TABLE OF CONTENTS

ARTICLES

NOTE

ARTICLES

QALYs and Policy Evaluation: A New Perspective

Matthew D. Adler, J.D.*

INTRODUCTION

The "quality-adjusted life year" (QALY) is a metric for health and longevity that is now widely used by health economists, public health scholars, and others researching the economics of health care.[1] QALYs work like this: Imagine a life

---

* Professor of Law, University of Pennsylvania. Many thanks to Craig Phillips for excellent research assistance, and to John Broome, Howard Chang, Robert Hahn, James Hammitt, Eric Posner, Chris Sanchirico, Amy Sinden, Cass Sunstein, and Reed Shuldiner for very helpful comments and conversations. This Article was written prior to the publication of JOHN BROOME, WEIGHING LIVES (2004), a rich and systematic philosophical treatment of the problem of measuring well-being and the impact of death on well-being. I have therefore, unfortunately, been unable to incorporate Broome's foundational discussion into my own, more applied, analysis.

1. Good overviews of the QALY approach to measuring health and longevity include OFFICE OF RES. & DEV., U.S. EPA, HUMAN HEALTH METRICS FOR ENVIRONMENTAL DECISION SUPPORT TOOLS: LESSONS FROM HEALTH ECONOMICS AND DECISION ANALYSIS (2001) [hereinafter U.S. EPA, HUMAN HEALTH METRICS]; Paul Dolan, *The Measurement of Health-Related Quality of Life for Use in Resource Allocation Decisions in Health Care*, *in* 1B HANDBOOK OF HEALTH ECONOMICS 1723 (A.J. Culyer & J.P. Newhouse eds., 2000); Robert Fabian, *The Qualy Approach*, *in* VALUING HEALTH FOR POLICY 118 (George Tolley et al. eds., 1994); James K. Hammitt, *QALYs Versus WTP*, 22 RISK ANALYSIS 985 (2002) [hereinafter Hammitt, *QALYs Versus WTP*]; Magnus Johannesson et al., *Outcome Measurement in Economic Evaluation*, 5 HEALTH ECON. 279 (1996); Robert M. Kaplan, *Utility Assessment for Estimating Quality-Adjusted Life Years*, *in* VALUING HEALTH CARE: COSTS, BENEFITS, AND EFFECTIVENESS OF PHARMACEUTICALS AND OTHER MEDICAL TECHNOLOGIES 31 (Frank A. Sloan ed., 1996); Graham Loomes & Lynda McKenzie, *The Use of QALYs in Health Care Decisionmaking*, 28 SOC. SCI. & MED. 299 (1989); and J. Brazier et al., *A Review of the Use of Health Status Measures in Economic Evaluation*, HEALTH TECH. ASSESSMENT, May 1999, at 1. A comprehensive guide is QUALITY OF LIFE AND PHARMACOECONOMICS IN CLINICAL TRIALS (Bert

history or "profile" of health states $h_1, h_2 \ldots h_n$, where each state h_j persists for t_j years. A health state can be death, perfect health, or any disease condition in-between: angina, bronchitis, lung cancer, depression, headaches, heart disease, and so on. Patients who have experienced the states, physicians familiar with the states, or members of the general population will have been surveyed and asked to rank each state h_j on a 0-1 scale of health quality, with 1 representing perfect health and 0 representing death. There are various techniques for eliciting the quality ranking, $q(h_j)$, the two most popular being the "time-tradeoff" method and the "standard-gamble" method. The first method seeks to determine the respondent's point of indifference between living y years with the condition h_j, and x years in perfect health (with x less than y), and assigns h_j the number x/y. The second method seeks to determine the respondent's point of indifference between living a given period of time with the condition, and a gamble with probability p of living in perfect health for that same period of time and probability $1-p$ of dying instantaneously. Health state h_j is then assigned the indifference probability, p, as its quality ranking.

The QALY number for a health profile is calculated as the sum of the quality-weighted years spent in each of its component health states. For example, if some person's life-history, absent a medical intervention, would consist of $h_1 \ldots h_n$, and after intervention would improve to $h_1^* \ldots h_n^*$, then the QALY measure of the first profile is $\sum_j q(h_j) \times t_j$, the QALY measure of the second is $\sum_j q(h_j^*) \times t_j^*$, and the QALY gain secured by the intervention is the difference between these two sums. Similarly, if the individual would live m years in perfect health absent intervention, and $m+x$ years in perfect health after intervention, the QALY gain is x QALYs. QALYs are evidently a powerful tool for measuring the impact of choices that affect morbidity, longevity, or both—not only the choices of physicians, hospitals, HMOs, and insurers, but also governmental choices, such as the FDA's pharmaceutical licensure decisions, the regulation of toxins by the EPA or OSHA, or HHS's choices about Medicare coverage.

To date, QALYs have been generally employed in cost-effectiveness studies.[2] In a cost-effectiveness study, both the health and non-health impacts of different health-affecting choices are determined. Non-health impacts are measured in dollars, but health impacts are measured using some nonmonetary scale (either a QALY scale, or some other scale, e.g., a disease-specific scale in the case where the health effects of the choices at issue are confined to a single disease). Cost-effectiveness ratios are then used to determine which choice should be undertaken. Alternatively, the choice which maximizes health given a

---

Spilker ed., 2d ed. 1996).

2. *See infra* text accompanying notes 25-30 (discussing cost-effectiveness analysis and use of QALYs in that context).

fixed budget for non-health costs is selected.

Though QALYS were invented in the early 1970s,[3] the use of QALYS—at least by researchers—has skyrocketed in recent years. Literally hundreds of health care cost-effectiveness studies now appear in academic journals every year.[4] An increasing fraction of these employ QALYs. One recent review found that only one or two such studies were published annually during the 1980s and less than ten annually during the early 1990s, but that since 1997 roughly fifty "cost-utility" studies have appeared each year.[5] This body of work has been accompanied by new surveys, including a massive survey of the general population performed a few years ago in England,[6] and a burgeoning secondary literature on QALYs.[7]

To be sure, QALYs remain largely an academic tool, at least in the United States. Fifteen years ago, Oregon infamously relied upon, then abandoned, QALYs in deciding what treatments it would cover under Medicaid.[8] Around the same time, HHS published, but never finalized, a proposal to use cost-effectiveness in Medicare coverage decisions.[9] More generally, "cost-

---

3. *See* Kaplan, *supra* note 1, at 35 (crediting the invention of QALYs to a 1970 article by S. Fanshel and J.W. Bush); George W. Torrance & David Feeny, *Utilities and Quality-Adjusted Life Years*, 5 INT'L J. TECH. ASSESSMENT IN HEALTH CARE 559, 568 (1989) (same).

4. *See* Anne Elixhauser et al., *Health Care CBA and CEA from 1991 to 1996: An Updated Bibliography*, 36 MEDICAL CARE MS1, MS6 (May Supp. 1998) (surveying health care literature and identifying 1792 published cost-effectiveness studies for the period 1991-96 and 1123 for the period 1979-90).

5. *See* Dan Greenberg & Joseph S. Pliskin, *Preference-Based Outcome Measures in Cost-Utility Analysis: A 20-Year Overview*, 18 INT'L J. TECH. ASSESSMENT 461, 463 (2002). Another study finds even greater numbers of published cost-utility studies. *See* Peter J. Neumann et al., *Growth and Quality of the Cost-Utility Literature, 1976-2001*, 8 VALUE IN HEALTH 3, 5 (2005).

6. *See* Paul Dolan et al., *The Time Trade-Off Method: Results from a General Population Study*, 5 HEALTH ECON. 141 (1996).

7. For bibliographies, see Dolan, *supra* note 1, at 1755-60; and U.S. EPA, HUMAN HEALTH METRICS, *supra* note 1, at 51. A useful list of references is provided at the end of each chapter in Brazier, *supra* note 1.

8. *See, e.g.*, Kaplan, *supra* note 1, at 53-59 (discussing the Oregon episode).

9. *See* Medicare Program; Criteria and Procedures for Making Medical Services Coverage Decisions that Relate to Health Care Technology, 54 Fed. Reg. 4302, 4308-09 (proposed Jan. 30, 1989); Medicare Program; Procedures for Making National Coverage Decisions, 64 Fed. Reg. 22,619, 22,620 (Apr. 27, 1999) (withdrawing proposed rule); *see also* Medicare Program; Criteria for Making Coverage Decisions, 65 Fed. Reg. 31,124, 31,127 (May 16, 2000) (notice of intent to publish proposed rule establishing criteria for Medicare coverage decisions, which suggests QALYs as a possible measure of health benefits); Medicare Program; Revised Process for Making Medicare National Coverage Determinations, 68 Fed. Reg. 55,634, 55,634-35 (Sept. 26, 2003) (withdrawing intent to pursue rulemaking).

effectiveness analysis has been used by [U.S.] public sector agencies on a very limited basis" in the health policy area,[10] by contrast with some foreign jurisdictions, such as Australia, Britain, Canada, and New Zealand, which now regularly evaluate the cost-effectiveness of pharmaceuticals or health technologies proposed for public reimbursement.[11] And private entities, such as hospitals, HMOs, or medical insurers, which in principle could incorporate QALYs into a wide range of their decisions, seldom do so at present.[12]

But there are signs of change. The FDA, over the last half-decade or so, has repeatedly relied on QALYs in its rulemakings, pioneering a new approach to QALY-based analysis that I will describe and defend below.[13] The EPA, in its

---

10. Frank A. Sloan & Henry G. Grabowski, *Introduction and Overview*, 45 SOC. SCI. & MED. 505, 508 (1997); *see also* Eric A. Posner, *Transfer Regulations and Cost-Effectiveness Analysis*, 53 DUKE L.J. 1067, 1105-06 (2003) (defending use of cost-effectiveness analysis to assess transfer regulations and stating that "[agencies] have rarely performed cost-effectiveness analysis, and their occasional efforts have been inadequate").

11. *See* Nancy Devlin & David Parkin, *Does NICE Have a Cost-Effectiveness Threshold and What Other Factors Influence Its Decisions? A Binary Choice Analysis*, 13 HEALTH ECON. 437, 437 (2004); Paul P. Glasziou & Andrew S. Mitchell, *Use of Pharmacoeconomic Data by Regulatory Authorities*, *in* QUALITY OF LIFE AND PHARMACOECONOMICS, *supra* note 1, at 114.

12. *See* Bryan R. Luce & Ruth E. Brown, *The Use of Technology Assessment by Hospitals, Health Maintenance Organizations, and Third-Party Payers in the United States*, 11 INT'L J. TECH. ASSESSMENT IN HEALTH CARE 79, 85 (1995); Peter Neumann, *Why Don't Americans Use Cost-Effectiveness Analysis?*, 10 AM. J. MANAGED CARE 308, 308 (2004); Elaine J. Power & John Eisenberg, *Are We Ready To Use Cost-Effectiveness Analysis in Health Care Decision-Making? A Health Services Research Challenge for Clinicians, Patients, Health Care Systems, and Public Policy*, 36 MED. CARE MS10, MS11-12 (May Supp. 1998). *But see* Bernard S. Bloom, *Use of Formal Benefit/Cost Evaluations in Health System Decision Making*, 10 AM. J. MANAGED CARE 329, 332-33 (2004) (reporting results of a 2002 survey of public and private health care organizations, and finding that a majority of the private respondents employ cost-benefit or cost-effectiveness analysis). For general discussions of the use of QALYs by governmental agencies or private actors, see COST-EFFECTIVENESS IN HEALTH AND MEDICINE 18-20 (Marthe Gold et al. eds., 1996); U.S. CONGRESS OFFICE OF TECH. ASSESSMENT, IDENTIFYING HEALTH TECHNOLOGIES THAT WORK 122-30 (1994); Bloom, *supra*; Glasziou & Mitchell, *supra* note 11; Luce & Brown, *supra*; Power & Eisenberg, *supra*; Sloan & Grabowski, *supra* note 10; Neumann et al., *supra* note 5; and Frank A. Sloan & Christopher J. Conover, *The Use of Cost-Effectiveness/Cost-Benefit Analysis in Actual Decision Making: Current Status and Prospects*, *in* VALUING HEALTH CARE, *supra* note 1, at 207. Two important works that were published subsequent to the drafting of this Article, and that discuss the use of QALYs or cost-effectiveness analysis, are: PETER J. NEUMANN, USING COST-EFFECTIVENESS ANALYSIS TO IMPROVE HEALTH CARE: OPPORTUNITIES AND BARRIERS (2005); and Symposium, *Cost-Effectiveness Analysis in U.S. Healthcare Decision-Making: Where Is It Going?*, 43 MED. CARE II-1 (July Supp. 2005).

13. *See infra* text accompanying notes 192-201.

major arsenic and radon rulemakings during the late 1990s, specifically declined to employ a QALY-based monetary valuation of the morbidity- and mortality-reduction benefits of the rules.[14] In other rulemakings around the same time, however, and more recently as well, the EPA has experimented with the so-called "value of statistical life year" (VSLY) approach to monetizing mortality, an approach closely related to QALYs.[15] The D.C. Circuit opinion in the *American Trucking* case, after striking down a provision of the Clean Air Act on nondelegation grounds, suggested that the EPA might cure those constitutional difficulties by measuring the benefits of air pollution regulations on a QALY scale.[16] The HHS, in a recent rulemaking facilitating flu and pneumonia vaccinations, cited the cost-effectiveness of these vaccinations in promoting "year[s] of healthy life"—a synonym for a QALY.[17] The Office of Management and Budget's (OMB) current guide for Executive Order 12,866, which requires formal regulatory analyses of major rules, stipulates that these documents must include a cost-effectiveness analysis in the case of rules targeted at public health and safety,[18] and it gives a qualified endorsement to QALYs as an appropriate effectiveness metric.[19] The Public Health Service's "Healthy People" initiative, an informational program which measures progress towards public health goals,

---

14. *See* National Primary Drinking Water Regulations; Arsenic and Clarifications to Compliance and New Source Contaminants Monitoring, 66 Fed. Reg. 6976, 7044 (Jan. 22, 2001); National Primary Drinking Water Regulations; Radon-222, 64 Fed. Reg. 59,246, 59,337 (Nov. 2, 1999).

15. *See* Control of Air Pollution from New Motor Vehicles: Heavy Duty Engine and Vehicle Standards and Highway Diesel Fuel Sulphur Control Requirements, 66 Fed. Reg. 5002, 5104 (Jan. 18, 2001); Control of Air Pollution from New Motor Vehicles: Tier 2 Motor Vehicle Emissions Standards and Gasoline Sulfur Control Requirements, 65 Fed. Reg. 6698, 6784-87 (Feb. 10, 2000); Findings of Significant Contribution and Rulemaking on Section 126 Petitions for Purposes of Reducing Interstate Ozone Transport, 65 Fed. Reg. 2674, 2721-22 (Jan. 18, 2000); U.S. EPA, TECHNICAL ADDENDUM: METHODOLOGIES FOR THE BENEFIT ANALYSIS OF THE CLEAR SKIES INITIATIVE 35-37 (2002) [hereinafter U.S. EPA, TECHNICAL ADDENDUM]; Laura J. Lowenstein & Richard J. Revesz, *Anti-Regulation Under the Guise of Rational Regulation: The Bush Administration's Approaches to Valuing Human Lives in Environmental Cost-Benefit Analyses*, 34 ENVTL. L. REP. 10,954 (2004) (describing the EPA's use of VSLYs).

16. *See* Am. Trucking Ass'ns v. U.S. EPA, 175 F.3d 1027, 1039-40 (D.C. Cir. 1999), *rev'd in part*, 531 U.S. 457 (2001) (finding that the Act was not an unconstitutional delegation).

17. Medicare and Medicaid Programs; Conditions of Participation: Immunization Standards for Hospitals, Long-Term Care Facilities, and Home Health Agencies, 67 Fed. Reg. 61,808, 61,813-14 (Oct. 2, 2002).

18. *See* OFFICE OF MGMT. & BUDGET, CIRCULAR A-4, at 9 (2003), *available at* http://www.whitehouse.gove/omb/circulars/a004/a-4.pdf.

19. *See id.* at 12-14.

employs QALYs as one of its key metrics.[20]

In short, large amounts of money, time, and brainpower are currently being expended on QALY research, and that research now promises (or threatens, if you prefer) to shape policy. Legal scholars and policy analysts outside the health field should therefore pay more attention than they have done to QALYs. Are QALYs more than a mere artifact of the resistance of physicians and public health officials to cost-benefit analysis? Is there a good reason, other than mere squeamishness about the monetary pricing of life and health that inheres in cost-benefit analysis, to incorporate QALYs in policy evaluation?

This Article addresses that question and answers it, affirmatively, in a novel way.[21] My focus is governmental rather than private sector decision-making. I shall adopt a welfarist approach to policy analysis but—building on my prior work with Eric Posner[22]—shall advance a conception of welfarism quite different from that held by many welfare economists, particularly the health economists who write about QALYs. Many welfare economists, and certainly many health economists, continue to structure their work around the criterion of Kaldor-Hicks efficiency[23]—a criterion that, Posner and I have argued, should be rejected. Instead, welfarists should care about overall well-being. Overall well-being and Kaldor-Hicks efficiency are distinct criteria.

Traditional cost-benefit analysis, understood as the sum of willingness-to-

---

20. *See* PENNIFER ERICKSON ET AL., U.S. DEP'T OF HEALTH & HUMAN SERVS., HEALTHY PEOPLE 2000 STATISTICAL NOTES NO. 7 (1995) (discussing "years of health life" measure), *available at* http://www.cdc.gov/nchs/data/statnt/statnt/statnt07.pdf; U.S. DEP'T OF HEALTH & HUMAN SERVS., HEALTHY PEOPLE 2010: UNDERSTANDING AND IMPROVING HEALTH 10 (2000) (same), *available at* http://www.healthypeople.gov/document/pdf/uih/2010uih.pdf.

21. Richard Pildes and Cass Sunstein have explored the role of QALYs in public policy, but their approach is different from mine. They see QALYs as an alternative to monetization, while I conceptualize QALYs as the input to a nontraditional, but monetized, cost-benefit analysis. *See* Richard H. Pildes & Cass R. Sunstein, *Reinventing the Regulatory State*, 62 U. CHI. L. REV. 1, 83-85 (1995).

22. *See* MATTHEW D. ADLER & ERIC A. POSNER, NEW FOUNDATIONS OF COST-BENEFIT ANALYSIS (forthcoming 2006) [hereinafter ADLER & POSNER, COST-BENEFIT ANALYSIS]; Matthew D. Adler & Eric A. Posner, *Implementing Cost-Benefit Analysis when Preferences are Distorted*, *in* COST-BENEFIT ANALYSIS: LEGAL, ECONOMIC, AND PHILOSOPHICAL PERSPECTIVES 269 (Matthew D. Adler & Eric A. Posner eds., 2000) [hereinafter Adler & Posner, *Implementing Cost-Benefit Analysis*]; Matthew D. Adler & Eric A. Posner, *Rethinking Cost-Benefit Analysis*, 109 YALE L.J. 165 (1999) [hereinafter Adler & Posner, *Rethinking Cost-Benefit Analysis*].

23. A policy is Kaldor-Hicks efficient, relative to some alternative, if there is a hypothetical costless redistribution of wealth, from those whose welfare is increased by the policy to those whose welfare is reduced, which would make the policy Pareto-efficient, i.e., no one would be worse off and some would be better off.

pay/accept (WTP/WTA) amounts, is often seen by applied economists and policy analysts as a way to implement Kaldor-Hicks efficiency. Whatever the connection here,[24] traditional cost-benefit analysis is at best an imperfect proxy for overall well-being. Under certain conditions, QALY measures can better track overall welfare than do WTP/WTA measures. In addition, the cognitive difficulties that ordinary individuals have in processing probabilities, and in trading off life and health for money, mean that the WTP/WTA amounts that individuals express in contingent-valuation studies or reveal through their behaviors may deviate substantially from the true WTP/WTA amounts—the amounts that genuinely track the individuals' welfare. QALY interviews and QALY aggregation can partly circumvent these cognitive failures.

So QALYs do have a place in welfarist policy evaluation, but not as the outcome measure in cost-effectiveness analysis. Rather, welfarist policy evaluation properly employs a nontraditional kind of cost-benefit analysis—one that monetizes certain impacts, such as health impacts, using valuation constructs other than WTP/WTA. Cost-benefit analysis, in the nontraditional form that I describe and defend in this Article, remains a technique that monetizes the welfare impact of governmental choices on affected individuals, then aggregates those monetary sums. But the use of QALY-to-dollar conversions, rather than WTP/WTA amounts derived from contingent valuation or revealed preference studies, is sometimes the best welfarist approach to monetizing welfare impacts.

Part I of the Article surveys existing scholarship about the role of QALYs in policy evaluation, split between extrawelfarists and welfarists. Extrawelfarists see QALYs as a measure of "healthiness" rather than well-being, while welfarists either support QALY-based cost-effectiveness analysis for the pragmatic reason that many in the health care field are averse to cost-benefit analysis, or reject QALYs altogether.

Part II and III present a novel defense of QALYs. Part II outlines the construct of overall well-being, understood (as per John Harsanyi) as the sum of utility numbers assigned by impartial spectators to individual life histories. I shall call these utility numbers "lifetime welfare units" (LWUs). Part III argues that QALYs are sometimes better proxies for LWUs than WTP/WTA amounts, for both conceptual and cognitive reasons.

---

24. Traditional cost-benefit analysis is not a perfect measure of Kaldor-Hicks efficiency given the Boadway paradox. *See* ROBIN BOADWAY & NEIL BRUCE, WELFARE ECONOMICS 263-71 (1984). Since I deny the moral significance of Kaldor-Hicks efficiency and see traditional cost-benefit analysis as a decision procedure that implements overall well-being rather than Kaldor-Hicks efficiency, it is not important for my purposes to determine whether the Boadway paradox, in practice, creates a large or only a small gap between cost-benefit analysis and Kaldor-Hicks efficiency.

But QALYs are hardly a perfect welfarist measuring rod. The LWU value of a life-history is a product of both its health and non-health attributes. QALYs capture only the first. Further, the overall health utility of a life-history may not be perfectly decomposable into the sum of the quality-adjusted durations of its component health states. These difficulties with QALYs are described in Part IV.

Part V describes the role that QALYs should play in cost-benefit analysis, given both the strengths and limitations of QALYs. It argues that QALY measures of health and longevity impacts should be converted into dollar amounts using a conversion factor, such as $100,000 per QALY. Part V discusses how to set an optimal conversion factor. And it specifies the conditions under which QALY-to-dollar conversions should be used as inputs to cost-benefit analysis in lieu of WTP/WTA amounts.

Part V also provides a concrete illustration of my approach, on a topic of much contemporary interest: the debate between those who argue that agencies should use a monetary value of life that is insensitive to (or at least not proportional to) the life expectancy of the persons whose lives are saved or lost, and those who contend that agencies should price life-years rather than lives. I argue for the pricing of life-years, at least under some circumstances. Many economists have criticized that approach, because it is at odds with traditional cost-benefit analysis. But once cost-benefit analysis is understood as a pragmatic technique that need not always rely on WTP/WTA measures, the life-years approach may be vindicated.

I. QALYS: THE CURRENT VIEW

This Part briefly reviews the current scholarly understanding of QALYs. Readers acquainted with this scholarship may want to proceed directly to Part II, but others are likely to find this review helpful as a point of departure. My own, novel approach to QALYs, presented in Parts II through V below, is best understood by contrast with the current literature.

Existing scholarship conceptualizes QALYs as an effectiveness metric for the purposes of a health policy cost-effectiveness analysis.[25] This is true both of scholars who adopt an "extrawelfarist" approach to health policy and of those who adopt a "welfarist" approach—a distinction to which I will return shortly. The literature on QALYs is characterized by various debates and divisions, but concurs in seeing QALY measures as a component of cost-effectiveness analysis.

Cost-effectiveness analysis of health policy is best described as a set of

25. On health care cost-effectiveness analysis generally, and cost-utility analysis specifically, see COST-EFFECTIVENESS IN HEALTH AND MEDICINE, *supra* note 12; and MICHAEL F. DRUMMOND ET AL., METHODS FOR THE ECONOMIC EVALUATION OF HEALTH CARE PROGRAMMES 96-204 (2d ed. 1997).

decision-analytic techniques that measure non-health effects in dollars but measure health and longevity on some nonmonetary scale, such as QALYs. Cost-effectiveness analysis has two main variants. The first variant assumes that non-health costs should be incurred up to some fixed budget, which is given exogenously, and picks the policy choice that maximizes health benefits within the budget. The second variant is not constrained by, nor intent on expending, an exogenous budget. Instead (roughly speaking) it excludes "dominated" choices (those that both produce smaller total health benefits than some alternative and are more expensive than that alternative); arrays the remaining choices in the order of their total health benefits, smallest to largest; and selects the choice furthest down the list whose incremental cost-effectiveness ratio does not exceed some cut-off ratio, for example $50,000 per QALY.[26]

Appendix I provides an extended example of both variants of cost-effectiveness analysis. The first variant, maximizing health benefits for a fixed budget, is easy to understand. The second variant, which increases health production until the incremental cost-effectiveness ratio reaches some cut-off level, is a bit harder to grasp. One way to characterize this second variant of cost-effectiveness analysis is as follows: This technique reaches the same result that traditional cost-benefit analysis would *if* individual willingness-to-pay/accept (WTP/WTA) for a QALY were a constant value and the cut-off ratio were set to equal this constant.[27]

---

26. This description is rough because the technique actually uses two different "dominance" notions to exclude choices: ordinary dominance, as described above, and "weak" or "extended" dominance. Further, it is structured to accommodate the possibility that health policies can be pursued simultaneously rather than being mutually exclusive. *See* Magnus Johannesson, *The Relationship Between Cost-Effectiveness Analysis and Cost-Benefit Analysis*, 41 SOC. SCI. & MED. 483, 483-84 (1995); Magnus Johannesson & Milton C. Weinstein, *On the Decision Rules of Cost-Effectiveness Analysis*, 12 J. HEALTH ECON. 459, 460-62 (1993); Göran Karlsson & Magnus Johannesson, *The Decision Rules of Cost-Effectiveness Analysis*, 9 PHARMACOECONOMICS 113 (1996); *infra* App. I.

The second variant of cost-effectiveness analysis, unlike the first, relies on a cut-off value. There are various methods, suggested in the literature, to generate this cutoff value. For example, the analyst might consult tables that show the cost-per-QALY of a range of existing programs, and might choose a cutoff value in the middle of that range. Or, recognizing that willingness-to-pay per QALY is heterogeneous, the analyst might try to determine what individuals are (in some sense) on average willing to pay. For a discussion, see Hans-Georg Eichler et al., *Use of Cost-Effectiveness Analysis in Health-Care Resource Allocation Decision-Making: How Are Cost-Effectiveness Thresholds Expected To Emerge?*, 7 VALUE IN HEALTH 518 (2004); and Milton C. Weinstein, *From Cost-Effectiveness Ratios to Resource Allocation: Where To Draw the Line?*, *in* VALUING HEALTH CARE, *supra* note 1, at 91-96.

27. *See* Mohan V. Bala et al., *Conditions for the Near Equivalence of Cost-Effectiveness and Cost-Benefit Analyses*, 5 VALUE IN HEALTH 338, 339-40 (2002); Johannesson, *supra* note 26, at

In reality, however, both variants of cost-effectiveness analysis deviate from traditional cost-benefit analysis (the sum of individual WTP/WTA amounts). The first variant obviously does, because it takes for granted that the fixed budget should be expended on health. It selects the choice that maximizes health within the budget, over the choices of not expending all or any of the budget, even if the monetized benefits of the health-maximizing choice are less than the monetized costs.

As for the second variant, the premise that makes this coincide with traditional cost-benefit analysis, namely the existence of a constant WTP/WTA per QALY, is counterfactual. WTP for a given QALY gain is in fact heterogeneous.[28] Wealthier individuals will tend to pay more for a given QALY gain.[29] If health does not change the welfare effect ("marginal utility") of material consumption, healthier individuals will tend to pay less for a given QALY gain. Where health does change the marginal utility of consumption, healthier individuals may pay more for a given QALY gain.[30] Individuals for whom consumption is relatively more important, and health less important, will tend to pay less for a given QALY gain than their less materialistic counterparts. All these points are true both if WTP is keyed to preferences and even if WTP is keyed to a conception of well-being which is allowed to deviate from preferences. Since health and consumption do not make linear contributions to well-being at a single, constant rate—on any plausible account of well-being, preferentialist or not—WTP per QALY is heterogeneous. Thus the second variant of cost-effectiveness analysis can reach different results than traditional cost-benefit analysis. Appendix I may be helpful here, as it provides a concrete example that shows how both variants of cost-effectiveness analysis can diverge from traditional cost-benefit analysis.

485-86.

28. *See* Paul Dolan & Richard Edlin, *Is It Really Possible To Build a Bridge Between Cost-Benefit Analysis and Cost-Effectiveness Analysis?*, 21 J. HEALTH ECON. 827, 837-38 (2002); Alan M. Garber & Charles E. Phelps, *Economic Foundations of Cost-Effectiveness Analysis*, 16 J. HEALTH ECON. 1, 28-29 (1997); Johannesson, *supra* note 26, at 486-87; *see also* Bala et al., *supra* note 27, at 344-45 (finding that constant WTP per QALY is sufficient but not necessary for the equivalence of cost-effectiveness and cost-benefit analysis, and identifying weaker necessary conditions that "are unlikely to hold in practice."). For a recent empirical study confirming heterogeneity of WTP per QALY, see Duska M. Franic et al., *Quality-Adjusted Life Years Was a Poor Predictor of Women's Willingess To Pay in Acute and Chronic Conditions: Results of a Survey*, 58 J. CLINICAL EPIDEMIOLOGY 291, 301 (2005).

29. *See, e.g.*, William N. Evans & W. Kip Viscusi, *Income Effects and the Value of Health*, 28 J. HUM. RESOURCES 497, 498-99, 516 (1993).

30. *See id.* at 499-500 (noting that ill health can increase or decrease the marginal utility of income).

Given the deviation between QALY-based policy evaluation, in the form of cost-effectiveness analysis, and traditional cost-benefit analysis, why use QALYs? The literature provides two answers to this question, "extrawelfarist" and "welfarist."[31] The extrawelfarist view sees health as a measurable characteristic of each person, distinct from that person's well-being; conceptualizes QALYs as the measure of health; and argues that health programs should aim at some health-related goal (be it maximizing aggregate population QALYs, equalizing lifetime QALYs across the population, or something in between) rather than welfarist goals such as Kaldor-Hicks efficiency or overall well-being.

The best-known proponent of the extrawelfarist view is A.J. Culyer, who—drawing on scholarship by Amartya Sen—sees health as a "capability."[32] Sen's work, in turn, is part of a larger philosophical school, including such luminaries as John Rawls and Ronald Dworkin, which suggests that the "currency of justice" consists in something other than welfare.[33] Rawls famously argues that distributive justice concerns the distribution of "primary goods," rather than well-being. Dworkin argues that genuine equality between persons involves their starting from a position of equal "resources," not attaining equal welfare levels. Sen's notion of a "capability" is a variation on Rawlsian primary goods and Dworkinian "resources."

Extrawelfarists such as Culyer have a ready answer for why health policy analysts should continue to use cost-effectiveness analysis or some other form of QALY-based analysis rather than cost-benefit analysis: namely that health, rather than welfare, ought to be the underlying concern of health policy.[34] Cost-effectiveness analysis, as seen by the extrawelfarist, is a technique for maximizing the population's health. The fact that the health-maximizing choice need not be the same as the welfare-maximizing choice or the Kaldor-Hicks efficient choice is no surprise, and no demerit for this technique, since health and welfare are different.

---

31. *See* Dolan, *supra* note 1, at 1727-29.

32. *See, e.g.*, A.J. Culyer, *Commodities, Characteristics of Commodities, Characteristics of People, Utilities, and the Quality of Life*, *in* QUALITY OF LIFE: PERSPECTIVES AND POLICIES 9 (Sally Baldwin et al. eds., 1990) [hereinafter Culyer, *Commodities*]; A.J. Culyer, *The Normative Economics of Health Care Finance and Provision*, 5 OXFORD REV. ECON. POL'Y 34, 50-55 (1989). Sen has developed the capability view in many works, including AMARTYA SEN, *Equality of What?*, *in* CHOICE, WELFARE AND MEASUREMENT 353 (1982). For a recent symposium discussing the application of the capability approach to health policy, see Symposium, *Equity, Capabilities and Health*, 60 SOC. SCI. & MED. 219 (2005).

33. This extrawelfarist school is surveyed and criticized in Adler & Posner, *Rethinking Cost-Benefit Analysis*, *supra* note 22, at 212, 215-16.

34. *See, e.g.*, Bala et al., *supra* note 27, at 345; Dolan & Edlin, *supra* note 28, at 838.

Is the extrawelfarist defense of QALY-based analysis persuasive? The extrawelfarist is plausibly correct that health is a measurable characteristic of individual lives which is distinct from welfare.[35] For example, restoring a paraplegic to full mobility would dramatically improve her health, but might not dramatically improve her welfare if she is happy as a paraplegic and her lifestyle with full mobility would remain sedentary.[36]

Whether health rather than welfare should be the concern of health care policy and, more generally, whether the maximization or fair distribution of "capabilities" or "resources" or "primary goods" rather than welfare should be the focus of social planning and policy evaluation, are thorny questions that I will not address in detail here. Richard Arneson and others have ably defended the view that welfare is the currency of justice,[37] and I have elsewhere made my own modest contribution to this deep normative debate.[38] Sometimes, extrawelfarism is linked to egalitarianism.[39] But that linkage is a mistake, since welfarist views can be as egalitarian as you like, and reciprocally extrawelfarist views need not be particularly egalitarian. More concretely, there is nothing egalitarian about cost-effectiveness analysis, which will choose a program that benefits healthy individuals over one that benefits sick individuals if further health improvements for the healthy are less expensive.[40]

Sometimes, instead, a linkage is drawn between extrawelfarism and the morally attractive notions of responsibility and desert.[41] Simple welfarism

---

35. *See* Culyer, *Commodities*, *supra* note 32, at 14-15 (distinguishing between arthritic's health need and her marginal or total utility).

36. *Cf.* Philip Brickman et al., *Lottery Winners and Accident Victims: Is Happiness Relative?*, 36 J. PERSONALITY & SOC. PSYCH. 917, 920-21 (1978) (finding that paraplegic and quadriplegic accident victims were surprisingly happy).

37. *See, e.g.*, Richard J. Arneson, *Welfare Should Be the Currency of Justice*, 30 CAN. J. PHIL. 497 (2000); Andrew Moore & Roger Crisp, *Welfarism in Moral Theory*, 74 AUSTRALASIAN J. PHIL. 598, 613 (1996) (defending welfarism against influential criticisms, and concluding that "welfarism [is] a powerful and attractive position").

38. *See* Matthew D. Adler, *Beyond Efficiency and Procedure: A Welfarist Theory of Regulation*, 28 FL. ST. U. L. REV. 241, 307-09 (2000) [hereinafter Adler, *Beyond Efficiency and Procedure*]; Adler & Posner, *Rethinking Cost-Benefit Analysis*, *supra* note 22, at 215-16.

39. *See* Dolan, *supra* note 1, at 1727 ("[Extrawelfarists] will typically focus attention on equality of health . . . . ").

40. *See* Peter A. Ubel et al., *Improving Value Measurement in Cost-Effectiveness Analysis*, 38 MED. CARE 892, 893-94 (2000) (noting that QALY-based cost-effectiveness analysis ignores the distribution of health benefits, and in particular has been criticized for not giving sufficient priority to interventions that are life-saving interventions, benefit those who are severely ill, or benefit those with limited treatment potential due to disability or chronic illness).

