significance of a detected physical abnormality. The physician in *Gates* found that pressure readings in the plaintiff's eyes registered in the borderline area for glaucoma. Without dilating the patient's pupils for increasing visibility, the physician examined the patient's optic nerves and diagnosed her problem as difficulties with her contact lenses. He treated her without mentioning either that he had found borderline-glaucoma pressure readings in both eyes or that this high pressure and her myopia increased her risk of glaucoma. Nor did he mention the existence of two simple, inexpensive, and risk-free tests that would have revealed her glaucoma. Later, the patient's symptoms worsened, and she became functionally blind from open-angle glaucoma.

The Supreme Court of Washington reversed a trial court verdict for the doctor.[166] First, the court noted the fiduciary nature of the physician-patient relationship, which required the physician to inform the patient of abnormalities in her body to allow her to make an informed decision about her course of medical care.[167] The court's holding articulated a new standard under which the doctor's *knowledge* of an abnormality triggers the duty to disclose:

> The existence of an abnormal condition in one's body, the presence of a high risk of disease, and the existence of alternative diagnostic procedures to conclusively determine the presence or absence of that disease are all facts which a patient must know in order to make an informed decision on the course which future medical care will take . . . . The physician's duty of disclosure arises, therefore, *whenever*

---

[164] *Gates*, 595 P.2d 919, and *Keogan*, 622 P.2d 1246, involved defendant physicians who omitted tests without even mentioning their existence to the patients, as compared to *Truman*, 155 Cal. Rptr. 752, where the patient knew of the test, but refused it without understanding the full ramifications of such refusal. The term "informed refusal" contemplates both situations.

[165] 595 P.2d 919 (Wash. 1979).

[166] *Id.*

[167] The court cited Miller v. Kennedy, 522 P.2d 852 (Wash. Ct. App. 1974), *aff'd*, 530 P.2d 334 (Wash. 1975). The court explained that a patient's right to know material facts regarding her condition is not limited to treatment choices after a disease has been conclusively diagnosed, but includes non-treatment situations as well. *Gates*, 595 P.2d at 922–23.

> *the doctor becomes aware* of an abnormality which may indicate risk or danger.[168]

This case could have led to developments in informed consent doctrine that supersede the touch-based limitations on protection of patient choice to affirmative proposed intervention.[169]

A majority of the same court indicated one year later in *Keogan v. Holy Family Hospital*[170] that the duty to inform does not arise when a doctor observes a symptom that could have a potentially fatal cause. The plaintiff in that case complained of exertional chest pain. The physician diagnosed the patient's condition as sternochondral inflammation. Although the doctor suspected angina, he disclosed neither this suspicion nor the possibility of receiving three readily available tests to diagnose angina. Two weeks later, the patient returned with postprandial pain and problems. Despite anomalous test results, the physician again did not mention the possibility of heart disease or diagnostic tests for angina. Shortly thereafter, the patient died of a heart attack.

The facts of *Keogan* were similar to those in *Gates*, yet five concurring justices indicated that they would not impose on the doctor a duty to disclose the detected abnormality.[171] Drawing a tenuous distinction from *Gates*, these justices insisted that the physician in *Keogan* had not arrived at a diagnosis or detected a bodily abnormality. The five justices characterized the doctor's knowledge of the patient's abnormal cardiac enzyme tests and abnormal EKG as "[a] suspicion of a possibility of an abnormality (angina pectoris)" that "hardly seem[ed] sufficient to trigger, as a matter of law, a duty to inform a patient of tests that could be given to diagnose the severity of such an evanescent bodily abnormality."[172] Ironically, the five justices noted

---

[168] *Gates*, 595 P.2d at 923 (emphasis added).

[169] Shultz, *supra* note 82, at 243.

[170] 622 P.2d 1246 (Wash. 1980).

[171] Joined by two other justices, Justice Horowitz delivered the opinion of the court which reversed both the trial court's refusal to give informed consent instructions on the primary care physican's failure to disclose available diagnostic procedures and the trial court's refusal to find that the emergency room physician was negligent as a matter of law. Five justices concurred only as to the emergency room physician's negligence as a matter of law. Thus, an actual majority of five justices refused to apply the *Gates* duty of disclosure to the primary care physician. Although the precedential effect of the opinion is uncertain, it would appear that *Keogan* did not overrule *Gates*, but did limit its holding.

[172] *Keogan*, 622 P.2d at 1261. The dissent appears to have confused the two issues of the doctor's negligence in failing to diagnose the angina and his failure to inform the patient of available diagnostic tests—a pattern that Shultz labels "transposition" of the

the potential financial impact of requiring disclosure of detected abnormalities and diagnostic tests to evaluate their significance: "[T]he cost of medical attention is escalating, seemingly at an ever increasing rate. As this court dictates what doctors must do to protect themselves in malpractice actions, it adds to that escalation."[173] Although the five justices may have been concerned that the court's opinion in this case would give physicians improperly motivated under the fee-for-service payment system an excuse to disclose and recommend unnecessary tests, a more disturbing possibility is that physicians in an economically competitive regime will withhold information regarding potentially beneficial diagnostic and therapeutic options in an effort to curb costs. Informed refusal offers to fill some of the gaps in informed consent doctrine and to preserve a degree of patient choice even within an economically competitive health care setting.

Cases in California have paralleled the increased disclosure requirements in Washington for medical decisions not to treat. In *Truman v. Thomas*,[174] the Supreme Court of California held that a physician breached his duty of care by failing to inform a patient of the potentially fatal consequences of allowing cervical cancer to progress undetected by a pap smear. During the six years in which he treated the patient, the physician failed to perform a pap smear test, which would have revealed the cancerous tumor at an operable stage. Although the patient had twice refused the test because of financial constraints, the doctor had never informed her of the risks in foregoing the pap smear test.

Applying the leading California case on informed consent, *Cobbs v. Grant*,[175] the *Truman* court noted that the amount of information a patient needs in order to make an informed choice determines the scope of a physician's duty to disclose.[176] Thus, the court held that "[i]f a patient indicates that he or she is going to *decline* [a] risk-free test or treatment, then the doctor has the

---

duty to disclose. *See id.* ("If Dr. Snyder was negligent because he should have discovered Keogan's diseased heart and failed to do so, that is what should be alleged and proved in this case . . . . This court with its benefit of hindsight should not now enter the fray on the plaintiffs' side with rulings as a matter of law as to what the doctor should have told the patient."); Shultz, *supra* note 82, at 238–41; *supra* notes 139–143 and accompanying text.

[173] *Keogan*, 622 P.2d at 1261.

[174] 611 P.2d 902 (Cal. 1980).

[175] 502 P.2d 1 (Cal. 1972).

[176] *Truman*, 611 P.2d at 905.

additional duty of advising of all material risks of which a reasonable person would want to be informed before deciding not to undergo the procedure."[177]

The three-judge dissent distinguished *Cobbs* where there had been a surgery, an intrusion, to which the patient's consent may have been invalid. In *Truman*, there was no such intrusion, and accordingly, the dissent argued against extending *Cobbs* to require that the doctor inform a patient of the risks of inaction.[178] The dissent objected to the new disclosure requirements because they imposed upon doctors "the intolerable burden of having to explain diagnostic tests to healthy patients . . . [and] to provide each such patient with a summary course covering most of his or her medical education."[179] Finally, the dissent noted another critical distinction from *Cobbs*: "The *Cobbs* duty to warn in cases where an adequately informed prudent person would have declined treatment shows a concern for preventing over-selling of services by physicians. By contrast, today's duty appears designed to increase selling of medical services."[180] As in *Keogan*, the dissent's point may be well taken in a fee-for-service system where physicians are paid more for increasing their services, but not in an economically competitive regime where the emphasis on cost-containment may motivate the physician to decide unilaterally to withhold treatments or tests.