41. *See* Arneson, *supra* note 37, at 504-05.

ignores the fact that a given individual might be responsible for a shortfall in her welfare. But welfarism need not be so simplistic.[42] Conversely, traditional cost-effectiveness analysis is no more sensitive to responsibility and desert than traditional cost-benefit analysis. Cost-effectiveness analysis will channel social resources to individuals who are responsible for their poor health states or high mortality risks, and away from more deserving types, if the poor health or high risk of the less deserving is cheaper to remedy.

I will not belabor these arguments here, but will simply take as given the welfarist framework for policy evaluation. To be sure, even if welfarists are correct as a moral matter—even if it is true that governmental choice morally ought to track some function of welfare, rather than some function of health or other capabilities/resources/primary goods—it remains the case that welfarism is legally precluded in certain domains. For example, the Clean Air Act requires the EPA to set air quality standards that "are requisite to protect the public health" and "allow[] an adequate margin of safety."[43] The Occupational Safety and Health Act stipulates that OSHA shall regulate toxic substances in the workplace so as to ensure, to the extent feasible, that no worker suffers "material impairment of health or functional capacity."[44] These provisions evidently give special priority to health. They would therefore seem to preclude cost-benefit analysis (both the traditional form and the nontraditional or "hybrid" variant this Article will defend), and indeed have been read to preclude cost-benefit analysis by the Supreme Court.[45] Yet, as Cass Sunstein has demonstrated, there are plenty of regulatory domains in which health, safety, and environmental agencies such as the EPA, National Highway Traffic Safety Administration (NHTSA), OSHA, and the FDA are statutorily permitted and perhaps even required to engage in cost-benefit analysis.[46] In these domains, the extrawelfarist has no particularly strong legal leg to stand on, nor (as I believe Arneson has shown) does she have a particularly strong moral leg either.

In short: Do QALYs have a role to play in *welfarist* policy evaluation? The remainder of this Article will focus on this question, and place the broader debate between welfarists and extrawelfarists to one side. The welfarists who have written about QALYs, mainly health economists, divide into two groups. One group recognizes that cost-effectiveness analysis can diverge from traditional cost-benefit analysis and therewith Kaldor-Hicks efficiency, but offers a

---

42. *See id.* at 506-08.

43. 42 U.S.C. § 7409(b)(1) (2000).

44. 29 U.S.C. § 655(b)(5) (2000).

45. *See* Whitman v. Am. Trucking Ass'ns, 531 U.S. 457, 464-71 (2001); Am. Textile Mfg. Inst., Inc. v. Donovan, 452 U.S. 490, 506-23 (1981).

46. *See* Cass R. Sunstein, *Cost-Benefit Default Principles*, 99 MICH. L. REV. 1651 (2001).

pragmatic defense of QALYs: The health care field will not accept cost-benefit analysis. For example, Alan Garber explains:

> [Cost-benefit] analysis requires placing dollar valuations on the outcomes of any program or intervention .... To many in the worlds of medicine and of public health, any attempt to place a value on a human life ... is anathema. Thus most 'economic' evaluations in health care have applied [cost-effectiveness] analysis, which limits the analyst's responsibility to providing information about the efficiency with which alternative strategies achieve health effects.[47]

Similarly, *Cost-Effectiveness in Health and Medicine*, the influential report by an expert panel of economists and public health scholars commissioned by the Public Health Service, states:

> Because of [cost-benefit analysis's] explicit grounding in welfare-economic principles, it is natural to ask why one would use cost-effectiveness rather than cost-benefit analysis if one wants to build from a welfare-economic foundation. Our interest in cost-effectiveness analysis derives largely from its broad acceptance within the health care field, in contrast to the skepticism that often greets cost-benefit analyses in that arena.
>
> It is the distinguishing feature of [cost-benefit analysis] that offends some sensibilities: In [cost-benefit analysis], the benefit of the health intervention is expressed in dollar terms rather than in terms of a nonmonetary effectiveness measure.[48]

Other health economists reject this pragmatic defense of QALYs and argue that, given the divergence between cost-effectiveness analysis and Kaldor-Hicks efficiency, QALYs should not be used in policy evaluation. Instead, they think, traditional cost-benefit analysis is the appropriate tool. For example, Mark Pauly explains how cost-benefit analysis implements Kaldor-Hicks efficiency ("the reason for preferring programs that maximize net benefits is [that] any such program can always be financed in such a way that everyone in society can be made better off"[49]), describes efficiency as the foundation of welfare economics ("[the] potential compensation test . . . is the one welfare economics generally

47. Alan M. Garber, *Advances in Cost-Effectiveness Analysis of Health Interventions* 4 (Nat'l Bureau of Econ. Research, Working Paper No. 7198, 1999).

48. COST-EFFECTIVENESS IN HEALTH AND MEDICINE, *supra* note 12, at 28.

49. Mark V. Pauly, *Valuing Health Care Benefits in Money Terms*, *in* VALUING HEALTH CARE, *supra* note 1, at 101.

uses"[50]) and concludes by rejecting cost-effectiveness analysis: "[In general] cost-effectiveness analysis is much less suitable . . . than cost-benefit analysis."[51] Similarly, Paul Dolan and Richard Edlin suggest that the only way to justify QALYs is by moving to extrawelfarism:

> In showing that there is currently no meaningful link between [cost-benefit analysis] and [cost-effectiveness analysis], we have also shown that [cost-effectiveness analysis] is not currently justifiable on strictly welfarist grounds. Instead, [cost-effectiveness analysis] would seem to be justifiable only on non-welfarist grounds where the output of health care is judged according to its contribution to health itself, rather than according to the extent to which it contributes to overall welfare.[52]

Or, to quote Don Kenkel: "[W]hen we accept the methodology of welfare economics, we should use cost-benefit analysis, not cost-effectiveness analysis."[53]

Both of these positions are problematic. Consider first the welfarist defense of QALY-based analysis as a second-best decision procedure justified by the resistance of the relevant community (doctors and public health officials) to the first-best procedure, cost-benefit analysis. At least at the regulatory level, any norm against pricing life that might once have existed has long been dissipated. Federal agencies have published numerous cost-benefit analyses incorporating an explicit, monetary valuation of human life.[54] More than thirty-five years ago, welfare economists demonstrated how the WTP/WTA methodology could be employed to monetize fatalities, by asking for WTP/WTA for a change in the risk of death rather than for certain death or the avoidance of certain death.[55] This theoretical scholarship generated a vast body of empirical work, typically employing wage-risk studies to estimate a "value of statistical life";[56] and these

---

50. *Id.*

51. *Id.* at 111.

52. Dolan & Edlin, *supra* note 28, at 838.

53. Don Kenkel, *On Valuing Morbidity, Cost-Effectiveness Analysis, and Being Rude*, 16 J. HEALTH ECON. 749, 755 (1997).

54. *See, e.g.*, Matthew D. Adler, *Fear Assessment: Cost-Benefit Analysis and the Pricing of Fear and Anxiety*, 79 CHI.-KENT L. REV. 977, 981 n.21 (2004) [hereinafter Adler, *Fear Assessment*] (finding that more than twenty-five percent of the cost-benefit analyses in the American Enterprise Institute database of major rulemakings for 1996-99 monetized death, illness, or injury).

55. *See* Michael W. Jones-Lee, *The Value of Changes in the Probability of Death or Injury*, 82 J. POL. ECON. 835 (1974); E.J. Mishan, *Evaluation of Life and Limb: A Theoretical Approach*, 79 J. POL. ECON. 687 (1971); T.C. Schelling, *The Life You Save May Be Your Own*, *in* PROBLEMS IN PUBLIC EXPENDITURE ANALYSIS 127 (Samuel B. Chase, Jr. ed., 1968).

56. For overviews of the theoretical and empirical scholarship on valuing life, see, for

empirical and theoretical literatures, in turn, have now decisively influenced agency practice.

Pragmatic constraints are surely relevant to policy evaluation. If a strong taboo on pricing life were in place, then that might justify governmental agencies in employing a non-cost-benefit procedure that, albeit second-best, did not offend the taboo. But there is no such taboo, now, at least in the governmental context. And the anti-pricing norms that might still obtain in other contexts (for example, among physicians or hospital administrators) would presumably weaken or dissolve once cost-benefit analysis started to be practiced there.

Consider next the welfarist position that rejects QALYs because cost-benefit analysis, not cost-effectiveness analysis, tracks Kaldor-Hicks efficiency. This position is doubly problematic. First, Kaldor-Hicks efficiency itself lacks moral significance. To recapitulate very quickly arguments that Posner and I have set forth in great detail elsewhere,[57] and that many others have made as well:[58] A Kaldor-Hicks efficient policy is merely one that could, hypothetically, be transformed into a Pareto-superior policy, through a costless redistribution from those who gain to those who lose. If the redistribution actually occurs, then the policy is actually (not just potentially) Pareto-superior and the criterion of Kaldor-Hicks efficiency is not needed to explain why the policy is attractive. If the redistribution doesn't occur, then the choice actually produces both winners and losers, and the link to Pareto-superiority has disappeared. Overall welfare, by contrast with Kaldor-Hicks efficiency, is a morally important notion; but a welfare-maximizing policy need not be potentially Pareto-superior, or vice versa. Because Kaldor-Hicks efficiency is no gold standard, the divergence between QALY-based analysis and the Kaldor-Hicks criterion is not an indictment of QALYs.

The welfarist position that rejects QALYs is problematic, too, because it hews to the widely shared but mistaken assumption that QALY-based analysis is necessarily a species of cost-effectiveness analysis. Once we shift from efficiency to the genuine gold standard—overall welfare—it will emerge that QALYs might function not as the measure of health in a cost-effectiveness

---

example, A. MYRICK FREEMAN III, THE MEASUREMENT OF ENVIRONMENTAL AND RESOURCE VALUES: THEORY AND METHODS 297-351 (2d ed. 2003); W. KIP VISCUSI, FATAL TRADEOFFS: PUBLIC AND PRIVATE RESPONSIBILITIES FOR RISK 17-74 (1992); W. Kip Viscusi & Joseph E. Aldy, *The Value of a Statistical Life: A Critical Review of Market Estimates Throughout the World*, 27 J. RISK & UNCERTAINTY 5 (2003).

57. *See* Adler, *Beyond Efficiency and Procedure*, *supra* note 38, at 248-59; Adler & Posner, *Implementing Cost-Benefit Analysis*, *supra* note 22, at 273-74; Adler & Posner, *Rethinking Cost-Benefit Analysis*, *supra* note 22, at 187-94.

58. *See* Adler, *Beyond Efficiency and Procedure*, *supra* note 38, at 249 (citing critics of Kaldor-Hicks efficiency).

analysis, but in a new analytic role.

I will ultimately argue below, in Part V, for a nontraditional, "hybrid" form of cost-benefit analysis in which health impacts are monetized by measuring those impacts on a QALY scale and then converting them to dollars using a QALY-to-dollar conversion factor.[59] Such QALY-to-dollar conversions should be used as a supplement to, and in some cases in lieu of, WTP/WTA measures of health impacts. The preferred tool for welfarist health policy evaluation should be neither traditional cost-benefit analysis alone nor cost-effectiveness analysis, but rather should include the "hybrid" cost-benefit technique just described.

This concrete recommendation will be my practical advice to policymakers. But it will not make much sense until we reconceptualize welfarist policy analysis in substantial, novel ways: by seeing overall welfare, not Kaldor-Hicks efficiency, as the underlying goal; and by understanding both QALYs and WTP/WTA amounts as workable but imperfect tools for measuring changes to overall welfare. Let us turn now to that reconceptualization.

II. OVERALL WELL-BEING AND LWUS

This Part provides the philosophical foundations for my revisionary account of QALYs. It clarifies the construct of overall well-being, drawing upon John Harsanyi's notion of "extended" lotteries, specifically "extended" lotteries over life-histories. Overall well-being, thus construed, should be representable as the sum of "lifetime welfare units" (LWUs). LWUs are the true metric for interpersonal welfare—the true welfarist gold standard for evaluating governmental decisional techniques.

The discussion here is quite theoretical. But it is, I believe, absolutely essential to a proper understanding of QALYs. In order to see how QALYs and WTP/WTA amounts are both imperfects proxies for LWUs—the thrust of Parts III and IV below—it is crucial to grasp the concept of an LWU.

Overall well-being is morally relevant, if not morally decisive.[60] Utilitarianism, the moral view that gives decisive weight to overall welfare, is often criticized because it ignores rights, distributive considerations, and "perfectionist" values such as the alleged value of preserving the environment quite apart from its welfare value. But none of these criticisms speaks to the

---

59. *See infra* Part V.

60. This Part draws heavily on my prior work, particularly Adler & Posner, *Rethinking Cost-Benefit Analysis*, *supra* note 22, and Adler, *Beyond Efficiency and Procedure*, *supra* note 38. I therefore generally cite to those works, rather than to the underlying scholarship in philosophy and economics, which they reference. My forthcoming book with Eric Posner, ADLER & POSNER, COST-BENEFIT ANALYSIS, *supra* note 22, will present a yet fuller account of the foundations of welfarist policy evaluation. The arguments in this Part are fleshed out in great detail in chapter 2 of the book.

weaker, pluralistic view that overall well-being is one criterion among others that together determine what ought morally to be done. Surely it is morally incorrect for government to reject a policy that increases overall welfare if the policy does not infringe any rights, distributive claims, or "perfectionist" values.[61]

There *is* one objection that applies to this weaker, pluralistic view as well as its more austere cousin, utilitarianism—namely, that interpersonal welfare comparisons are impossible and that talk about "overall well-being" is therefore meaningless. But this objection is misconceived. The scholarly skepticism about interpersonal comparisons that was widespread a half-century ago, tied to a broader skepticism about moral truth, has dissipated—certainly among philosophers and, to a substantial extent, among theoretical welfare economists too.[62] Moral truth does not entail the existence of unnatural moral objects and properties.[63] Rather, moral assertions can plausibly be unpacked as assertions about the judgments or reactions of impartial and well-informed observers, and are true just in case such observers would indeed judge or react in the manner asserted.[64] Folk moral discourse presupposes the existence of moral truths, not just coincidentally, but because it ends up being extremely difficult to maintain a thoroughgoing moral skepticism. Consider the welfare economist who purports to be a moral skeptic but then claims that government "should" implement Pareto-superior policies. What is the status of that "should"?

Once moral skepticism goes by the wayside, skepticism about interpersonal comparisons also naturally disappears. Why would it be the case that murdering, raping or torturing me truly violates my moral rights, but there can never be any truth of the matter about whether a welfare gain for me exceeds a welfare loss for you? Note also that if interpersonal welfare comparisons are impossible, then a wide range of moral theories must be rejected out of hand—not just utilitarianism, and the softer pluralistic view I am defending here, but also the egalitarian view currently popular among philosophers known as "prioritarianism,"[65] as well as any other partly or wholly egalitarian moral view

---

61. *See* Adler, *Beyond Efficiency and Procedure*, *supra* note 38, at 302-19; Adler & Posner, *Rethinking Cost-Benefit Analysis, supra* note 22, at 209-16, 243-45.

62. *See* Adler & Posner, *Rethinking Cost-Benefit Analysis, supra* note 22, at 204-09.

63. For an overview of "cognitivist" accounts of morality (accounts that recognize the existence of moral truths, moral beliefs, and moral facts) and their criticisms of noncognitivism, see generally ALEXANDER MILLER, AN INTRODUCTION TO CONTEMPORARY METAETHICS (2003).

64. *See id.* at 195-201.

65. Prioritarians claim that distributive and utilitarian considerations are fused in a single moral criterion, namely overall weighted welfare. Greater weight is given to welfare changes affecting individuals whose welfare levels are low, and less weight to changes affecting those whose welfare levels are high. *See, e.g.*, SHELLY KAGAN, NORMATIVE ETHICS 52-54 (1998). I have previously argued against prioritarianism, in favor of a moral view that incorporates ordinary, unweighted,

that requires a comparison of the welfare levels of different persons.[66] Welfarism, in all its plausible variants, just goes out the window.

How then should interpersonal welfare comparisons be understood? What does it mean to say that overall welfare in outcome O is truly greater than overall welfare in outcome O^*? One plausible approach derives from Adam Smith and has been endorsed in modern times by such scholarly luminaries as Kenneth Arrow, R.M. Hare, and Donald Davidson and formalized by John Harsanyi.[67] Harsanyi suggests that the two outcomes might be viewed as "extended" lotteries. Assume, for simplicity, that the same N individuals exist in each outcome. Then the extended lottery $(O; p_1, p_2 \ldots p_N)$ offers a p_1 chance of assuming the identity of individual *1* in outcome O, a p_2 chance of assuming the identity of individual *2* in outcome O, and so on. The extended lottery $(O^*; q_1, q_2 \ldots q_N)$ offers a q_1 chance of being individual *1* in outcome O^*, a q_2 chance of being individual *2* in outcome O^*, and so on. Harsanyi argues that individuals can have "extended" preferences—preferences over extended lotteries—and that humans who were fully informed, fully rational, and otherwise idealized (let us call these idealized humans "spectators" or "observers") would have the same extended preferences. Harsanyi proposes to analyze interpersonal welfare comparisons as involving an equiprobability extended lottery—one where the probabilities of assuming the identities of different persons are equal. Intuitively, overall welfare embeds a certain kind of equality among persons: Your welfare has no greater weight than mine, just because it is yours, in determining overall welfare.[68] In Harsanyi's schema, this deep impartiality is realized by stipulating the following: O is better for overall welfare than O^* just in case all spectators would prefer the extended lottery $(O; 1/N, 1/N \ldots)$ to $(O^*; 1/N, 1/N \ldots)$. There are N persons in each outcome and the chance of being any one is $1/N$.

Harsyani's schema is an idealization. Asking actual humans, with all their various cognitive limitations, to rank life-histories, let alone lotteries over life-histories, might be thought a ludicrous undertaking. In fact the undertaking may not be so ludicrous: There is a small but growing survey literature, an offshoot

---

overall welfare as one among a plurality of moral criteria, and Posner and I do so at greater length in our forthcoming book. ADLER & POSNER, COST-BENEFIT ANALYSIS, *supra* note 22, ch. 2; Adler, *Beyond Efficiency and Procedure*, *supra* note 38, at 309-11.

66. *See* ADLER & POSNER, COST-BENEFIT ANALYSIS, *supra* note 22, ch. 2.

67. *See* Adler & Posner, *Rethinking Cost-Benefit Analysis*, *supra* note 22, at 206-08; Adler, *Beyond Efficiency and Procedure*, *supra* note 38, at 292-302; John C. Harsanyi, *Morality and the Theory of Rational Behavior*, *in* UTILITARIANISM AND BEYOND 39 (Amartya Sen & Bernard Williams eds., 1982); John A. Weymark, *A Reconsideration of the Harsanyi-Sen Debate on Utilitarianism*, *in* INTERPERSONAL COMPARISONS OF WELL-BEING 255 (Jon Elster & John E. Roemer eds., 1991).

68. *See* KAGAN, *supra* note 65, at 49.

from the main body of QALY surveys, which seeks to elicit individuals' preferences for temporally extended health profiles consisting of a sequence of different health states.[69] In any event, the Harsanyi schema is not offered here as a decision-procedure that governmental decision-makers seeing to maximize overall welfare should actually employ. Rather, its function is analytic: It tells us what overall well-being *means*. Harsanyi invokes the preferences of idealized spectators contemplating extended lotteries in an attempt to *define* overall welfare.[70] This should not seem so strange to anyone familiar with contemporary welfare economics or contemporary philosophical scholarship about welfare; appeals to "fully-informed" or otherwise idealized preferences are very common in these literatures.[71]

But how, then, *should* actual policymakers go about their business? The quick answer is that maximizing overall welfare in the Harsanyi sense poses the predicament of choice under uncertainty (uncertainty about which extended lotteries the spectators would prefer) by decision-makers (the actual humans who make policy choices) with various cognitive limits. The technique of cost-benefit analysis is one response to this predicament. It may not be the ideal technique, but it is (or seems to be) the best available technique given the state of policy science at the dawn of the twenty-first century.[72] The trick is to design cost-benefit analysis so as to be sensitive to the Harsanyi construct and, more concretely, so as to incorporate whatever evidence we might have about what the preferences of fully-informed humans would be, without exceeding the cognitive capacities of actual policymakers or system-designers. So the field of policy analysis faces a difficult, but not insoluble, problem of specifying the optimal, feasible variant of cost-benefit analysis. Part V of the Article addresses this problem at somewhat greater length. The point I wish to emphasize here is that

---

69. *See* Paul F.M. Krabbe & Gouke J. Bonsel, *Sequence Effects, Health Profiles, and the QALY Model*, 18 MED. DECISION MAKING 178 (1998); Miriam Kuppermann et al., *Can Preference Scores for Discrete States Be Used To Derive Preference Scores for an Entire Path of Events?*, 17 MED. DECISION MAKING 42 (1997); Jeffrey Richardson et al., *The Measurement of Utility in Multiphase Health States*, 12 INT'L J. TECH. ASSESSMENT IN HEALTH CARE 151 (1996); Anne Spencer, *A Test of the QALY Model When Health Varies over Time*, 57 SOC. SCI. & MED. 1697 (2003); Jonathan R. Treadwell, *Tests of Preferential Independence in the QALY Model*, 18 MED. DECISION MAKING 418 (1998). Indeed, the so-called HYE ("healthy years equivalent") methodology for health policy analysis, a competitor to QALYs that some scholars have vigorously defended, requires individual valuations of temporally extended profiles. *See* Dolan, *supra* note 1, at 1729 (discussing HYEs); Johannesson, *supra* note 1, at 286-88 (same).

70. *See* Adler & Posner, *Rethinking Cost-Benefit Analysis*, *supra* note 22, at 216-25 (distinguishing between moral criteria and morally justified decision procedures).

71. *See id.* at 203 & n.100.

72. *See id.* at 225-43.

Harsanyi's schema defines a moral notion, overall well-being, as an esoteric standard (one framed in terms of spectators' extended preferences) which will never be transparent to us, but which we can do our best to grasp and apply in the design of actual governmental choice-procedures.

A more troubling feature of Harsanyi's schema is his claim that spectators would have the same extended preferences. What would justify this claim? Our ordinary preferences might diverge, even with full information. I might prefer to eat chocolate ice cream, while you might prefer to eat vanilla ice cream. If so, it is hard to see why idealized extended preferences must converge: I might prefer, *ceteris paribus*, a life history in which the subject eats chocolate to one in which the subject eats vanilla, while you might have the opposite preference.[73]

This flaw in Harsanyi's account can be remedied, I have argued elsewhere, by appealing to *convergent* extended preferences.[74] O is better for overall welfare than O^* just in case the spectators would *all* prefer a lottery that delivers outcome O and a $1/N$ chance of being each individual in O, over a lottery that delivers outcome O^* and a $1/N$ chance of being each individual in O^*. On this modification of Harsanyi's schema, the extent to which extended preferences converge under full information is an empirical issue, dependent on the facts about human nature; full convergence is not presupposed a priori.[75]

In their seminal work on decision theory, von Neumann and Morgenstern proposed axioms for the rationality of preferences over lotteries which, if satisfied, will permit those preferences to be represented by "utility" numbers.[76] Given those axioms, an individual will prefer a lottery to another just in case it has greater utility, in turn calculated as the expected utility of its component parts. [77] In the Harsanyi setup, the component parts of the lotteries are what I have referred to as life-histories: having the identity of some individual i in outcome O. An equiprobability extended lottery can be represented as a lottery over life histories $\{L_1, L_2 \ldots L_N\}$, each of which has probability $1/N$. The axioms of expected utility theory are, famously, violated in actual practice by untrained

---

73. *See* Adler, *Beyond Efficiency and Procedure*, *supra* note 38, at 293.

74. *See id.* at 297-300.

75. Indeed, respondents to QALY surveys do not fully converge on the weights they assign to health states. *See* U.S. EPA, HUMAN HEALTH METRICS, *supra* note 1, at 23; *infra* text accompanying notes 136-137.

76. *See* JOHN VON NEUMANN & OSKAR MORGENSTERN, THE THEORY OF GAMES AND ECONOMIC BEHAVIOR 15-31, 617-32 (3d ed. 1953).

77. On expected utility theory, see, e.g., SIMON FRENCH, DECISION THEORY: AN INTRODUCTION TO THE MATHEMATICS OF RATIONALITY 149-209 (1986); DAVID M. KREPS, NOTES ON THE THEORY OF CHOICE (1988); and MICHAEL D. RESNIK, CHOICES: AN INTRODUCTION TO DECISION THEORY 81-120 (1987).

humans with their various biases and limitations.[78] But a strong case can be made that the axioms, if not descriptive, are normative—they are partly constitutive of rational choice—and thus would be satisfied by idealized spectators. So each spectator's extended preferences can be represented by a utility function W, such that the spectator prefers an equiprobability lottery over life-histories $\{L_1 \ldots L_N\}$ to an equiprobability lottery over life-histories $\{L_1^* \ldots L_N^*\}$ just in case: $1/N\ W(L_1) + 1/N\ W(L_2) + \ldots 1/N\ W(L_N) > 1/N\ W(L_1^*) + 1/N\ W(L_2^*) + \ldots 1/N\ W(L_N^*)$.[79] Since the same number of individuals exist in each outcome, and since each life history is given an equal probability, the $1/N$ factors drop out and it emerges that a given spectator prefers one extended lottery to a second just in case the sum of utility numbers for the life histories involved is greater in the first case: $W(L_1) + W(L_2) + \ldots W(L_N) > W(L_1^*) + W(L_2^*) + \ldots W(L_N^*)$.

Because (*pace* Harsanyi) extended preferences might not converge, the utility numbers tracking the various spectators' extended preferences might not be identical either. Formally, there is a family of interpersonal utility functions $W_1 \ldots W_M$, corresponding to the M spectators; and one outcome O has greater overall welfare than another O^* just in case, for each of the utility functions $W_1 \ldots W_M$, the sum of the utility numbers assigned to the component life-histories of O is greater than the sum of the utility numbers assigned to the component life-histories of O^*.[80] The technical term for this is a "supervaluation": Overall well-being is a supervaluation over the extended utility functions of fully-informed spectators, conceiving each outcome as a lottery over life-histories with an equal chance of living each one. For simplicity, in the analysis of QALYs, WTPs, and cost-benefit analysis below, I will omit the supervaluationist caveat, assume convergence of extended preferences (as does Harsanyi), and refer to a single extended utility function W. But the argument carries over, I conjecture, to the more plausible case of partly divergent extended preferences. Because the analysis will prove to be long and complicated even on the assumption that spectators' extended preferences fully converge, I will not attempt to verify the conjecture in this Article.

To sum up: Overall well-being has at least prima facie moral relevance. Talk of "overall well-being" presupposes the possibility of interpersonal welfare comparisons—but comparisons of welfare levels or differences are presupposed by a wide variety of moral views, skepticism about them is no more plausible than general moral skepticism, and Harsanyi's construct of extended preferences

78. *See, e.g.*, SCOTT PLOUS, THE PSYCHOLOGY OF JUDGMENT AND DECISION MAKING 84-188 (1993).

79. *See* Weymark, *supra* note 67, at 289-97.

80. *Cf.* AMARTYA SEN, INEQUALITY REEXAMINED 46-49 (1992) (discussing an approach to social choice that focuses on convergent judgments and tolerates incompleteness).

and utility functions is a plausible way to give content to the notion of overall welfare. In the simple, limiting case, the spectators—our idealized counterparts—have convergent extended preferences, representable by a single utility function *W*. *W* assigns numbers to life-histories, and the overall well-being in some outcome is the sum of the utility numbers assigned to its component life-histories. The *W* function constitutes a cardinal and interpersonally comparable scale of welfare: a numerical representation of the degree to which changes in individual lives change the overall level of well-being.

Let us call the *W*-numbers "lifetime welfare units" (LWUs). I use this term, rather than "utility," because "utility" is a very general concept that subsumes a host of numerical representations of preferences, including but not limited to the extended preferences of our idealized counterparts. Further, the word "lifetime" underscores that overall welfare is first and foremost a composite of whole lives, not momentary time-slices. Individual moments or periods within an individual life contribute to overall well-being by contributing to the goodness of the individual life-history of which they are a part. To assume otherwise—to stipulate that overall well-being is necessarily decomposable into the sum of momentary or periodic well-being—is to rule out the very possibility of sequencing effects.[81] That would be deeply counterintuitive. Certain constituents of well-being, such as pains and pleasures, may not be subject to significant sequencing effects;[82] but to insist a priori that this is true of all constituents seems wrongheaded.[83] As David Velleman explains:

> Consider two different lives that you might live. One life begins in the depths but takes an upward trend . . . . Another life begins at the heights but slides downhill . . . . Surely, we can imagine two such lives as containing equal sums of momentary well-being. Your retirement is as blessed in one life as your childhood in the other; your nonage is as blighted in one life as your dotage in the other.
>
> Yet even if we were to map each moment in one life onto a moment of equal well-being in the other, we would not have shown these lives to be equally

81. *See infra* text accompanying notes 160-163 (discussing empirical studies of sequencing effects).

82. *See* Daniel Kahneman, *Experienced Utility and Objective Happiness: A Moment-Based Approach*, *in* 1 THE PSYCHOLOGY OF ECONOMIC DECISIONS: RATIONALITY AND WELL-BEING 187, 191-94 (Isabelle Brocas & Juan D. Carrillo eds., 2003) (presenting an "objective happiness" account of the experiential component of well-being in which total experiential utility is a function of momentary experiential utility without sequencing effects).

83. *See id.* at 205 ("Objective happiness is not proposed as a comprehensive concept of human well-being, only as a significant constituent of it.").

> good. For after the tally of good times and bad times had been rung up, the fact would remain that one life gets progressively better while the other gets progressively worse . . . . To most people, I think, the former story would seem like a better life story . . . in the sense that it is the story of a better life.[84]

The notion of preferences over whole lives, rather than moments or periods, might seem odd. Yet this notion is accepted by the welfare-economic literature on inter-temporal choice, where the decomposability of utility functions over lives into the sum of utility functions over periods is understood as a contingent rather than necessary fact.[85] The same holds true of the QALY literature, where once more the additive formula $\sum_j q(h_{i,j}) \times t_j$ (the sum total of time alive, with each time span weighted for the quality of health in that period) is understood to be an accurate representation of a lifetime health history only if certain axioms are satisfied.[86]

III. QALYS VERSUS WTP/WTA AMOUNTS AS PROXIES FOR LWUS: THE ADVANTAGES OF QALYS

What are QALYs? What are WTP/WTA amounts? And how do they relate to LWUs, the welfarist gold standard?

Section A answers these questions. Sections B and C then demonstrate how QALYs can, under some conditions, be better proxies for LWUs than WTP/WTA amounts. Section B brackets issues of measurement and shows that QALYs, if accurately measured, can in some special contexts perfectly track LWUs, while accurately measured WTP/WTA amounts will not. Section C describes various cognitive phenomena that make the measurement of WTP/WTA amounts particularly difficult and that may not interfere as substantially with QALY measurement.

It should be stressed that Sections B and C are *not* meant to demonstrate the overall superiority of QALYs to WTP/WTA amounts as proxies for LWUs. QALYs have their own difficulties, which will be discussed in Part IV. But it *is* crucial to see that the comparison of QALYs and WTP/WTA is a mixed bag. WTP/WTAs are not perfect welfarist measures. They have disadvantages, as well as advantages, vis-a-vis other practicable measures such as QALYs. This point is wholly overlooked in the existing welfarist literature on QALYs, where the only reason advanced to prefer QALY-based policy analysis to WTP/WTA-based

84. J. David Velleman, *Well-Being and Time*, *in* THE METAPHYSICS OF DEATH 329, 331 (John Martin Fischer ed., 1993).

85. *See, e.g.*, Han Bleichrodt & Amiram Gafni, *Time Preference, the Discounted Utility Model and Health*, 15 J. HEALTH ECON. 49, 53-58 (1996).

86. *See infra* Section IV.B.

analysis is a taboo on monetization.[87]

This Part therefore breaks new ground. Showing that WTP/WTAs do not dominate QALYs as a welfarist tool is a vital point, which policy-analytic scholarship has yet to grasp. My intention is not to ignore the limitations of QALYs, but rather to shine a bright light on the welfarist imperfections of the WTP/WTA measure, in this part of the Article, and then to do the same for QALYs in Part IV below.

A. What Are WTP/WTA Amounts? What Are QALYs?

QALYs and WTP/WTA amounts, like LWUs themselves, can be understood as idealized constructs. They constitute different ways to measure the welfare goodness of a life-history. These idealized constructs should serve to orient the actual practice of policy analysis: Responses to QALY surveys provide evidence that policy analysts can employ in estimating (idealized) QALYs, and similarly, responses to contingent-valuation surveys, or market behavior, are evidence of (idealized) WTP/WTA amounts.

Traditionally, cost-benefit analysis is defined as the sum of WTP/WTA: the money amounts that individuals are willing to pay or accept in return for choices that affect them. Posner and I have suggested that WTP/WTA amounts may need to be "laundered" in various ways, to correct for perceptual and evaluative biases that may cause individuals to be mistaken about what truly benefits them.[88] This suggestion motivates the following definition of (idealized) WTP/WTA amounts:[89] If O is the status quo and O^* is an alternative outcome, then individual i's WTP/WTA for O^* is the amount of wealth subtracted from or added to her life-history L_i^* in O^* (at the present time) such that her life-history L_i^*, with this wealth subtracted or added, is just as good as her baseline life-history L_i in O. By "just as good," I mean that the incremental resources just compensate for the difference in LWUs between L_i and L_i^*. The "laundering" here occurs because it is the preferences of idealized spectators—not the

87. *See supra* Part I.

88. *See* ADLER & POSNER, COST-BENEFIT ANALYSIS, *supra* note 22, ch. 5; Adler & Posner, *Rethinking Cost-Benefit Analysis*, *supra* note 22, at 220-22; Adler & Posner, *Implementing Cost-Benefit Analysis*, *supra* note 22, at 289-300; Matthew Adler, *Incommensurability and Cost-Benefit Analysis*, 146 U. PA. L. REV. 1371, 1381-83 (1998) [hereinafter Adler, *Incommensurability and Cost-Benefit Analysis*].

89. In prior work, I have used the term "welfare equivalent" (WE) to mean an idealized or "laundered" WTP/WTA amount. So as to avoid burdening the reader with too much unfamiliar jargon, I will stick to the WTP/WTA terminology here. But the substance of my analysis is fully consistent with my prior claims that cost-benefit analysis, ideally, should be understood as the sum of WEs rather than the sum of actual WTP/WTA.

individual's actual preferences—that determine the LWU values of the individual's various possible lives.

What about QALYs? How do they fit into my framework of life-histories and LWUs? An individual's longevity and morbidity profile is part of her life-history—part of the agglomeration of facts that affect her well-being—but not the only part. A given life history L_i might be thought of as a combination of a health-and-longevity history H_i, plus many other background features B_i, including individual consumption, recreational activities, social life, professional accomplishment, sexuality, and so on.[90] What do QALYs mean, given these other features in the background? Note that this question about the interaction between QALYs and background characteristics does not arise for the extrawelfarist, for whom two individuals with identical health profiles, but different consumption, leisure, social, professional and sexual profiles, are equally "healthy." The extrawelfarist social planner focused on the capability/resource/primary good of "health" can, quite correctly, insist that these lives should be assigned the same QALY value. But how can the welfarist social planner insist on that? Indeed, is it

---

90. The philosophical literature on objective welfare goods specifies a range of plausible dimensions of human well-being, such as life, bodily health, bodily integrity, the use of the "senses, imagination and thought," emotions, practical reason, affiliation, interaction with other species, play, and control over one's environment (Martha Nussbaum's list). *See* MARTHA C. NUSSBAUM, WOMEN AND HUMAN DEVELOPMENT: THE CAPABILITIES APPROACH 78-80 (2000); *see also* JOHN FINNIS, NATURAL LAW AND NATURAL RIGHTS 59-99 (1980) (identifying life itself, knowledge, play, aesthetic experience, sociability, practical reasonableness, and religion as objective welfare goods); JAMES GRIFFIN, VALUE JUDGEMENT: IMPROVING OUR ETHICAL BELIEFS 29-30 (1996) (identifying accomplishment, agency, understanding, enjoyment, and deep personal relations); DEREK PARFIT, REASONS AND PERSONS 499 (1984) (identifying moral goodness, rational activity, the development of one's abilities, having children and being a good parent, knowledge, and the awareness of true beauty). These philosophers' lists are, in effect, scholarly speculation about which feature of life-histories idealized spectators would intrinsically prefer. *See* Adler, *Beyond Efficiency and Procedure*, *supra* note 38, at 297-300. The World Health Organization has recently undertaken a more rigorous and comprehensive process, involving focus groups, surveys, and expert consultation, to arrive at an index of quality of life, the World Health Organization Quality of Life (WHOQOL) assessment, that encompasses twenty-four attributes. Some of these involve health, but many do not (for example, positive feelings, self-esteem, body image, personal relationships, social support, sexual activity, home environment, financial resources, leisure, and spirituality). *See* The World Health Organization Quality of Life Group, *The World Health Organization Quality of Life Assessment (WHOQOL): Development and General Psychometric Properties*, 46 SOC. SCI. & MED. 1569, 1576-78 (1998) [hereinafter The WHOQOL Group]; *infra* text accompanying note 170 (discussing the WHOQOL).

In short, there is much disagreement about what the welfare-relevant dimensions of life-histories are, but no disagreement that there are a considerable number of dimensions and that welfare transcends mere health and longevity.

even a coherent enterprise for the welfarist to measure the welfare contribution of health to a life that depends not only on health, but much else?

In answering this question, I shall draw upon important recent scholarship in utility theory by Miyamoto, Wakker, Bleichrodt, and Peters.[91] A very general problem in utility theory involves so-called "multidimensional" objects: objects with multiple types of attributes, falling into different domains or "dimensions." The general question is whether the utility numbers for the objects can in turn be decomposed into a sum, a product, or some other relatively tractable function of "subutility" numbers representing the different dimensions.[92] The simplest decomposition is additive, but it seems clear that the additive approach will not work for the problem at hand. A life-history L_i, as I have conceived it, is comprised of health-and-longevity attributes H_i plus background characteristics B_i. It would be nice if $W(L_i)$, the LWU measure of a given life-history, were representable as the sum of two subutility functions $Q(H_i) + V(B_i)$, but the requisite conditions for additive decomposition are too stringent to be met here. Those conditions say that two dimensions are additively separable if and only if a given change in one dimension has the same effect on overall utility, regardless of where the object (here, life-history) is located in the other dimension.[93] That surely is not true of the interaction between health and longevity, on the one hand, and leisure, consumption, social interaction, professional accomplishment and the other non-health constituents of welfare, on the other. For example, it is typically and plausibly supposed by welfare economists that health can change the welfare benefit of consumption.[94] If that is true, then presumably health can also change the welfare benefit of sex, socializing and so on. An additive representation of lifetime welfare as the sum of health-and-longevity subutility plus background subutility would preclude all of these sorts of interactions.

What Miyamato et al. demonstrate is that a *multiplicative* decomposition of utility is possible if two conditions are met: the "zero condition" and "standard-gamble invariance." The "zero condition" says that one of the dimensions must have a zero level, such that all objects at that level in that dimension have equal utility regardless of where they are located in the other dimension. Standard-

---

91. John M. Miyamato et al., *The Zero-Condition: A Simplifying Assumption in QALY Measurement and Multiattribute Utility*, 44 MGMT. SCI. 839 (1998).

92. *See generally* FRENCH, *supra* note 77, at 102-201 (discussing multiattribute value theory and utility theory); RALPH L. KEENEY & HOWARD RAIFFA, DECISIONS WITH MULTIPLE OBJECTIVES: PREFERENCES AND VALUE TRADEOFFS (1976) (same); DETLOF VON WINTERFELDT & WARD EDWARDS, DECISION ANALYSIS AND BEHAVIORAL RESEARCH 259-350 (1986) (same).

93. *See, e.g.*, FRENCH, *supra* note 77, at 182-91; VON WINTERFELDT & EDWARDS, *supra* note 92, at 302-08, 334-41.

94. *See, e.g.*, Evans & Viscusi, *supra* note 29, at 499-500; Hammitt, *QALYs Versus WTP*, *supra* note 1, at 991-92 & n.11.

gamble invariance says that lotteries in which the first dimension (the one with the zero level) is varied and the second held fixed must be ranked the same way, regardless of where the second dimension is fixed.[95] These axioms, albeit technical and a bit difficult to grasp, turn out to have considerable relevance for QALYs, LWUs, and policy evaluation.[96] For I suggest that health-and-longevity and background characteristics do interact in just the way required by the Miyamato conditions. At a minimum, the assumption that they do thus interact is a reasonable working premise, one that I will consider relaxing later on but, for now will take as true.[97]

The health-and-longevity dimension satisfies the "zero condition." Because being alive is a precondition for consumption, leisure, sex, socializing, and so on, a life-history whose health profile H_i has zero longevity must be given the same ranking by extended spectators, and have the same LWU number, regardless of whatever background characteristics we might link to that life-history. As for standard-gamble invariance: Imagine that idealized spectators are indifferent between a given life history $L_i = (H_i, B_i)$, and a lottery over health characteristics holding fixed the background characteristics of that life-history, where there is a chance p of getting (H_i', B_i) and a chance $(1 - p)$ of getting (H_i'', B_i). Then idealized spectators must also be indifferent between the same lottery over the health dimension with a different set of background characteristics. $L_i^* = (H_i, B_i^*)$ must be just as good as a chance p of getting (H_i', B_i^*) and a $(1 - p)$ chance of getting (H_i'', B_i^*). Being wealthier or poorer, more or less social, more or less accomplished, cannot change the relative attractiveness of different health

95. *See* Miyamato et al., *supra* note 91, at 845-48. The Miyamato conditions allow for any finite number of attributes or "dimensions." I have stated the conditions as they apply to the two-attribute case, since I am here conceptualizing life-histories as involving two broad dimensions of welfare (each of course encompassing a variety of more specific subdimensions): health/longevity and background characteristics.

96. Han Bleichrodt and John Quiggin have also employed the standard-gamble invariance condition and zero condition to analyze the interaction between health and consumption, and James Hammitt has employed the standard-gamble invariance condition. *See* Han Bleichrodt & John Quiggin, *Life-Cycle Preferences over Consumption and Health: When Is Cost-Effectiveness Analysis Equivalent to Cost-Benefit Analysis?*, 18 J. HEALTH ECON. 681, 688-90 (1999); James Hammitt, How Much Is a QALY Worth? Admissible Utility Functions for Health and Wealth 3-6 (May 2002) (unpublished manuscript, prepared for Association of Environmental and Resources Economists at the Allied Social Sciences Associations meeting), *available at* www.feem.it/NR/Feem/resources/conferences/PRE2004-01-03-01.Hammitt.pdf) [hereinafter Hammitt, How Much Is a QALY Worth?]. My analysis, with its focus on LWUs and the relative accuracy of QALYs and WTP/WTA in tracking LWUs, is quite different from Bleichrodt and Quiggin's and Hammitt's, but I am indebted to their work.

97. *See infra* Section IV.A.

profiles and gambles over health profiles. This health-gamble-invariance condition, while surely less compelling than the proposition that all zero-longevity lives are equally good for welfare, seems just as plausible as standard restrictions on the structure of preferences that welfare economists employ so as to achieve tractable formalization and, therewith, cheap and feasible policy-evaluation procedures.[98]

If spectators' extended preferences over life-histories indeed satisfy the Miyamato conditions, then $W(L_i) = Q(H_i) \times V(B_i)$. The welfare value of a life is representable as the multiplicative product of the value of its health-and-longevity characteristics and the value of its background characteristics. This brings us to QALYs. My proposal is that the QALY value of a given life be defined as this first value: $Q(H_i)$, the numerical value of the health-and-longevity profile partly constitutive of that life, which when multiplied by a numerical value $V(B_i)$ representing the consumption, leisure, professional, and other non-health attributes of the life, equals the overall utility or LWU value $W(L_i)$.

To reiterate: Assuming the Miyamato conditions hold true, we can define a "subutility" number $Q(H_i)$ encapsulating the health-and-longevity characteristics of a given life-history, and a "subutility" number $V(B_i)$ encapsulating the background characteristics of that life-history. The overall utility or LWU value of the life-history will be the multiplicative product of these two "subutilities." H_i is a nonnumerical description of the nature and duration of the health states making up a life history. $Q(H_i)$ is the "subutility" that represents that health profile. Similarly, B_i is a nonnumerical description of the background characteristics of a life history. $V(B_i)$ is the subutility that represents that non-health profile. I define the QALY value of a life history as $Q(H_i)$. This QALY value is one input, along with $V(B_i)$, in a multiplicative formula that determines the LWU value associated with that life-history.

At this point, the reader might wonder how my definition of the QALY value of a life history, as $Q(H_i)$, relates to the standard formula for calculating QALYs, namely $\sum_j q(h_{i,j}) \times t_j$, where $h_{i,j}$ is the health of individual i in period j and t_j is the duration of period j. The answer is that $Q(H_i)$ is an idealized construct: the true health subutility of a life. $\sum_j q(h_{i,j}) \times t_j$ is a practicable formula for estimating $Q(H_i)$. If certain axioms hold true, the estimate will be perfect: $\sum_j q(h_{i,j}) \times t_j$ will necessarily equal $Q(H_i)$. If those axioms do not hold true, then $\sum_j q(h_{i,j}) \times t_j$ will not necessarily equal $Q(H_i)$, and must be understood as a rough

98. For example, it is standardly assumed that preferences are complete and transitive and often assumed that they can be represented by well-behaved (continuous, differentiable, monotonic, quasi-concave) functions. *See, e.g.*, DAVID M. KREPS, A COURSE IN MICROECONOMIC THEORY 17-37 (1990) (discussing economic theory of consumer choice); CARL P. SIMON & LAWRENCE BLUME, MATHEMATICS FOR ECONOMISTS 544-57 (1994) (same).

estimate, perhaps quite rough. The axioms, and the possible divergence between the QALY value of a life and $\sum_j q(h_{i,j}) \times t_j$, are discussed below in Part IV.[99]

B. How QALYs Can Improve on WTP/WTA Amounts (Bracketing Measurement Problems)

This Section brackets measurement issues and shows that accurately measured QALYs, i.e., $Q(H_i)$ values, will, under certain conditions, be perfect proxies for LWUs. Under the same conditions, accurately measured WTP/WTA amounts will not be perfect proxies for LWUs. This analysis is admittedly artificial, because policymakers do not have direct access to $Q(H_i)$ values, any more than they do to LWUs themselves. Measurement problems are in fact omnipresent. Still, the analysis will help highlight one of the potential advantages of nontraditional, QALY-based cost-benefit analysis, as compared to traditional, WTP/WTA-based cost-benefit analysis.

The proposition that accurately measured QALYs will perfectly track LWUs, under some conditions, is a direct upshot of the Miyamato conditions. The qualifying phrase "under some conditions" is crucial. If one life is different in the value of its background characteristics than another, then a health improvement in the first life will have a different impact on overall welfare than a health improvement in the second life. This is a point I will return to below.[100] But the reciprocal point is also true: Where the lives being compared have the same, or equally valuable, background characteristics, differences in their QALY values will perfectly track differences in their welfare goodness.

Imagine a class of life-histories $\{L_1, \ldots L_m\}$, composed of health profiles $\{H_1, \ldots H_m\}$ and profiles of consumption, leisure, and other background characteristics $\{B_1, \ldots B_m\}$. Assume that all the life-histories are equal in the value of these non-health profiles: that is, $V(B_1) = V(B_2) = \ldots = V(B_m) = K$. As a shorthand, let us call this a "background-equivalent" class of lives. The QALY values of the lives may vary. $Q(H_1)$ does not necessarily equal $Q(H_2)$ or $Q(H_3)$ and so on. The Miyamato model entails that $W(L_i) = Q(H_i) \times V(B_i)$, which reduces to $W(L_i) = Q(H_i) \times K$. In other words, the LWU value of a life within the background-equivalent class equals its QALY value multiplied by a constant K representing the background value of all the lives in the class. It follows that QALY changes are a perfect indicator of the welfare impacts of policy choices whose effects are limited to changing, either for certain or probabilistically, which life-histories within the class occur.

Consider first a policy that affects morbidity rather than mortality. In the

99. *See infra* Section IV.B.

100. *See infra* Section IV.C.

status quo, individual *1* has life-history L_1, and individual *2* has life-history L_2. The policy would improve the first individual's health, changing her life-history to L_1', and degrade the second individual's health, changing her life-history to L_2'. Assume, crucially, that the status quo and alternative possible lives, L_1, L_2, L_1', and L_2', fall within the same background-equivalent class. Then the overall change in QALYs achieved by implementing the policy would be $Q(H_1') + Q(H_2') - Q(H_1) - Q(H_2)$. And the overall change in welfare, represented by LWUs, would be $Q(H_1') \times K + Q(H_2') \times K - Q(H_1) \times K - Q(H_2) \times K = K[Q(H_1') + Q(H_2') - Q(H_1) - Q(H_2)]$. So the policy improves overall welfare if and only if it increases total QALYs.

Consider next a policy that affects mortality risks, again assuming that all the lives involved fall in the same background-equivalent class, with scaling constant K representing the value of the lives' background characteristics. In the status quo, individual *1* has life-history L_1, and individual *2* has life-history L_2. The policy would create a risk p for individual *1* of life-history L_1', and a risk $(1 - p)$ of life-history L_1'', where L_1' and L_1'' involve a longer or shorter life-span than the baseline L_1. Similarly, the second individual would incur a r risk of life-history L_2', and a $(1 - r)$ risk of L_2'', again with different life-spans than baseline L_2. As before, each life history L_i has a component health history H_i; in this example, the various health histories differ in the length of time that the subject lives. The expected QALY change from the policy is $p \times Q(H_1') + (1 - p) \times Q(H_1'') - Q(H_1) + r \times Q(H_2') + (1 - r) \times Q(H_2'') - Q(H_2)$. And the expected change in LWUs produced by the policy turns out to be that amount multiplied by the scaling constant K. Once more, the policy produces an expected increase in overall welfare if and only if it produces an expected increase in QALYs.

What about WTP/WTA amounts? There are various ways in which WTP/WTA measures of welfare effects on individual lives can fail to be perfectly correlated with utility measures of those effects. Our concern here is extended preferences and "laundered" or "idealized" WTP/WTA amounts, but the imperfections in money measures of utility carry over to this context. WTP/WTA amounts, again, are present wealth changes that compensate for welfare changes. Specifically, a present wealth increase compensates for a subject's loss in welfare by increasing the subject's expected material consumption over his lifetime. A present wealth reduction compensates for a subject's increase in welfare by decreasing the subject's expected material consumption over his lifetime.

One reason why WTP/WTA amounts do not perfectly track LWUs is that changes in consumption need not translate into changes in LWUs at an interpersonally constant rate. To see this point, consider the following case. In the status quo, individual *1* has life L_1; if the policy were implemented, his life would instead be L_1'. In the status quo, individual *2* has life L_2, replaced by L_2'

under the policy. Each of these lives is a combination of health characteristics and background characteristics including consumption. So $L_1 = (H_1, B_1) = (H_1, (C_1, S_1, P_1 \ldots))$; and $L_2 = (H_2,B_2) = (H_2, (C_2, S_2, P_2 \ldots))$. The change in LWUs resulting from the policy would be $W(L_1') + W(L_2') - W(L_1) - W(L_2)$. Assume that present wealth changes directly translate into present changes in consumption. In other words, the first individual's laundered WTP/WTA for the policy is that present change in his consumption profile C_1', $\Delta C_1'$, such that the lifetime welfare units of the amended L_1' equals $W(L_1)$. The same holds for the second individual's WTP/WTA.

If LWU changes are proportional to WTP/WTA amounts—that is, if $[W(L_1') - W(L_1)] / \Delta C_1' = [W(L_2') - W(L_2)] / \Delta C_2'$—then the sum of WTP/WTA does track LWUs. But there are a variety of reasons why, in general, the proportionality constraint just articulated need not hold true. Individual *1* may have a much higher level of lifetime consumption than individual *2*, so that changes in his consumption have a smaller effect on lifetime welfare. Consumption has diminishing "marginal utility," to use the standard economic lingo.[101] Or sequencing effects might kick in: Although the two individuals have roughly the same level of lifetime consumption, the $\Delta C_1'$ would occur at a point in individual *1*'s consumption sequence where his welfare is particularly sensitive, or insensitive, to consumption changes. Finally, consumption might interact with health, or with other background characteristics.[102]

A different reason why the sum of WTP/WTA amounts need not perfectly track the sum of LWUs is that present wealth changes need not always induce the same expected consumption changes in different individuals. WTP/WTA amounts involve changes to present *wealth*, and the linkage between that change and the change in the individual's expected consumption (and therewith his welfare), need not be interpersonally constant. The so-called "dead anyway" effect, first discovered by Pratt and Zeckhauser,[103] and evident in various economic models of WTP/WTA for longevity,[104] might be understood in these

101. *See, e.g.*, Adler, *Incommensurability and Cost-Benefit Analysis*, *supra* note 88, at 1398-1401.

102. *See* sources cited *supra* note 94.

103. *See* John W. Pratt & Richard Zeckhauser, *Willingness To Pay and the Distribution of Risk and Wealth*, 104 J. POL. ECON. 747, 750-53 (1996). For an accessible discussion of this effect, see Hammitt, *QALYs Versus WTP*, *supra* note 1, at 992-93.

104. *See, e.g.*, Anna Alberini et al., *Does the Value of a Statistical Life Vary with Age and Health Status? Evidence from the US and Canada*, 48 J. ENVTL. ECON. & MGMT. 769, 771-73 (2004); Maureen Cropper & Frances Sussman, *Valuing Future Risks to Life*, 19 J. ENVTL. ECON. & MGMT. 160, 162-65 (1990); Hammitt, *QALYs Versus WTP*, *supra* note 1, at 992-93; *cf.* Friedrich Breyer & Stefan Felder, *The Dead-Anyway Effect Revisited* 9 (CESifo, Working Paper No. 805, 2002), *available at* http://www.cesifo.de/~DocCIDL/805.pdf (analyzing dead-anyway effect in

terms. To see the effect most simply, ignore non-consumption background characteristics, assume that the value of consumption is a constant function of lifetime consumption, and assume that all the individuals involved are healthy. So, for all lives, $W(L_i) = Q(H_i) \times V(C_i) = y_i \times k \times c_i$ (y_i is years, c_i is total lifetime consumption.) In the status quo, individual *1* is currently y_1 years old. He has a baseline risk r of dying immediately, in which case his lifetime consumption will have been c_1, and $(1 - r)$ of living until age y_2, in which case his lifetime consumption will be c_2, which is larger than c_1. He is asked for his willingness to pay for a policy that will reduce the risk of dying immediately to q, which is less than r.

The welfare change induced by the policy, measured in LWUs, turns out to be: $(r - q) \times k \times [y_2c_2 - y_1c_1]$. This is, as it should be, solely a function of the risk reduction $(r - q)$ and the difference in value of the lives involved ($k \times (y_2c_2 - y_1c_1)$). The LWU change is the same regardless of the absolute level of initial mortality risk r or residual risk q. But the amount of wealth—potential future consumption—that the individual is willing to sacrifice for the risk reduction depends on the absolute level of risk. In this case, $WTP = (r - q) \times (y_2c_2 - y_1c_1) / [(1 - q) \times y_2]$. So WTP is partly a function of q, the residual risk of dying that remains after the risk reduction. Note that, the larger the residual risk q, holding constant the risk reduction $(r - q)$, the larger WTP becomes. Why? The larger the residual risk, the larger the chance that the present resources that the individual is sacrificing will not eventuate in consumption: The individual will "die anyway." The expected change in lifetime consumption for a present change in wealth depends, not just on the wealth change, but on the probability that death will intervene and preclude consumption. That is the essence of the "dead anyway" effect.[105]

A different source of slippage in the wealth-consumption nexus is the fact that individuals might face different investment opportunities, for example investment horizons, so that (even bracketing differences in the risk of death) present wealth changes can produce different changes in lifetime consumption for different individuals.[106] Individuals *1* and *2* have identical health histories, leisure profiles, and so on, but the first is older than the second. At present,

---

models with and without bequests and under perfect and imperfect insurance markets and finding that "for individuals without a bequest motive the value of a statistical life always increases with the level of risk exposure if and only if they are risk-averse with respect to wealth").

105. *See* Olivier Armantier & Nicholas Treich, *Social Willingness To Pay, Mortality Risks and Contingent Valuation*, 29 J. RISK & UNCERTAINTY 7, 8 (2004) (showing that heterogeneity in individuals' survival probabilities can lead sum-of-WTP criterion to deviate from overall welfare).

106. This effect is evident in Yew-Kwang Ng's model of WTP/WTA for longevity. *See* Yew-Kwang Ng, *The Older the More Valuable: Divergence Between Utility and Dollar Values of Life as One Ages*, 55 J. ECON. 1, 4-11 (1992).

individual *1* will live for five more years, while individual *2* will live for twenty years. Both can invest their resources at a nonzero real interest rate, and spectators have no time preference for present over future consumption.[107] Then the two individuals will have different, "laundered" WTP/WTA amounts for the very same morbidity change. Specifically, the older individual will be willing to pay more to avoid increased morbidity, and demand more to accept it, because he has fewer years in which to invest present dollars in the market, so that present wealth increments translate into smaller consumption increments for him than for the younger individual. This "horizon" effect, along with the "dead anyway" effect, helps explain why the elderly might have inflated WTP/WTA amounts for morbidity or mortality changes, relative to the welfarist gold standard for measuring those changes, namely LWUs.

I have discussed different ways in which WTP/WTA amounts might be imperfect indicators of LWUs.[108] I have not yet shown that the WTP/WTA amounts can, under some conditions, be worse proxies than QALYs. To see that, consider the conditions where QALYs are perfect proxies: where all the lives affected by the policy choice fall in the same background-equivalent class. It is straightforward to demonstrate that the various imperfections of WTP/WTA can occur under these conditions, and I do that in the margin.[109] To be sure,

---

107. *See infra* text accompanying notes 153-154 (suggesting that spectators might lack an intrinsic time preference).

108. *Cf.* Armantier & Treich, *supra* note 105, at 17 (using simulation analysis and finding that, where wealth, baseline risk, and risk reduction are heterogeneous and uncorrelated, the sum-of-WTP method overestimates overall welfare for a risk reduction project by fifteen percent).

109. Here are some simple models illustrating that WTP/WTA amounts can fail to be perfect proxies for LWUs even where all the lives affected by a policy choice fall in the same background-equivalent class. This is an important difference from QALYs, which *cannot* fail to be perfect proxies under these conditions. (1) *Diminishing marginal utility of consumption.* Imagine that each $L_i = (H_i, B_i)$, where background characteristics are decomposable into consumption C_i and other characteristics Z_i. By hypothesis, $W(L_i) = Q(H_i) \times V(B_i) = Q(H_i) \times V(C_i, Z_i)$. Assume that $V(C_i, Z_i)$ is increasing in C_i and Z_i; that C_i has a diminishing marginal impact on $V(C_i, Z_i)$; and that the marginal impact of C_i does not decrease with increasing Z_i. Consider now a class of lives $\{L_i\}$ with the same V value $= K$. Some of these lives are high consumption, others low consumption, in each case with corresponding Z_i characteristics such that overall $V(C_i, Z_i) = K$. A larger ΔC_i will be required to compensate for a given health change (meaning here both the initial level of $Q(H_i)$ and the change in health value) and concomitant LWU change in a high-consumption life than in a low-consumption life. (2) *Interaction between consumption and health.* Assume that $L_i = (H_i, C_i)$ and that $W(L_i) = Q(H_i) \times c_i$. Consider the background equivalent class of lives with the same total consumption c. Imagine that the lives in the status quo are L_1 and L_2 and that a policy would improve their health to L_1' and L_2', such that $Q(H_1') - Q(H_1) = Q(H_2') - Q(H_2) = q$ for both pairs of lives. It follows that $\Delta W = qc$ for both pairs of lives. But the Δc amount required to balance the change in the first life is $cq/Q(H_1')$, while the Δc amount required to balance the change in the

background equivalence is a special set of conditions; under other conditions, WTP/WTA may well track LWUs better than QALYs do. But my aim here is not to claim general superiority for QALYs. Rather, it is the more modest—but still important—aim of demonstrating that, contrary to the general wisdom in the welfarist literature on QALYs, QALYs are not dominated by WTP/WTA.

C. Measuring QALYs and WTP/WTA Amounts

The previous Section bracketed the problem of measuring QALYs and WTP/WTA and argued that a well-informed respondent's ("spectator's") QALY valuations for various life-histories, if perfectly measured, might under some conditions more accurately track her preferences over those possible lives than her WTP/WTA values, again perfectly measured. This Section suggests that QALY values might be easier to measure than WTP/WTA.[110] Certain cognitive difficulties interfering with the elicitation of WTP/WTA for mortality and morbidity may not afflict QALY measurement, at least not as substantially.

In this Section, I will assume that the QALY formula $\sum_j q(h_{i,j}) \times t_j$ is a reasonably good estimate of the true QALY value of a life, namely $Q(H_i)$. It may not be—as mentioned above, and further elaborated in the next Part. But if the axioms implying the equivalence of $Q(H_i)$ and $\sum_j q(h_{i,j}) \times t_j$ are not grossly violated, then $Q(H_i)$ may well be easier to measure—via the additive formula $\sum_j q(h_{i,j}) \times t_j$—than WTP/WTA amounts.

Consider first the measurement of WTP/WTA amounts. One technique for determining these amounts is "contingent valuation": an interview-based methodology. Various interviewing techniques have been developed.[111] Respondents might be asked to state the maximum amount of money they would

---

second life is $cq/Q(H_2')$. (3) *Dead anyway effect.* Imagine healthy individuals who live y_i years, with consumption level c_i and further characteristics z_i such that $W(L_i) = y_i \times c_i \times z_i$. Within a background equivalent class, $c_i \times z_i = K$, or $z_i = K/c_i$. In the status quo, an individual has a risk r of dying now after living y_1 years and consuming c_1, and risk $(1 - r)$ of living longer to y_2 years, consuming c_2. He is asked for WTP to reduce r to q. The LWU change is $(r - q) \times (y_2 - y_1) \times K$. But the WTP amount equals $(r - q) \times (y_2 - y_1) \times c_2 / [y_2 \times (1 - q)]$.

110. *But see* Richard D. Smith, *The Relative Sensitivity of Willingness-To-Pay and Time-Trade-Off to Changes in Health Status: An Empirical Investigation*, 10 HEALTH ECON. 487, 495-96 (2001) (finding WTP survey to be a more sensitive indicator of changes in health status than QALY survey).

111. For overviews of this methodology, see IAN BATEMAN ET AL., ECONOMIC VALUATION WITH STATED PREFERENCE TECHNIQUES: A MANUAL (2002); FREEMAN, *supra* note 56, at 161-87; VALUING ENVIRONMENTAL PREFERENCES: THEORY AND PRACTICE OF THE CONTINGENT VALUATION METHOD IN THE US, EU, AND DEVELOPING COUNTRIES (Ian J. Bateman & Kenneth G. Willis eds., 1999).

be willing to pay in return for a good, or the minimum amount they would be willing to accept in exchange for a bad. Or they might be asked about their willingness to trade goods or bads for particular amounts of money, with the pattern of "yes" and "no" answers then used to infer a maximum WTP or minimum WTA. Whatever the technique employed, contingent-valuation studies in the area of health and mortality have collided with certain characteristic biases.

These biases emerge most dramatically in contingent-valuation studies of mortality or morbidity risks.[112] Expected utility theory implies that WTP for reducing the risk of death or harm should be nearly proportional to the change in risk, if it is small. "[I]f a reduction in annual mortality risk from 20 in 100,000 to 18 in 100,000 is valued at $20, then a larger reduction from 20 to 16 in 100,000 should be valued at about $40 (ignoring a tiny income effect)."[113] Hammitt and Graham surveyed published contingent-valuation studies of mortality or health risks, focusing on the studies that allowed either "internal" tests of proportionality or "external" tests. If each respondent is asked for WTP for multiple risk reductions, and her responses are proportional, then that provides "internal" evidence of proportionality. Asking one group of respondents for their WTP for a particular risk reduction, and a different group for their WTP for a different reduction, permits an "external" test of proportionality.

Hammitt and Graham found that, of the ten studies permitting an "internal" test of proportionality, not a single study confirmed it.

> [T]he average respondent [in these studies] does state a larger willingness to pay for larger risk reductions—i.e., the direction of change in payment size is in accordance with expectations. Yet a significant minority of respondents often report the same willingness to pay, regardless of the size of risk reduction . . . When the proportionality assumption is tested through internal tests, it generally fails. Mean willingness to pay is less—usually much less—than proportional to risk reduction. It is not uncommon for a doubling of risk

112. To be sure, the problems in measuring WTP/WTA for morbidity risks might be avoided by measuring WTP/WTA for disease states that occur with certainty, but the same is not true of mortality. WTA for certain death may well be infinite; WTP to avoid certain death may equal the subject's entire stock of wealth. The money amounts that traditional cost-benefit analysts employ to value lifesaving, the so-called "value of statistical life" (VSL), are therefore derived from contingent-valuation or revealed-preference studies examining WTP/WTA for the risk of death. *See* Matthew D. Adler, *Against "Individual Risk": A Sympathetic Critique of Risk Assessment*, 153 U. PA. L. REV. 1121, 1198 n.300 (2005) [hereinafter Adler, *Against "Individual Risk"*] (citing sources discussing the VSL method).

113. James K. Hammitt & John D. Graham, *Willingness To Pay for Health Protection: Inadequate Sensitivity to Probability?*, 8 J. RISK & UNCERTAINTY 33, 35 (1999).

reduction to be associated with far less than a 50% increase in payment.[114]

Proportionality was similarly disconfirmed by all nine of the "external" studies.[115] Indeed, "[e]ven the less demanding test that willingness to pay be larger among respondents who are offered larger risk reductions, while satisfied in some studies . . . is not always satisfied in others."[116]

What explains these disheartening results? In other contexts, for example contingent-valuation studies of environmental preservation, the insensitivity of WTP/WTA to the magnitude of the good produced or destroyed might reflect the fact that respondents are voicing moral rather than self-interested preferences.[117] That is a less compelling explanation, however, of the magnitude-insensitivity that Hammitt and Graham describe.[118] More plausibly, respondents queried about WTP/WTA for mortality and morbidity risks are affected by well-known biases that interfere with the evaluation of risk—for example, a misunderstanding of the basic rules of probability, or a departure from expected-utility theory in valuing lotteries, such as lotteries over personal health and longevity.[119]

A different set of biases, also presumably at play in the contingent-valuation studies about mortality and morbidity risk, involves tradeoffs between different kinds of goods—here money, on the one hand, and health and longevity, on the other. "[M]aking tradeoffs is a cognitively demanding task that people will try to minimize."[120] There can also be emotional resistance to commensuration, for example where one good in the tradeoff is understood (pre-theoretically) by the

---

114. *Id.* at 39; *see also id.* at 37-38 (listing the studies that the authors relied upon, showing that proportionality failed in every study where it could be tested).

115. *See id.* at 39-40.

116. *Id.* at 40. For similar findings that WTP/WTA is not sensitive to the magnitude of risk reduction, see Jane Beattie et al., *On the Contingent Valuation of Safety and the Safety of Contingent Valuation: Part 1-Caveat Investigator*, 17 J. RISK & UNCERTAINTY 5 (1998), and the sources cited in U.S. EPA, HUMAN HEALTH METRICS, *supra* note 1, at 18; Jonathan Baron, *Biases in the Quantitative Measurement of Values for Public Decisions*, 122 PSYCHOL. BULL. 72, 74 (1997); and Hammitt, *QALYs Versus WTP*, *supra* note 1, at 997.

117. *See, e.g.*, Baron, *supra* note 116, at 75.

118. *See also* Alan Shiell & Lisa Gold, *Contingent Valuation in Health Care and the Persistence of Embedding Effects Without the Warm Glow*, 23 J. ECON. PSYCHOL. 251 (2002) (finding "embedding effects" in a health care contingent-valuation study designed to exclude moral preferences).

119. *See* PLOUS, *supra* note 78, at 84-188 (discussing these biases). *But see* Hammit & Graham, *supra* note 113, at 35 ("There is a variety of descriptive models of choice that predict [sic] responses to risk that are nonlinear in the probabilities [such as prospect theory] . . . but even these models are locally linear . . . .").

120. John W. Payne et al., *Measuring Constructed Preferences: Towards a Building Code*, 19 J. RISK & UNCERTAINTY 243, 257 (1999).

respondent as "protected" relative to the second good—as qualitatively more valuable than the second.[121] Life and health are often thought of this way, relative to money.[122]

The literature on tradeoff difficulty suggests that respondents to contingent-valuation studies, facing cognitively or emotionally demanding questions about the relative significance of wealth and nonmonetary goods for their personal welfare, will refuse to answer entirely or (less dramatically) will employ various strategies for answering the questions—strategies that will produce stated WTP/WTA amounts that do not reflect their well-informed self-interested preferences for the goods. For example, the respondent may articulate a WTP/WTA amount that expresses the general "importance" of the type of good at issue, not the welfare-significance of the particular quantity involved. He may construct a mental budget that permits a limited expenditure on the good, and refuse to spend more than the budget even where additional expenditures would be worthwhile. Or the respondent may say what a fair price for the good would be, in effect focusing on the cost of supply rather than the personal benefits.[123]

These sorts of tradeoff biases, together with risk biases, help explain the insensitivity of stated WTP to the magnitude of mortality or morbidity risk reduction. They also imply that stated WTP may inaccurately measure health benefits even where risk processing is not an issue. That implication seems to be confirmed by a study conducted by Alan Shiell and Lisa Gold, who asked respondents for WTP for two vaccines to treat two different infectious diseases and for a composite vaccine for both diseases.[124] In a majority of cases, the respondent's WTP for the composite vaccine was less than the sum of WTP for the two individual vaccines. Similar evidence of magnitude-insensitivity apart from risk[125] shows up in other health contingent-valuation studies.[126]

My discussion thus far has focused on the contingent-valuation format and

---

121. *See id.* at 257. On the cognitive and emotional obstacles to trading off different kinds of goods, see also MARY FRANCES LUCE ET AL., EMOTIONAL DECISIONS: TRADEOFF DIFFICULTY AND COPING IN CONSUMER CHOICE 1-10 (2001); and Baron, *supra* note 116, at 83-84.

122. *See, e.g.*, Cass R. Sunstein, *Incommensurability and Valuation in Law*, 92 MICH. L. REV. 779, 834-40 (1994).

123. *See* Baron, *supra* note 116, at 75-77, 79-80, 83-84; Shiell & Gold, *supra* note 118, at 258-60.

124. *See* Shiell & Gold, *supra* note 118, at 253-58.

125. To be sure, risk comes into play in the Shiell and Gold study because the vaccinations would merely reduce the risk of the infectious diseases, but given the study design it is hard to see how difficulties in processing risk would fully explain the magnitude-insensitivity observed.