Gates, *Keogan*, and *Truman* have been viewed as cases that elevate the patient's role in the decision-making process by extending that role to include participation in diagnostic decisions—traditionally the exclusive province of physician discretion.[181] However, these three cases do more than mandate patient participation in the diagnostic phase of decision-making. By placing an affirmative duty on physicians to disclose the risks of refusing or omitting tests, they afford some level of protection against unilateral decisions on the part of physicians to withhold some potentially beneficial test or procedure. As developed in these cases, the additional duty of disclosure can

---

[177] *Id.* at 906. The physician had raised the argument that *Cobbs* did not apply to him because the duty to disclose arises only where the patient *consents* to a recommended procedure. The court rejected this argument on the basis that imposing a burden on patients to inquire into the potential consequences of their decisions is inconsistent with *Cobbs*. *Id.*

[178] *See id.* at 911 ("When no intrusion takes place, no need for consent—effective or otherwise—arises.").

[179] *Id.* at 909–10.

[180] *Id.* at 911.

[181] *See* Rider, *supra* note 143, at 824–25.

be justified in terms of patient autonomy.[182] The reasoning behind these cases, which requires disclosure of potential treatments and their benefits to make patients' refusal of these options informed, would also apply in the managed competition setting where the physician has decided not to act, in part because of financial considerations.

A. *The Dilemma for Doctors*

Although informed refusal offers patients some protection against their physicians' unilateral withholding of tests and procedures, the doctrine offers little guidance for physicians in their clinical practice. In the first place, the scope of the physician's duty to disclose is unclear.[183] For example, when will a physician's decision not to mention a procedure be considered a breach of her duty to inform her patient of an "abnormal condition,"[184] and when will that same decision be deemed a mere "suspicion of a possibility of an abnormality"?[185] It may be that physicians will have to err on the side of disclosure to avoid liability; doctors may feel obligated to discuss every potentially beneficial procedure and test. This is hard to reconcile with their new and simultaneous role as rationers of health care. The problem may be an example of what Alan Stone has identified as "uncoordinated and contradictory legal regulations of health care."[186]

B. *Trust Within the Doctor-Patient Relationship*

Application of informed refusal to require increased disclosure in a managed competition setting also has uncertain implications for the trust in doctor-patient relationships. It has been argued that in entering a partnership of "pervasive trust and good faith," the physician can deliver legitimate services and

[182] *See id.* at 829–30. Rider argues that informed refusal can also be justified by the logical relationship between refusal and consent. When a patient assents to one treatment, she simultaneously refuses alternative treatments and non-treatment. Thus, both informed consent and refusal involve a patient's exercise of control over diagnostic choices.

[183] *See id.* at 831–32.

[184] Gates v. Jensen, 595 P.2d 919, 923 (Wash. 1979).

[185] Keogan v. Holy Family Hosp., 622 P.2d 1246, 1261 (Wash. 1980).

[186] Stone, *supra* note 9, at 310.

justifiably deny unnecessary ones.[187] However appealing this ideal situation is, the reality is that increased disclosure may force the physician into yet another dilemma: how to respond to the patient who, upon being informed of the availability of an additional test, and after conducting her own cost-benefit analysis, demands that it be performed.[188] A patient may have a right to know her doctor's financial incentives, but full disclosure of these motivations may devastate the physician-patient relationship: "[I]f it is ethically wrong to conceal these new economic incentives and the medical profession's responses to them, to reveal them may threaten the trust and confidence of patients even in 'caring physicians.' 'Caveat emptor' will be more relevant than 'primum non nocere'[189] in doctor-patient relationships."[190]

VI. CONCLUSION

As a model of reform for solving the nation's health care cost crisis, managed competition proposes to restructure the health care market in such a way that consumers will become more conscious of the cost implications of their health care decisions. The model encourages physicians to remain economically competitive. The extent to which managed-care plans will be held liable for the adverse consequences of their cost-containment measures designed to influence physicians' clinical decisions is uncertain. There is a danger that doctors will react to new economic pressures by withholding information regarding potentially beneficial treatments or tests. Developed in a regime where doctors had incentives to do everything remotely beneficial for their patients, informed consent has, for the most part, been limited to situations where there is some affirmative treatment or procedure to which the patient did not consent in an informed manner. Thus, a physician's unilateral decision to withhold treatment and information about treatment because of

---

[187] James L. Schroeder et al., *Prepaid Entitlements: A New Challenge for Physician-Patient Relationships*, 254 JAMA 3080, 3082 (1985); *see also* Kapp, *supra* note 75, at 251 (reasoning that informed consent under DRGs may necessitate candid discussion between doctor and patient regarding quality, cost, and resource allocation; suggesting that dialogue between doctor and patient may actually be cost-effective).

[188] *See* Gregory, *supra* note 82, at 496.

[189] "First of all, do no harm."

[190] Stone, *supra* note 9, at 312; *see also* Morreim, *supra* note 76, at 35 (referring to prospective payment settings).

financial considerations is especially problematic because it corresponds to the gap that informed consent fails to address, inaction. Thus, the physician incentives created by managed competition might cause patients to lose diagnostic or therapeutic choices without their knowledge. The development of informed refusal might afford some protection against the removal of patients from the decision-making process, but it is far from a panacea: the problems of the doctor's uncertainty under the doctrine and the patient's diminished trust in her physician remain to be addressed.

—*Elaine Lu**

* I would like to thank Professor Alan A. Stone for his invaluable advice, guidance, and support during the initial stages of this Recent Development.

BOOK REVIEWS

SERIOUS AND UNSTABLE CONDITION: FINANCING AMERICA'S HEALTH CARE. By *Henry J. Aaron*. Washington, D.C.: Brookings Institution, 1991. Pp. 158, index, appendices. $26.95 cloth, $9.95 paper.

The U.S. health care system is in critical condition. The United States spends more money on health care than any other industrialized nation,[1] yet millions of Americans live without adequate insurance coverage. This anomaly has fueled a debate on how to extend insurance coverage to the unemployed and increase both acute and long-term illness coverage, while curbing and actually reversing the climbing costs of health care. In *Serious and Unstable Condition: Financing America's Health Care*, Henry J. Aaron[2] outlines the economic problems facing our health care system and offers possible solutions to this health care dilemma.

Aaron believes the current U.S. health care system is plagued by a three-fold problem (p. 12). First, despite efforts to streamline insurance plans, health care expenditures by businesses and government have more than tripled in the past twenty years.[3] Second, many Americans may lose acute-care health benefits as a result of the increasing unemployment rate and employers' decisions to terminate insurance coverage. Only a fraction of insured Americans carry long-term insurance which covers care for illnesses requiring extended nursing care in either the private home or a nursing home (pp. 57–59). Third, Aaron credits the technological boom for causing a number of new experimental procedures to be adopted without a full cost-benefit analysis. These newer procedures tend to be overused, while other effective and less costly methods of care are underutilized.[4] Aaron contends that although these problems have existed for years,

---

[1] *See* George J. Schreber & Jean-Pierre Poullier, *Overview of International Comparisons of Health Care Expenditures*, 11 HEALTH CARE FINANCING REV. 1, 1–121 (1989) (based on percentage of GDP per capita).

[2] Henry J. Aaron is the director of the Economic Studies program at the Brookings Institution.

[3] Figures are based on an adjusted dollar figure that compensates for inflation. The amounts compare per capita outlays over a 40-year period (pp. 38–39).

[4] Aaron's view is unusual given the current trend to approve experimental drugs and procedures quickly in an effort to benefit terminally ill, and more often HIV-positive patients.

the United States must overhaul its financing policy soon to preserve a workable health care system.