126. *See* Jan Abel Olsen et al., *The Insensitivity of "Willingness-To-Pay" to the Size of the Good: New Evidence for Health Care*, 25 J. ECON. PSYCHOL. 445 (2004); *id.* at 447 (citing literature); Shiell & Gold, *supra* note 118, at 258-59 (same).

the evidence that risk and tradeoff biases interfere with the measurement of WTP/WTA for mortality and morbidity in that context. To be sure, a variety of so-called "revealed preference," noninterview techniques for estimating WTP/WTA are also widely used by applied economists.[127] Here the idea is to infer valuation from behavior. There is a large empirical literature that seeks to correlate wage differences with differences in occupational risks.[128] A smaller literature uses nonoccupational data to estimate WTP/WTA for mortality risks (for example, consumer purchases of safety devices), or morbidity (for example, using the variation in housing prices to estimate WTP/WTA for air pollution and related morbidity).[129] Revealed-preference techniques may eliminate some biases associated with the contingent-valuation method—for example, the so-called "hypothetical bias," namely that asking respondents what they would be willing to pay or accept in hypothetical choice situations may not elicit sincere and considered statements of the respondents' real preferences[130]—but it is hard to see how the shift from discursive to observational preference-measurement methods would eliminate risk or tradeoff biases. Individuals who process probabilities irrationally, or have cognitive or emotional difficulties trading life or limb for money, should exhibit insensitivity to the magnitude of risk or morbidity reduction in their purchasing or precautionary behavior.[131]

What about QALYs? My suggestion is that risk and tradeoff biases may pose less of a problem for the measurement of QALY values—assuming $\sum_j q(h_{ij}) \times t_j$ is a good estimate of the genuine QALY value of a life, i.e., $Q(H_i)$—than for the measurement of WTP/WTA. Consider first the QALY measurement of pure mortality or mortality risk—that is, measuring the QALY loss for each

127. For an overview of these techniques, see FREEMAN, *supra* note 56, at 95-136.

128. *See, e.g.*, *id.* at 317-19, 401-06; VISCUSI, *supra* note 56, at 34-65.

129. These literatures are discussed in VISCUSI, *supra* note 56, at 65-67; F. Reed Johnson et al., *Valuing Morbidity: An Integration of the Willingness-To-Pay and Health-Status Index Literatures*, 16 J. HEALTH ECON. 641, 644 (1997); and Richard Clemmer et al., *Household Health Production, Property Values, and the Value of Health*, *in* VALUING HEALTH FOR POLICY, *supra* note 1, at 105.

130. *See, e.g.*, Kevin J. Boyle & John C. Bergstrom, *Doubt, Doubts, and Doubters: The Genesis of a New Research Agenda?*, *in* VALUING ENVIRONMENTAL PREFERENCES, *supra* note 111, at 184-86 (discussing concerns about the hypothetical nature of contingent-valuation questions).

131. I am aware of no study that tests magnitude sensitivity in the revealed preference context. It may be difficult to design such a study. *See* E-mail from James K. Hammitt, Professor of Economics and Decision Sciences, Harvard School of Public Health, to Matthew D. Adler, Professor of Law, University of Pennsylvania Law School (Oct. 29, 2004, 11:38:53 EST) (on file with author). For a general discussion of difficulties with revealed-preference studies of WTP/WTA for morbidity and mortality, see U.S. EPA, HUMAN HEALTH METRICS, *supra* note 1, at 18-19; Johnson et al., *supra* note 129, at 644; Hammitt & Graham, *supra* note 113, at 33-34; and Hammitt, *QALYs Versus WTP*, *supra* note 1, at 997-98.

affected individual of a policy which causes each to lose ΔT_i years of healthy life or (more realistically) imposes an incremental r_i risk of losing ΔT_i years of healthy life. In these cases, proportionality to the amount of longevity at stake (ΔT_i) and the risk (r_i) is guaranteed by the additive formula $\sum_j q(h_{i,j}) \times t_j$. *If* $Q(H_i) = \sum_j q(h_{i,j}) \times t_j$ then, in the case of an individual who lives T_i years in perfect health, the QALY value of her life-history is simply T_i. If she loses or gains ΔT_i years, the QALY measure of the change is ΔT_i. If healthy individual *1* loses ΔT_1 years, and healthy individual *2* loses ΔT_2 years, and ΔT_2 is twice ΔT_1, then the QALY value of the second loss is automatically twice the QALY measure of the first. Tradeoff, risk and other biases are, in this limiting case, wholly circumvented because QALY values are derived using the additive formula, not from individual statements of preference. Ditto for the case in which individual *1* incurs an r_1 risk of losing ΔT_1 years of life and individual *2* incurs an r_2 risk of losing ΔT_2 years. The ratio of the expected QALY losses for the *2* individuals is necessarily equal to the ratio of the longevity risks that the individuals face ($(r_2 \times \Delta T_2) / (r_1 \times \Delta T_1)$), regardless of what the individuals might say about the dollar or nondollar value of these risks.

Of course, the QALY method is a tool for measuring morbidity and longevity on an integrated scale, and individual statements of preference are crucial in determining morbidity values—thus the massive efforts undertaken by QALY researchers to survey doctors, patients, or members of the public about their ranking of health states. But even where morbidity comes into play, biases that might induce individual insensitivity to the magnitude of morbidity change are still partly avoided through the additive formula $\sum_j q(h_{i,j}) \times t_j$. Consider a policymaker faced with the options of a temporary or medium-term reduction of some form of air pollution, thereby producing a temporary (one-year, say) or medium-term (five-year) abatement of asthma among certain asthmatics whose symptoms are caused by the pollution. Assume, for the sake of illustration, that the pollution causes asthma only on days when weather patterns lead to a particularly high amount of inhaled pollutant, which occur on average every tenth day. Asthma has a QALY value of 0.683.[132] Then the per-individual QALY value of the medium-term reduction ($0.317 \times 5 \times 0.1 = 0.1585$ QALY) will automatically be five times the per-individual value of the temporary reduction ($0.317 \times 1 \times 0.1 = 0.0317$ QALY). Bracketing income effects, WTP for the medium-term reduction should also be five times WTP for the temporary reduction; but in fact contingent-valuation studies eliciting WTP for longer versus shorter relief from the same symptoms generally fall far short of proportionality.[133]

---

132. *See* Johnson et al., *supra* note 129, at 651.

133. *See* Johnson et al., *supra* note 129, at 650-51 (summarizing results of contingent-valuation

As for the measurement of morbidity values themselves: Risk and tradeoff biases will, to some extent, affect these measurements. The two most widely accepted methods for eliciting morbidity values, as already mentioned, are the standard-gamble (SG) and time-tradeoff (TTO) methods.[134] The SG format, again, seeks to determine respondent's indifference probability p between a given health state, and a lottery with probability p of perfect health and $(1 - p)$ of dying immediately; while the TTO format focuses on the indifference ratio x/y between living for y years in the health state, and x years in perfect health. There is a substantial meta-literature that looks at the practicability, reliability, and validity of the two methods.[135] The methods appear to be practicable as well as "internally" reliable and valid, in the sense that a high proportion of respondents complete the surveys with internally consistent values that are stable over time. Elicited values diverge, to some extent, among respondents,[136] but this itself does not impugn the QALY technique, since (as I have already discussed) convergence in preferences or extended preferences is a contingent matter.[137] Other questions about the validity of SG and TTO values have to do with the decomposability of preferences for lifetime health histories into the sum of the durations of the component health states weighted by valuations of those states—

---

studies of short-term health conditions); Thomas Klose, *The Contingent Valuation Method in Health Care*, 47 HEALTH POL'Y 97, 106 (1999) (discussing magnitude insensitivity in these studies); U.S. EPA, HUMAN HEALTH METRICS, *supra* note 1, at 25 (same). *But see* Richard M. O'Conor et al., *Urge Incontinence: Quality of Life and Patients' Valuation of Symptom Reduction*, 14 PHARMACOECONOMICS 531, 536-37 (1998) (finding that WTP is sensitive to the degree of reduction of urinary incontinence). For a recent survey of the health care contingent valuation literature, see Jan Abel Olsen & Richard D. Smith, *Theory Versus Practice: A Review of 'Willingness-To-Pay' in Health and Health Care*, 10 HEALTH ECON. 39 (2001).

134. Good reviews of the SG, TTO, and other methods employed to elicit QALY valuations of morbidity are: U.S. EPA, HUMAN HEALTH METRICS, *supra* note 1, at 16-18; Baron, *supra* note 116, at 80-82; Brazier et al., *supra* note 1, at 23-56; Dolan, *supra* note 1, at 1732-36; and Johannesson et al., *supra* note 1, at 283-85. The visual analogue scale ("VAS") approach to eliciting QALYs, like SG and TTO, is widely used, *see* Brazier, *supra*, at 24, but it is quite controversial, because of concerns that VAS valuations have no theoretical foundations and are merely ordinal, rather than cardinal measures, *see* Brazier, *supra* at 34-35; Dolan, *supra*, at 1733. Two other methods, magnitude estimation and the person-tradeoff, are less often used. *See* Brazier, *supra*, at 24-27.

135. The meta-literature is exhaustively reviewed in Brazier et al., *supra* note 1, at 30-34, 36-39.

136. *See* Dolan, *supra* note 6, at 150.

137. *See supra* text accompanying notes 73-75. To be sure, the fact that some divergence in QALY values is to be expected, given the possibility that different individuals might have divergent preferences over life-histories, is hardly the end of the story. When survey respondents voice divergent QALY values, which ones should policymakers use? This is an important but difficult question which, along with the related problem of divergent extended preferences, I will not attempt to address in this Article.

an issue I will consider below.[138] Here, the problem is not really that biases are interfering with the measurement of SG and TTO values, but rather that the equivalence of $\sum_j q(h_{i,j}) \times t_j$ and $Q(H_i)$ presupposed by techniques for eliciting SG and TTO values is breaking down.

That said, there is some evidence that biases can interfere with the measurement of SG and TTO values.[139] The SG format involves both risk (since respondents are asked for indifference probabilities) and tradeoffs among different dimensions (since a lottery with a chance of dying is being compared to a health state). Note, however, that the different dimensions are health and longevity, not health or longevity and money. So the problem of trading off "protected" for "unprotected" goods is less pressing—certainly when it comes to trading off longevity for improvements in grave health conditions. Note also that risk biases can be eliminated by shifting to the TTO format.[140] Tradeoff biases remain, here,[141] but respondents do generally voice TTO values that are less than one—indicating that they have traded off some longevity for an improvement in the health state—and voice lower TTO values for more serious conditions. Dolan, in a very large TTO survey of the general public (3395 respondents) in the United Kingdom, found that:

> 46% of respondents were willing to sacrifice life expectancy to avoid *all* of the [13] dysfunctional states they were presented with . . . . A further 29% were willing to sacrifice life expectancy for all but one or two of the states. In such cases, the unwillingness to trade-off time was almost exclusively associated with one or both of the very mild states. In all, 95% of respondents were prepared to sacrifice life expectancy for 6 or more states.[142]

In a different, smaller survey, Dolan looked at each respondent's "consistency rate"—the percentage of pairings of more and less serious conditions in which the respondent gave the more serious condition a lower TTO value—and found that the median respondent had a consistency rate exceeding ninety percent.[143]

---

138. *See infra* Section IV.B.

139. *See* Brazier et al., *supra* note 1, at 30-34, 36-39.

140. It should be emphasized that, *if* the conditions for additive decomposition of preferences over lifetime health histories as per the QALY formula obtain, *then* the TTO and SG methods should produce identical valuations of health states. *See* Johannesson et al., *supra* note 1, at 285; Hammitt, *QALYs Versus WTP*, *supra* note 1, at 995.

141. *See* Brazier, *supra* note 1, at 38-39.

142. Dolan, *supra* note 6, at 149.

143. *See* P. Dolan et al., *Valuing Health States: A Comparison of Methods*, 15 J. HEALTH ECON. 209, 217-20 (1996).

IV. THE LIMITATIONS OF QALYS AS PROXIES FOR LWUS

Although QALYs have certain advantages over WTP/WTA amounts as proxies for LWUs, QALYs are hardly a perfect welfarist measuring rod. This Part discusses their limitations, in particular: violations of the basic presupposition that the health and non-health characteristics of welfare are multiplicatively separable, such that $W(L_i) = Q(H_i) \times V(B_i)$; violations of the further conditions requisite for an additive decomposition of the overall QALY value of a health-profile, $Q(H_i)$, into the sum of the quality-weighted durations of its component health states; the dependence of LWUs on both QALYs and background characteristics; and the related fact that QALYs cannot be used to measure changes in the non-health determinants of well-being, by contrast with the WTP/WTA method, which in principle is applicable to all aspects of welfare.

A. Is Lifetime Utility Separable into Health-Related Subutility and Background Subutility?

QALY values are typically elicited, in QALY surveys, without any discussion of the background characteristics (wealth or other attributes) that the subjects whose health states are being valued should be assumed to have.[144] This procedure is justified, for the welfarist, only if the Miyamato "standard-gamble invariance" condition is satisfied: Namely, respondents (or at least idealized spectators) should have preferences among lotteries over different lifetime health histories that are the same for every fixed level of background characteristics. If this condition holds true, along with the "zero condition" (less contestable), then respondents' utilities for life-histories are expressible as the multiplicative product of a health subutility and a background subutility, and stated QALY values should be the same regardless of the background characteristics, as long as they are held fixed.[145] If the Miyamato "standard-gamble invariance" condition does not hold true, then the current elicitation procedures are problematic. More fundamentally, if $W(L_i)$ is not equal to $Q(H_i) \times V(B_i)$, it is not clear what the QALY value of a life-history means for the welfarist. At a minimum, if the Miyamato condition fails and $W(L_i)$ does not equal $Q(H_i) \times V(B_i)$, the stated valuations that respondents provide in QALY surveys will be less useful to policymakers as proxies for LWUs.

Keeny and Raiffa, in their seminal work on multiattribute utility theory, write that different attributes satisfy "independence" conditions such as standard-

144. *See* Hammitt, How Much Is a QALY Worth?, *supra* note 96, at 5; Mark J. Sculpher & Bernie J. O'Brien, *Income Effects of Reduced Health and Health Effects of Reduced Income: Implications for Health-State Valuation*, 20 MED. DECISION MAKING 207, 209, 211 (2000).

145. *See supra* text accompanying notes 91-98.

gamble invariance "[i]n a surprisingly large number of contexts."[146] I have suggested that, intuitively, the Miyamato condition is at least approximately true of health versus non-health attributes. That suggestion is, in effect, just a guess about what spectators' preferences would be. The guess is not particularly well-informed. There appears to be very little empirical scholarship, at least in the QALY field, examining the structure of preferences over combinations of health profiles and background profiles.[147]

QALY scholars have examined a related issue: What predicts variations among respondents in QALY values? "[M]ost studies show[] that the values are independent from socio-economic factors or professional level."[148] This finding might be adduced as evidence for the standard-gamble invariance condition. If (1) preferences for lotteries over health characteristics do depend on the level of background characteristics, and (2) respondents to QALY surveys assume, absent instruction, that the health states being valued are packaged with the respondents' actual background characteristics, it would follow that (3) variation in QALY values would correlate with variation in respondents' background characteristics—which has not been observed. However, the finding of non-correlation is at best circumstantial evidence for the standard-gamble invariance condition.[149] More direct testing remains to be done.[150]

B. Can QALYs Be Decomposed as per the Standard Additive Formula?

A different issue concerns the additive decomposition of $Q(H_i)$ into $\sum_j q(h_{i,j}) \times t_j$, where $h_{i,j}$ is the health of individual i in period j and t_j is the duration of period j. The vast bulk of the QALY literature assumes that the health value of a lifetime health-history or "profile" can, at least approximately, be represented as the sum of years in its component health states adjusted by values for those

146. KEENY & RAIFFA, *supra* note 92, at 226.

147. *See* Hammitt, How Much Is a QALY Worth?, *supra* note 96, at 5 ("The literature on QALYs is virtually silent on the extent to which [valuations of health states] depend[] on wealth, income, or consumption.").

148. U.S. EPA, HUMAN HEALTH METRICS, *supra* note 1, at 21; *see also* Dolan, *supra* note 1, at 1747; Paul Dolan & Jennifer Roberts, *To What Extent Can We Explain Time Trade-Off Values from Other Information About Respondents?*, 54 SOC. SCI. & MED. 919, 927-28 (2002).

149. For example, it may be that respondents assume that the subject's background characteristics are population-average characteristics, rather than the respondents' own characteristics. In this case we would not expect QALY values to vary depending with respondents' background characteristics even if the Miyamato condition fails.

150. For an unusual example of a direct test of the independence of valuations of health and background characteristics, see Antonio Ciampi et al., *Measurement of Individual Preferences: The Importance of "Situation-Specific" Variables*, 2 MED. DECISION MAKING 483 (1982).

health states.[151] Cost-effectiveness studies use this additive formula for determining the effectiveness of different interventions; and typical QALY surveys, which attempt to elicit SG or TTO values for states rather than whole histories, assume additivity as well. But additivity is *not* entailed by the more basic conditions—the Miyamato conditions—that allow the LWU assigned to each life-history to be represented as the product of an overall health subutility for its component health history, $Q(H_i)$, and an overall non-health subutility for its component background characteristics, $V(B_i)$. Rather, theoretical work on QALYs has demonstrated that three further conditions that are "internal" to the valuation of health are required for this overall health utility $Q(H_i)$ to be additively decomposable into $\sum_j q(h_{i,j}) \times t_j$: (1) No discounting of future health; (2) risk neutrality with respect to longevity; and (3) no sequencing effects.[152]

The no-discounting condition is the least problematic. No discounting means that the spectator's present valuation of a life history in which he incurs a given health state does not vary depending on whether that state occurs in the near or more distant future. Although actual individuals often exhibit a time preference,[153] preferring present to future pleasures and future to present pains, and although policymakers certainly might use discounting to reflect the opportunity cost of current expenditures given the alternative of investing the resources in inter-temporal markets, a strong case can be made that the absence of an intrinsic time preference is normative.[154] Idealized spectators, one might suppose, would not give greater weight to the temporally proximate aspects of the lives they might lead just because of the proximity. In any event, if the risk-neutrality and no-sequencing conditions hold true, the formula $\sum_j q(h_{i,j}) \times t_j$ can be adjusted to incorporate a temporal discount factor.[155]

---

151. To be sure, the HYE approach drops this assumption; but HYEs have not been used much in practice, *see* Hammitt, *QALYs Versus WTP*, *supra* note 1, at 989.

152. For a discussion of these conditions or equivalent ones, see generally Han Bleichrodt, *QALYs and HYEs: Under What Conditions Are They Equivalent?*, 14 J. HEALTH ECON. 17, 20-25 (1995); Han Bleichrodt et al., *Characterizing QALYs by Risk Neutrality*, 15 J. RISK & UNCERTAINTY 107 (1997); Dolan, *supra* note 1, at 1729-31,1740-43; Hammitt, *QALYs Versus WTP*, *supra* note 1, at 986-88; Magnus Johannesson et al., *A Note on QALYs, Time Tradeoff, and Discounting*, 14 MED. DECISION MAKING 188 (1994); Johannesson et al., *supra* note 1, at 285-86; Miyamato et al., *supra* note 91, at 839-45; and Joseph S. Pliskin et al., *Utility Functions for Life Years and Health Status*, 28 OPERATIONS RES. 206, 207-15 (1980).

153. *See, e.g.*, Bleichrodt et al., *supra* note 152, at 110. Surprisingly, actual individuals may also exhibit a negative time preference. *See* Dolan, *supra* note 1, at 1742.

154. *See* David O. Brink, *Prudence and Authenticity: Intrapersonal Conflicts of Value*, 112 PHIL. REV. 215 (2003).

155. *See, e.g.*, Johannesson et al., *A Note on QALYs*, *supra* note 152. Discounting is not consistent with risk neutrality over life years, but it is consistent with risk neutrality over

Risk-neutrality and no sequencing are more serious difficulties for this formula. The risk-neutrality condition, here, means that longevity has constant rather than increasing or decreasing marginal health utility. Formally, for any health profile consisting of a constant health state, risk neutrality requires that a given increment in longevity produce the same increase in the overall $Q(H_i)$ value of the health profile regardless of the baseline longevity. For example, doubling the lifespan of an individual with chronic bronchitis from thirty-five to seventy years doubles the $Q(H_i)$ value. It is obvious that risk-neutrality in this sense is presupposed by the $\sum_j q(h_{i,j}) \times t_j$ formula, which makes estimated QALY values a linear function of longevity, holding health constant. But $Q(H_i)$ need not be a linear function of longevity, holding health constant, if the idealized spectators whose preferences ground LWU and $Q(H_i)$ values depart from risk-neutrality.

Would they? Empirical tests find that actual respondents are not always risk-neutral with respect to longevity.[156] Nor can it be said here—by contrast with discounting—that risk-neutrality with respect to longevity is normative. If money can have a declining marginal impact on welfare, then presumably so can lifespan. One solution, proposed in the literature, is to adjust the longevity component of a health history by a risk-aversion factor. In the case of a chronic health state h_i*, $Q(H_i)$ would equal the value of that state $q(h_i$*$)$ multiplied by the longevity raised to a risk-aversion factor.[157] But it is unclear how one might extend this formula to accommodate risk aversion in the more realistic case of health profiles where health varies over time.[158] Further, even in the case of unvarying health profiles, there are empirically documented instances of preferences that cannot be represented by either the ordinary QALY formula *or* the risk-adjusted version: namely the preference to live more than zero time, but less than an unbearable amount of time, in a painful health state.[159] Once more, it is hard to see why general considerations of rationality would preclude spectators from having such preferences. In short, departures from risk neutrality can create

---

discounted life years. *See id.*

156. Risk neutrality can be tested directly (by asking for the number of years in a health state that the respondent views as equivalent to a gamble over the chance of immediate death and longer duration in the health state) or indirectly, by testing the other conditions that are implied by risk neutrality plus the zero condition. *See* Bleichrodt et al., *supra* note 152, at 112-13. Both sorts of test show that risk neutrality can fail. *See* Dolan, *supra* note 1, at 1740-42.

157. *See* Johannesson et al., *supra* note 152, at 188-90; Miyamato et al., *supra* note 91, at 842-45.

158. *See* Johannesson et al., *supra* note 1, at 285.

159. *See, e.g.*, Miyamato et al., *supra* note 91, at 844-45; Paul Dolan & Peep Stalmeier, *The Validity of Time Trade-Off Values in Calculating QALYs: Constant Proportional Time Trade-Off Versus the Proportional Heuristic*, 22 J. HEALTH ECON. 445 (2003).

a divergence between the $\sum_j q(h_{i,j}) \times t_j$ formula and the genuine QALY value of a life $Q(H_i)$, and there appear to be no simple modifications to the QALY formula to circumvent the problem.

The same is true of sequencing. In a few studies, researchers have tested the no-sequencing condition directly by asking respondents to value sequences of states, and comparing the values with the sum of QALY values for the states involved.[160] Other research employs more indirect tests.[161] The literature is small and conclusions are mixed, but sequencing does emerge in some cases. For example, Richardson asked women to value separately three differentially serious breast cancer states, then a deteriorating holistic scenario composed of a progression from less to more serious, and found that the values could not be reconciled even allowing for discounting. A better explanation, he suggests, is that the "knowledge of future suffering and death casts a shadow over, or devalues, the enjoyment of earlier life years."[162] Note how this coheres with the philosopher David Velleman's suggestion that a life history where momentary welfare has a deteriorating trajectory is worse than a counterpart life history with the same overall sum of momentary welfare but an increasing trajectory.[163]

C. Background Characteristics Redux

Even if LWUs can be decomposed into the product of health and non-health subutility, and even if health subutility can in turn be decomposed as per the additive QALY formula into the cumulative time in different health states adjusted for the value of those states, QALYs are not generally a perfect proxy for LWUs. The problem of background characteristics reemerges here. I demonstrated earlier that QALY aggregation does perfectly track overall well-being where all the lives affected by the policy choices being considered fall in the same background-equivalent class. In general, of course, that need not be true. If life history L_1 has a higher level of consumption, or leisure, or socializing, or professional accomplishment, as compared to L_2, then a change in L_1's health history has a greater impact on overall well-being than the very same change (as measured in QALY units) in L_2's health history. This follows immediately from the multiplicative representation of LWUs implied by the Miyamato conditions, i.e., the standard-gamble invariance and zero conditions.[164] If those conditions

160. *See* Kuppermann et al., *supra* note 69; Richardson et al., *supra* note 69.

161. *See* Krabbe et al., *supra* note 69; Spencer, *supra* note 69; Treadwell, *supra* note 69.

162. Richardson et al., *supra* note 69, at 157.

163. *See supra* text accompanying note 84.

164. Bleichrodt and Quiggin make a similar point in the context of their own model, which (like mine) has a multiplicative structure deriving from standard-gamble invariance and the zero condition:

obtain and the multiplicative model therefore does indeed accurately represent LWUs, health *cannot* generally have equal welfare value for different persons; rather, its value is necessarily scaled up or down by the value of the non-health attributes, $V(B_i)$, with which health partly interacts.

Rejecting the standard-gamble invariance condition, which would mean that $W(L_i)$ need not equal $Q(H_i) \times V(B_i)$, will not solve the problem. In that event, LWUs will neither be additively nor multiplicatively separable into health and non-health subutilities, and health and non-health attributes would continue to interact, albeit in a more complicated way. Rejecting the zero condition and adding other conditions to ensure the additive separability of health and non-health attributes *would* ensure that equal QALYs have equal welfare value across persons. An additive decomposition, here, would mean that $W(L_i) = Q(H_i) + V(B_i)$, precluding any interaction between health and non-health attributes. But it is implausible, given our sense of what well-informed individuals can prefer, that spectators' preferences would separate health and background characteristics so completely. The additive form would mean that the marginal utility of consumption cannot depend on health, nor vice versa.[165]

Within the context of the multiplicative model, $W(L_i) = Q(H_i) \times V(B_i)$, is there any way to limit the scaling effect and make QALYs a reasonable proxy for LWUs even where all the lives involved do not fall in the same background-equivalence class? One way to do that might be to expand the definition of health. If health merely subsumes the physical condition of the subject, excluding his hedonic, emotional, or cognitive state, then hedonic, emotional, and cognitive attributes become background characteristics. QALYs, as a measure of health thus narrowly defined, would be a poorer proxy for welfare than if health were defined more inclusively—since a narrower definition of health, or equivalently a broader definition of background characteristics, implies a smaller number of choice situations in which the lives involved fall in the same background-equivalent class, and presumably a greater average range of the scaling factor $V(B_i)$ in other choice situations.

Fortunately, health for QALY purposes is not normally defined so narrowly. QALY surveys are often conducted using so-called "health state classification systems," which seek to regiment the evaluation task by describing health states

---

> [In our model] the utility of health status is multiplied by the utility of consumption. Consequently, a given gain in quality of life will be more appreciated at higher levels of consumption. This implies that in the allocation of health care resources, larger welfare gains can be obtained by devoting resources to those individuals who have a high level of general consumption. . . . This result is ethically troubling . . . . However, the need for a multiplicative utility structure shows that . . . such implications cannot be escaped.

Bleichrodt & Quiggin, *supra* note 96, at 685.

165. *See supra* text accompanying notes 93-94 (discussing additive decomposition).

as packages of health attributes.[166] Respondents are then asked to use the SG, TTO, or some other method to place the packages on a 0-1 scale, with 0 meaning death and 1 meaning the very best package. For example, the Health Utilities Index, one of the most widely used health state classification systems,[167] conceptualizes health states as a combination of vision, hearing, speech, ambulation, dexterity, emotion, cognition, and pain. Each of these eight attributes has five or six levels. The very best health state in the HUI system, i.e., the state with the QALY value of 1, is a state where the subject is at the best level with respect to all eight attributes. He is able to see well enough to read newsprint and recognize a friend across the street without glasses (vision); is able to hear what is said in a group conversation with at least three other people, without a hearing aid (hearing); is able to be understood completely when speaking with strangers or friends (speech); is able to walk around the neighborhood without difficulty and without walking equipment (ambulation); has the full use of two hands and ten fingers (dexterity); is happy and interested in life (emotion); is able to remember most things, think clearly, and solve day-to-day problems (cognition); and is free of pain and discomfort (pain).[168] The crucial point, for our purposes, is that the HUI system does have a more inclusive definition of health than the merely physical; and this is true of the other health classification systems as well, as Table 1 shows.

166. *See* Dolan, *supra* note 1, at 1731-32, 1744-45 (discussing generally health state classification systems); QUALITY OF LIFE AND PHARMACOECONOMICS, *supra* note 1, at 161-362 (describing specific systems in detail).

167. *See* Chaim Bell et al., *An Off-the-Shelf Help List: A Comprehensive Catalog of Preference Scores from Published Cost-Utility Analyses*, 21 MED. DECISION MAKING 288, 290 (2001) (finding that the Rosser Index, Quality of Well-Being Scale, and HUI are the most widely used health-state classification systems in published cost-utility analyses).

168. *See* David H. Feeny et al., *Health Utilities Index*, *in* QUALITY OF LIFE AND PHARMACOECONOMICS, *supra* note 1, at 239 (describing the HUI system). The HUI system has been updated several times, and the version described in the text is the most recent one.

TABLE 1: PRINCIPAL CONCEPTS COVERED IN EXISTING HEALTH CLASSIFICATION SYSTEMS[169]

Concept	DDR	EQL	15D	HUI_1	HUI_2	HUI_3	IHQOL	HP2000	QWB
Health Perceptions			X					X	
Social Function		X		X			X		X
Speech/ Communication			X		X				
Mental Function (Cognitive, Emotional, and/or Affective)		X	X	X	X	X	X		
Physical Function (Mobility, Physical Activity, and/or Self-Care)	X	X	X	X	X	X	X	X	X
Sensory Function or Other Impairments	X	X	X	X	X	X	X		X

Still, there are limits to the inclusiveness of QALYs. The QALY health-classification systems do not define health nearly as inclusively as the World Health Organization (WHO), which has developed a conception of the "quality of life" subsuming twenty-four attributes grouped into six domains—physical, psychological, "independence," social, environmental, and spiritual.[170] The first three domains cover the territory of QALY health-classification systems, but also include a self-esteem attribute, a body-image attribute, and an attribute for positive as well as negative feelings—all attributes that the QALY systems, and traditional QALY research, treat as background characteristics. The same is true

169. DDR = Disability Distress Ratio; EQL = EuroQOL; 15D = Fifteen Dimension Scale; HUI = Health Utility Index Mark I, II, or III; IHQOL = Index of Health-Related Quality of Life; HP2000 = Years of Healthy Life; QWB = Quality of Well-Being Scale. This is based on a table in Donald L. Patrick & Pennifer Erickson, *Applications of Health Status Assessment to Health Policy*, *in* QUALITY OF LIFE AND PHARMACOECONOMICS, *supra* note 1, at 721. *See also* COST-EFFECTIVENESS IN HEALTH AND MEDICINE, *supra* note 12, at 95 (listing "core concepts and domains of health-related quality of life," including health perceptions, social function, psychological function, physical function, and impairment).

170. On the development of the WHOQOL, see The WHOQOL Group, *supra* note 90; and Adler, *Fear Assessment*, *supra* note 54, at 1051 n.197.

of the attributes covered by the last three WHO domains. The "social" domain asks about the quality of the subject's personal relationships, social support, and sex life. The "environment" domain covers personal security, housing quality, wealth, access to information and education, access to social services, recreational activities, pollution, and transport. The spiritual domain asks about the perceived meaningfulness of the subject's life. Traditional QALY research does not see these domains as constitutive of "health."

In fact, there is very little that seems relevant to well-being which is not covered by one of the WHO quality of life domains. "Quality of life" ends up being a synonym for "welfare," and not merely "health," normally understood as a proper subset of the welfare-affecting features of life-histories. But why not scrap or redefine the concept of "health," for QALY purposes, so that the set of background characteristics shrinks to zero? Respondents would be asked to use the SG or TTO formula to place states on a 0-1 scale, where 0 is death and 1 is an ideal state not just with respect to physical, emotional, cognitive, and hedonic attributes, but also with respect to body image, self-esteem, personal relationships, social support, sex life, personal security, housing quality, and everything else that makes life worth living.

There are some apparent difficulties with making QALYs this inclusive. Many if not all respondents would presumably find it cognitively overwhelming to make cardinal comparisons (using TTOs or SGs) among states that vary with respect to the totality of welfare-relevant attributes. Current QALY surveys hold fixed most attributes, the non-health ones, and give respondents the more manageable task of making cardinal comparisons of states that vary only with respect to health attributes. Relatedly, QALY valuations elicited from many (if not all) respondents relative to a maximally inclusive QALY scale would be less sensitive to small changes than traditional QALYs.[171] States that are not radically different with respect to the physical, cognitive, emotional, and hedonic attributes subsumed by traditional QALY measures, and do not differ at all with respect to other attributes, might be lumped together and assigned the same number on a maximally inclusive QALY scale, but differentiated by traditional QALYs. For example, a recent synthesis of traditional QALY research suggests that 0.75 is a plausible QALY score for angina, and 0.81 for pancreatitis.[172] If 0 is death and 1

171. *See generally* Brazier et al., *supra* note 1, at 13 (discussing desirable properties of health valuation instruments, including "responsiveness," which is defined as "the ability of an instrument to measure clinically significant changes in health"). *See also* Cam Donaldson et al., *Should QALYs Be Programme-Specific?*, 7 HEALTH ECON. 239 (1988) (suggesting that program-specific health scales may be more sensitive than general scales).

172. *Cf.* Harvard Center for Risk Analysis, Cost-Effectiveness Analysis Registry, http://www.hpsh.harvard.edu/cearegistry (database of cost-utility analyses, including 0.75 as one QALY valuation for angina and 0.808 as the valuation of recurrent pancreatitis).

is a state that is ideal with respect to all the dimensions of welfare, would the typical respondent using the TTO or SG method be able to differentiate numerically between an ideal welfare state marred only by angina and an ideal welfare state marred only by pancreatitis, assigning the first (say) a score of 0.92 and the latter a score of 0.93?[173] It is quite plausible that he would not.

Given these measurement problems, the optimally inclusive QALY scale for purposes of standard surveys eventuating in policy evaluation might well be less than fully inclusive, and exclude some of the attributes that the WHO conception, philosophers of welfare, and ordinary intuition see as relevant to welfare. In any event, existing QALY surveys are much less than fully inclusive, and policy evaluation using these valuations must be sensitive to the fact that they are less than perfect proxies for LWUs, given background characteristics.

D. Non-Health Changes

The final difficulty with the QALY scale is intimately related to the one just described. If some determinants of welfare fall outside the domain of "health," then some welfare changes will not be measurable using QALYs (equivalently, the measure of the change will be zero). Changes in consumption provide the most practically compelling example. These sorts of welfare effects are the heartland of standard cost-benefit analysis, and are pervasively counted as costs or benefits in actual policy-evaluation practice, but do not register, as such, on the traditional QALY scale.[174] Extravagant consumption plus angina has the same QALY score, *ceteris paribus*, as moderate consumption plus angina—given a QALY scale that counts consumption and wealth as background attributes. A different example comes from environmental regulation. Environmental economists and, increasingly, agencies such as the EPA incorporate the following sorts of nonmarket benefits into cost-benefit studies: the enjoyment experienced by visitors to parks or other protected areas, the recreational benefits of hunting and fishing, the improved visibility that accompanies better air quality, smell- or noise-avoidance, the "scenic" benefit of viewing a nice landscape, and the sheer satisfaction of knowing that a site, ecosystem, or species exists.[175] All these effects will show up in WTP/WTA amounts, as elicited using contingent-valuation or revealed preference techniques, but none fall within the domain of

173. This calculation assumes that health is one-third of overall welfare and rounds valuations to two digits. If health is less important to overall welfare, the problem is exacerbated.

174. *See* Hammitt, How Much Is a QALY Worth?, *supra* note 96, at 5; Schulpher & O'Brien, *supra* note 144, at 214. *But see* COST-EFFECTIVENESS IN HEALTH AND MEDICINE, *supra* note 12, at 122 (recommending that financial consequences related to health status be measured on a QALY scale and be included in the denominator, rather than the numerator, of cost-effectiveness ratios).