Aaron offers several reasons why costs are spiraling upwards. For example, since insurance requires only a nominal financial outlay at the time of illness, individuals have access to virtually all procedures which are marginally related to a medical condition (pp. 39–40). The insured individual has no incentive to limit the services provided by health care professionals (p. 9). Sick people with insurance will not balance the costs and benefits of medical procedures, and therefore the demand for high-cost, low-benefit care increases. Additionally, insurance encourages preventive care and services which benefit society as a whole (i.e., vaccinations) which would not be utilized as much if insurance did not exist to cover these provisions (p. 12).

Aaron suggests that these factors contribute to higher costs and present several obstacles to reform. First, social values favor this attitude towards health care which is devoid of cost analysis. A basic survival instinct drives individuals to seek all care which may aid in recovery, and many argue that each individual should have control over the amount and type of care used to preserve his own life. Furthermore, health care should aim to prevent illness, saving the insured the mental anguish, lost work time, and possible further complications which may result from a serious illness.[5] Additionally, Aaron stresses that there is no clear index to evaluate the benefit of medical procedures and not enough knowledge exists to adequately judge the appropriate quantity of care necessary for any given ailment. The "fuzzy benefits curve" and the variation in care imply that inefficiencies exist in allocating resources. Since society has not reached a consensus on which procedures are of negligible value and what benefits are actually achieved from procedures, one cannot determine if expenditures are too large or small (pp. 19–21).

A health care policy that "fashions incentives that preserve meritorious increases in demand but curtail the low or no-benefit increases" might be achieved through rationing and prioritizing health care and eliminating procedures that are not cost effective (p. 13). Acknowledging the argument that any procedures which

---

[5] Aaron contends that the potential to save costs for medical care through prevention is seldom realized. His calculation, though, does not factor in economic and emotional harms which accompany illnesses which may be preventable (p. 12).

can extend or improve life should be pursued, Aaron counters that society makes daily choices which jeopardize lives in the name of cost or convenience, and thereby "contradict this pious sentiment" (p. 13). Adults keep chemicals in their garages and forgo tamper-proof caps on bottles to which children may have access (p. 13). However, these risky actions can be distinguished from health care rationing in two ways. First, health care rationing is a public action which limits individual choice, not a private action in which the individual chooses whether to ignore a risk. Second, the elimination of a health remedy, unlike risk-taking behavior, ignores a possible real and immediate benefit to the insured. Leaving household chemicals in an enclosed, non-ventilated area does not necessarily pose an immediate threat. Even if it did, the person affected would have the opportunity to determine for himself whether keeping the paint in the garage was worth the possible risk. The individual maintains the power to choose.

Aaron counters this assumption of individual choice in health care by arguing that, in fact, individuals do not control their health care choices. Only the physicians, as experts in diagnosis and care, truly have this power. Physicians theoretically have the power to determine demand for their own services (p. 15). In addition, they often practice "defensive medicine" to protect themselves from malpractice litigation (p. 45). These unnecessary tests and services are performed to protect the doctor rather than the patient, and therefore not only decrease the patient's level of individual choice, but also increase the costs of the service.

Aaron further points out that the aging American public and rising relative wages in the health care professions also contribute to the rise in health costs.[6] However, high demand, fear of lawsuits, age, and wage increases do not compose the bulk of health care costs. Aaron believes that technology contributes most greatly to the rising cost of health care. While innovation and continued research in biomedical areas can provide substantial societal benefits (p. 24), technology can also have the adverse effect of producing techniques which are only marginally beneficial and which are not thoroughly evaluated for effi-

[6] The Bureau of the Census provides information proving that the "consumption of health care" increases with age (p. 43). The salary for nurses and other health care professionals doubled on average between mid-1963 and 1989 (p. 41).

ciency (p. 49). In order to reduce health care costs on a long-term basis, low-benefit and high-cost procedures must be rationed (p. 53). Since health care is concentrated on the very ill and elderly, Aaron's plan would focus on curtailing services to these groups.

To fully analyze the health care problem, Aaron also considers the system's source of funding. Our current health care system, which insures six of seven Americans (p. 53), is a mix of private and governmental efforts. The private market provides coverage to individuals insured through their employment, and the federal government provides health care to target groups, influences private insurance companies through taxes and regulations, and funds biomedical research (p. 60).

The private or competitive system is praised for the choice of package plans it offers customers, and some suggest that this competition has encouraged preferred provider organizations to reduce costs and streamline efficiency (pp. 28–29). Aaron observes that some disadvantages include the enormous overhead of administering multiple coverage claims, the lack of incentive for providing long-term care, and the use of "experience ratings." Experience ratings as used in the current system result in inequities in insurance costs based upon personal or job-related characteristics. Insurance companies may set premiums based on an amount of care anticipated by "experience" factors such as age, sex, race, or degree of hazard in a job. While allowing insurance companies to effectively punish employers who subject their employees to high risk by assessing a higher rate, companies also have the flexibility to charge a higher rate to persons with a genetic tendency towards a disease or to deny coverage to persons with a pre-existing condition requiring extensive care (pp. 30–36). Aaron suggests that varying premiums with workplace conditions or controllable behaviors (i.e., smoking) would be ideal, but concludes that in practice experience ratings would be ineffective.[7]

Under the current market system, 8.6 million Americans became uninsured between 1979 and the mid-1980s (pp. 75–76). Aaron suggests that several factors have caused this trend towards decreasing coverage. For example, employers have

[7] Aaron suggests that low-risk groups would enjoy competitive rates while high-cost groups would be subject to poor marketing and poor plans. In order to insure fair ratings, government monitoring and regulation would be required, thereby undercutting the benefit of a free market system (p. 36).

shifted premiums to employees who subsequently dropped coverage. High-risk businesses may have had trouble finding or affording coverage (p. 76). Perhaps the most discomforting data presented suggests that the uninsured not only suffer economically when illness strikes, but they also suffer from an inferior quality of health care service (p. 77).[8]

Aaron goes on to compare the U.S. system to foreign health care systems. He concedes that while it compares "poorly by broad indicators with European and Canadian systems, [the U.S. system] performs well in many specific areas" (p. 101). By granting universal health care, the European and Canadian systems induce rationing of high-tech services, spark shortages in equipment, fail to provide incentives for medical improvements, and render patients hostage to large bureaucracies (pp. 101–02). Aaron believes that the United States needs to achieve the universal access to health care which is common worldwide, while reducing low-benefit, high-cost care.

After establishing the need for reform and the desired outcomes of that reform, Aaron presents three broad approaches to achieving these goals: voluntary incrementalism, mandatory employer-sponsored insurance, and mandatory restructuring (p. 110). Voluntary incrementalism modifies the current system of employer-sponsored insurance and expands government programs to include a large number of those currently uninsured (pp. 110–18). Tax credits would be awarded to low-income households, and small businesses would join forces to form "risk pools" which would lower the surcharges billed to small businesses. In addition, Medicaid-like programs would be extended to the uninsured.

Critics of voluntary incrementalism note that very large subsidies would have to be granted to provide coverage to all uninsured or underinsured people. If government subsidies were large enough to accomplish this goal, then employers would have little incentive to privately insure and would instead opt to join a risk pool. Using public funds as a crutch, employers may eliminate the search for competitive private insurance, and the free market core of this approach would be compromised

---

[8] Aaron looks to a study showing that uninsured patients are often more ill when being admitted to a hospital and are more likely to die while in hospital care (p. 77). *See* Jack Hadley et al., *Comparison of Uninsured and Privately Insured Hospital Patients: Condition on Admission, Resource Use, and Outcome*, 265 JAMA 374–79 (1991).

(pp. 118–19). However, if insurance companies actively market their plans and offer competitive long-term insurance plans, the health care system may experience cost savings over a period of time. However, these savings may be difficult to discern against the backdrop of rising absolute costs of health care.

Aaron's second approach, mandatory employer-based insurance, would use government resources to underwrite the health care system. Employers would be required either to sponsor employees' acute-care insurance or pay a tax to finance a public plan which would serve the unemployed and employees of businesses that opted for the tax payment. This program would limit the financial responsibility of businesses by limiting costs to a certain percentage of the payroll, imposing a cap on health care expenses (where public funds would pick up the difference between the care needed and the care covered), or by providing public coverage to certain subgroups of the population (pp. 120–21). The National Leadership Commission on Health Care has proposed and the state of Hawaii has implemented two versions of this approach (pp. 122–23).

The mandatory employer-based proposal would effectively cover the uninsured but does little to increase competition or curb current expenditures. Moreover, the plan mandates government intervention into the private sector, requires a large funding expenditure, and may cause a reduction in other benefits for those currently employed by non-insurance-subscribing companies that would have to begin expending funds on coverage.

The third general approach is a comprehensive restructuring of the health care system. One option is socialized medicine. It seems to offer both universal coverage and a large reduction in health care expenditures. Canada provides such government-sponsored insurance, yet cost-containment measures have apparently caused a reduction in patient services (p. 130).[9] Additionally, taxes would need to be increased significantly and a large number of private jobs affiliated with the current system would be eliminated. An alternative reform option does not require full public funding. A draft plan published by the Department of Health of the State of New York would create a single-payer authority that would collect payments from insurers

[9] However, opinion polls indicate that Canadians feel their system is better than the United States' and Americans feel the Canadian plan "compares favorably with their own" (p. 130).

to pay providers. The proposal establishes a cap for private liability and gives the state authority to establish fee limits for medical services (pp. 131–32). A similar plan by Alain Enthoven and Richard Kronick in the *New England Journal of Medicine*[10] would create "sponsors" to serve as go-betweens for employers, individuals, and providers (p. 133). These plans have the same basic faults—government expenditures would increase and it would be difficult to establish reasonable fees and budgets for health care procedures.

Aaron presents his own mandatory employment-based plan, providing complete acute and long-term health care with provisions to curb health care spending. His universal-coverage, single-payer plan is founded on four assumptions. Aaron assumes that: (1) Congress is unwilling to restructure the entire insurance system by shifting all private insurance to public plans; (2) the current system will never provide coverage for unemployed persons; (3) "control over health care spending must be concentrated in one or few hands" through overall budget restraints instead of unit service restrictions; and (4) federal action will be needed to provide a foundation for the coverage (pp. 140–41).

Aaron's proposal incorporates six components:

1. Guaranteed employment-based coverage would be afforded to all employees working twenty-five hours per week and to all dependents. Employers must pay approximately seventy-five percent of the insurance premium or a tax of approximately nine percent of the total business payrolls (pp. 141–42).[11]

2. Public coverage would be extended to families without an employed member and to workers of companies that opt to pay the nine percent tax instead of offering health insurance (p. 142).

3. Cost control would be administered through regional "financial agents," which would function as quasi-independent state regulatory agencies. The agents would be the single payer for acute care and serve as the administrator and budgeter for hospital spending. The federal government would determine the limits on spending, and the financial agents would determine how to distribute that amount. The financial agents would also

[10] Alain C. Enthoven & Richard Kronick, *A Consumer-Choice Health Plan for the 1990's: Universal Health Insurance in a System Designed to Promote Quality and Economy*, 320 NEW ENG. J. MED. 29–37, 94–101 (1989).

[11] These provisions are similar to those offered in the National Leadership Committee on Health Care proposal (p. 122).

be responsible for transforming the experience-rated insurance policies to community rates (pp. 142–43).[12]

4. Financial agents would be responsible for "high-cost episodes of care" (p. 143). This change should allow private insurers to virtually pay only for routine care and should therefore insulate them from large losses.

5. Long-term health care would be administered by financial agents through one of three options. The first option encourages private purchase of insurance through tax concessions and subsidies. The second implements a far-reaching national social insurance program. The third option would provide grants to nursing homes and home-care businesses (p. 144).

6. Aaron's comprehensive plan would be financed through Medicare and Medicaid funds, the employer tax, and an increase in public taxes (p. 147).

Aaron provides an in-depth critical analysis of his own proposal. One potential problem he discusses is that required sponsoring of health care will force businesses to incur a large increase in expenditures. These businesses may try to offset that cost by reducing employees' wages and other benefits or by eliminating low-wage staff positions (pp. 149–50). This loss of jobs would undermine any public benefit this plan seeks to deliver. Aaron accurately points out that businesses may opt instead to pay the tax. However, nine percent of the total payroll is a significant amount of money for a small business to suddenly be required to pay.

A second concern is whether effective administration by financial agents is feasible. The financial agents would require a new bureaucracy which would take time to establish. This bureaucracy must issue numerous complex regulations and policy decisions, such as rules governing how dependents' insurance is to be handled when a household has two working parents (pp. 149–50), or determining which medical procedures provide benefits too small to justify their costs. These inquiries will meet with "difficulty [in] finding analytically and politically defensible policies" (p. 150). Finally, as with all single-payer plans, rationing of health care may inhibit new innovations and technologies,

[12] The idea of "financial agents" bears a strong similarity to the proposed Enthoven-Kronick plan (pp. 132–33). However, these plans differ significantly on whether the goal should be to eliminate experience-rated policies which consider risk characteristics. The Enthoven-Kronick plan supports these characteristic distinctions (p. 133).

and create an unnecessary backlog of patients waiting for non-emergency care.

A third problem with Aaron's proposal is that it gives three options for long-term care, but it fails to specify or suggest the preferable policy. Aaron suggests that the best option would be selected by the "national decision about the nature of the public role in supporting long-term care" (p. 144). It is unclear how the financial agents will evaluate the "national decision."

Lastly, this plan will undercut the competitive nature of the private insurance system in several ways. First, the nine percent payroll option essentially caps the amount of insurance cost-effective businesses will purchase. Second, insurance companies will need to lower rates and cut plans to make the option of purchasing insurance attractive to medium-sized and small companies. Third, one may reasonably conclude that the average quality of health care will decrease to the minimum standards as demand for enhanced benefits from employers dwindles. Therefore, unless insurers can market enhanced packages to employers or allow for individual family upgrades, the insurance industry will collapse into the public program.

Serious and Unstable Condition: Financing America's Health Care presents a concise overview of economic health care issues, and Aaron's clear outline of both the problems and the possible solutions should be welcomed in this complex national debate. Aaron presents the information "less to argue an approach to the reform of health care finance than to fuel the debate on how to change" the financing system (p. 7). He does not claim to have a perfect solution, but rather presents the advantages and disadvantages of each plan, allowing each reader to personally determine which theory of reform seems most likely to produce the desired results.

There is no perfect solution to the health care dilemma. No matter what policy or combination of policies is implemented, the American public, the government, and the insurance industry will have to take medicine that is hard to swallow. Bold measures are needed to remedy the problem and Aaron's presentation and analysis should promote interest in the debate and provide an understanding of the current health care crisis. The debate now lies only in determining the strength and content of the medicine.

—*Josephine Aiello*

DIVIDED WE GOVERN: PARTY CONTROL, LAWMAKING, AND INVESTIGATIONS 1946–1990. By *David R. Mayhew*. New Haven, Conn.: Yale University Press, 1991. Pp. 199, appendices, index. $25.00 cloth.