175. *See* Adler, *Fear Assessment*, *supra* note 54, at 981-82.

health, at least as defined by the traditional QALY health-classification systems.[176]

V. QALYs, NONTRADITIONAL COST-BENEFIT ANALYSIS, AND WELFARIST POLICY EVALUATION

What role should QALYs play in welfarist policy evaluation, given both their advantages relative to a WTP/WTA metric and their limitations? This Part answers the question in two different ways. First, I sketch a theoretically appealing, but currently unrealistic, decisional approach where LWUs themselves serve as the policy metric, and QALY surveys, WTP/WTA surveys, revealed preference data, and other information is used by the decision-maker in a Bayesian fashion to "update" her estimate of LWUs. Second, I describe an approach that is less elegant but more realistic, given current practices: a nontraditional kind of cost-benefit analysis that incorporates dollar amounts derived from QALY measurements, not merely WTP/WTA amounts, to measure health effects. The FDA has, over the last half-decade or so, followed precisely this approach in some of its important rulemakings—converting QALYs to dollars using a conversion factor ranging from $100,000 to $500,000, and then plugging those dollar amounts into a monetized cost-benefit analysis.

The FDA, here, is traversing the frontiers of policy science, and in a way that (I believe) constitutes a true advance. This Part will argue that the FDA's nontraditional cost-benefit procedure can improve on traditional policy analysis, at least in some choice situations, and will analyze, in a preliminary way, how the procedure is optimally structured. At what rate should QALYs be converted to dollars? When should QALYs be substituted for traditional WTP/WTA measurements of health effects? I conclude by discussing a concrete case where the applicability of traditional cost-benefit analysis has been hotly contested and where QALY-based cost-benefit analysis may well be an improvement: the use of WTP/WTA amounts to measure the cost of premature death.

A. First-Best Policy Evaluation: LWU Maximization with Some Help from Bayes

I have argued, to this point, that overall welfare is relevant to policy evaluation; that overall welfare is best understood in terms of the convergent preferences of idealized spectators contemplating extended lotteries; and that LWUs, a numerical scale of these preferences, constitute the welfarist gold-standard. One outcome has greater overall well-being than a second just in case the sum of LWUs is greater in the first case. Correctly measured WTP/WTA

176. *See* Klose, *supra* note 133, at 115 (noting that WTP/WTA amounts, by contrast with QALYs, "provide[] a more comprehensive measure of the effects of a health care technology").

amounts are not perfect proxies for LWUs, and in any event the WTP/WTA valuations revealed by behavior or expressed in surveys diverge from true WTP/WTA amounts. Correctly measured QALYs, too, are not generally perfect proxies for LWUs, given the problem of background characteristics; further, the additive formula for estimating QALYs will be inaccurate, if the conditions for additive decomposition break down; and cognitive failures may interfere with QALY surveys, just as they interfere with the measurement of WTP/WTA amounts.

What does this analysis imply for welfarist policy evaluation? It suggests that—placing to one side problems of deliberation costs, including computational limits—the optimal welfarist procedure would use neither dollar amounts, nor QALY amounts, but rather the LWU scale itself to evaluate different policy options. The proposal may seem outlandish—but remember that measurement techniques now intimately familiar to us, such as monetized cost-benefit analysis, are themselves quite new on the scale of human time, invented only one or two generations ago.[177] The science of measurement evolves, and my notion of an LWU scale is in fact well-grounded in current measurement theory, representing the application of utility theory—a theory very widely accepted by contemporary welfare economists—to a particular set of preferences, the spectators' preferences over life-histories.

To be sure, the spectators are idealized, and this raises a large problem of uncertainty. Who knows what life-histories, and lotteries over life-histories, we would favor, if we were fully informed, fully deliberative about the different life-histories and lotteries being considered, and unhampered by errors in cognition or judgment? Humans' actual preferences for life-histories and lotteries over life-histories, as evidenced by their verbal or physical behavior, are only very imperfect evidence of their idealized preferences. But policy-analytic techniques for handling uncertainty continue to develop apace, just as measurement techniques do.[178] The best, general approach to uncertainty is the "Bayesian" approach. Given some item of interest whose numerical value is uncertain, the policymaker starts with a "prior" probability distribution over possible numerical values of the item, and then updates her probability distribution as new information arrives.[179] Practicable, statistical techniques for implementing the

177. *See* Adler & Posner, *Rethinking Cost-Benefit Analysis*, *supra* note 22, at 167-76 (summarizing the history of cost-benefit analysis).

178. *See generally* M. GRANGER MORGAN & MAX HENRION, UNCERTAINTY: A GUIDE TO DEALING WITH UNCERTAINTY IN QUANTITATIVE RISK AND POLICY ANALYSIS (1990).

179. On Bayesian approaches to probability generally, see Matthew D. Adler, *Risk, Death and Harm: The Normative Foundations of Risk Regulation*, 87 MINN. L. REV. 1293, 1312 n.73 (2003) (citing sources). On Bayesian statistics, see, e.g., JOSE M. BERNARDO & ADRIAN F.M. SMITH, BAYESIAN THEORY (1994).

Bayesian idea are now available,[180] and these are increasingly employed by policy analysts. For example, such techniques are employed to determine the optimal design of nuclear reactors or other large, dangerous structures in the teeth of uncertainty about the external stresses that the structures will be subjected to and the processes that will occur inside them;[181] or to predict the number of deaths that will result from pollution or other environmental releases of toxic chemicals, despite fairly deep uncertainty about dose-response relationships, the environmental "fate and transport" of toxins, and demographic patterns over time.[182]

I suggest that Bayesian techniques could, in principle, be used to estimate LWUs. For a given type of welfare impact (a risk of death, a headache, a scenic view, angina, and so on), a prior probability distribution over the LWU values of that impact would be defined. That distribution might be quite "diffuse"—the analyst might have a very poor initial sense of how idealized spectators would value the impact—or her priors might be less diffuse, incorporating for example her own intuitions about welfare, or philosophical wisdom, or her knowledge about the psychological processes of preference formation (all supporting probabilistic guesses about what the spectators would want).[183] QALY surveys, contingent-valuation surveys, revealed-preference studies, as well as LWU surveys[184] focused directly on measuring the LWU value of the impact could then

---

180. *See, e.g.*, ANDREW GELMAN ET AL., BAYESIAN DATA ANALYSIS (2d ed. 2004).

181. *See, e.g.*, ROGER M. COOKE, EXPERTS IN UNCERTAINTY: OPINION AND SUBJECTIVE PROBABILITY IN SCIENCE 27-41 (1991) (discussing use of Bayesian approaches in analyzing nuclear reactor safety).

182. *See* Adler, *Against "Individual Risk," supra* note 112, at 1208-10 (discussing Bayesian approaches to risk assessment).

183. *See, e.g.*, GELMAN ET AL., *supra* note 180, at 33-72 (discussing informative and noninformative priors).

184. LWU studies would seek to determine respondents' utilities for whole life-histories. Unlike QALY surveys, these studies would characterize life-histories in terms of background characteristics as well as health. Unlike contingent-valuation or revealed-preference studies they would seek to measure the respondent's preferences among different life-histories on a utility scale, not a dollar scale. I see nothing to preclude this sort of LWU study, and indeed it would not be radically different from some existing survey work (for example, the surveys in the QALY literature that ask about preferences over entire health-histories rather than assuming temporal decomposition, see *supra* note 69 and accompanying text; the survey work leading up to WHO's inclusive quality-of-life index, see *supra* note 170 and accompanying text; and the survey work looking to preferences among different life-saving programs, see *infra* text accompanying notes 238-240). To be sure, as suggested earlier, see *supra* text accompanying notes 171-172, the ordinary respondent might find it more difficult to complete an LWU survey than a standard QALY survey. A successful LWU survey might therefore need to incorporate cognitive aids (for example, visual aids to help the respondents grasp the different life-histories being compared; tutorials in the

be used to "update" the analyst's priors—just as, for example, the Bayesian scientist's prior beliefs about the toxicity of a given chemical compound, grounded in the existing scientific literature, are updated by the results of a particular experiment, such as feeding the chemical to a group of rats and seeing how many die.[185]

QALY studies, contingent-valuation studies, revealed-preference data, and LWU studies themselves—"observational" evidence analogous, for the Bayesian, to the rat experiment—will have a substantial effect in changing the analyst's priors if the individuals who are surveyed, or whose behaviors are examined, are particularly well-informed and deliberative. If the individuals' deliberational and informational characteristics are less proximate to the spectators' characteristics, then the studies and data may have a less dramatic impact on the analyst's probability distribution. In either event the analyst will integrate the sources of information available to her to arrive at a probability distribution over LWUs.

The crucial point, here, is that LWUs need not be known for certain for them to figure in policy choice. Uncertainty about LWUs can, in principle, be expressed probabilistically and updated systematically with observations about what people do or say, and the welfarist rule for policy analysis could be: Pick the alternative with the greatest expected LWUs.

I will not elaborate on the approach just sketched, which asks analysts and decision-makers to use the LWU scale itself as their basic decisional tool, rather than QALYs, WTP/WTA amounts, or some other proxy measure. First, once deliberation costs and the problems of administrative error and opportunism are taken into consideration, the approach may be welfare-suboptimal[186]—although it should be pointed out that increases in computing power have made Bayesian techniques feasible in a much wider range of contexts, and much cheaper, than a generation ago.[187] Second, my primary aim here is to inform current welfarist policy-evaluation practices. Cost-benefit analysis is the dominant technique, at least at federal agencies, and describing a different, nonmonetized, LWU-based technique is not of much use to analysts constrained by current practices,

---

axioms of rational choice; interviewer interventions to point out inconsistencies in valuations), or perhaps to be limited to particularly able respondents.

185. *See, e.g.*, Ryan A. Hill, *From Science to Decision-Making: The Applicability of Bayesian Methods to Risk Assessment*, 2 HUM. & ECOLOGICAL RISK ASSESSMENT 636 (1996) (discussing use of new data to update prior beliefs about the carcinogenicity of a chemical).

186. *See* ADLER & POSNER, COST-BENEFIT ANALYSIS, *supra* note 22, chs. 3, 4 (arguing that welfare-maximizing procedures are sensitive to deliberation costs and decision-maker error and opportunism); Adler & Posner, *Rethinking Cost-Benefit Analysis*, *supra* note 22, at 217-18 (same).

187. *See, e.g.*, GELMAN ET AL., *supra* note 180, at xx (noting that increases in computing speed and improvements in computational methods have made Bayesian methods feasible for more complicated models and larger datasets).

institutional routines, and expectations.

B. Second-Best Policy Evaluation: Nontraditional Cost-Benefit Analysis

Current welfarist policy-analytic practices are proxy-based. The best-developed measures of welfare are money and QALYs, neither of which directly represent spectators' preferences over complete life-histories, as an LWU scale would. Further, because money has a much broader range than QALYs, subsuming not just health impacts but all manner of non-health effects,[188] the fact that regulatory agencies generally use a money rather than QALY scale to implement the welfarist mandates of presidential cost-benefit orders and of statutes that require cost-benefit analysis is not surprising.[189] This Section will show how QALYs can play a role within, and improve on, monetized cost-benefit analysis.

Monetizing QALYs and incorporating them into cost-benefit analysis is not a technique much discussed by the scholarly literature on QALYs (where QALYs are generally seen as an outcome measure for cost-effectiveness analysis[190]), or by the cost-benefit literature (where Kaldor-Hicks efficiency, not overall welfare, is often taken as the gold standard[191]). Here, practice outruns scholarship—for, as already mentioned, federal agencies have recently started to convert QALY measures of various health impacts to dollar amounts and to add those sums to WTP/WTA measures of non-health effects in evaluating policy choices. For short, I will call this approach "hybrid" or nontraditional cost-benefit analysis.

The leader in this area is the FDA, which to date has used hybrid cost-benefit analysis in almost twenty rulemakings.[192] The practice apparently began

---

188. *See supra* Section IV.D.

189. *See* Adler & Posner, *Rethinking Cost-Benefit Analysis*, *supra* note 22, at 167-76 (describing use of monetized cost-benefit analysis by regulators); *supra* text accompanying notes 8-20 (noting infrequent use of QALYs by regulators, with the exception of the FDA).

190. *See supra* text accompanying notes 25-30.

191. *See, e.g.*, Richard Revesz & Robert N. Stavins, *Environmental Law and Policy*, *in* THE HANDBOOK OF LAW AND ECONOMICS 5 (A. Mitchell Polinsky & Steven Shavell eds., forthcoming 2006) ("The general view from economics is that other criteria in addition to efficiency [in particular, distributive criteria] can and should be employed by policymakers, but that the existence of such criteria does not invalidate the efficiency criterion, which should remain part of social decision-making."); sources cited *supra* notes 49-53.

192. *See infra* notes 194-195. This count, to be conservative, excludes the handful of rulemakings in which the FDA has used VSLYs. As mentioned earlier, the EPA has also flirted with VSLYs. *See supra* text accompanying notes 14-15; *see also* Cass R. Sunstein, *Lives, Life-Years, and Willingness To Pay*, 104 COLUM. L. REV. 205, 252 (2004) (listing regulatory impact

at the FDA in the early 1990s,[193] was employed in the huge tobacco rulemaking a decade ago,[194] and has accelerated since the late 1990s.[195] To give one illustrative

---

statements using life-years or QALYs).

193. In a 1993 rulemaking concerning food standards of identity, the FDA appeared to use the VSLY approach as one measure of mortality costs. *See* Food Standards: Amendment of the Standards of Identity for Enriched Grain Products To Require Addition of Folic Acid, 58 Fed. Reg. 53,305, 53,310 (Oct. 14, 1993). As discussed below, *see infra* text accompanying notes 233-234, this approach is identical to the use of QALY-to-dollar conversions to value mortality if the VSLYs are not age-adjusted or adjusted for other characteristics such as wealth, and if it is assumed the years lost are perfectly healthy.

194. *See* Regulations Restricting the Sale and Distribution of Cigarettes and Smokeless Tobacco To Protect Children and Adolescents, 61 Fed. Reg. 44,396, 44,576 (Aug. 28, 1996).

195. The FDA has employed QALY-to-dollar conversions to monetize morbidity or mortality for purposes of cost-benefit analysis in the following rulemakings. (Where both final and proposed rulemakings employed conversions, only the final rulemaking is cited.) Performance Standard for Diagnostic X-Ray Systems and Their Major Components, 70 Fed. Reg. 33,998, 34,019-22 (June 10, 2005); Establishment and Maintenance of Records Under the Public Health Security and Bioterrorism Preparedness and Response Act of 2002, 69 Fed. Reg. 71,562, 71,621-22 (Dec. 9, 2004); Prevention of Salmonella Enteritidis in Shell Eggs, 69 Fed. Reg. 56,824, 56,853-55 (Sept. 22, 2004); Requirements Pertaining to Sampling Services and Private Laboratories Used in Connection with Imported Food, 69 Fed. Reg. 23,460, 23,469 (Apr. 29, 2004); Bar Code Label Requirement for Human Drug Products and Biological Products, 69 Fed. Reg. 9120, 9159-61 (Feb. 26, 2004); Final Rule Declaring Dietary Supplements Containing Ephedrine Alkaloids Adulterated Because They Present an Unreasonable Risk, 69 Fed. Reg. 6788, 6837 (Feb. 11, 2004); Food Labeling; Trans Fatty Acids in Nutrition Labeling; Consumer Research To Consider Nutrient Content Health Claims, 68 Fed. Reg. 41,434, 41,487-89 (July 11, 2003); Administrative Detention of Food for Human or Animal Consumption, 68 Fed. Reg. 25,242, 25,261 (May 9, 2003); Medical Devices; Patient Examination and Surgeons' Gloves; Test Procedures and Acceptance Criteria, 68 Fed. Reg. 15,404, 15,411-13 (Mar. 31, 2003); Current Good Manufacturing Practice in Manufacturing, Packing, or Holding Dietary Ingredients and Dietary Supplements, 68 Fed. Reg. 12,158, 12,229-30 (Mar. 13, 2003); Prior Notice of Imported Food, 68 Fed. Reg. 5428, 5455 (Feb. 3, 2003); Registration of Food Facilities, 68 Fed. Reg. 5378, 5410 (Feb. 3, 2003); Marking Requirements for and Prohibitions on the Reimportation of Imported Food Products That Have Been Refused Admission to the United States, 66 Fed. Reg. 6502, 6508 (Jan. 22, 2001); Hazard Analysis and Critical Control Point (HAACP); Procedures for the Safe and Sanitary Processing and Importing of Juice, 66 Fed. Reg. 6138, 6179-84 (Jan. 19, 2001); Current Good Manufacturing Practice for Blood and Blood Components, 65 Fed. Reg. 69,378, 69,398 (Nov. 16, 2000); Surgeon's and Patient Examination Gloves; Reclassification, 64 Fed. Reg. 41,710, 41,732-36 (July 30, 1999); Preliminary Regulatory Impact Analysis and Initial Regulatory Flexibility Analysis of the Proposed Rule To Require Refrigeration of Shell Eggs at Retail and Safe Handling Labels, 64 Fed. Reg. 36,516, 36,522-24 (July 6, 1999); Preliminary Regulatory Impact Analysis and Initial Regulatory Flexibility Analysis of the Proposed Rules To Ensure the Safety of Juice and Juice Products, 63 Fed. Reg. 24,254, 24,258-61 (May 1, 1998); Quality Mammography Standards, 62

example: The FDA in 2003 promulgated a regulation requiring food nutrition labels to disclose information about trans fats.[196] The rule was economically "significant" for purposes of Executive Order 12,866, the current Presidential cost-benefit order,[197] and the FDA therefore published and sent to OMB a lengthy cost-benefit document, including both monetary estimates of the rule's costs (the costs of re-labeling foods, testing them to determine trans fats levels, and reformulating them) and monetary estimates of the rule's benefits (avoided cases of fatal and nonfatal coronary heart disease).

The FDA monetized these benefits in two different ways, once with QALYs and once without, both showing the rule to have greater monetized benefits than costs.[198] In its QALY-based analysis, the FDA calculated the cost of each nonfatal case as the sum of medical costs, functional disability costs, and pain-and-suffering costs. A QALY value for functional disability and pain-and-suffering was estimated, then converted to dollars using a conversion factor of $100,000 per QALY. As the agency explained:

> [A] recent study . . . estimated . . . that the quality adjusted life year for a [coronary heart disease] survivor was 0.71, which indicates that the annual loss to the victim is 0.29 quality adjusted years. This loss represents the combined effects of functional disability and pain and suffering. FDA assumed that the loss lasts for 13 years, or 8.4 discounted years.[199]

The agency concluded that the monetized pain-and-suffering and functional

---

Fed. Reg. 55,852, 55,963 (Oct. 28, 1997). In addition, the FDA in a few rulemakings has employed VSLYs to monetize mortality without using the term "QALY." *See* Labeling Requirements for Systemic Antibacterial Drug Products Intended for Human Use, 68 Fed. Reg. 6062, 6076 (Feb. 6, 2003); Iron-Containing Supplements and Drugs, 62 Fed. Reg. 2218, 2243 (Jan. 15, 1997); sources cited *supra* note 193; *infra* text accompanying 233-234 (discussing equivalence of QALY and VSLY approaches under certain conditions).

196. *See* Food Labeling; Trans Fatty Acids in Nutrition Labeling; Consumer Research To Consider Nutrient Content Health Claims, 68 Fed. Reg. 41,434.

197. *See* Exec. Order No. 12,866 §§ 3(f), 6(a)(3), 3 C.F.R. 638 (1993), *reprinted in* 5 U.S.C. § 601 (2000); *see also* Unfunded Mandates Reform Act of 1995, 2 U.S.C. § 1532 (2000) (requiring the preparation of cost-benefit analyses for rules resulting in annual expenditures of $100 million or more).

198. *See* Food Labeling; Trans Fatty Acids in Nutrition Labeling; Consumer Research To Consider Nutrient Content Health Claims, 68 Fed. Reg. at 41,487-90. In the proposed rulemaking, the FDA relied solely on QALY-to-dollar conversions to quantify the benefit. *See* Food Labeling; Trans Fatty Acids in Nutrition Labeling, Nutrient Content Claims, and Health Claims, 64 Fed. Reg. 62,746, 62,771-75 (proposed Nov. 17, 1999).

199. Food Labeling; Trans Fatty Acids in Nutrition Labeling; Consumer Research To Consider Nutrient Content Health Claims, 68 Fed. Reg. at 41,488.

disability cost of each nonfatal case was 0.29 × 8.4 × $100,000 = $243,600.[200] This amount, plus the per-case medical costs, multiplied by the annual number of nonfatal cases prevented by the rule, was the rule's annual morbidity-reduction benefit. The FDA's QALY-based approach to monetizing the mortality-reduction benefit of the rule was to estimate the discounted average years of life lost in each fatal case of chronic heart disease (eight or eleven years, depending on the discount rate) and then multiply this number by the $100,000 QALY-to-dollar conversion factor, to arrive at a monetized benefit figure per fatal case prevented. That figure, multiplied by the annual number of fatal cases prevented by the rule, yielded the annual lifesaving benefit.[201]

The FDA practice underscores that hybrid cost-benefit analysis is a genuine policy-analytic option. The practice puts this novel decision-procedure on the welfarist's table, as it were. But how should the procedure be structured, from the point of view of overall welfare? There are two fundamental issues here. First, what conversion rate should be used to translate QALYs into dollars? The FDA in the trans fats rule used $100,000 per QALY as its conversion factor, and also considered what total benefits would be at a conversion rate of $300,000 and $500,000 per QALY.[202] Elsewhere the agency has used a conversion rate of $373,000 per QALY,[203] and in valuing short-term morbidity it has repeatedly used a "Quality Adjusted Life Day" value of $630,[204] which implies a conversion

200. *See id.*

201. *See id.*

202. See Food Labeling; Trans Fatty Acids in Nutrition Labeling; Consumer Research To Consider Nutrient Content Health Claims, 68 Fed. Reg. at 41,487-90. For other recent rulemakings in which the FDA has used the trio of $100/$300/$500,000 per QALY to determine possible benefits, see Prevention of Salmonella Enteritidis in Shell Eggs During Production, 69 Fed. Reg. at 56,855; and Final Rule Declaring Dietary Supplements Containing Ephedrine Alkaloids Adulterated Because They Present an Unreasonable Risk, 69 Fed. Reg. at 6842.

203. *See* Medical Devices; Patient Examination and Surgeons' Gloves; Test Procedures and Acceptance Criteria, 68 Fed. Reg. at 15,411; Surgeon's and Patient Examination Gloves; Reclassification, 64 Fed. Reg. at 41,734.

204. *See* Requirements Pertaining to Sampling Services and Private Laboratories Used in Connection with Imported Food, 69 Fed. Reg. at 23,469; Current Good Manufacturing Practice in Manufacturing, Packing, or Holding Dietary Ingredients and Dietary Supplements, 68 Fed. Reg. at 12,230; Marking Requirements for and Prohibitions on the Reimportation of Imported Food Products That Have Been Refused Admission to the United States, 66 Fed. Reg. at 6508; Hazard Analysis and Critical Control Point (HAACP); Procedures for the Safe and Sanitary Processing and Importing of Juice, 66 Fed. Reg. at 6180; Preliminary Regulatory Impact Analysis and Initial Regulatory Flexibility Analysis of the Proposed Rule To Require Refrigeration of Shell Eggs at Retail and Safe Handling Labels, 64 Fed. Reg. at 36,523; Preliminary Regulatory Impact Analysis and Initial Regulatory Flexibility Analysis of the Proposed Rules To Ensure the Safety of Juice and Juice Products, 63 Fed. Reg. at 24,261.

factor of $230,000 per QALY.[205] Which of these conversion factors should an agency such as the FDA employ? More fundamentally, what principles should guide the choice of factor? Second, what function should the converted amounts play in cost-benefit analysis? Specifically, under what conditions should they displace WTP/WTA amounts, where these are available?

1. How Should QALYs Be Converted to Dollars?

The FDA derives a QALY-to-dollar conversion factor from VSL (the "value of statistical life"), a number based on WTP/WTA to avoid the risk of death. For those not familiar with traditional cost-benefit analysis of mortality impacts, the following very quick summary might be helpful. Imagine that subjects in a contingent-valuation or revealed-preference study assign a WTP/WTA amount (for example, $40) to some small risk of death (for example, a 1-in-100,000 risk). That WTP/WTA amount, divided by the risk, is the "VSL" implied by the study (in this instance $4 million). It is the cumulative amount that a large population of individuals with the subjects' preferences would be willing to pay so as to avoid, or willing to accept as compensation for, a single death that would occur for certain, but with an uncertain victim.

The standard cost-benefit technique that agencies currently employ to monetize the mortality effect of a policy is to estimate the total lives saved or lost and multiply that number times a VSL figure inferred from a large group of contingent-valuation and/or revealed preference studies of WTP/WTA for the risk of death.[206] The VSL figures actually employed by agencies are in the vicinity of $6 million.

Now for the question of QALY-to-dollar conversion factors. Imagine that we calculate the average VSL from a group of contingent-valuation or revealed-preference studies. The average life expectancy of the individuals in the studies is X years; life expectancy is discounted at some rate so that the average discounted life expectancy is Y; and the average health of the group is at QALY level $q < 1$. VSL/qY seems, intuitively, like an appropriate QALY-to-dollars conversion factor—and in any event this has been the approach generally articulated by the FDA,[207] as well as the approach taken in scholarly work by some health

205. *See* Preliminary Regulatory Impact Analysis and Initial Regulatory Flexibility Analysis of the Proposed Rules To Ensure the Safety of Juice and Juice Products, 63 Fed. Reg. at 24,261.

206. For example, the agency might determine the VSL implied by each study (average WTP/WTA of the subjects divided by the risk involved), and then average those study-specific VSLs. On the VSL method and the use of VSL by administrative agencies, see Adler, *Against "Individual Risk," supra* note 112, at 1198 n.300.

207. *See, e.g.*, Prevention of Salmonella Enteritidis in Shell Eggs During Production, 69 Fed. Reg. at 56,855; Bar Code Label Requirement for Human Drug Products and Biological Products,

economists.[208] *VSL/qY* is the average WTP/WTA measure of the loss of longevity that the subjects in these studies were at risk of losing, divided by the average (discounted) QALY measure of that longevity.

But intuitions can mislead. The approach I will propose for deriving a QALY-to-dollar conversion factor is, at least conceptually, quite different from the FDA's. My approach flows directly from my view of cost-benefit analysis as a decision procedure maximizing LWUs.

To begin, it makes little sense, I suggest, to think of decision-makers setting a QALY-to-dollar conversion factor on a one-off basis, for some particular choice situation. To identify the factor that maximizes LWUs, the decision-maker needs to have some sense of how QALYs and WTP/WTA amounts translate into LWUs. If she can do *that* on a case-by-case basis with relative ease and accuracy, why not simply analyze the choices at hand directly in terms of LWUs? Why do a cost-benefit analysis of each option, which translates its various welfare effects into money amounts and then aggregates, rather than an LWU analysis, which translates those effects into LWU amounts and then aggregates?

Rather, the choice of the QALY-to-dollar conversion factor is best understood as a problem at the level of systems design. An agency head or an oversight body (the "system-designer") is anticipating that a conversion factor will or might be used in a range of choice situations confronted by agency

---

69 Fed. Reg. at 9160; Food Labeling; Trans Fatty Acids in Nutrition Labeling; Consumer Research To Consider Nutrient Content Health Claims, 68 Fed. Reg. at 41,489; Medical Devices; Patient Examination and Surgeons' Gloves; Test Procedures and Acceptance Criteria, 68 Fed. Reg. at 15,411; Current Good Manufacturing Practice in Manufacturing, Packing, or Holding Dietary Ingredients and Dietary Supplements, 68 Fed. Reg. at 12,230; Current Good Manufacturing Practice for Blood and Blood Components, 65 Fed. Reg. at 69,368; Surgeon's and Patient Examination Gloves; Reclassification, 64 Fed. Reg. at 41,734; Preliminary Regulatory Impact Analysis and Initial Regulatory Flexibility Analysis of the Proposed Rules To Ensure the Safety of Juice and Juice Products, 63 Fed. Reg. at 24,261. The $100,000 per QALY conversion factor that the FDA periodically uses is based on research by Garber and Phelps, who employ a different approach to estimating the factor. *See, e.g.*, Prevention of Salmonella Enteritidis in Shell Eggs During Production, 69 Fed. Reg. at 56,855; Food Labeling; Trans Fatty Acids in Nutrition Labeling; Consumer Research To Consider Nutrient Content Health Claims, 68 Fed. Reg. at 41,489 (citing Garber & Phelps, *supra* note 28). A less refined version of the approach described in the text would assume that q is 1. A more refined version would allow q to vary over time. A related approach would use average population longevity rather than the average life expectancy of individuals in WTP/WTA studies for the risk of death. All these variations of the *VSL/qY* approach are quite different from the method I advocate, in the text below, and are vulnerable to the criticisms of the *VSL/qY* approach articulated below.

208. *See, e.g.*, Richard A. Hirth et al., *Willingness To Pay for a Quality-Adjusted Life Year: In Search of a Standard*, 20 MED. DECISION MAKING 332, 335 (2000); George Tolley et al., *State-of-the-Art Health Values*, *in* VALUING HEALTH FOR POLICY, *supra* note 1, at 328-29.

decision-makers, and is determining what the optimal factor would be. The decision-makers themselves will evaluate choices by performing cost-benefit analysis, not LWU analysis. As noted above, there are various plausible reasons, having to do with deliberation costs and with the competence and trustworthiness of the decision-makers,[209] why it might be LWU-maximizing for the decision-makers to employ cost-benefit analysis rather than LWU analysis as their decision-procedure, at least at present.

However, the system-designer *will* employ LWUs in setting the QALY-to-dollar conversion factor. At a minimum, she will need to have some very rough sense of the expected LWU gains and losses associated with different conversion factors. In the set of choice situations that the system-designer is considering, health effects will be measured by the decision-maker in QALYs and converted to dollar amounts through some conversion factor f, while non-health effects will be measured on a WTP/WTA scale.

For simplicity, think of each choice situation as binary, presenting the decision-maker with a choice between the status quo and some regulatory intervention or "project," which has health benefits and non-health costs. Ideally, for any choice situation in the set, the system-designer would be able to express probabilistically her judgments about the total amount of the project's health benefits and non-health costs, in LWUs, plus her judgments about the likely ratio between the total WTP and QALY amounts that the agency decision-maker will observe and the total LWU amounts.[210] For simplicity, I will also assume that the

---

209. *See supra* text accompanying note 186.

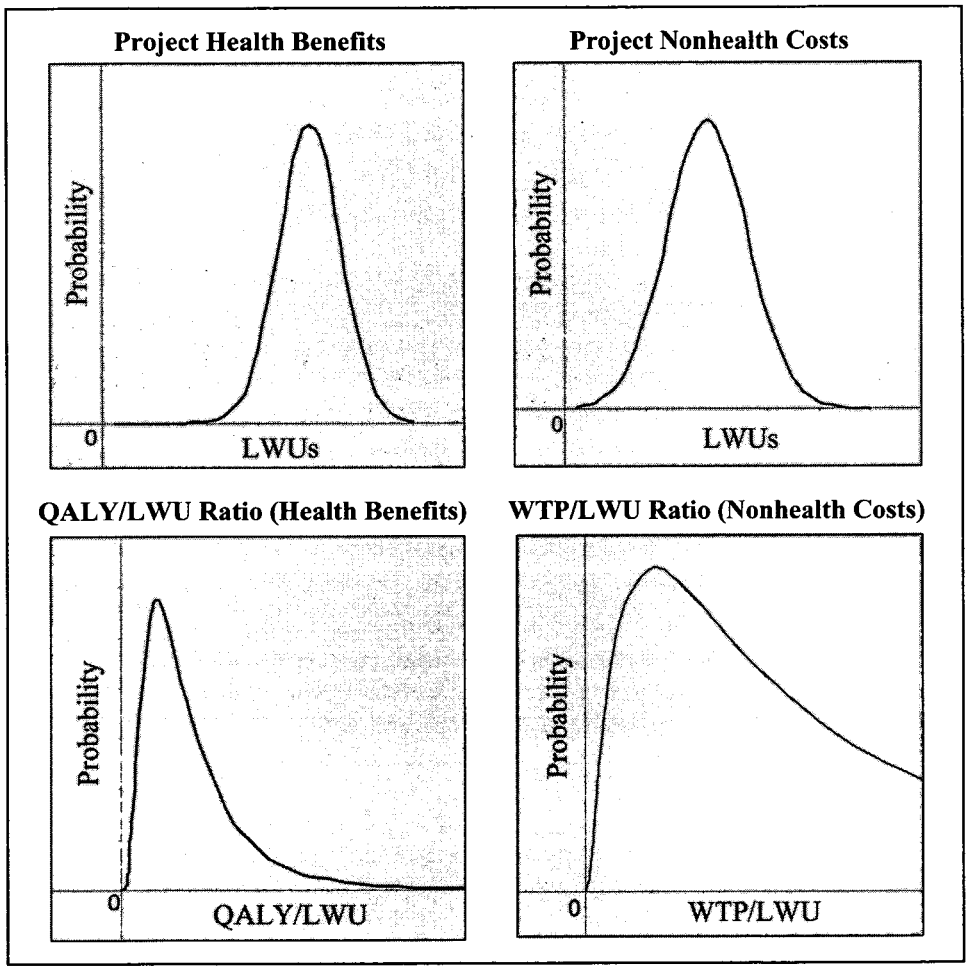
210. To be clear: The WTP/LWU and QALY/LWU ratios here—the ratios that drive the system-designer's choice—are the ratios between the aggregate WTP or QALY amounts for project costs or benefits that the decision-maker will calculate, and the total amount in LWUs of project costs or benefits. These are ratios of total project impacts, not the average of the QALY/LWU and WTP/LWU ratios for the various individuals affected by the project (although under some conditions the average of individual ratios may equal or approximate the ratio of totals).

The graphs in the text are easiest to grasp if it is assumed that the QALY/LWU ratio for health benefits is independent of the amount of health benefits in LWUs, and similarly that the WTP/LWU ratio for non-health costs is independent of the amount of non-health costs in LWUs. In that event, the system-designer has a single subjective probability distribution for the QALY/LWU ratio and WTP/LWU ratio conditional on any given level of benefits and costs (in LWUs), and the ratio graphs display this distribution. However, the system-designer's maximization problem as modeled here does not require this sort of independence. *See infra* App. II.

The analysis does make a different sort of independence assumption, namely that the system-designer's probability distributions with respect to LWUs, the QALY/LWU ratio, and the WTP/LWU ratio for a given choice situation remain the same regardless of which choice the decision-maker selects in any other choice situation. For example, the designer's probability distributions with respect to the second choice that the decision-maker will confront are the same

system-designer's subjective probability distributions over project LWU benefits and costs, over the QALY/LWU ratio, and over the WTP/LWU ratio are the same for every choice situation in the set, represented by the following sorts of graphs. A more comprehensive analysis would permit the designer to have different subjective distributions for different subsets of choice situations; but that analysis would be overly complicated for my purposes here, which is to explain a general approach to setting the QALY-to-dollar conversion factor.

FIGURE 1: SETTING A QALY-TO-DOLLAR CONVERSION FACTOR



regardless of whether the decision-maker has picked the project or status quo in the first-choice situation.

Note that both the QALY/LWU ratio for health benefits, and the WTP/LWU ratio for non-health costs, are variable. This reflects the fact that the actual QALY or WTP amounts measured by the decision-maker will not be perfect proxies for LWUs—for all the reasons already discussed. A perfect proxy would have some single fixed ratio to LWUs with probability one.

With these probability distributions in hand, the system-designer can pick an optimal QALY-to-dollar conversion factor *f**. This is the factor *f** that maximizes expected LWUs. To reiterate: The system-designer anticipates that decision-makers will perform hybrid cost-benefit analysis across a range of choice situations, with health effects measured in QALYs and converted to dollars using a conversion factor that the system-designer will announce, and non-health effects measured in WTP/WTA. Her underlying objective is to maximize overall (expected) welfare,[211] i.e., total expected LWUs. The optimal conversion factor *f** is the factor that does just that. Cost-benefit analysis, as I see it, is simply a tool for advancing overall well-being, and should be calibrated (e.g., in the choice of the QALY-to-dollar conversion factor) with that objective in mind.