Since its inception, it has been assumed that the United States government would function most effectively when the Legislative and Executive Branches were controlled by the same party. "Divided government" exists when one or both of the houses of Congress are dominated by the party opposite the President. This divided government is perceived to frustrate communication and interaction between the two branches, making it difficult to accomplish meaningful work, and eventually reducing the governmental process to an unproductive atmosphere of chaos. However, divided government, long considered a rare phenomenon, has become somewhat of a norm in recent years, causing many academic and political observers to examine the long-term aspects of divided government.

In *Divided We Govern*, David R. Mayhew[1] challenges these common beliefs about the evils of divided government. After analyzing the activities of every President and Congress from 1946 to 1990, Mayhew contends that in terms of the number of major laws enacted and congressional committee investigations of the Executive Branch conducted, periods of divided government have not differed significantly from times when the same party controlled the Legislative and Executive Branches.

Mayhew collected data relating to congressional investigations using a series of limiting criteria, to include only the most highly publicized and controversial. While many of these investigations have had significant political and historical implications, this is not universally true. Each one, however, concerns an issue that was significant at the time and attracted at least a reasonable share of national attention.

Mayhew included investigations that were reported on the front page of the *New York Times* for at least twenty days (not necessarily consecutively) during any Congress between 1946 and 1990 (p. 9). Relevant investigations began as allegations made by congressional committees, segments or individual members of a committee, or outsiders who made charges to a committee member who then made them public (p. 9). The target

[1] David R. Mayhew is Alfred Cowles Professor of Government at Yale University.

of these allegations was a "unit or a past or present official(s) or employee(s) of the executive branch" (p. 10). Reports featuring a variety of executive responses to such allegations, such as speeches or news conferences, were also considered part of the history of an allegation when determining its longevity (p. 10).

Using these criteria, Mayhew located thirty-one investigations (p. 11).[2] Fifteen of these charges were at least partly concerned with alleged corruption, conflict of interest, or favoritism (p. 12). Under Richard Nixon, for example, there were alleged suspicious dealings between the Justice Department and the International Telephone and Telegraph Corporation ("ITT") (p. 12).[3] Under Ronald Reagan there were charges of conflict of interest and mismanagement in the Environmental Protection Agency ("EPA") (p. 12), and in 1989 President George Bush faced allegations of corruption under Reagan's Department of Housing and Urban Development ("HUD") (pp. 12, 26). Six probes dealt with issues of loyalty or security, including the inquiry in 1950 into Senator McCarthy's charges of disloyalty in the State Department (p. 26).[4] Eight investigations dealt with military, foreign policy, or intelligence matters. This category includes post-Watergate inquiries into covert operations[5] and the Iran-Contra hearings in 1987 (pp. 26–27).[6] Finally, the Watergate inquiries themselves, conducted in 1973–74 by a special Senate committee and the House Judiciary Committee, were placed in a separate category.[7]

One striking characteristic in the data is the relative importance accorded by the press to the different investigations. For example, McCarthy's probe of the State Department in 1953–54 appeared for 203 days, the Korean War hearings were pub-

---

[2] The general topics of these investigations are discussed on p. 12 and pp. 26–27, with more detailed analysis of the results continuing on pp. 27–33. Table 2.1, which lists and provides information about all of the relevant investigations, can be found on pp. 13–25.

[3] *See* ANTHONY SAMPSON, THE SOVEREIGN STATE OF ITT 217–58 (1973).

[4] *See* ROBERT GRIFFITH, THE POLITICS OF FEAR: JOSEPH R. MCCARTHY AND THE SENATE 65–114 (1970).

[5] *See* LOCH K. JOHNSON, A SEASON OF INQUIRY: THE SENATE INTELLIGENCE INVESTIGATION (1985).

[6] *See* Frederick M. Kaiser, Causes and Conditions of Inter-Branch Conflict: Lessons from the Iran-Contra Affair (paper delivered at the 1989 Annual Meeting of the American Political Science Association).

[7] *See* JAMES HAMILTON, THE POWER TO PROBE: A STUDY OF CONGRESSIONAL INVESTIGATIONS (1977); J. ANTHONY LUKAS, NIGHTMARE: THE UNDERSIDE OF THE NIXON YEARS 469–569 (1976).

licized for fifty-three days, and a couple of matters such as the use of President Carter's brother as an intermediary to the Libyan government were publicized for only twenty-one days, barely enough to meet Mayhew's threshold for inclusion (p. 27).

It is also worthy of notice how sharply the number of investigations has declined. Sixteen of the thirty-one probes took place under Truman and Eisenhower, while only five took place under Carter, Reagan, and Bush (p. 28). There are several possible reasons for this. With the development of investigative journalism, sources other than congressional committees have assumed greater importance. Journalists have gained greater access to documents, including analytical sources. This, along with the decreasing use of congressional committees as a source, has resulted in a trend toward fewer, albeit more thorough reports on scandals, rather than a series of shorter tracking reports over time (pp. 28–29). Mayhew also notes that the *New York Times* decreased its front page from eight news columns to six. Since some stories had to be cut, it may be that some of the investigative stories that might otherwise have been on the front page would no longer fit there (p. 29 n.20). Finally, Mayhew suggests that the Executive Branch has improved its ability to "[fend] off congressional inquiries" by conducting its own investigations. Examples of this include the response to the Challenger disaster and the first Iran-Contra revelations (p. 29).

Although there were markedly fewer investigations in the later years of Mayhew's analysis, it does not hurt the integrity of his inquiry because there were periods of unified and divided government both early and late in the span, so there is plenty of room for analysis.

Mayhew finds no greater likelihood of congressional investigation of the Executive Branch in periods of divided government than when there is unified government. Focusing either across all of Mayhew's data or on individual administrations, the only variation that correlates with the existence of divided government occurs in House committees, which held eight of eleven investigations (seventy-three percent) in times of divided government (which existed only fifty-nine percent of the time) (p. 32, tbl. 2.2). However, these numbers are too small to be truly significant. Even looking at the most important investigations, the balance between divided and unified administrations comes out relatively equal. This is not to say that party does not play a role in the initiation of these investigations or in their

conduct, but these influences are the same regardless of which party controls which branches, so that in the end, there is virtual equality (pp. 31–33).

Next, Mayhew inventories the major laws enacted between 1946 and 1990. His goal is to analyze large amounts of legislation quickly and in such a summary form that a "yes-no" answer indicates whether a given law has been considered important. Using a two-step analysis, Mayhew distills the most relevant and important legislation of each congressional session. He then compares legislation enacted by divided government to key laws enacted by unified government. His first analysis uses the end of session reports of the *New York Times* and the *Washington Post*, which highlight the major legislative enactments (pp. 37–41).

The second analysis starts with the results of the first. Mayhew asks policy experts which laws they believe have been important in specific policy areas. The goal is to validate the enactments uncovered in the first analysis. The result is that of the 211 enactments found, 147 are validated and 56 new ones are added (pp. 44–50).

Mayhew lists the 203 laws under analysis in an expansive table (pp. 52–73).[8] His methodology and inquiries here are similar to those used earlier with congressional investigations. However, since there is so much more data involved in this segment of his analysis, it is more complex, and therefore more difficult to summarize. He notes that as with investigations, some administrations provided little numerical data for his analysis. For example, all of the contributions of a given time period may be in one area, such as during times of war. These instances were found to be equally instructive, however, because they reflected other phenomena in society at that time. For example, one of the lower producers was the Reagan administration during 1983–84 (p. 75). It should be remembered that from 1981 to 1986, the two houses of Congress were controlled by different parties, so the situation was further complicated. There was also a tendency toward omnibus legislation, which was not among the types of measures considered in this analysis (p. 76).

Mayhew finds support for his contention that there will be little difference between periods of divided and unified govern-

[8] Referring to this section is helpful in understanding the remainder of this discussion. For a more extensive analysis, see pp. 74–99.

ment. He finds no glaring differences in the effect of the two governmental structures. In fact, Mayhew finds greater differences between different periods of unified government. In general, unified government does not do better in terms of the quantity or impact of laws passed (p. 79). Mayhew argues that the legislative patterns are functions of factors other than simply whether the government is unified or divided. Instead, a variety of factors affect the way the legislative history unfolds. Among the factors having the greatest impact are presidential leadership, public mood, and electoral climate. One of Mayhew's primary goals in discussing these ideas is to show that there is more to the legislative system than whether government is divided or unified. Therefore, according to Mayhew, divided government should not be seen as the single evil which condemns our system to operate inefficiently, but rather as one of a number of potential contributors, and perhaps not even one of the most significant.

In his concluding chapter, Mayhew uses a detailed analysis of congressional investigations and legislation to analyze five arguments that have been used to justify the assertion that divided government causes inefficiencies in our governance system and should be avoided whenever possible. The first idea challenged is that divided government yields "worse" laws. This view holds that since the enacting coalitions in times of divided government are not naturally united, they are more likely to write laws with vague goals and inefficient means. The electorate may also check partisan coalitions less than when the government is unified. However, Mayhew contends that worse laws are passed in periods of unified control, in a frenzy to pass more laws while the controlling party is still in power. Citing the Great Society as an example of this phenomenon, Mayhew admits that a great deal was accomplished, but argues that ultimately the Johnson administration reduced its effectiveness by rushing to pass laws before clearly understanding the problems the laws were designed to solve (pp. 180–81).

Mayhew argues that unified government produces statutes which share greater ideological and budgetary coherence. Ideological coherence will be better in a unified government because the ideology will take into account the public mood. He also discusses budgetary coherence, but asserts that there is insufficient research in the area to support a conclusion (pp. 184–85).

Mayhew then challenges the belief that the country suffers administratively from divided government. He claims that the nation's problems in enacting and implementing the most effective policies arise out of congressional attempts to "micro-manage" the Executive Branch through the expansive use of monitoring staff, oversight hearings, and the legislative veto (p. 191). While divided government may be conducive to this micro-management, it is not the only factor causing it. In addition, Mayhew argues that micro-management is an insufficient basis to condemn the entire idea of divided government (pp. 191–93).

Some also argue that foreign policy suffers in divided government. Mayhew responds with examples that demonstrate a lack of a pattern in foreign policy regardless of divided government (pp. 195–96).

The final argument contested is that divided government hurts the poor. Mayhew contends that this is false because the safety net of social programs in the Nixon and Ford administrations was developed in a time of divided government. Many of these same programs were attacked by the Reagan administration during divided government. Thus, the moods of the two time periods clashed, but parties did not.

Mayhew's thesis, method of analysis, and conclusion are interesting but by no means definitive. If he had set different parameters in collecting his data, the results could have come out quite differently. Also the scope of the study is so broad that he is able to explain away facts that do not support his thesis. The method of data selection employed, particularly in the section on important legislation, is designed to produce a vast amount of data. As a result, the analysis required is extremely complex. Anyone seeking a full understanding of Mayhew's results would find a firm background in statistics helpful.

Mayhew admits that there are several disadvantages to his methodology. For example, his formula for analyzing important legislation is particularly vulnerable because it looks only to major enactments and ignores the impact of smaller or incremental legislation. It works better in policy areas, such as tax or labor, where a few major enactments make a big difference, and is less efficient in other areas, such as foreign relations. Mayhew's analysis also includes many acts which looked important at the time of enactment, but later failed to have a significant impact, while missing others which had greater im-

pact than expected. The analysis therefore emphasizes highly public and controversial provisions.

Mayhew assigns importance across Congresses and assumes that each Congress produced some important legislation. His approach does not take into account omnibus bills, which eliminated large bills in single policy areas to a certain extent. It focuses excessively on the President in the period prior to 1970. Finally, while the methodology is generally free of ideological bias, liberal initiatives dominate the lists of enactments. (pp. 41–44).

The best feature of this book, particularly in the more statistically dense sections, is the use of historical examples to illustrate his ideas and conclusions. Most of the examples are major events with which Mayhew's audience will likely be familiar. This breaks up the complexity of the analysis, makes the text more readable, and provides an entry point into the intimidating charts, which are also event-based.

Mayhew's thesis that divided government does not have a significant effect, and his ability to find data and draw conclusions to support this thesis, presents important questions for the future of the party system in American government. These issues are not substantially addressed within the scope of the book, which is primarily devoted to describing Mayhew's study, how he conducted it, and what he found. Looking ahead, however, the implications of divided and unified government will be particularly relevant over the next four years, since for the first time since the Carter administration, we will not have a divided government. Some governmental leaders and political analysts expect this change to be helpful to the country. President Clinton will undoubtedly be hoping that these commentators are right and Mayhew is wrong.

—Colleen Brennan

FORBIDDEN GROUNDS: THE CASE AGAINST DISCRIMINATION LAWS. By *Richard A. Epstein*. Cambridge, Mass.: Harvard University Press, 1992. Pp. 530, appendix, table of cases, indices. $39.95 cloth.

The debate over how to best provide equal opportunity for all Americans has pervaded almost all aspects of our current political and legal culture. Its persistence in fomenting a divisive debate will be the inevitable result of Richard Epstein's[1] most recent work, *Forbidden Grounds: The Case Against Discrimination Laws*. This book adds a new and provocative element to the ongoing debate as it sweepingly condemns antidiscrimination laws and their application to the private sector. As a libertarian, Epstein is opposed to state intervention in private employment practices because he believes they are not only costly and inefficient but also contrary to individual liberty and the societal welfare. It is this belief which compels Epstein in *Forbidden Grounds* to boldly call for the repeal of the discrimination laws as applied to private employment contracts.

Forbidden Grounds wages what the author calls a "frontal intellectual assault" against the current political consensus which advocates legal proscriptions against private discrimination (p. 6). The challenging question the book poses to civil rights proponents is whether the antidiscrimination principles and their current sanction in law results in net burdens or net benefits for the parties directly affected and society as a whole. Epstein concludes that an unregulated competitive market is the best way to preserve individual liberties and maximize market efficiency.

The basic premise of *Forbidden Grounds* is that freedom of contract principles at work in a competitive market would minimize costs and maximize efficiency, thereby ensuring net social benefits. Under this view, Title VII of the 1964 Civil Rights Act,[2] which, among other things, outlawed racial discrimination by private employers, adversely affects society due to the cost of compliance. This theme runs throughout *Forbidden Grounds* and is based on Epstein's conception of the employment anti-

[1] Richard A. Epstein is James Parker Hall Distinguished Service Professor of Law at the University of Chicago.

[2] Pub. L. No. 88-352, 78 Stat. 241, 253–66 (1964) (codified as amended at 42 U.S.C. § 2000e-1 to -17 (1988)).

discrimination law as the antithesis of freedom to contract (p. xii). For Epstein, the resulting costs to the private sector leave both employers and employees worse off than they would have been without the antidiscrimination law.

Epstein attacks each area of current law where the antidiscrimination principle has found legal sanction, most notably in discrimination laws governing race, sex, age, and disability. Epstein also offers his insights on affirmative action and its effects on the economy and society. In each of these areas, the main inquiry in the book is whether competitive markets operate better in creating benefits for all without the intrusion of antidiscrimination laws.

As a prelude to Epstein's critique, Part I of *Forbidden Grounds* examines the social theories of Hobbes and Locke and the ways in which they prescribe normative rules to govern human relations (pp. 15–27). The emphasis in this part of the book is on the theoretical foundations of law and its function in protecting the individual's labor and property. More importantly, Part I explains why Epstein adopts a libertarian view of the state's function. He believes that the state's purpose is to control the use of force and fraud against the person and the property of others (p. 19).