What can we say about *f**, the optimal QALY-to-dollar conversion factor? An Appendix to this Article derives some formal results about *f**, which are informally summarized here. First, *f** is context-dependent. It depends on the shapes of all the probability distributions displayed here: on the system-designer's probability distributions with respect to the QALY/LWU ratios for health benefits and WTP/LWU ratios for non-health costs, as well as her probability distributions with respect to the total amount of project health benefits in LWUs, and total project non-health costs in LWUs.[212] This may seem like a negative result, but in fact it underscores the crucial point that specifying the QALY-to-dollar conversion factor is a pragmatic decision, a matter of optimizing hybrid cost-benefit analysis with respect to the underlying criterion of expected overall welfare, representable in expected LWUs. The optimal factor for one set of choice situations may be *f*+, given the system-designer's QALY/LWU, WTP/LWU, and LWU probability distributions for that set of situations. The optimal factor for a different set of choice situations, e.g., a different administrative agency, (or for the same set of choice situations but a different designer, with different probability distributions) may be *f*++. There is no single, natural, acontextual "rate" at which QALYs convert into dollars.[213]

---

211. The system-designer does not know for certain what choices will be presented to the decision-makers, and what the WTP/LWU and QALY/LWU ratios will be. Proceeding, therefore, under uncertainty, she aims to maximize expected LWUs.

212. *See infra* App. II.

213. In a recent article, Gyrd-Hansen proposes a "pragmatic" approach to determining WTP per QALY, which is similar in spirit to my proposal. He suggests that "seeking to apply a unique WTP

Second, one can show that, with some simplifying assumptions, the optimal conversion factor in a given context, for a given set of QALY/LWU, WTP/LWU and LWU probability distributions, is a kind of *average*, which can be calculated by taking the average WTP/LWU ratio for non-health costs, and dividing that by the average QALY/LWU ratio for health benefits.[214]

Conceptually, this approach to setting the QALY-to-conversion factor is quite different from the FDA's *VSL/qY* approach. My approach seeks to calibrate the conversion factor in terms of LWUs; the FDA's does nothing of the sort, at least not explicitly.

Now, it might be objected that FDA's approach does implicitly incorporate LWUs. Leaving aside discounting, the FDA in effect takes VSL, which is the WTP/WTA measure of a certain kind of health impact, namely death, and divides that by a kind of estimate of the average QALY loss that occurs in death, namely *qX*. But this WTP/QALY ratio, *VSL/qX*, mathematically, is identical to: *(VSL/L)/(qX/L)*, where *L* is the LWU loss that on average occurs with death. In other words, the FDA's conversion factor can be seen as a rough kind of approximation for the number that would emerge if a system-designer calculated the average WTP/LWU ratio for the mortality-reduction benefits of policy choices, and divided that by the average QALY/LWU ratio for those choices.

Even so, there is a crucial difference between the FDA's approach and mine. My analysis shows that, with simplifying assumptions, the optimal conversion factor is the average WTP/LWU ratio for the *non-health* impacts of agency choices, e.g., reduced consumption, or recreation, or employment, divided by the average QALY/LWU ratio for health impacts. By contrast, the FDA takes the average WTP/LWU ratio for a particular *health* impact (the risk of death) and divides that by the average QALY/LWU ratio for that impact. Think of the point this way: Hybrid cost-benefit analysis converts QALYs into dollars and then adds these dollar sums to WTP/WTA amounts for non-health impacts. So what is crucial in optimizing the conversion is how WTP/WTA amounts for non-health impacts translate into LWUs and how QALYs translate into LWUs, *not* how WTP/WTA amounts for health impacts translate into LWUs.

A final attempt to salvage the FDA's approach: Assume that the distribution of WTP/LWU ratios for health impacts is roughly the same as the distribution of WTP/LWU ratios for non-health impacts. On that assumption, the FDA's approach is perhaps a rough and ready way to approximate the optimal

---

for a QALY should not be seen as defining the theoretical link between CEA and CBA, but rather as an aid to decisionmakers," and notes that decision-makers might use "situation-specific" WTP per QALY values. Dorte Gyrd-Hansen, *Willingness To Pay for a QALY: Theoretical and Methodological Issues*, 23 PHARMACOECONOMICS 423, 428, 430 (2005).

214. *See infra* App. II.

conversion factor—although even here one would want to consider WTP/LWU ratios for a range of health impacts, not simply the risk of death (as the FDA does). Doing that suggests that the range of numbers the FDA has employed ($100,000 to $500,000) per QALY is on the high side; a conversion factor of $100,000 per QALY looks closer to optimal, and lower factors such as $50,000 or even $10,000 should be considered.[215] But in any event the assumption may not be true, which means that the numbers emerging from this quite rough and ready analysis would need to be adjusted. For example, the biases that affect WTP/WTA measurement may tend to inflate WTP/LWU ratios for health as opposed to non-health impacts, or vice versa. I will not attempt to estimate the

---

215. Hirth et al. examined a wide range of VSL estimates, based on contingent valuation studies, wage-risk studies, and other revealed preference studies. Their estimates of the conversion factor were in the same range as the FDA's numbers: $93,402 based on the revealed preference studies other than wage-risk studies, $161,305 based on the contingent valuation studies, and $428,286 based on the wage-risk studies, or an overall estimate of $265,345. *See* Hirth et al., *supra* note 208, at 338. By contrast, the conversion factor that results from comparing QALY values and WTP/WTA amounts for light morbidity is substantially lower. I estimated this factor using a review article by Johnson, which collects WTP/WTA and QALY values for a range of light symptoms, such as angina, throat congestion, coughs, runny noses, and headaches. *See* Johnson et al., *supra* note 129. The estimation procedure was straightforward. If, for example, a study determines that average WTP to avoid seven days of a severe cough and sneeze is $87.35, and the QALY loss associated with a severe cough and sneeze is 0.318, then the conversion factor implied by this study is $(365 \times 87.35)/(7 \times 0.318)$, or $14,323. Averaging the conversion factors implied by the fifty-some WTP/WTA and QALY valuations for light morbidity produces an overall estimate of $37,663 per QALY. This is an order of magnitude lower than the overall Hirth et al. result of $265,345 per QALY, and (interestingly) much closer to the number traditionally used by many public health scholars as the cut-off ratio for cost-effectiveness analysis: $50,000 per QALY. *See* Hirth, *supra*, at 333. It is also very close to the cut-off ratio that, in practice, the Australian government uses in deciding whether to list a pharmaceutical for public funding, which one study estimates to be $40,400. *See* Dorte Gyrd-Hansen et al., *Willingness To Pay for a QALY*, 12 HEALTH ECON. 1049, 1049 (2003).

Gyrd-Hansen and colleagues come up with a yet lower estimate of the conversion factor. They surveyed 3201 Danish individuals and elicited WTP values for changes in health states described using a standard health state classification system. These money valuations were integrated with preexisting QALY valuations for the changes, yielding a mean WTP per QALY of roughly $10,000. *See id.* at 1058. A simple average of this figure, the $37,663 ratio implied by the Johnson article, and the $265,345 overall estimate based on VSL set forth by Hirth et al. gives a value of about $100,000. This rough, heuristic calculation suggests that $100,000, currently the low end of the FDA's range, might well be an appropriate central QALY-to-dollar conversion factor for hybrid cost-benefit analysis—absent further information, for example information concerning the distribution of LWU costs and benefits across agency choice situations, that would bear on the optimal conversion factor—and that lower factors, such as $50,000 or even $10,000, should be considered.

degree or direction of the appropriate adjustment here.

It might be objected that my approach requires the system-designer to do something impossible: to estimate project impacts in an esoteric and unobservable metric, namely LWUs. But that endeavor is *not* impossible—although it may be expensive and time-consuming, which is why it may well be welfare-maximizing for deliberation about LWUs to occur at the level of systems design, rather than in the evaluation of individual projects. The graphs in this Section, again, represent the designer's subjective probability distributions—her degrees of belief about the different possible quantities of some item. An individual's subjective probability distributions for *anything* (the temperature in the middle of the sun, the number of electrons in George Washington's finger, the number of mistakes that students taking a hypothetical exam under hypothetical conditions will make) can be generated using Bayesian probability-elicitation techniques, at least if the individual is smart and patient enough to go through the exercise.[216]

LWUs are utility numbers representing the preferences of idealized spectators with respect to the different combinations of health, consumption, and other attributes that make up different possible human life-histories. LWUs are not observable; but neither are QALYs. LWU surveys, asking respondents to think about, and express numerically, their preferences over possible lives, can and certainly should be conducted.[217] Even without such surveys in hand, system-designers can arrive at a sense of how strongly spectators would prefer different packages of health and non-health attributes by consulting their own preferences, and by drawing on the rich body of economic and philosophical scholarship about well-being.[218] Sketching a subjective probability distribution over LWUs means sketching a subjective distribution over the utilities of hypothetical, idealized spectators. Those utility numbers—like physical magnitudes in inaccessible places, or the quantities of miniscule items or items in the past or future—cannot be perceived. But we can at least quantify our beliefs about what the numbers might be, in the form of subjective probabilities. Once it is recognized that prevalent policy metrics, such as QALYs and WTP/WTA amounts, are simply rough proxies for well-being, it becomes plausible—indeed compelling—to undertake more systematic efforts to estimate LWUs, and to use such estimates in structuring cost-benefit analysis. That is what I am advocating here.

---

216. On Bayesian techniques, see *supra* text accompanying notes 178-185.

217. *See supra* note 184 and accompanying text.

218. *See supra* text accompanying notes 183-185.

2. *What Role Should QALYs Play in Nontraditional Cost-Benefit Analysis?*

What precise role should QALYs play in cost-benefit analysis? At a minimum, QALYs can provide alternate estimates of health benefits. Where both a QALY-based money estimate and a traditional WTP/WTA estimate are available, the agency should undertake cost-benefit analyses using both figures.[219] If the parallel analyses reach convergent policy recommendations, then the agency can be especially confident in that course of action. If they don't, then the agency has reason to scrutinize the WTP/WTA numbers, perhaps conducting additional contingent-valuation or revealed-preference studies. In effect, cost-benefit analysis in parallel provides a simple heuristic of the value of further expenditures to estimate WTP/WTA amounts. It is a kind of rough-and-ready value-of-information analysis.[220] The OMB now recommends that agencies conduct a kind of hybrid cost-benefit analysis as a source of alternate cost-benefit estimates, at least where longevity is at issue.[221]

Second, agencies should use QALY-based money estimates of health benefits where WTP/WTA estimates are not available. This is not a minor point.[222] There is a large empirical literature on WTP/WTA for mortality risks,[223] but much less research on WTP/WTA for morbidity;[224] and many, perhaps most, health conditions lack even a single contingent-valuation or revealed preference study. By contrast, as evidenced by the Harvard Center for Risk Analysis's comprehensive compilation of QALY estimates,[225] QALY surveys have been conducted for a large number of conditions. Even where no survey exists for a

---

219. The FDA sometimes follows this approach, undertaking both traditional and hybrid cost-benefit analysis. *See, e.g.*, Food Labeling; Trans Fatty Acids in Nutrition Labeling; Consumer Research To Consider Nutrient Content Health Claims, 68 Fed. Reg. 41,434, 41,488. (July 11, 2003).

220. On value of information analysis, see, e.g., Maxine Dakins, *The Value of the Value of Information*, 5 HUM. & ECOLOGICAL RISK ASSESSMENT 281 (1999). On the usefulness of heuristics, see generally GERD GIGERENZER ET AL., SIMPLE HEURISTICS THAT MAKE US SMART (1999).

221. OFFICE OF MGMT. & BUDGET, *supra* note 18, at 30.

222. A dramatic example is provided by the EPA's cost-benefit analyses in the major arsenic and radon rulemakings, which employed WTP to avoid chronic bronchitis as an estimate of the cost of nonfatal cancers, since no WTP data for nonfatal cancers was available. *See* National Primary Drinking Water Regulations; Arsenic and Clarifications to Compliance and New Source Contaminants Monitoring, 66 Fed. Reg. 6976, 7012 (Jan. 22, 2001); National Primary Drinking Water Regulations; Radon-222, 64 Fed. Reg. 59,246, 59,325 (Nov. 2, 1999).

223. *See* sources cited *supra* note 56.

224. *See* Johnson et al., *supra* note 129, at 642 ("[T]he literature providing monetary health values is deficient in both breadth and quality.").

225. *See* Bell et al., *supra* note 167.

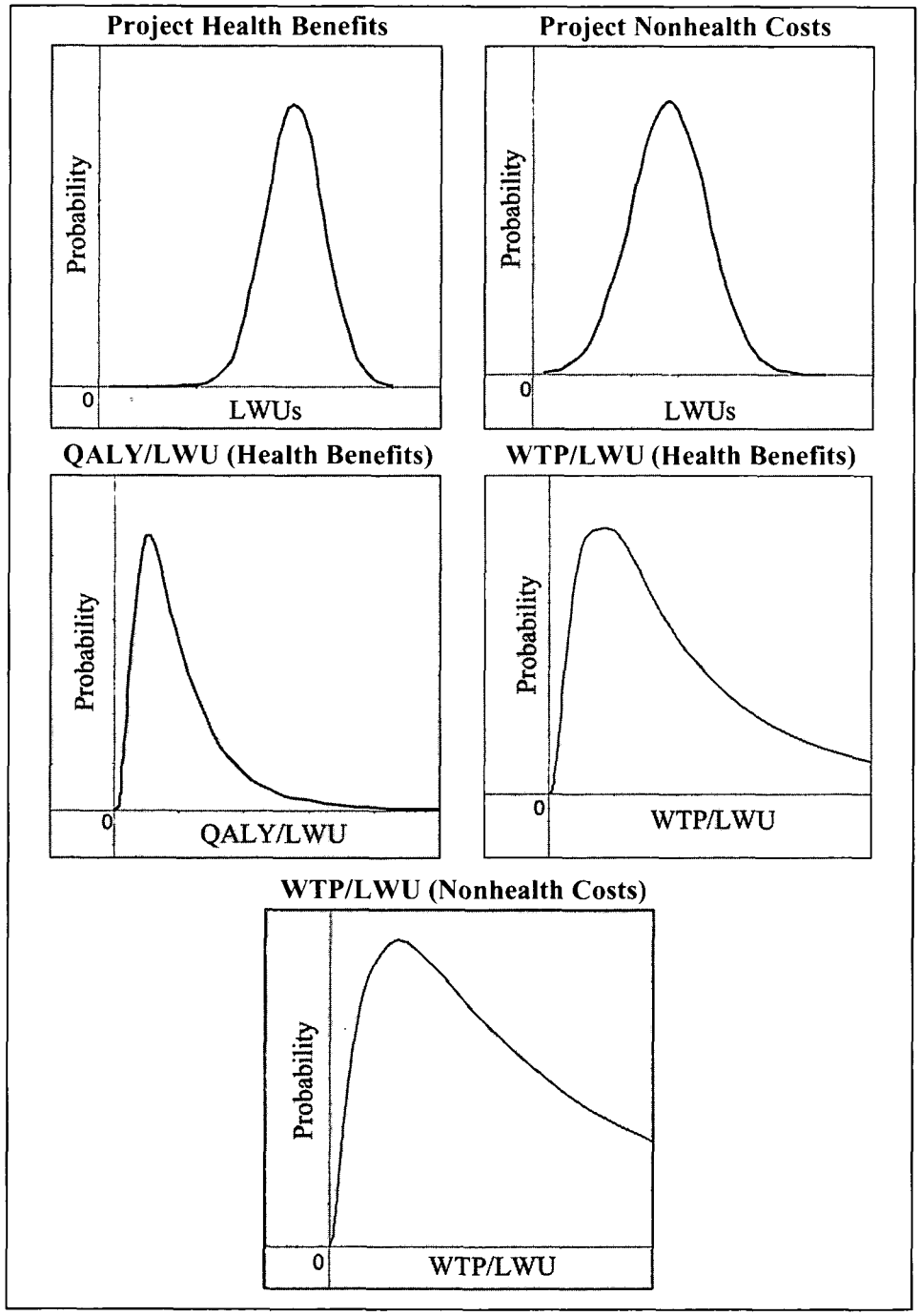
particular health state, health classification systems often permit researchers to extrapolate a valuation for that state from other, surveyed states.[226]

What about a more robust role for QALYs? When, if ever, should an agency prefer a QALY-based money estimate to a WTP/WTA amount? The general welfarist strategy for answering this question is the same as the general strategy for identifying an appropriate QALY-to-dollar conversion factor: Hybrid cost-benefit analysis should be preferred to traditional cost-benefit analysis, over some range of choice situations, when that increases expected welfare (in LWUs). Again, the conditions under which this is true will depend on the sorts of choice situations involved: specifically, the system-designer's probability distributions over QALY/LWU ratios, WTP/LWU ratios, and LWU amounts for these choice situations.

But it is possible to make some general observations. Let us return to the simple scenario discussed above. A system-designer is considering a range of binary choice situations that agency decision-makers will face, each involving a regulatory project with health benefits relative to the status quo and non-health costs. The designer supposes that the non-health costs will be measured using WTP/WTA amounts and is now choosing between the following two decision procedures: traditional cost-benefit analysis, where the health benefits are measured on a WTP/WTA scale and added to the non-health costs, and hybrid cost-benefit analysis, where the health benefits are measured on a QALY scale, converted to dollars at the optimal conversion rate f^*, and then added to the non-health costs. The designer chooses between these two procedures by determining which one maximizes expected LWUs, and instructs the decision-makers to use that procedure.

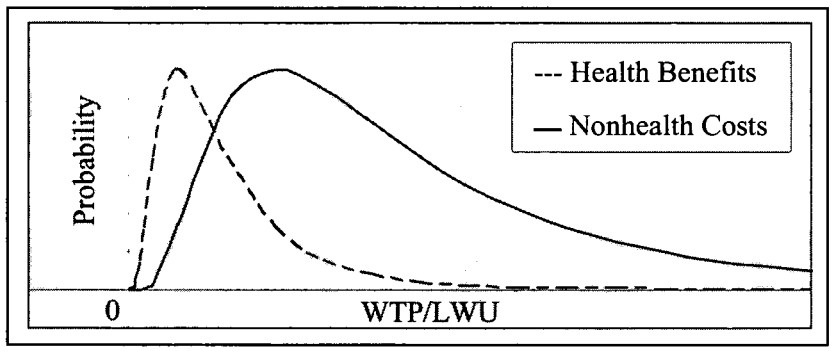
226. *See* sources cited *supra* note 166 (discussing health state classification systems).

FIGURE 2: CHOOSING BETWEEN TRADITIONAL & HYBRID COST-BENEFIT ANALYSIS



There are two general reasons why hybrid cost-benefit analysis might produce higher expected LWUs than traditional cost-benefit analysis. First, the distribution of the WTP/LWU ratio for health benefits might be skewed to the right or left of the optimal point. In other words, a variant of the traditional approach in which health benefits are first measured in WTP/WTA, then multiplied by a constant k, and finally added to non-health costs, also measured in WTP/WTA, might lead to better results than straight traditional cost-benefit analysis. This might occur, for example, if the average WTP/LWU ratio for health benefits is greater or less than the average WTP/LWU ratio for non-health costs.

FIGURE 3: WTP/LWU RATIOS FOR PROJECT COSTS & BENEFITS

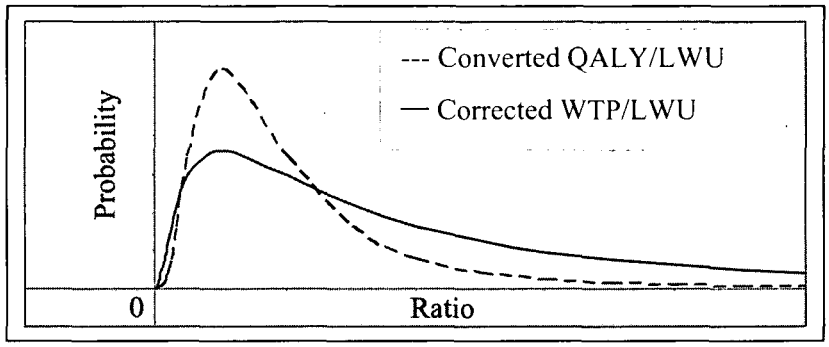


In principle, the skewing of the distribution of WTP/LWU for health benefits, relative to the optimal point, can be corrected by applying a scaling factor to the WTP amounts.[227] Thus, at least in principle, the system-designer should compare (1) a cost-benefit procedure that measures health in QALYs, and converts those amounts to dollars using the optimal conversion factor f^* to (2) a cost-benefit procedure that measures health on a WTP/WTA scale, which is then corrected using the optimal correction factor k^*.

In this competition, the QALY-based metric might win out because it has smaller variance or spread.

227. *Cf.* Hirth et al., *supra* note 208, at 340 (discussing use of "calibration factors" to correct WTP amounts).

FIGURE 4: CONVERTED QALY/LWU & CORRECTED WTP/LWU RATIOS, PROJECT HEALTH BENEFITS



Take the simplest case in which all non-health costs measured in WTP convert to LWUs at the same rate and in which the designer's QALY/LWU distribution and WTP/LWU distribution for health benefits are independent of the absolute level of benefits and costs, in LWUs. In this simple case, if the ratio of optimally converted QALYs to LWUs has less "area under the tails" (roughly, less variance) than the ratio of optimally corrected WTPs to LWUs, then the QALY metric is better.[228] The Appendix demonstrates this rigorously. For each pairing of health benefits, in LWUs, and non-health costs, in LWUs, the lower variance of the QALY/LWU distribution means a lower probability that the QALY metric will reach the incorrect result (that is, choosing the project where health benefits are less than costs, or choosing the status quo where health benefits exceed costs).

The idea of variance provides a unifying rubric under which to group the diverse ways in which QALYs and WTP amounts can fail to track LWUs. Both (1) the cognitive difficulties that drive a wedge between the QALY or WTP valuations available to regulators and idealized QALYs or WTP amounts, and (2) the factors that drive a wedge between idealized QALYs or WTP amounts and LWUs (such as the wealth effect, dead-anyway effect, and the problem of background characteristics) can be conceptualized as sources of increased variance in the QALY/LWU and WTP/LWU ratios.[229] This observation also underscores the point that the choice between hybrid cost-benefit analysis and

228. *See infra* App. II.

229. To be sure, these various factors and cognitive difficulties determine not only the variance of the QALY/LWU and WTP/LWU distributions, but also the location of these distributions along the x-axis, i.e., the absolute level of the average QALY/LWU and WTP/LWU ratio. But that effect, in principle, can be compensated for through the conversion factor f^* and correction factor k^*.

traditional cost-benefit analysis is not clear-cut. It represents an exercise in optimization, hinging in part on the variance in QALY/LWU and WTP/LWU ratios in the relevant choice situations. Because the factors producing variance are different for the two sorts of health metrics (QALYs and WTP), there is no reason to think that the QALY/LWU variance will always be smaller or larger than the WTP/LWU variance, and thus no reason to think that hybrid cost-benefit analysis will always dominate traditional cost-benefit analysis, or vice versa.

C. Why Not Cost-Effectiveness Analysis?

I have argued that a policy-analytic procedure which employs a "hybrid" cost-benefit analysis, supplementing or displacing traditional cost-benefit analysis, may increase overall welfare (LWUs) as compared to traditional cost-benefit analysis. The observant reader may wonder whether cost-effectiveness analysis should also be back on the table at this point. As discussed earlier, welfarists often criticize cost-effectiveness analysis because it deviates from traditional cost-benefit analysis and Kaldor-Hicks efficiency. Now that our underlying goal is overall welfare, not Kaldor-Hicks efficiency, might not some version of cost-effectiveness analysis be the optimal procedure for health policy?

The answer is pretty clearly no. Cost-effectiveness analysis, again, is a set of policy-analytic techniques that monetize non-health impacts but not health, instead measuring health on some non-monetary scale such as QALYs. The first variant of cost-effectiveness analysis maximizes health for a given budget. Here, the shift from Kaldor-Hicks efficiency to overall well-being as the underlying goal does nothing to salvage the procedure, because it is quite possible that the net welfare benefits of a health-maximizing exhaustion of the budget are less than the net benefits of a smaller or even null expenditure. Cost-benefit analysis tests for that, while cost-effectiveness analysis does not.

The second variant, recall, uses a fixed cut-off ratio to select the best choice. The problem with this approach, as compared to hybrid cost-benefit analysis, is *not* the use of a single cut-off number. Hybrid cost-benefit analysis also does that, in the form of a single QALY-to-dollar conversion factor. Indeed, hybrid cost-benefit analysis employing a conversion factor f^* and the cut-off ratio variant of cost-effectiveness analysis employing that same number f^* as the cut-off would, I believe, always reach the same results.

The problem with this second variant of cost-effectiveness analysis, rather, is the abandonment of WTP/WTA valuations of health. I have suggested that WTP/WTA valuations of health impacts and QALY-to-dollar conversions *both* have a useful role in guiding policy choice. First, as stated above, the fact that traditional cost-benefit analysis and hybrid cost-benefit analysis produce divergent recommendations in some policy situation has informational value: It may point to the need for further studies to produce better estimates of

WTP/WTA or QALY values. Second, given the system-designer's probability distributions over WTP/LWU, QALY/LWU, and LWU amounts, traditional cost-benefit analysis using WTP/WTA to value health may end up being the welfare-maximizing procedure in some contexts, as compared to a hybrid procedure that employs QALY-to-dollar conversions to value health.

In short, I suggest that the welfare-maximizing procedure will use QALY-based analysis alongside traditional cost-benefit analysis and, where they conflict, will sometimes (not always) prefer the results of the latter analysis. Cost-effectiveness analysis relies solely on QALYs to value health, but since QALYs are not perfect proxies for LWUs, any more than WTP/WTA amounts, that seems overly rigid from the point of view of overall welfare.

D. An Example: The Valuation of Lifesaving

What version of cost-benefit analysis, traditional or hybrid, should be employed to value regulatory measures that cause deaths or save lives? This question has recently been a matter of considerable political controversy and concomitant scholarly discussion, triggered by the EPA's cost-benefit analysis in connection with the "Clear Skies" legislative initiative. The EPA in an alternative estimate used an age-adjusted VSL figure: specifically, $3.7 million for deaths of individuals under seventy, and $2.3 million for deaths of individuals over seventy.[230] Vociferous protests by senior citizen groups ensued,[231] prompting the OMB to issue a memorandum instructing the EPA and other agencies not to use age-adjusted VSL.[232] In the same memorandum, the OMB discusses valuation of death or life-saving through "VSLYs," that is, calculating the numbers of life-years lost or saved and multiplying by a conversion factor to yield a dollar

---

230. *See* U.S. EPA, TECHNICAL ADDENDUM, *supra* note 15, at 35-37. Recall that VSLs are derived from WTP/WTA to avoid small risks of death. If an individual is WTP $3 to avoid a 1 in 1 million risk of death, then the VSL figure implied by that valuation is $3 million. Cost-benefit analysis with age-invariant VSLs looks at the average VSL, observed in the entire universe of contingent valuation and revealed preference studies (involving subjects of different ages), and employs that single figure (say, $6 million) to value each death. The age-adjusted VSL methodology seeks to determine the average VSL of individuals in a particular age group, and then uses that age-specific VSL number to value deaths in that age group. For a good discussion of VSLs, age adjustments, and VSLYs, see Lowenstein & Revesz, *supra* note 15.

231. For a description of the controversy, see Lowenstein & Revesz, *supra* note 15, at 10,957; Robert W. Hahn & Scott Wallsten, *Whose Life Is Worth More? (And Why Is It Horrible To Ask?)*, WASH. POST, June 1, 2003, at B03.

232. Memorandum from John D. Graham, U.S. Office of Mgmt. & Budget, to the President's Mgmt. Council (May 30, 2003) [hereinafter Graham Memorandum], *available at* http://www.whitehouse.gov/omb/inforeg/pmc_benefit_cost_memo.pdf; *accord* OFFICE OF MGMT. & BUDGET, *supra* note 18, at 30.

amount. If the life-years are priced at a *constant* conversion factor, and if the years lived would be years of perfect health, then this "VSLY" method is just the kind of QALY-based or hybrid cost-benefit analysis that I have been discussing.[233] However, the OMB memorandum encourages agencies to use an age-adjusted VSLY approach (in particular, to use a dollar conversion factor that would price life-years saved or lost at a higher amount for senior citizens than for younger individuals).[234]

In short, four different cost-benefit approaches to valuing life are now on the table: (1) using an age-invariant VSL figure (e.g., $6 million), which is standard practice; (2) using an age-adjusted VSL figure; (3) using QALY-to-dollar conversions, i.e., converting each life-year lost or saved by regulation to dollars at an age-invariant conversion rate; and (4) using age-adjusted VSLYs, i.e., converting life-years to dollars at an age-specific rate. The OMB memorandum in response to the Clear Skies controversy, and more recently its authoritative general guide to agency cost-benefit analysis, discourages QALY-to-dollar conversions as well as age-adjusted VSLs.[235]

What is the right approach? If Kaldor-Hicks efficiency is the underlying criterion, then age-adjusted VSLs are ideal, at least if they can be measured accurately and cheaply. A decision procedure that sums fully individualized WTP/WTA amounts will track potential Pareto improvements.[236] Measurement problems may push in the direction of age-invariant VSLs, which represent an average of VSLs across age groups. But VSLs of some kind are preferable to VSLYs (either age-invariant, i.e., QALYs, or age-adjusted VSLYs), since the use of VSLYs to value mortality has no grounding in Kaldor-Hicks efficiency or the WTP/WTA methodology. In short, conventional economic wisdom prefers a VSL measure and is skeptical of QALY-to-dollar conversions and age-adjusted VSLYs.[237]

---

233. *See generally* Lowenstein & Revesz, *supra* note 15 (discussing life-years approach); Sunstein, *supra* note 192 (same).

234. Graham Memorandum, *supra* note 232; *accord* OFFICE OF MGMT. & BUDGET, *supra* note 18, at 30.

235. *See* sources cited *supra* note 232. By recommending age-adjusted VSLYs, OMB discourages age-invariant VSLYs, i.e., straight QALY-to-dollar conversions.

236. This statement is only roughly true—a point that traditional proponents of cost-benefit analysis often overlook—because of the Boadway paradox. *See* source cited *supra* note 24.

237. *See, e.g.*, Lowenstein & Revesz, *supra* note 15, at 10,963-69. The VSLY method, at least in its simplest variant, values increments to longevity at a linear rate: either a single rate for the entire population (age-invariant VSLYs), or a rate specific to an age group (age-adjusted VSLYs). *See infra* note 246. In either event, a policy that adds five years of longevity (say) to the lives of some individuals in a particular age bracket will be assigned a monetary value five times that of a policy that adds one year. But WTP/WTA for the risk of losing five years of life need not be five

But what if overall welfare, not Kaldor-Hicks efficiency, is the gold standard? The answer, then, is trickier.

Consider, first, the case in which the system-designer (OMB) is choosing an optimal procedure for valuing life across a range of choice situations in which all the individuals whose lives are at stake fall in roughly the same "background equivalent class": They have equally valuable non-health attributes. A crucial question for this case (and for the more general analysis too) is whether spectators are risk-prone, risk-neutral, or risk-averse with respect to longevity[238]—that is, whether the LWU value of an incremental year increases, stays constant, or decreases as affected individuals become older. It seems intuitively plausible that longevity, like wealth, has a diminishing marginal impact on overall well-being; and this assumption is bolstered by some survey work. Johannesson and Johansson surveyed 1000 Swedes for their policy preferences as between equally costly programs that differed in the number of lives saved and the age of the persons saved. Integrating the survey responses with life expectancy data, they determined that: "[T]hree life-years gained for 50-year-olds are judged equivalent to one life-year gained for 30-year-olds, and ten life-years gained among 70-year-olds are judged equivalent to one life-year gained for 30-year-olds."[239] Cropper et al. reached similar results in an earlier U.S. study: "[T]he median respondent in our surveys places more weight on saving young persons than he would if people were weighted strictly by life expectancy."[240]

If longevity has a diminishing marginal impact on overall well-being, then

---

times WTP/WTA for the risk of losing one year of life, given wealth effects, horizon effects, the dead-anyway effect, and so on. Therein lies the basic criticism of VSLYs by those who see cost-benefit analysis as a tool to implement Kaldor-Hicks efficiency.

238. *See supra* text accompanying notes 156-159.

239. Magnus Johannesson & Per-Olov Johansson, *Is the Valuation of a QALY Gained Independent of Age? Some Empirical Evidence*, 16 J. HEALTH ECON. 589, 595 (1997).

240. Maureen L. Cropper et al., *Preferences for Life Saving Programs: How the Public Discounts Time and Age*, 8 J. RISK & UNCERTAINTY 243, 244 (1994). Cropper does not find that the utility of lifesaving decreases continuously with the age of the person saved. Rather, "[e]ight 60-year-olds are judged equivalent to saving one 20-year-old" and "eleven 60-year-olds are judged equivalent to saving one 30-year-old," suggesting "that the utility attached to saving an anonymous life is a hump-shaped function of the age of the person saved." *Id.* at 244-45. For a parallel survey focused on the social value of health improvements at different ages, see Jan J.V. Busschbach et al., *The Utility of Health at Different Stages in Life: A Quantitative Approach*, 37 SOC. SCI. & MED. 153 (1993). A recent review of studies asking respondents to prioritize health benefits for different members of the population finds that "most studies suggest that health gains to the old are weighted less." Paul Dolan et al., *QALY Maximisation and People's Preferences: A Methodological Review of the Literature*, 14 HEALTH ECON. 197, 202 (2005).

the LWU value of increments to longevity is smaller as individuals get older, and is a sublinear function of years saved. The monetized QALY value of increments to longevity is constant as individuals get older, and is a linear function of years saved. The traditional age-invariant VSL value is, of course, constant as individuals get older and does not vary with years saved. The age-adjusted VSL value may not vary much as individuals get older, and does not vary with years saved.

The fact that age-adjusted VSLs do not vary much is initially surprising, but not on reflection. If age-adjusted VSLs were a perfect proxy for LWUs, then these values would decrease with age, as long as life expectancy is decreasing.[241] But cognitive errors, the dead-anyway effect, shorter investment horizons for the aged, and other distorting effects (relative to LWUs) all mean that age-adjusted VSLs need not decrease with age, even when life expectancy does. Alberini et al., in a contingent-valuation study, found that WTP to avoid mortality risk did not decline with age among Americans, even after age seventy.[242] Kerry Smith et al., in one estimation based on wage data, found that VSL *increases* with age.[243] Aldy and Viscusi, in a more recent wage study, estimate an "inverted U" relationship, with VSL increasing and then decreasing with age. For example, individuals aged eighteen to twenty-two have a VSL of $3.13 million; VSL *increases* until age twenty-eight to thirty-two (where it equals $5.76 million) and then decreases, reaching $2.51 million for individuals aged fifty-eight to sixty-two.[244]

Finally, if age-adjusted VSLYs are calculated by dividing the age-adjusted VSL by discounted life-expectancy, these amounts, too, will be an "inverted U" function of age;[245] and, for a given age, will be a linear function of years saved.[246]

---

241. More precisely, the LWU value of life remaining should decrease with age, assuming life expectancy decreases and the value of background characteristics does not change.