The author believes the reason for rules of ownership, transfer, and the protection of property is to ensure the individual's right to the free use of his labor and property, free from coercion and fraud (p. 25). According to this view, the state should intervene in human relations only when it is necessary to prevent fraud or coercion in either contractual relations or the use of property (p. 25). Therefore, in Epstein's ideal society a privately ordered employment market would benefit both employers and employees since neither class would have to bear the costs associated with state intervention for antidiscrimination purposes.

Part II of *Forbidden Grounds* shifts from the theoretical foundations of freedom of contract and the role of the state to the legal history of racial discrimination in America. The book traces the historical role that the state played in racial discrimination, examining in particular the Jim Crow laws in the South. Not only were discriminatory rules and regulations enacted by the state during this era, most notably in public transportation and schools, but the excessive use of the police power enabled local governments in the South to trample the property and

private contract rights of its citizens (p. 115). Epstein chronicles the history of state-sponsored racism in the South in an attempt to demonstrate the need for a limited government restricted to those areas where its presence is essential, such as taxation and law enforcement (p. 94).

Within this framework, Epstein begins Part III of *Forbidden Grounds* by attacking the antidiscrimination principle and its application to private employment contracts. Epstein's analysis initially evaluates Title VII and its subsequent judicial extensions governing race in the private employment context. While the book does acknowledge that the Civil Rights Act of 1964 was a monumental breakthrough in American life, reversing all racial barriers imposed by law (p. 252), it also presents and analyzes the purported adverse consequences of implementing Title VII.

Forbidden Grounds lays blame for these adverse consequences on the development of the two legal standards used to establish Title VII liability. These judicial extensions of Title VII are known as the disparate treatment and disparate impact standards.

In a disparate treatment suit, the plaintiff attempts to prove that an employer had a discriminatory intent when it made an employment decision adversely affecting that plaintiff. The standard, first applied by the Supreme Court in *McDonnell Douglas Corp. v. Green*,[3] requires that, after the plaintiff has met an initial burden of proof, the defendant must articulate a legitimate, nondiscriminatory reason for its employment practices in order to avoid liability.

The disparate treatment standard is characterized by Epstein as an inefficient extension of Title VII. Epstein is critical of this standard since it only identifies a single motive as the basis for an employer's decision to hire or fire (p. 170). Epstein believes that an employer may have a dual motive when making employment decisions and establishing the precise relative magnitude that discrimination played in the employer's decision is too difficult and too costly to litigate (p. 174).

Epstein is even more opposed to the use of the disparate impact doctrine in employment discrimination suits. The doctrine allows courts to infer discrimination in the absence of any discriminatory intent, from the effect that the employer's use of

[3] 411 U.S. 792 (1973).

tests or other hiring procedures has on hiring decisions (p. 160). In a disparate impact suit, an employer must establish that an employment practice which adversely affects a protected class under Title VII is guided by business necessity. Under this standard, an employment practice that is unrelated to job performance is illegal if the employer fails to demonstrate that it is based on business necessity.

Epstein first points out that the disparate impact doctrine has no support in either statutory contruction or the legislative history of the Civil Rights Act (p. 186). Epstein then proceeds to criticize the Supreme Court decision in *Griggs v. Duke Power Co.*,[4] which established the disparate impact doctrine for race discrimination cases.

Griggs dealt with employee testing, a practice Epstein sees as an efficient and reliable indicator of worker aptitude and potential job performance (p. 214). A job test that has a disparate impact on the rates of hiring for whites and blacks places a burden on the employer to establish that the test relates to business necessity. According to Epstein, the disparate impact doctrine is even worse than the disparate treatment doctrine because disparate impact is a virtually universal consequence of many employment practices, thereby broadening the scope of potential liability well beyond the levels established under the disparate treatment doctrine (p. 200).

Forbidden Grounds also questions whether Title VII has benefited those workers it was intended to protect. For instance, Epstein disagrees with the view that an increase in the relative wage level for blacks may be attributed to the Civil Rights Act. He suggests that aggressive affirmative action programs operating during the late 1960s and early 1970s might be the reason for the income gains blacks have made (p. 256). Epstein notes that, despite the existence of Title VII, the current unemployment rate for blacks has remained at twice the level for whites. He emphasizes that black male participation in the labor market has fallen significantly since the introduction of Title VII (p. 258).

---

[4] 401 U.S. 424 (1971). The disparate impact doctrine announced in *Griggs* was modified by Wards Cove Packing Co. v. Atonio, 490 U.S. 642 (1989) (holding that if an employer proves the employment practice at issue serves a legitimate employment goal, the plaintiff must then show the availability of an alternative practice with less of a disparate impact that the employer could have used to achieve the same goal).

From these statistics Epstein concludes that Title VII has done little, if anything, to satisfy its redistributive goals and objectives. Epstein claims that Title VII may have even made matters worse on the redistributive end. He seeks to illustrate this point by citing case studies which suggest that the current effect of Title VII is to make employers value highly educated, highly skilled blacks more than low-skilled blacks. In his view, therefore, educated blacks disproportionately reap the benefits of Title VII protection, thus redistributing resources from the worse-off to the better-off within this racial group (p. 263).

Part IV of *Forbidden Grounds* comprises Epstein's critique of Title VII as applied to sex discrimination. In this section, Epstein describes the trends that the disparate treatment and disparate impact doctrines have evinced in private employment markets when applied to gender discrimination.

Forbidden Grounds begins this discussion by first intimating that sex roles outside the workplace have a profound impact on what occurs within the workplace (p. 270). Epstein offers several reasons for opposing Title VII protection for women in the private employment context. Since the general prohibition against discrimination extends to all terms and conditions of the employment contract, an employer is legally required to provide a package of equal benefits, or at least a package of equal value, to each employee in the same work position, irrespective of gender (p. 314). Since the prohibition against discrimination extends to all terms and conditions of the employment contract, it might also cover pensions and insurance.

According to Epstein, Title VII protection for women imposes high costs on employers, particularly in light of the Supreme Court ruling in *Los Angeles Department of Water & Power v. Manhart*.[5] In *Manhart*, the Supreme Court held that it is illegal under Title VII to use the sex of an employee as a factor in determining contribution levels into a standard pension plan. Epstein criticizes the *Manhart* ruling for making the mistake of lumping together individuals in the workplace without any consideration of their value to the employer (p. 321). Epstein asserts that a sex-blind approach to all terms and conditions of employment may be a financial disaster, particularly with regard to pension and insurance plans.

---

[5] 435 U.S. 702 (1978).

In the area of pensions, Epstein believes that the *Manhart* decision will result in the common practice of hiring women solely to avoid liability, thereby reducing the overall expected pension of all the other workers at a firm. This would be due primarily to the fact that women remain in the workforce substantially longer than men (p. 325). Also, he suggests that the ruling in *Manhart* enables women to manipulate market conditions to their advantage by choosing benefit packages that will maximize their private returns. For these reasons, Epstein claims that the *Manhart* decision in conjunction with Title VII makes matters worse for all market participants.

In Part V of *Forbidden Grounds*, Epstein addresses the antidiscrimination principle in the context of affirmative action programs and policies. This section is a significant turning point in Epstein's analysis, as most affirmative action programs within the private market are voluntarily implemented. Given this fact, Epstein is prepared to defend affirmative action programs in the private sector since they are consistent with freedom-of-contract principles. Further, Epstein concedes that race-based preferences in hiring are tolerable and sometimes even beneficial, particularly when an employer has a legitimate reason to believe there is a positive correlation between job performance and race, ethnicity, or sex (p. 424).