242. Alberini et al., *supra* note 104, at 771. WTP declined among Canadians after age seventy.

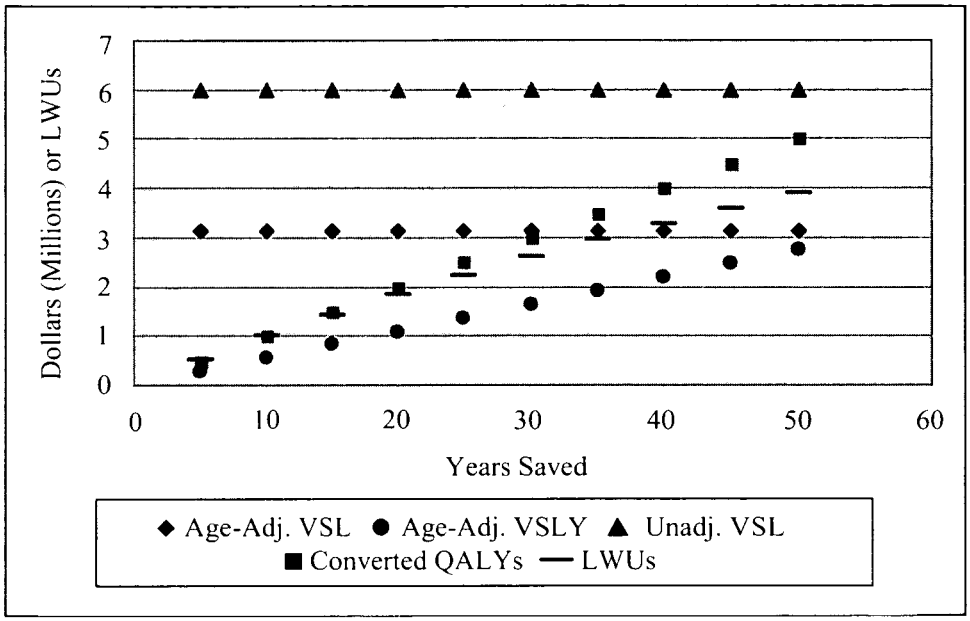
243. V. Kerry Smith et al., *Do the Near-Elderly Value Mortality Risks Differently?*, 86 REV. ECON. & STAT. 423, 427-28 (2004).

244. *See* Joseph E. Aldy & W. Kip Viscusi, *Age Variations in Workers' Value of Statistical Life* 19-23, 42, 49-50 (Nat'l Bureau of Econ. Research, Working Paper No. 10,199, 2003), *available at* http://www.nber.org/papers/w10199. For reviews of the literature on how VSL varies with age, see *id.* at 1-4; Hammitt, *QALYs Versus WTP*, *supra* note 1, at 992-94; and Revesz & Stavins, *supra* note 191, at 21. Two recent studies are Thomas J. Kniesner et al., *Life-Cycle Consumption and the Age-Adjusted Value of Life* (Harvard John M. Olin Ctr. for Law, Econ., & Bus., Discussion Paper No. 459, 2004), *available at* http://ssrn.com/abstract=580761; and Anna Alberini et al., *Willingness To Pay To Reduce Mortality Risks: Evidence from a Three-Country Contingent Valuation Study* (Fondazione Eni Enrico Mattei, Working Paper No. 2004.111, 2004), *available at* http://www.feem.it/NR/rdonlyres/8904A715-57A3-4FDD-A7E9-52318537EEFF/1258/11104.pdf.

245. *See* Aldy & Viscusi, *supra* note 244, at 23-24.

The following graphs suggest how LWUs, and the different proxies for LWUs, might correlate with longevity and age.[247]

FIGURE 5: VALUATIONS OF INCREASING INCREMENTS TO LONGEVITY FOR INDIVIDUALS OF CONSTANT AGE (20 YEARS OLD)



246. The age-adjusted VSLY technique, in its simplest variant, determines an age-specific VSLY for each age, and then monetizes the change to someone's longevity induced by a policy choice by multiplying the life-years saved or lost by a single age-adjusted VSLY determined by the individual's age in the status quo. It is this variant of the approach that is linear in changes to longevity, and that I focus on in this Section. More complicated variants would incorporate a discount rate into the policy analysis or use a different age-specific VSLY for each year added or lost (so that the first incremental year of a policy that adds three years to the life of a fifty year-old would be valued at the age fifty VSLY, the second year at the age fifty-one VSLY, and the third year at the age fifty-two VSLY). I do not evaluate these refinements to the approach here.

247. In the first graph, the individuals are aged twenty, and five- to fifty-year increments to longevity are valued using converted QALYs ($100,000 per QALY); unadjusted VSL ($6 million); age-adjusted VSL (from Aldy & Viscusi, *supra* note 244, at 42); age-adjusted VSLYs (calculated by dividing the age-adjusted VSL by the age-specific life expectancy without discounting—since nondiscounted QALYs and LWUs are employed—and then multiplying by the increment to longevity); and LWUs (on the assumption that the LWU value of a life is the square root of its length). In the second graph, the individuals range from age twenty through sixty, and a ten-year increment to longevity is valued using the same measures.

FIGURE 6: VALUATIONS OF A CONSTANT (10 YEAR) INCREMENT TO LONGEVITY

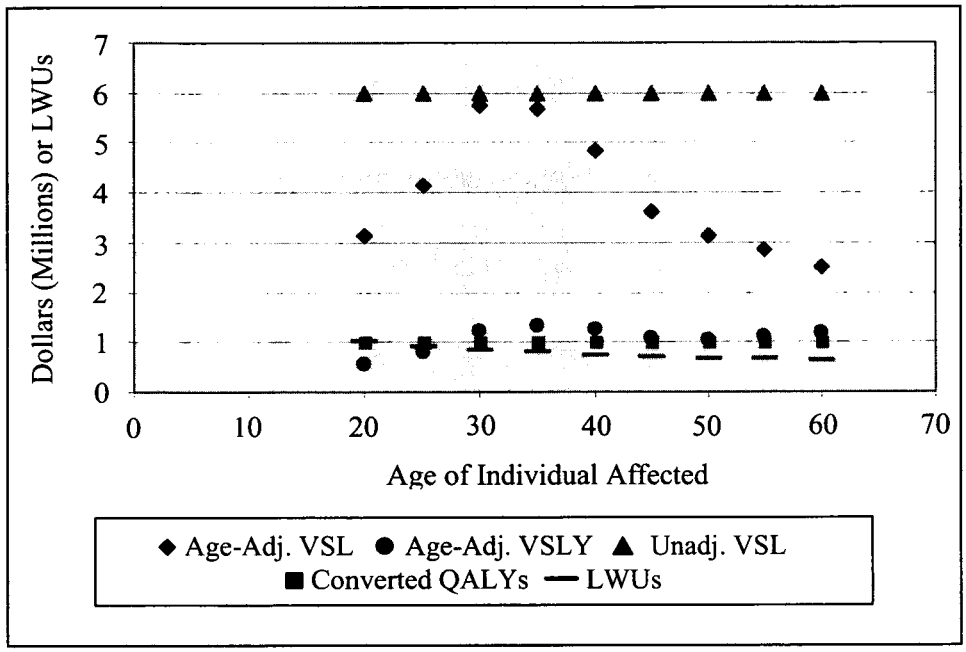


TABLE 2: VARIANCE IN THE DOLLARS/LWU RATIO[248]

Varying Increments to Longevity (20 Years), Individual of Constant Age		Constant Increment to Longevity (10 Years), Individuals of Varying Age	
Dollar Measure	*Variance Log (Dollars/LWU) in Ascending Order*	*Dollar Measure*	*Variance Log (Dollars/LWU) in Ascending Order*
Converted QALY	0.0017	Converted QALY	0.0045
Age-Adj. VSLY	0.0017	Unadj. VSL	0.0045
Unadj. VSL	0.0683	Age-Adj. VSL	0.0124
Age-Adj. VSL	0.0683	Age-Adj. VSLY	0.0282

These analyses suggest that, for sets of choice situations where the age of persons whose longevity is affected by agency decisions is constant, but increments to longevity vary, QALYs and age-adjusted VSLYs will correlate more closely with LWUs than do age-invariant VSLs or age-adjusted VSLs. In other words, the variance in the QALY/LWU and VSLY/LWU ratio will be

248. For each set of longevity changes (constant age and increasing increments to longevity, or constant increments and increasing age), I calculated the log QALY/LWU, log VSL/LWU, log adjusted VSL/LWU, and log VSLY/LWU ratio for each change, and then calculated the variance.

lower than the variance in the VSL/LWU or adjusted VSL/LWU ratio.[249] For sets of choice situations where increments to longevity are held constant but age varies, QALYs and age-invariant VSLs will correlate more closely with LWUs than do age-adjusted VSLYs and age-adjusted VSLs.

What if background characteristics are allowed to vary? Consider the simplest version of this case, where the individuals are healthy and vary in the value of their background characteristics but the longevity changes do not change the value of those characteristics. The LWU value for the loss or addition of a life-year to an individual will be the product of a person-specific scaling factor (for background characteristics) and an amount that decreases as the individual gets older.[250] The QALY value for the loss or addition of a life-year to an individual will be constant, regardless of the individual's background characteristics and age. QALYs, in this context, are clearly imperfect proxies for LWUs. Age-invariant VSLs, age-adjusted VSLs, and age-adjusted VSLYs could clearly be better proxies for LWUs if they were adjusted for background characteristics: if individuals with greater consumption, or better sex lives, were given greater VSLs, age-adjusted VSLs, or age-adjusted VSLYs. Agencies do not thus adjust VSLs, and certainly should consider doing so.[251] Until they do, however, it would seem that hybrid cost-benefit analysis incorporating monetized QALYs is a better way to value lifesaving than the alternatives.

This conclusion, I should stress, is no more than an educated guess. Much more analysis remains to be done. My rough and ready treatment in this Section considered only the two limiting cases where the number of years saved or lost by agency action varies across choice situations but the age of those affected does not, or vice versa. More realistically, the instructions that OMB or other system-designers provide to agency decision-makers about the valuation of longevity will cover a more heterogeneous range of choice situations—where both the number of years saved or lost and the age of those affected, as well as their background characteristics, can vary. The most important point of this Section is not the substantive recommendation to use QALYs in valuing longevity, but is rather methodological: *None* of the alternatives on the table (QALYs, age-adjusted VSLYs, age-invariant VSLs, or age-adjusted VSLs) are perfect proxies for LWUs when it comes to changes in longevity, and the choice

249. This statement about variance assumes, of course, that QALYs have been converted to dollars using an appropriate conversion factor. Otherwise the QALY/LWU ratio could have a different variance from the VSLY/LWU or VSL/LWU variance simply because of the difference in units. In the variance tables above I have used the variance in the logarithm of the dollar/LWU ratio to wash out the scaling effect.

250. *See supra* text accompanying notes 100, 164-165.

251. *See* Cass R. Sunstein, *Valuing Life: A Plea for Disaggregtion*, 54 DUKE L.J. 385, 386-89 (2004).

between them is a complicated exercise in optimization for OMB or other bodies ("system-designers") that guide agency decision-makers in performing cost-benefit analysis.

E. Recommendations and Summary

The analysis in this Part has been quite complicated, so let me summarize the main prescriptions that emerge from it. Although, in principle, administrative decision-makers could seek to maximize welfare directly—by measuring the welfare effects of their choices on an LWU scale, rather than a dollar scale—this approach would amount to a radical change in current policy-analytic practice. I have therefore focused on a less radical possibility: nontraditional or "hybrid" cost-benefit analysis, where the scale for measuring policy impacts *is* a monetary scale, but the dollar amounts for certain welfare effects, such as health, are calculated by converting QALYs to dollars (or in some other nontraditional way) rather than by employing the traditional WTP/WTA methodology.

"Hybrid" cost-benefit analysis is no pipe dream. The FDA has used the approach in almost twenty rulemakings. This Part has described the FDA's activities, and has taken a first stab at analyzing how hybrid cost-benefit analysis should be structured. To begin: At what rate should QALYs be converted to dollars? On this issue, I have criticized the FDA's approach (which is to derive the conversion factor from a VSL value) and have argued that the choice of conversion factor is, rather, a pragmatic matter of maximizing expected LWUs. The system-designer who specifies the conversion factor (OMB, or the Administrator or policy office of a particular agency) will need to estimate, at least in a rough and ready way, how QALY measures of health effects and WTP/WTA measures of non-health effects correlate with LWUs, and should choose a QALY-to-dollar conversion factor that maximizes expected LWUs given these estimates. There is no natural rate at which QALYs convert to dollars. The conversion factor should be seen not as a mirror of some spurious economic reality, but rather as a numerical setting chosen to optimize the functioning of a decision-making tool, namely cost-benefit analysis.

What role should QALY-to-dollar conversions play in cost-benefit analysis? First, these numbers should be used in lieu of WTP/WTA amounts where WTP/WTA data are not available. Second, where both QALY and WTP/WTA measures of a particular type of health effect are available, agencies should conduct cost-benefit analysis using both measures. If the policy prescriptions resulting from the parallel analyses differ, the agency should reexamine its WTP/WTA data. (In effect, then, the QALY measure functions as a rough test of the value of acquiring further WTP/WTA information.) If the divergence persists, then the agency will need to choose between the WTP/WTA and converted QALY measure. That choice, like the initial choice of conversion factor, is a

pragmatic one—a matter of maximizing expected LWUs. System-designers should instruct agencies to be guided by hybrid over traditional cost-benefit analysis, or vice versa, depending on the designers' estimates or rough guesses of the correlation between WTP/WTA, QALYs, and LWUs.

For all the reasons discussed in Parts III and IV, neither QALYs nor WTP/WTA amounts are perfect proxies for LWUs. In some contexts, traditional cost-benefit analysis will be a superior tool for maximizing overall welfare. In other contexts—particularly where background characteristics such as wealth do not vary much among those affected by policy choice, or where available WTP/WTA data are not sensitive to background characteristics—nontraditional cost-benefit analysis will be superior. For reasons discussed in Section V.D., measuring the cost of death may well be a policy context where nontraditional cost-benefit analysis *is* superior. The existing literature on the choice between QALYs, age-invariant VSLs, age-adjusted VSLs, and age-adjusted VSLYs completely misses the crucial point that *none* of these measures are perfectly correlated with the welfare value of longevity. Longevity, like money and many other goods, quite likely has a diminishing marginal impact on individual and overall welfare. Thus the LWU value of a given life expectancy decreases as individuals age, but the QALY and age-invariant VSL values remain constant, and the age-adjusted VSL or VSLY values may increase.[252] And both age-invariant and age-adjusted VSLs, by contrast with QALYs and LWUs, are not sensitive to changes in the amount of life expectancy holding age constant. This admittedly preliminary analysis suggests that, absent adjustments to VSL or VSLY values for wealth or other background characteristics, QALYs are probably a better measure of the value of lifesaving than alternatives.

CONCLUSION

This Article has provided a novel, *welfarist* view of QALYs. Although the academic literature on QALYs is huge, encompassing a wealth of survey data, cost-effectiveness research, and ancillary analysis,[253] QALYs have (until recently) been little used by governmental bodies in the United States[254]—in part because their policy role has been poorly understood. Welfare economists and

252. As mentioned earlier, age-adjusted VSLs and age-adjusted VSLYs both may have an "inverted U" shape, first increasing and only later decreasing with age. *See supra* text accompanying notes 244-246. If, for example, the age-adjusted VSL and VSLY for a fifty year-old are greater than for a forty year-old, both methods will place a larger money value on an X-year increment to the fifty year-old's expected longevity than on the same, X-year increment to the forty year-old's expected longevity.

253. *See supra* text accompanying notes 3-7.

254. *See supra* text accompanying notes 8-20.

other welfarists, who think that well-being-related constructs such as Kaldor-Hicks efficiency or overall welfare should play a large role in determining governmental choices, will be unpersuaded by the "extrawelfarist" perspective on QALYs that dominates the public health literature. Welfarists wonder: Why should a policymaker ever use QALYs rather than a WTP/WTA scale to measure health benefits?[255] I have offered an innovative answer to that question, making claims along the way that will undoubtedly surprise many welfarists.

In particular (here building on my prior work with Eric Posner), I have argued that policymakers should focus on overall well-being, not Kaldor-Hicks efficiency, and that cost-benefit analysis is no normative gold standard, but simply a decision-making tool that can help policymakers maximize overall welfare.[256] I have suggested that overall well-being is, in principle, measurable in LWUs (lifetime welfare units), and that QALYs and WTP/WTA amounts are both imperfect, practicable estimates of LWUs.[257] WTP/WTA valuations deviate from LWUs for a variety of reasons—wealth effects, the "dead anyway" effect, risk and tradeoff biases, and others—and although QALYs certainly have their own flaws they can, in some contexts, furnish an improved scale of welfare. I have proposed that QALYs should function, not as the nonmonetary maximand in a cost-effectiveness analysis, but rather as a valuation of health that is converted into dollars using some conversion factor and then incorporated into a monetized cost-benefit analysis. And I have analyzed how this nontraditional or "hybrid" cost-benefit analysis should be structured: in particular, what the optimal QALY-to-dollar conversion rate is, and when QALY-to-dollar conversions should displace WTP/WTA amounts as the monetary measure of health.[258]

255. *See supra* text accompanying notes 31-59 (discussing welfarist and extrawelfarist views of QALYs).

256. *See supra* text accompanying notes 57-59.

257. *See supra* Parts II-IV.

258. *See supra* Part V.

Appendix I: Cost-Effectiveness Analysis, and How It Can Deviate from Traditional Cost-Benefit Analysis[259]

The following example illustrates the two variants of cost-effectiveness analysis (fixed budget and cut-off ratio), using QALYs as the metric of effectiveness, and demonstrating how both variants can deviate from traditional cost-benefit analysis.

Imagine that a governmental body can implement two general programs, which are not mutually exclusive. Each program can be implemented through a variety of mutually exclusive subprograms; if one subprogram within the group is picked, another cannot be.[260]

A subprogram is "dominated" by another within the same general program if the first has smaller total effectiveness (QALYs) and greater total costs.[261] A dominated subprogram can be eliminated from consideration; nothing is lost, along either the health or cost dimension, by replacing a dominated subprogram with the one that dominates it.

Table 1 shows the total dollar cost of each subprogram as well as its total QALY benefit, with "dominated" subprograms eliminated. The subprograms are listed in order of total effectiveness. To be clear, the numbers in this table are the total costs for each possible subprogram, relative to the status quo of inaction. For example, if subprogram *A* is picked, the total cost will be $1 million and the total QALY benefit will be 10 QALYs. If subprogram *B* is picked, the total cost will be $2 million and the total QALY benefit will be 14 QALYs. *A* and *B* are mutually exclusive, and thus cannot be jointly picked. However, *A* and *F* are not mutually exclusive. If they are jointly picked, the total cost of that policy, relative to the status quo, will be $3 million and the total QALY benefit will be 22 QALYs.

---

259. The numerical example used in this section is based on Karlsson & Johannesson, *supra* note 26.

260. By contrast with standard presentations of cost-effectiveness analysis, my example assumes, for simplicity, that the subprograms are indivisible—they cannot be partially implemented. For each program, the decision-maker has the choice of not implementing any of its subprograms or fully implementing only one of its subprograms.

261. Strictly, this should be: Subprogram *P1* is dominated by subprogram *P2* if (1) *P1*'s health benefits are less than *P2*'s and *P1*'s costs are greater than or equal to *P2*'s, or (2) *P1*'s health benefits are less than or equal to *P2*'s and *P1*'s costs are greater than *P2*'s.

TABLE 1: TOTAL COSTS AND QALYS OF POSSIBLE SUBPROGRAMS

Program I			*Program II*		
Subprograms	Total Cost (millions)	Total Benefit (QALYs)	Subprograms	Total Cost (millions)	Total Benefit (QALYs)
A	$1	10	*F*	$2	12
B	$2	14	*G*	$4	16
C	$3	16	*H*	$5.5	18
D	$4	19			
E	$5	20			

The next table shows the incremental cost, effectiveness, and cost-effectiveness ratios of each subprogram, relative to the subprogram above it in the table.

TABLE 2: INCREMENTAL COST-EFFECTIVENESS OF SUBPROGRAMS

Program I				*Program II*			
Sub-programs	ΔCost (millions)	ΔQALYs	ΔCost/ ΔQALYs	Sub-programs	ΔCost (millions)	ΔQALYs	ΔCost/ ΔQALYs
A	$1	10	$100,000	*F*	$2	12	$166,667
B	$1	4	$250,000	*G*	$2	4	$500,000
C	$1	2	$500,000	*H*	$1.5	2	$750,000
D	$1	3	$333,333				
E	$1	1	$1,000,000				

The fixed-budget variant of cost-effectiveness analysis tells the decision-maker to choose that mix of subprograms which maximizes QALYs for the given budget. Assume that, in the case at hand, the budget to be maximized is $4 million. Then the QALY-maximizing mix of subprograms is *B* and *F*. No other mix of subprograms that costs less than or equal to $4 million produces more QALYs.[262]

What about the cut-off ratio variant of cost-effectiveness analysis? The decision rule here is as follows. "Weakly dominated" subprograms are eliminated from consideration. A subprogram is "weakly dominated" if its incremental cost-effectiveness ratio is greater than the incremental cost-effectiveness ratio of the subprogram immediately below it in the table. In the current example, subprogram *C* is weakly dominated. Then, incremental cost-effectiveness ratios for the remaining subprograms are recalculated, as shown in Table 3. Finally,

262. In the simplified case at hand, the budget-maximizing mix of subprograms can be identified by inspection. In more complicated cases, if certain assumptions are made about the divisibility of subprograms and the constancy of returns to scale, the budget-maximizing mix can be identified using a decision rule that looks to incremental cost-effectiveness ratios. *See* Johanesson, *supra* note 26, at 484-85.

"the [subprogram] within each cluster that has the highest incremental cost-effectiveness ratio that is equal to or below the [cut-off ratio] should be implemented."[263] For example, if the cut-off ratio is $200,000 per QALY, then the government applying this decision rule picks subprograms *A* and *F*.

TABLE 3: INCREMENTAL COST-EFFECTIVENESS OF SUBPROGRAMS, EXCLUDING WEAKLY DOMINATED (SUBPROGRAM *C*)

Program I				*Program II*			
Sub-programs	Δ Cost (millions)	ΔQALYs	ΔCost/ ΔQALYs	Sub-programs	Δ Cost (millions)	ΔQALYs	ΔCost/ ΔQALYs
A	$1	10	$100,000	*F*	$2	12	$166,667
B	$1	4	$250,000	*G*	$2	4	$500,000
				H	$1.5	2	$750,000
D	$2	5	$400,000				
E	$1	1	$1 million				

To see how both variants of cost-effectiveness analysis can deviate from traditional cost-benefit analysis, consider the following table, which shows the total costs and the total QALYs as well as total WTP for the different subprograms in Table 1—on the assumption that Program I benefits a population all members of which are willing to pay a constant $300,000 per QALY, while Program II benefits a population all members of which are willing to pay only a constant $100,000 per QALY.

TABLE 4: PROGRAM COSTS AND PROGRAM BENEFITS (BENEFICIARIES OF PROGRAM I ARE WILLING TO PAY $300,000 PER QALY, WHILE BENEFICIARIES OF PROGRAM II ARE WILLING TO PAY $100,000 PER QALY)

Program I				*Program II*			
Sub-programs	Total Cost (millions)	Total Benefit (QALYs)	Total Benefit (WTP, millions)	Sub-programs	Total Cost (millions)	Total Benefit (QALYs)	Total Benefit (WTP, millions)
A	$1	10	$3	*F*	$2	12	$1.2
B	$2	14	$4.2	*G*	$4	16	$1.6
C	$3	16	$4.8	*H*	$5.5	18	$1.8
D	$4	19	$5.7				
E	$5	20	$6.0				

The fixed-budget variant of cost-effectiveness analysis, with a budget of $4 million, tells the government to pick subprograms *B* and *F*. But aggregate WTP

263. Johannesson, *supra* note 26, at 484.

for the health benefits would be greater if the government spent the $4 million on subprogram *D* instead ($5.7 million versus $5.4 million). The cut-off ratio variant, with a ratio of $200,000 per QALY (the average of the two populations) tells the government to pick subprograms *A* and *F*—spending a total of $3 million. But aggregate WTP would be increased by spending the $3 million on subprogram *C* ($4.8 million versus $4.2 million). The best choice of all, as per traditional cost-benefit analysis, would be to spend only $2 million on subprogram *B*. All of the Program II subprograms have greater total money costs than benefits. In Program I, by moving from subprogram *D* to *C* we save $1 million in costs and give up only $0.9 million in benefits; and by moving again from *C* to *B*, we save $1 million in costs and give up only $0.6 million in benefits.

Assume now that all individuals in both populations are willing to pay a constant $200,000 per QALY. In that case, the costs and benefits of the various subprograms are as follows:

TABLE 5: PROGRAM COSTS AND PROGRAM BENEFITS (ALL BENEFICIARIES ARE WILLING TO PAY $200,000 PER QALY)

Program I				*Program II*			
Sub-programs	Total Cost (millions)	Total Benefit (QALYs)	Total Benefit (WTP, millions)	Sub-programs	Total Cost (millions)	Total Benefit (QALYs)	Total Benefit (WTP, millions)
A	$1	10	$2	F	$2	12	$2.4
B	$2	14	$2.8	G	$4	16	$3.2
C	$3	16	$3.2	H	$5.5	18	$3.6
D	$4	19	$3.8				
E	$5	20	$4.0				

As before, the fixed-budget variant of cost-effectiveness analysis maximizes QALYs for a budget of $4 million by picking subprograms *B* and *F*. As before, the cut-off ratio variant using the $200,000 per QALY cut-off selects subprograms *A* and *F*. But this time, with WTP constant at $200,000 per QALY, traditional cost-benefit analysis also selects *A* and *F*, as can be seen by inspection.

Appendix II: Hybrid Cost-Benefit Analysis[264]

I. THE OPTIMAL QALY-TO-DOLLAR CONVERSION FACTOR

When choosing between the status quo and a project that will yield q QALYs in health benefits while costing w dollars,[265] a decision-maker given a QALY-to-dollar conversion factor f will implement the project if and only if $fq > w$. The system-designer, having a subjective probability distribution with respect to the levels of health benefits in QALYs and non-health costs in WTPs, as well as their respective ratios to LWUs, should try to determine a fixed value of f that, when used by the decision-maker, will yield the greatest overall increase in LWUs.

A. The General Case

Let A_f be defined for all $f \in R^+$ as the set of pairs $\langle q, w \rangle \in (R^+)^2$, such that $fq > w$. In a choice situation where a project will cost w dollars and yield q QALYs in benefits, a decision-maker using a fixed QALY-to-dollar conversion factor, f, will select the project if and only if $\langle q, w \rangle \in A_f$.

Let $p(r,s,q,w)$ be the probability density function of continuous random variables R, S, Q, and W where

R represents the QALY/LWU ratio for health benefits,

S represents the WTP/LWU ratio for non-health costs,

Q represents the health benefits, measured in QALYs, and

W represents the non-health costs, measured in WTP.

Then, the expected change in LWUs from the implementation of the project using any conversion factor, f, is

$$\iint_{A_f} \iint_{R^2} \left(\frac{q}{r} - \frac{w}{s} \right) p(r,s,q,w)\, dr\, ds\, dq\, dw;$$

264. Many thanks to Craig Phillips for preparing this Appendix.

265. Assume here that all of the costs are non-health costs, directly measurable in WTPs, and that all of the benefits are health benefits, directly measurable in QALYs.

and so the optimal QALY-to-dollar conversion factor will be the value of f that maximizes the above function.

B. If r *and* s *are independent of* q *and* w, *and if* r *is independent of* s, *then the optimal value of* f *will be* $f_0 = \left(\frac{E(\frac{1}{r})}{E(\frac{1}{s})}\right)$.

Because of the above independence assumptions, p can be represented as the product of p_1, the density function of R, p_2, the density function of S, and p_3, the density function of Q and W.

$$\iint_{A_f}\iint_{R^2}\left(\left(\frac{q}{r}\right)-\left(\frac{w}{s}\right)\right)p_1(r)p_2(s)p_3(q,w)\,dr\,ds\,dq\,dw$$

$$= \iint_{A_f}\left(q\int_R \frac{1}{r}p_1(r)dr - w\int_R \frac{1}{s}p_2(s)ds\right)p_3(q,w)\,dq\,dw$$

$$= \iint_{A_f}\left(qE\left(\frac{1}{r}\right)-wE\left(\frac{1}{s}\right)\right)p_3(q,w)\,dq\,dw\,.$$

When the conversion factor is set at f_0, the increase in overall LWUs is determined by integrating over the set A_{f_0}, which is the set of all $\langle q,w\rangle \in (R^+)^2$ such that $q\left(\frac{E(\frac{1}{r})}{E(\frac{1}{s})}\right) > w$ or, equivalently, where $qE(\frac{1}{r}) - wE(\frac{1}{s}) > 0$. The expected increase in overall LWUs thus reaches its maximum when the decision-maker uses f_0 as the conversion factor.

C. When (1) the probability density function, u, *of the log of the QALY/LWU ratio is symmetric about its mean, log* r_0, *increasing only when* $r < r_0$; *(2) the WTP/LWU ratio for costs is fixed; and (3) the value of the probability density function for costs and benefits,* v(c,b), *is such that* v(x,y) = v(y,x), *the optimal conversion factor,* f_*, *is* $\frac{s}{r_0}$, *where* s *is the fixed WTP/LWU ratio for costs.*

Let

s be the fixed WTP/LWU ratio for costs,
r_0 be such that u has its mean at log r_0,
f be the QALY-to-dollar conversion factor,
R be the continuous random variable representing the QALY/LWU ratio,
$u(\log r)$ be the probability density function of log R,
u be independent of v and increasing when $r < r_0$ and decreasing otherwise,
$U(x)$ be the cumulative distribution function for log R, and

the probability density function of the costs and benefits in LWUs, $v(c,b)$, be such that, for all $x,y \in R^+$, $v(x,y) = v(y,x)$.

Let $x,y \in R^+$, such that $x \geq y$.

When $b = x$ and $c = y$, there will be a loss of $x - y$ whenever $frx < sy$, or $r < \frac{sy}{fx}$.

When $b = y$ and $c = x$, there will be a loss of $x - y$ whenever $fry > sx$, or $r > \frac{sx}{fy}$.

If there is a single QALY-to-dollar conversion factor that minimizes the sum of the probabilities that $r < \frac{sy}{fx}$ and that $r > \frac{sx}{fy}$ for all $x, y \in R^+$, then that factor will be the optimal conversion factor.

The probability that $r < \frac{sy}{fx}$ is

$$\int_{-\infty}^{\log \frac{sy}{fx}} u(\log r)d(\log r) = \int_{-\infty}^{1} u(\log r)d(\log r) + \int_{1}^{\log \frac{sy}{fx}} u(\log r)d(\log r),$$

and, similarly, the probability that $r > \frac{sx}{fy}$ is

$$\int_{\log \frac{sx}{fy}}^{+\infty} u(\log r)d(\log r) = \int_{\log \frac{sx}{fy}}^{1} u(\log r)d(\log r) + \int_{1}^{+\infty} u(\log r)d(\log r).$$

In order to find the value of f that minimizes these probabilities, we differentiate the sum of these two probabilities with respect to f, so that

$$\frac{d}{df}\left(\int_{\log \frac{sx}{fy}}^{1} u(\log r)d(\log r) + \int_{1}^{\log \frac{sy}{fx}} u(\log r)d(\log r) \right)$$

$$= \frac{d}{df}\left(\int_{\log \frac{sx}{fy}}^{\log \frac{sy}{fx}} u(\log r)d(\log r) \right)$$

$$= \frac{d}{df}\left(U\left(\log \frac{sy}{fx}\right) - U\left(\log \frac{sx}{fy}\right) \right) = -\frac{1}{f}\left(u\left(\log \frac{sy}{fx}\right) - u\left(\log \frac{sx}{fy}\right) \right).$$

Let f_* be such that $u\left(\log \frac{sy}{f_* x}\right) = u\left(\log \frac{sx}{f_* y}\right)$. It follows from the above that the sum of (1) the probability of error when costs are x and benefits are y and (2) the probability of error when costs are y and benefits are x reaches its minimum for all $\langle x, y \rangle \in (R^+)^2$ when $f = f_*$. Because $u(\log r)$ is symmetrical about its mean, $\log r_0$, it follows that if $f_* = \frac{s}{r_0}$, then:

$$u\left(\log\frac{sy}{f_\bullet x}\right)=u\left(\log\frac{r_0 y}{x}\right)=u\left(\log(r_0)+\log\left(\frac{y}{x}\right)\right)=u\left(\log(r_0)-\log\left(\frac{y}{x}\right)\right)=u\left(\log\frac{sx}{f_\bullet y}\right).$$

The optimal WTP/QALY conversion factor is therefore $f_\bullet=\frac{s}{r_0}$.

II. Choosing Between QALYs and WTPs as the Measure of Health Benefits

Let the WTP/LWU ratio for costs of a set of projects be fixed at s and let the random variables R_w and R_q represent the WTP/LWU and converted QALY/LWU ratios for the health benefits, respectively. Assume that R_w and R_q are independent of the levels of costs and benefits in LWUs. Then, if the probability density functions, $p_w(r)$ of R_w and $p_q(r)$ of R_q, are such that $\int_{-\infty}^{x} p_q(r)dr < \int_{-\infty}^{x} p_w(r)dr$ for every $x<s$ and $\int_{x}^{+\infty} p_q(r)dr < \int_{x}^{+\infty} p_w(r)dr$ for every $x>s$, then using a converted QALY measurement of benefits, rather than a direct dollar measurement, will yield greater expected overall welfare in LWUs.

If $c > b$, then the project will be implemented at a loss if $sc < rb$, or $r > \frac{sc}{b}$, when r is the ratio of the monetized benefit measure (WTP or converted QALYs) to LWUs. By hypothesis, because $\frac{sc}{b} > s$, the probability that the project will be implemented at a loss is greater when measuring benefits directly in WTP than with converted QALYs. Analogous reasoning yields the same result where $c < b$, and so measuring benefits in converted QALYs will yield greater expected overall LWUs than a direct dollar measurement.

Conversely, if $\int_{-\infty}^{x} p_w(r)dr < \int_{-\infty}^{x} p_q(r)dr$ for every $x < s$ and $\int_{x}^{+\infty} p_w(r)dr < \int_{x}^{+\infty} p_q(r)dr$ for every $x > s$, then, by analogous reasoning, using a direct dollar measurement of benefits will yield greater overall welfare in LWUs than a converted QALY measurement will yield.

Managed Process, Due Care: Structures of Accountability in Health Care

Nan D. Hunter, J.D.*

INTRODUCTION

Almost unnoticed, a new kind of adjudication system has appeared in American law. In forty-one states and the District of Columbia, special entities have been established to resolve contract and tort claims. State law created and mandates each system; these are not arbitrations agreed to by contract between the parties. Despite their public nature, however, these systems are not offered or operated by courts; the public function of adjudication is entirely outsourced to private actors. The decision-makers are neither elected nor appointed, nor are they public sector employees; they work in private companies. Most do not write opinions, and they neither establish nor follow precedent.

These new entities are the external review systems set up to resolve disputes between patients and managed care organizations (MCOs), which arise when such organizations deny coverage for medical treatment, services, or equipment that the patient, generally upon the recommendation of a physician, believes to be medically necessary. When pre-authorization for care is required, the coverage decision about whether to pay also becomes in effect a treatment decision that determines whether the care will ever be rendered. The payor's decision merges with and trumps what used to be solely the treating physician's decision. It is this massive shift in the ramifications of these disputes that has caused lawmakers to pay much greater attention to the processes for resolving them.

This Article examines how legal, political, and economic change produced a new adjudicatory mechanism for resolving disputes between patients and MCOs. It is in essence a case study of the multiple determinants of procedure in a particular field—health law—at a time of rapid change in the underlying

* Professor of Law and Co-Director, Center for Health, Science and Public Policy, Brooklyn Law School. I received support for the writing of this Article from the Brooklyn Law School Summer Research Program. Robin Fukuyama, Melissa Gable, and Emily Kern provided excellent research assistance. Many thanks for feedback and support to Sylvia Law and, especially, Chai Feldblum.

industry. We will see how these changes, plus a broader cultural shift toward less acceptance of professional omniscience, are reshaping our core concepts of medical decision-making in an increasingly corporate environment. Through close analysis of this example, we can also learn a significant amount about newly-emerging models of procedural justice.