While Epstein commends affirmative action policies within the private sector, he is opposed to the pursuit of affirmative action objectives in the public sector. One of the government-sponsored affirmative action initiatives he opposes is the minority set-aside program, since he believes that it creates a costly bureaucratic mechanism that is unresponsive to the price or quality of the product the government is purchasing (p. 436). According to Epstein, minority set-asides may bring about instances in which the government pays more for a product that could be purchased elsewhere at a cheaper price.

Part VI of *Forbidden Grounds*, the final section of the book, presents an unfavorable view of the latest areas in which the antidiscrimination principle has been applied, age and disability discrimination. The author calls for the repeal of the 1967 Age Discrimination in Employment Act ("ADEA")[6] and the 1990 Americans with Disabilities Act ("ADA").[7]

---

[6] 29 U.S.C. §§ 621–634 (1988).
[7] 42 U.S.C.A §§ 12101–12213 (West Supp. 1992).

The ADEA is modeled closely on the 1964 Civil Rights Act and incorporates both the disparate treatment and disparate impact standards. The major difference between the two acts is that a plaintiff in an age discrimination case may request a jury trial, whereas a Title VII case must be before tried before a court (p. 442).

Epstein comments that the ADEA is even more detrimental than the antidiscrimination statutes that govern race or sex. He believes that age discrimination, unlike discrimination based on race or sex, is a common practice at virtually all major American firms (p. 447). For Epstein, the ubiquity of age discrimination is proof that there is no market imperfection which justifies the extension of the antidiscrimination principle to age in private employment contracts.

Also, Epstein is certain the ADEA makes it more difficult for employers to use efficient and reliable tests and other standards to determine the future productivity of older members of the workforce. The employment tests are only legally permissible when they can be proven to relate to business necessity. The ADEA thus forces employers to abandon the use of a reliable technique to predict an individual's work potential, which in turn results in higher costs to employers (p. 452).

Epstein's final critique of the antidiscrimination principle is directed at the ADA. His overall impression of the ADA is that it imposes burdensome costs on employers as they seek to comply with its provisions. This act mandates that an employer may not discriminate against an individual with a disability if he is qualified to perform the essential functions of the job in question. The ADA also requires an employer to provide "reasonable accommodations" for employees with disabilities, unless the employer can prove that providing such accommodations would create an "undue hardship."[8]

Similar to the ADEA and the antidiscrimination statutes governing race and sex, the ADA makes it illegal to use an employment practice that discriminates against disabled persons unless the employer can demonstrate that the practice is job-related and constitutes a business necessity.

Epstein justifies his dissatisfaction with the ADA in several ways. For the most part, he believes that the "reasonable accomodations" standard set forth in the ADA will force employ-

[8] *Id.* § 12111(9)-(10).

ers to provide universal access to avoid the risk of liability. The costs of compliance will have two effects. On the one hand, the ADA will increase costs per individual worker. On the other hand, the ADA will also increase costs for those, including disabled persons, who are outside the labor market, and who receive none of the benefits of its enforcement (p. 491).

Epstein's analysis of the antidiscrimination principle and private employments contracts should be lauded for its thorough treatment of some of the losses American society must incur in its attempt to provide equal opportunity for its citizens. Epstein's analysis of the antidiscrimination principle elicits new questions that are critical to the debate over whether current legal trends will prove to be beneficial. Perhaps the true strength of *Forbidden Grounds* is its re-evaluation of the ways American society has opted to pursue its ideals and principles while also asking whether these ideals are worth pursuing, and if so, at what cost to the individual and to society.

Yet *Forbidden Grounds* does not convincingly answer these questions, mainly because many of Epstein's conclusions cannot be empirically supported. First and foremost, Epstein criticizes the antidiscrimination principle by using a cost-benefit analysis that only speculates as to what the real costs to society actually are. There is little information in the book that might help to gauge the extent of the burden. Epstein is careful to remind the reader of the costs of compliance and the costs of liability, but neither of these costs is quantified so that a meaningful cost-benefit analysis can be made. Perhaps the first step should be to critically evaluate the available empirical data so that the costs of antidiscrimination laws may be ascertained, rather than to accept Epstein's speculative thoughts as to what the costs of compliance and liability might be.

Also, *Forbidden Grounds* does not acknowledge the trends in employment discrimination suits since the passage of the 1964 Civil Rights Act. The nature of the litigation has, in fact, shifted dramatically. While most cases formerly attacked discrimination in hiring, today the vast majority of suits challenge discrimination in discharge.[9] By some estimates, the likelihood of suit when an employer fires a protected applicant is thirty times

---

[9] *See* John J. Donohue & Peter Siegelman, *The Changing Nature of Employment Discrimination Litigation*, 43 STAN. L. REV. 983, 984 (1991).

greater than the likelihood of suit if the employer simply fails to hire the worker.[10]

Forbidden Grounds makes no mention of this current trend. If employers are less likely to be sued on account of hiring practices, they might be less inclined to comply with the law during the early stages of their employment practices and procedures. Employers may be more likely to forgo the costs of compliance altogether when hiring. Thus the current trends in discrimination suits suggest that Epstein may overstate not just the hiring costs he associates with compliance but the searching costs as well.

Another significant element that is missing from *Forbidden Grounds* is the difference between discrimination suits filed by private individuals and those filed by the government on behalf of citizens. The Equal Employment Opportunity Commission ("EEOC") brought only 4356 cases between 1972 (the year it was granted the power to bring suit) and 1989. This amounts to less than four percent of all employment discrimination litigation between the years 1969 and 1989.[11] In view of this data, the EEOC might be categorized as an inert government bureaucracy with little, if any, broad influence to compel employers to comply with antidiscrimination statutes.

Epstein should have offered a more direct assessment of the role, or lack thereof, the EEOC plays in regulating private employment. *Forbidden Grounds* should not claim to weigh the net benefits or net burdens for the private market and society if it does not accurately describe the costs employers bear in complying with the antidiscrimination statutes, particularly when enforcement efforts are lax and infrequent.

Moreover, Epstein's broad assumption that employers bear significant costs when seeking to avoid liability under the antidiscrimination laws is presented without regard to the differences among employees and among employers. Private enforcement of the antidiscrimination principle will likely be higher among those who are not only cognizant of their rights under these laws, but also capable of detecting the subtleties of discrimination in employment. Those workers with higher levels of income and education are more likely to recognize violations

---

[10] *Id.* at 1027.
[11] *Id.* at 1000 n.66.

of the antidiscrimination laws, and to possess the legal sophistication and resources to initiate a discrimination suit.[12]

Epstein's analysis also fails to recognize that employers who generally hire workers with lower levels of income and education will probably have substantially less incentive to comply with antidiscrimination laws. *Forbidden Grounds* thereby overstates compliance costs to employers by ignoring the differences among employers and employees and the degree to which the probability of private discrimination claims is dependent upon these differences.

Nevertheless, Epstein does raise questions that are critical to the ongoing debate over equal opportunity, and whether it is a goal worth pursuing by regulating private employment practices. The reassessment for which Epstein calls, albeit in critical terms, is an invaluable exercise. But more importantly, the shortcomings of *Forbidden Grounds* make clear the need to develop a more thorough and comprehensive empirical basis for answering the unresolved questions concerning the effects of antidiscrimination laws.

—Anthony Q. Fletcher

[12] *See generally* Richard E. Miller & Austin Sarat, *Grievances, Claims and Disputes: Assessing the Adversary Culture*, 15 LAW & SOC'Y REV. 525 (1980)(concluding that differences between income and education have a profound effect on the probability of an individual bringing a discrimination claim).