The Article proceeds in the following steps: Part I describes the complex cluster of issues implicated in the adjudicatory procedures for health-care-related disputes. These cases are extraordinarily rich in consequences, because they concern two issues in addition to procedures for dispute resolution: the allocation of health care resources and the quality of medical care. Part II analyzes the external review systems in the context of what I call health process exceptionalism: the unexamined seesaw between principles of due process and the tradition of deference to medical authority in health law decision-making systems. It explores the curious way in which deference to medical judgment substituted for the structural protections of procedural due process and then, in a kind of doctrinal blowback, how due process norms came to trump deference. It places the new external review laws in that context, as reflecting the ascendancy of a due process model. Part III describes the new external review adjudication systems that have emerged from these reform campaigns, focusing on their procedural law components (such as scope of jurisdiction, standards of review, and remedies) and their legislative and political histories. It then applies due process criteria to these systems. I argue that these laws fully satisfy neither the due process nor the deference paradigm, nor are they likely to enhance the quality of medical care, but instead reflect the trend to hybrid public-private institutions in law and governance.

Part IV of the Article asks how changes in health law relate to issues raised by broader theories of process. Traditionally, a focus on the right to a hearing prior to a deprivation of liberty or property held center court in process discourse. I argue that external review systems signal the emergence of new process values in procedural due process theory. In the context of private sector disputes, rather than challenges to actions of the government, accountability carries greater significance as a process value than does the right to a hearing. This Part seeks to reinvigorate the theoretical foundations for procedural due process, an area largely abandoned since debates over dignitarian values in process petered out in the late 1970s. Finally, Part V applies process values to external review systems and concludes that there is no persuasive evidence that they have achieved the primary function for which they were created, i.e., providing accountability. I then return to the phenomenon of health process exceptionalism to analyze its future directions.

I. HEALTH, LAW, AND PROCESS

When one considers the process aspects of health law, important questions abound. Should the resolution of disputed medical issues in law be delegated to professional experts, or resolved by traditional adjudicatory methods? Who should determine which medical issues get decided in which way? Should it depend on the medical characteristics of the particular issue or on the nature of the legal questions presented? Are the individuals whose treatment is at issue to be conceptualized as rights-bearing autonomous citizens or as semi-dependent patients to whom fiduciary duties are owed? Each of these questions arises at the individual case level, and also carries broader ramifications. At the individual level, resolutions of health care disputes implicate bodily integrity and autonomy concerns as well as major financial consequences. At the societal level, the concerns resonate with broad normative echoes, as well as huge costs.

Conceptualizing the process aspects of health law at the systemic level, however, presents an even thornier set of problems. Perhaps more so than in any other field, health law process questions reverberate with policy ramifications in multiple and divergent realms. The design and operation of an optimal process for resolving health-related disputes implicates three arenas of policy-making:

(1) The processes by which health-related disputes are decided requires articulation of, and choices among, various norms of procedure. One finds in procedure theory a rich debate among proponents of models emphasizing dignitarian, accuracy, or efficiency goals. Considering the relationship among these models in a health care context is a daunting challenge because of the technical nature of many of the questions involved.

(2) Because the context is health care, enormous social concern attaches to quality control issues, a concern which affects the structure of decision-making procedures. Not only must a procedural system be designed to achieve maximum accuracy in any given case, but the overall impact of how disputed questions about what is proper treatment are decided, across many cases, should operate to improve standards of care across the board.

(3) Finally, the impact of the economics of health care far exceeds that of most kinds of cost-benefit outcome assessments. Determinations about whether certain treatments fall within the range of what is considered standard medical care constitute direct allocation of health care resources. The law of medicine not only governs standards for the care of individual sick people, but also controls the distribution of massive benefits and burdens. Review of managed care treatment denial decisions amounts to review of rationing

choices made at the level of individual cases.

To engage in a full-scale consideration of the process side of health law leads one into an extraordinary juggling act. Decisions about whether a certain health care service is medically necessary, and therefore will be covered under an insurance contract, are multiply mixed decisions. Not only are they mixed coverage and treatment decisions (because treatments which are unauthorized are unlikely to be provided), as the Supreme Court has realized,[1] but they also function as mixed treatment and rationing decisions.

Procedural theorists do not usually engage the issues outlined in points two and three. Nor does the main body of work on either health care financing or quality address questions of the procedures of adjudication. Even within health law, analysis of process models paying simultaneous attention to these three complex sets of issues is rare. Debates on health care reform in 1993 and 1994 generated a burst of writing about the procedural mechanisms of health care systems.[2] Since then, however, there has been relatively little attention paid to the decision-making procedures for resolving competing claims about whether particular health care treatments are medically necessary, the universal standard for coverage decisions.[3]

---

1. Pegram v. Herdrich, 530 U.S. 211, 229 (2000).

2. Although procedural issues received little attention in the press reports about health care reform proposals, the moment presented an opportunity "to construct an analytic framework based in legal process theory and health policy analysis within which to evaluate mechanisms for resolving individual claims to health care and treatment." Margaret G. Farrell, *The Need for a Process Theory: Formulating Health Policy Through Adjudication*, 8 J.L. & HEALTH 201, 202 (1994); *see also* Mark R. Fondacaro, *Toward a Synthesis of Law and Social Science: Due Process and Procedural Justice in the Context of National Health Care Reform*, 72 DENV. U. L. REV. 303 (1995); Timothy S. Jost, *Administrative Adjudication Under Health Care Reform*, 47 ADMIN. L. REV. 425 (1995); Eleanor D. Kinney, *Protecting Consumers and Providers Under Health Reform: An Overview of the Major Administrative Law Issues*, 5 HEALTH MATRIX 83 (1995); Sallyanne Payton, *Medical Rationing and the Allocation of Adjudicatory Responsibility Under Comprehensive Health Care Reform in the 103rd Congress: An Administrative Lawyer's Postmortem*, 47 ADMIN. L. REV. 381 (1995); Rand E. Rosenblatt, *Equality, Entitlement, and National Health Care Reform: The Challenge of Managed Competition and Managed Care*, 60 BROOK. L. REV. 105 (1994). Unfortunately, the moment passed.

3. The primary exception is a series of articles by Eleanor Kinney. *See* Eleanor D. Kinney, *Tapping and Resolving Consumer Concerns About Health Care*, 26 AM. J.L. & MED. 335 (2000); Eleanor D. Kinney, *Consumer Grievance and Appeal Procedures in Managed Care Plans*, 3 HEALTH LAW. 17 (1998); Eleanor D. Kinney, *Resolving Consumer Grievances in a Managed Care Environment*, 6 HEALTH MATRIX 147 (1996); Eleanor D. Kinney, *Procedural Protections for Patients in Capitated Health Plans*, 22 AM. J.L. & MED. 301 (1996). These articles have been compiled into a book: ELEANOR DEARMAN KINNEY, PROTECTING AMERICAN HEALTH CARE CONSUMERS (2002). *See also* William M. Sage, *Managed Care's Crimea: Medical Necessity,*

Moreover, the dynamics of private markets do not mitigate the problems of design, because the health care market is famously dysfunctional in multiple ways.[4] Complaints about access to health care arise within what is usually a closed delivery system. The nature of health insurance in the United States is such that virtually everyone of working age who has private health insurance purchases it through an employer, as a member of a defined group consisting of that employer's workforce.[5] This situation creates a market context which is fundamentally different from the situation of a consumer of automobiles or houses or appliances, who can make purchases from among a large and changing assortment of sellers.

Even if a plan provides relatively free choice of providers, only very large employers enable individuals to choose from among different insurers.[6] Among all working age persons with health insurance, roughly half will be limited to a single insurance plan.[7] They will not have the option to choose a plan offered by a different insurance company, which might have different procedures and policies for determining what is medically necessary. Even those who do have a choice may be unable to obtain the information needed to assess differences between plans in claim adjudication systems. Such a limited market is particularly needful of strong procedural rules to assure that imbalances in power are not abused.[8]

These peculiarities of the health care delivery system produce two important

---

Therapeutic Benefit, and the Goals of Administrative Process in Health Insurance, 53 DUKE L.J. 597 (2003).

4. The classic and earliest statement is Kenneth J. Arrow, *Uncertainty and the Welfare Economics of Medical Care*, 53 AM. ECON. REV. 941 (1963). For a succinct point and counter-point analysis, see Frank A. Sloan & Mark A. Hall, *Market Failures and the Evolution of State Regulation of Managed Care*, 65 LAW & CONTEMP. PROBS. 169, 172-83 (2002). Scholars with widely divergent policy philosophies agree on the diagnosis of dysfunction. *See, e.g.*, CLARK C. HAVIGHURST, HEALTH CARE CHOICES: PRIVATE CONTRACTS AS INSTRUMENTS OF HEALTH REFORM (1995); PAUL STARR, THE LOGIC OF HEALTH CARE REFORM (1992).

5. Wendy K. Mariner, *Can Consumer Choice Plans Satisfy Patients? Problems with Theory and Practice in Health Insurance Contracts*, 69 BROOK. L. REV. 485, 489 (2004).

6. James Maxwell et al., *Corporate Health Care Purchasing Among Fortune 500 Firms*, 20 HEALTH AFF. 181 (2001).

7. JEANNE M. LAMBREW, THE COMMONWEALTH FUND, ISSUE BRIEF: "CHOICE" IN HEALTH CARE: WHAT DO PEOPLE REALLY WANT? 8 (2005), *available at* http://www.cmwf.org/usr_doc/lambrew_853_choice_ib.pdf; David A. Hyman & Mark Hall, *Two Cheers for Employment-Based Health Insurance*, 2 YALE J. HEALTH POL'Y L. & ETHICS 23, 27 (2001); Thomas Rice et al., *Workers and Their Health Care: Free To Choose?*, 21 HEALTH AFF. 182 (2002).

8. See discussion of agency costs caused by employer purchasing in Gail B. Agrawal, *Resuscitating Professionalism: Self-Regulation in the Medical Marketplace*, 66 MO. L. REV. 341, 370-72 (2001).

results. First, the typical market-based accountability mechanism of supply and demand does not operate well. In other markets, consumers who are dissatisfied with a given product or service can elect to purchase from another manufacturer or vendor if they decide to buy the same kind of item again. If your Dell laptop produces sub-par performance, you can always buy a ThinkPad the next time out, or even switch to Apple. If you encounter problems with your health plan covering a treatment that you will need repeatedly, however, you cannot simply opt out of the plan.

Second, patients' lack of an exit option compounds the vulnerability that individuals involved in such disputes are likely to experience. Disputes over whether insurers will cover a particular course of treatment often arise in the context of expensive treatments, when denial of pre-authorization has the greatest impact on cost containment.[9] Patients facing such disputes tend to be sufficiently ill that they are not in command of their full faculties and can be especially susceptible to manipulation. Moreover, they are often simply frightened and do not want to expend their limited energy arguing. Even when health care consumers are not ill, deficits in information limit their ability to exercise maximum rationality in making health care decisions.[10]

To some extent, as consumer dismay has translated into market pressures, MCOs have responded to ameliorate these problems. The demand for greater choice of providers has led to the dominance of preferred-provider models, which allow reimbursement of out-of-network physicians with less of a penalty than some other forms of MCOs.[11] To settle litigation brought by the Texas Attorney General, Aetna agreed to limit the financial incentives it offered to doctors who recommended fewer higher-cost treatments.[12] Yet it is doubtful that these kinds of adjustments will adequately address the most serious process and equity questions, given that the structure of the market and the power imbalances among market participants remain unchanged.

---

9. Issues related to hospitalization—including surgery, length of hospital stay, and inpatient mental health treatment—are among the most common bases for external reviews, as well as disputes related to durable medical equipment and pharmaceuticals. Leatrice Berman-Sandler, *Independent Medical Review: Expanding Legal Remedies To Achieve Managed Care Accountability*, 13 ANNALS HEALTH L. 233, 254-55 (2004).

10. Russell Korobkin, *The Efficiency of Managed Care "Patient Protection" Laws: Incomplete Contracts, Bounded Rationality, and Market Failure*, 85 CORNELL L. REV. 1 (1999). For a comparison of "patients" and "consumers," see Mariner, *supra* note 5, at 491-95.

11. Jon Gabel et al., *Health Benefits in 2004: Four Years of Double-Digit Premium Increases Take Their Toll on Coverage*, 23 HEALTH AFF. 200, 205 (2004); James C. Robinson, *The End of Managed Care*, 285 JAMA 2622, 2623-24 (2001).

12. M. Gregg Bloche & Peter D. Jacobson, *The Supreme Court and Bedside Rationing*, 284 JAMA 2776, 2778 n.27 (2000).

Decisions about whether a treatment will be covered constitute micro-rationing, with ramifications at multiple levels: for the particular patient, for the profession of health care, and for the society's allocation of resources.[13] The procedures by which such decisions are made merit much greater attention and analysis than they have yet received. This Article seeks to articulate a new analytic framework for assessing and understanding this area where procedural theory, professional responsibilities, and social values overlap.

The exact dimensions of the problem of denials by MCOs are unknown. According to a public opinion survey conducted by the Kaiser Family Foundation and the Harvard University School of Public Health in 2001, approximately twelve percent of adults with private health insurance report that they have experienced denials of coverage or care.[14] In a 1998 survey, twenty-two percent of respondents said that they had wanted—but had been unable—to appeal a denied claim to an independent reviewer, and thirteen percent said that they had wanted to sue a health plan for malpractice.[15] A nationally representative sample of managed care enrollees in 1999 found that less than twenty percent of patients who experienced serious health declines or incurred large unreimbursed expenses because of perceived problems with their health plans filed a formal appeal or grievance.[16] One federal judge, known for his expertise in procedural justice, described "disputes about the new breed of Health Maintenance Organizations' denials of service and Medicare and Medicaid payments" as being "for many millions, among the most critical of their grievances over what they consider to be denials of substantive rights."[17]

The Supreme Court's decision in *Rush Prudential HMO, Inc. v. Moran*[18]

---

13. The use of medical necessity decisions as a rationing device occurred prior to the widespread adoption of managed care. *See* Timothy P. Blanchard, *"Medical Necessity" Denials as a Medicare Part B Cost-Containment Strategy: Two Wrongs Don't Make It Right or Rational*, 34 ST. LOUIS U. L.J. 939 (1990). However, the pricing and profit structure of managed care exacerbate the incentives for rationing.

14. KAISER FAMILY FOUND. & HARVARD SCH. OF PUB. HEALTH, NATIONAL SURVEY ON CONSUMER EXPERIENCES WITH AND ATTITUDES TOWARD HEALTH PLANS: KEY FINDINGS 1 (2001), *available at* http://www.kff.org/kaiserpolls/pomr111704pkg.cfm.

15. KAISER FAMILY FOUND., KAISER PUBLIC OPINION UPDATE 2 (2000).

16. Mark Schlesinger et al., *Voices Unheard: Barriers to Expressing Dissatisfaction to Health Plans*, 80 MILBANK Q. 709, 731-32 (2002).

17. Jack B. Weinstein, *Adjudicative Justice in a Diverse Mass Society*, 8 J.L. & POL'Y 385, 400-01 (2000) [hereinafter Weinstein, *Adjudicative Justice*]. Judge Weinstein is well known for his thoughtful responses to the challenges of adjudicating mass tort cases. *See* JACK B. WEINSTEIN, INDIVIDUAL JUSTICE IN MASS TORT LITIGATIONS: THE EFFECTS OF CLASS ACTIONS, CONSOLIDATIONS, AND OTHER MULTI-PARTY DEVICES (1995).

18. 536 U.S. 355 (2002).

marked a crucial point in the establishment of external review systems as new structures for adjudication of these disputes. In *Moran* the Court upheld an Illinois law which established a system of independent external review when a managed care organization denied coverage for a treatment. The decision turned on whether the state law was preempted by the Employee Retirement Income Security Act (ERISA), a federal statute governing benefit systems offered by private employers.[19] The Court ruled that the Illinois law was not preempted by ERISA because it functioned as "insurance regulation"—a categorical exemption from ERISA's preemptive scope for insurers (except for self-insured plans).[20] The result not only legitimated Illinois's law, but sent at least a conditional go-ahead signal for other states to pursue external review schemes.[21]

At one level, these state-level external review systems represent an attempt to restore medical authority as the trump card in medical decision-making. The means used, however, is an adjudicatory system, and the consequences of these decisions parallel those of other adjudicatory systems. External review organizations can be the crucial venue in deciding, for example, the medical status of administering a particular treatment for a particular condition, whether denying such treatment would fall below an acceptable standard of care, or whether patients have been adequately apprised of their rights to review within the system.

External review systems offer a fascinating case study of both the jurisprudence of health law and a contemporary undertaking to create rules of procedure. Because they are almost universal, because they are the first new procedural mechanism produced by the ascendancy of managed care, and because recalibrating the role of professional authority is the core of their function, such systems provide a powerful window into the jurisprudence of health law. In addition, external review schemes implicate many of the recurring issues of procedure debates: the tension between expertise and democratic values, the role of specialized courts, the decrease in written decisions, and the truncation of collective action, to name a few.

What has resulted from this undertaking is a public-private adjudication system. External review laws channel appeals from MCO denials to private judging companies via a statutorily established process. They signal the ascension of a market model of adjudication embedded within the architecture of

19. 29 U.S.C. §§ 1101-1461 (2000).

20. 29 U.S.C. § 1144(b)(2) (2000); *see* Corporate Health Ins., Inc., v. Tex. Dep't of Ins., 314 F.3d 784, 786 (5th Cir. 2003) (relying on *Moran* to limit preemption of external review to self-funded ERISA plans).

21. See *infra* text accompanying notes 166-169 for a discussion of why other external review laws may still be vulnerable to ERISA challenges.

public law, a fundamentally new form of adjudication, based on a different conception than that underlying the more familiar parallel universe of arbitration.

II. THE HISTORY OF HEALTH PROCESS EXCEPTIONALISM

Because health law is a field in which decisions often carry enormous consequences, one would expect that a legally enforceable and externally binding decision about whether certain treatments were medically appropriate would trigger a parallel degree of heightened concern with the procedures by which the decisions were made. But that has not been the case. Instead of especially careful procedures, deference to the judgment of physicians has become instantiated in a variety of legal contexts, creating in effect a distinctly different process paradigm for medical disputes. This has occurred largely because physicians' decisions are presumed to be based solely on the interests of the patient and to have built-in protections against error, thus obviating the need for traditional procedural due process protections. This norm of deference to physician judgment constitutes what I call "health process exceptionalism."

Deference originated as a substantive standard first in tort and then in contract law subfields of health law, and later came to also short-circuit process protections in health law. A range of factors—the economic costs of health care, the institutional structure of managed care systems, patient empowerment movements,[22] and physicians' stands on controversial issues[23]—have now coalesced to flip the model, so that the due process concepts associated with procedural protections now operate as a substitute for deference to medical authority.

A. Deference to Medical Authority as Substance and Substitute for Due Process

Across multiple fields of law, and across public and private law, courts have ceded to medical professionals the authority to decide whether a given treatment for an individual patient is (or was) medically necessary and appropriate. From this practice grew a principle of substantive law, initially in tort, that the law would defer to the reasonable judgments of physicians. However, that substantive

22. Groups of patients, such as AIDS and cancer patients, have begun demanding greater autonomy vis-à-vis the medical establishment, and the Internet has enabled a massive increase in access to medical information. Both of these developments have contributed to widespread medical populism. *See generally* GEORGE J. ANNAS, THE RIGHTS OF PATIENTS: THE AUTHORITATIVE ACLU GUIDE TO THE RIGHTS OF PATIENTS 3-24, 365-67 (3d ed. 2004).

23. The Supreme Court's view of physicians who perform abortions has shifted dramatically. *See* Nan D. Hunter, *Medicine as Politics: Justice Blackmun, Abortion and the Myth of Independence*, 71 BROOK. L. REV. (forthcoming 2006).

principle spread beyond its initial scope, as courts came to invoke it routinely as a substitute for those procedures considered, as both legal and cultural norms, to constitute "due process."[24]

But while judges frequently defer to such medical authority in resolving disputes, they seldom provide a satisfactory analysis as to why such deference is warranted. When one closely examines the justifications that are provided for such deference—a recognition of scientific expertise and a reliance on physicians acting as fiduciaries for their patients—they do not suffice as a rationale for the substitution of deference for due process that has occurred. The point I wish to make is not that judges or juries should be making medical decisions, but that the process aspects of health law deserve careful attention when such decisions extend beyond the realm of medical expertise.

The foundation for deference to medical authority is a well-told story, originating in malpractice law. Beginning in the British common law, the principle that courts should defer to the judgment of doctors became a commonplace feature of torts jurisprudence.[25] The standard measure for the assessment of malpractice was whether the doctor deviated from practices which were the professional custom in that situation.[26] It became, as Prosser noted, "not the middle but the minimum common skill which is to be looked to."[27]

This principle of deference to physician judgment migrated to contract law as well. Private health insurance did not become widespread in the United States

---

24. Of course, procedural due process is required only in settings where the government is an actor or in public institutions such as courts. Decisions by private sector physicians or insurers are not generally considered to be state action. *E.g.*, Am. Mfrs. Mut. Ins. Co. v. Sullivan, 526 U.S. 40 (1999); Blum v. Yaretsky, 457 U.S. 991 (1982). In addition, however, there is a broader cultural concept of due process.

25. KENNETH ALLEN DE VILLE, MEDICAL MALPRACTICE IN NINETEENTH CENTURY AMERICA: ORIGINS AND LEGACY 157-66 (1990); Charles Markowitz, *Medical Standard of Care Jurisprudence as Evolutionary Process: Implications Under Managed Care*, 2 YALE J. HEALTH POL'Y L. & ETHICS 59 (2001); Theodore Silver, *One Hundred Years of Harmful Error: The Historical Jurisprudence of Medical Malpractice*, 1992 WIS. L. REV. 1193.

26. Markowitz, *supra* note 25, at 62-64; Silver, *supra* note 25, at 1213. Philip Peters argues that "judicial deference to physicians' customs [has been] quietly eroding" in the law of malpractice in the 1980s and 1990s, a trend that he tracks as coinciding with a sharp decrease in the level of public trust for physicians. Philip G. Peters, Jr., *The Quiet Demise of Deference to Custom: Malpractice Law at the Millennium*, 57 WASH. & LEE L. REV. 163, 170, 196-97 (2000). Many of the same factors that, according to Peters, diminished trust and created a new trend in malpractice law also fueled the creation of external review systems and drove the system toward a due process paradigm.

27. W. PAGE KEETON, ET AL., PROSSER AND KEETON ON THE LAW OF TORTS § 32, at 187 (5th ed. 1984).

until World War II, when its adoption spread through contractual arrangements between insurers and employers, with employees as third-party beneficiaries.[28] Prior to the widespread use of managed care organizations as insurers, physicians provided the treatment which they in their sole judgment considered to be medically necessary, and insurers almost invariably reimbursed on that basis.[29] Doctors and other providers thus implicitly determined the scope of the terms of insurance contracts, which were and still are almost universally written to cover care which is medically necessary.[30]

Several factors resulted in the anomaly of market players, whose economic interests lay in minimizing medical expenses, acceding to determinations of medical necessity by care providers who would profit from larger expenditures. One was the relative modesty of health care costs, as a portion of the economy and as compared to the much higher levels of expenditure today.[31] Another was the control by physicians and hospitals of Blue Shield and Blue Cross, respectively, the two entities which created most of the early group health plans.[32] Thus provider-dominated institutions reviewed determinations of medical necessity made by providers.

Congress reinforced provider control in 1965 when it adopted customary professional charges as the basis for Medicare reimbursement, a system that dwarfed in size any individual insurance company.[33] Private reimbursement entities found it economically efficient to adopt the Medicare standard.

---

28. PAUL A. STARR, THE SOCIAL TRANSFORMATION OF AMERICAN MEDICINE 200-06 (1982); Hyman & Hall, *supra* note 7, at 25-26.

29. E. HAAVI MORREIM, HOLDING HEALTH CARE ACCOUNTABLE: LAW AND THE NEW MEDICAL MARKETPLACE 45-46 (2001); Marcia Angell, *The Doctor as Double Agent*, 3 KENNEDY INSTITUTE OF ETHICS J. 279, 281 (1993) (noting that "the third parties [payors] happily paid the charges, with few questions asked"); Mark A. Hall & Gerard F. Anderson, *Health Insurers' Assessment of Medical Necessity*, 140 U. PA. L. REV. 1637, 1644-45 (1992).

30. Sara Rosenbaum et al., *Who Should Determine When Health Care Is Medically Necessary?*, 340 NEW ENG. J. MED. 229, 230 (1999).

31. Health care expenses accounted for less than 6% of the gross national product in 1965. BRADFORD H. GRAY, THE PROFIT MOTIVE AND PATIENT CARE: THE CHANGING ACCOUNTABILITY OF DOCTORS AND HOSPITALS 246 (1991). The proportion rose to 9.1% in 1980 and then to 12.1% in 1990. RAND ROSENBLATT ET AL., LAW AND THE AMERICAN HEALTH CARE SYSTEM 16 (1997). Recent figures project an increase in the health share of gross domestic product from 15.3% in 2003 to 18.7% in 2014. Stephen Heffler et al., *Health Spending Projections for 2004-2014*, HEALTH AFF., Feb. 23, 2005, at W5-74, http://content.healthaffairs.org/webexclusives/index.dtl?year=2005.

32. STARR, *supra* note 28, at 306-09; SYLVIA A. LAW, THE HEALTH LAW PROJECT OF THE UNIVERSITY OF PENNSYLVANIA, BLUE CROSS: WHAT WENT WRONG 7-11, 19-20, 26-27 (1974).

33. JUDITH M. FEDER, MEDICARE: THE POLITICS OF FEDERAL HOSPITAL INSURANCE 2-3, 37 (1977); LAW, *supra* note 32, at 59-65, 117; STARR, *supra* note 28, at 375, 385.

Physicians fought early efforts to reign in costs.[34] Even as both public and private payors shifted farther and farther away from a charges-based or fee-for-service system,[35] the foundational fee-for-service structure limited the possibilities for any significant change.

In payment disputes and in private law matters such as malpractice suits, courts asked whether the doctor's acts met the professional custom test, or whether the insurer had properly applied statutory or contract standards of medical necessity. In tort, the collective norm of physicians created the standard by which malpractice decisions were rendered. In contract, public and private insurers acceded to individual physicians' decisions as binding, creating a custom of deference by the market to provider decision-making.

An even more powerful form of deference to physician judgment became a feature of constitutional law as well. When medical decision-making occurs in a public institution or as part of a governmental benefits (as distinct from insurance) program, a doctor's decision can become the decision of the state. The state action required for the procedural due process clause to apply may come from the fact that the decision-making process is a statutory creation[36] or from the fact that the doctor is an employee of the state.[37] In such cases the law has extended direct deference, framing the question as whether a qualified physician made this decision within professional bounds.[38]

In procedural due process cases, the core question focuses on certain indicia of decision-making: notice, opportunity to contest, an impartial decision-maker, a fully-articulated rationale, the opportunity to appeal, and other desiderata of

34. For example, in 1975 the American Medical Association (AMA) blocked implementation of hospital utilization review committees for Medicare and Medicaid enrollees. Am. Med. Ass'n v. Weinberger, 395 F. Supp. 515 (N.D. Ill. 1975). The AMA asserted the constitutional right of physicians to practice medicine by exercising their best medical judgment, *id.* at 520, and also asserted that the regulations at issue violated a statutory provision which prohibited government officials from "exercis[ing] any supervision or control over the practice of medicine or the manner in which medical services are provided," *id.* at 524 (quoting 42 U.S.C. § 1395).

35. The Medicare program began the professional standards review organization (PSRO) system in the early 1970s. Although physician-controlled, PSROs developed criteria for assessing appropriateness of care, thus ending automatic acceptance of customary charges set by hospitals. Congress replaced PSROs with peer review organizations (PROs) in 1984, in part to encourage systems with greater independence from hospitals. GRAY, *supra* note 31, at 246-48; FEDER, *supra* note 33, at 117-35; LAW, *supra* note 32, at 130-31.

36. *See, e.g.*, Mathews v. Eldridge, 424 U.S. 319 (1976); *infra* text accompanying notes 52-59.

37. West v. Adkins, 487 U.S. 42 (1988) (holding that decisions by prison doctor are state action).

38. *See, e.g.*, Youngberg v. Romeo, 457 U.S. 307, 321 (1982).

fairness.[39] Although a robust adherence to the principles of procedural due process may be necessary to keep our constitutional democracy honest, the concept of procedural due process itself is thin. Its protection attaches based on an uncertain trigger: when the individual faces "jeopardy of serious loss."[40] Moreover, the Supreme Court has decided that only three component rights of procedural due process are essential: adequate notice,[41] a hearing at a meaningful time,[42] and a hearing before an impartial decision-maker.[43] Procedural due process does not necessarily include, for example, the right to a written decision[44] or to an appeal.[45] The Court has made clear that the contours of procedural due process should not be "technical [or] . . . unrelated to time, place and circumstances;"[46] but rather "flexible[,] and [include] . . . such procedural protections as the particular situation demands."[47] Yet, however porous, the components of procedural due process provide the fundamental metric by which we assess adjudicatory systems.

Many aspects of health law process, however, became an exception to the normal operations of procedural due process requirements. In the decision that set the existing standard for procedural due process analysis, medical authority functioned as a substitute for procedural protections. In *Mathews v. Eldridge*,[48] a case involving the termination of disability benefits under the Supplemental Security Income (SSI) program, the Supreme Court used a radically different approach to assessing the need for prior evidentiary hearings than it had reached six years earlier in the seminal procedural due process case of *Goldberg v.*

---

39. ERWIN CHEMERINSKY, CONSTITUTIONAL LAW: PRINCIPLES AND POLICIES 557-58 (2d ed. 2002).

40. Joint Anti-Fascist Refugee Comm. v. McGrath, 341 U.S. 123, 171 (1951) (Frankfurter, J., concurring).

41. Mullane v. Cent. Hanover Bank & Trust Co., 339 U.S. 306 (1950).

42. Fuentes v. Shevin, 407 U.S. 67 (1972); Goldberg v. Kelly, 397 U.S. 254 (1970).

43. United States v. James Daniel Good Real Prop., 510 U.S. 43, 55 (1993); Schweiker v. McClure, 456 U.S. 188, 195 (1982). Several commentators have suggested that an independent decision-maker should be the most important single prerequisite of a constitutionally acceptable proceeding. Henry J. Friendly, *"Some Kind of Hearing,"* 123 U. PA. L. REV. 1267, 1279 (1975); Martin H. Redish & Lawrence C. Marshall, *Adjudicatory Independence and the Values of Procedural Due Process*, 95 YALE L.J. 455, 475-76 (1986).

44. John Leubsdorf, *Constitutional Civil Procedure*, 63 TEX. L. REV. 579 (1984); Martha I. Morgan, *The Constitutional Right To Know Why*, 17 HARV. C.R.-C.L. L. REV. 297 (1982).

45. M.L.B. v. S.L.J., 519 U.S. 102 (1996); Lindsey v. Normet, 405 U.S. 56, 77 (1972); Leubsdorf, *supra* note 44, at 628; Kenneth E. Scott, *Two Models of Due Process*, 27 STAN. L. REV. 937, 945 (1975).

46. Mathews v. Eldridge, 424 U.S. 319, 334 (1976).

47. *Id.*

48. 424 U.S. 319 (1976).

Kelly.[49] In *Goldberg* the Court ruled that the Constitution required evidentiary hearings prior to a cut-off of welfare benefits, and that the subsequent hearings provided for by the relevant statute, the Aid to Families with Dependent Children (AFDC), were insufficient.[50] By contrast, in *Eldridge*, the Court found that no evidentiary hearing was required prior to the termination of disability benefits under the SSI program.[51]

The defining difference for the Court between the two cases was that the SSI statute required a medical assessment as to whether an individual's disability made that individual unable to engage in any substantial gainful activity.[52] Such an assessment, reasoned the Court, "is a more sharply focused and easily documented decision than the typical determination of welfare entitlement."[53]

The Court then adopted, in *Eldridge*, what remains its test for what process is due: a balancing of factors that embodies a rough cost-benefit assessment of whatever enhanced procedure is being sought. The Court held that the assessment would depend on a comparison of the nature of the personal interest at risk and its importance to the individual, the cost to the government of the additional procedural protections being sought, and the risk of an erroneous decision balanced against the likelihood that the additional protections would produce a more accurate outcome.[54] In the three-part test adopted in *Eldridge*, the question of how much greater accuracy in outcome a hearing, for example, would produce works as the tie-breaker between the competing interests of the two parties.[55]

The Court thus allowed the final decision, in cases where evidence was conflicting, to turn on physicians' reports, with medical judgment serving as the

---

49. 397 U.S. 254 (1970).

50. *Id.* at 264, 266.

51. 424 U.S. at 340. The SSI program provides income benefits to an individual who, because of his or her disability, is unable to engage in any substantial gainful employment. Such individual must also meet certain income limitations. 42 U.S.C. § 423 (2000).

52. The Court also noted that the earlier case had involved cash payments to enable indigents to secure necessities such as food and shelter. This distinction is unpersuasive for three reasons. First, it essentially limited *Goldberg* to its facts, when the Court's language in *Goldberg* had been sweeping. Second, because SSI is an income replacement program, it differs little in function from the AFDC benefits at issue in *Goldberg*. Third, the Court had already explicitly declined to limit the *Goldberg* principle to cases involving the necessities of life. Fuentes v. Shevin, 407 U.S. 67, 88 (1972).

53. 424 U.S. at 343.

54. *Id.* at 334-35.

55. Jerry L. Mashaw, *The Supreme Court's Due Process Calculus for Administrative Adjudication in* Mathews v. Eldridge*: Three Factors in Search of a Theory of Value*, 44 U. CHI. L. REV. 28, 29 (1976).

tie-breaking factor.[56] Justice Brennan's dissent in *Eldridge*[57] referred to his dissenting opinion in *Richardson v. Wright*, in which he rejected the use of medical authority as a substitute for a pre-termination hearing: "I see no reason to suppose . . . that the 'credibility and veracity' of doctors . . . can never be in issue."[58]

In *Eldridge* the Court asserted it was drawing on medical authority as a source of neutral evidence in determining whether an individual's physical condition met the statutory criteria for disability. In reality, however, the Court was invoking deference to physician judgment as an alternative process paradigm. The Court imported deference to medical authority from tort law, and then deployed it as the primary rationale for its conclusion that the additional protection of a prior evidentiary hearing would result in no greater likelihood of an accurate determination. Medical authority thus helped mark the boundaries of procedural due process.[59]

---

56. 424 U.S. at 344. To some extent, the Court's reliance on medical judgment fudged the issue, since the dispute turned on a disagreement *between* doctors, with the enrollee's physician contesting the assessment of the government's medical expert. Thus, it was also a contest over which doctor, or whose doctors, would be believed. The profession itself had already recognized the ethical problems implicit in physicians working for government and private sector entities in a capacity in which the goal was to curb expenses. Robert M. Veatch, *Ethical Dilemmas of For-Profit Enterprise in Health Care, in* THE NEW HEALTH CARE FOR PROFIT: DOCTORS AND HOSPITALS IN A COMPETITIVE ENVIRONMENT 125, 131, 143-44 (Bradford H. Gray ed., 1983). A 1969 American Medical Association Judicial Council opinion called such arrangements "absolutely destructive of that personal responsibility and relationship which is essential to the best interest of the patient." *Id.* at 131.

57. 424 U.S. at 349-50.

58. Richardson v. Wright, 405 U.S. 208, 219 (1972). The *Wright* case had raised the same issue as to pre-termination hearings as in *Eldridge*; the Court had remanded it for consideration of what were then pending new regulations. *Eldridge* eventually came before the Court challenging the revised regulations.

59. Even stronger strains of deference emerged in cases involving treatment decisions by physicians employed by state institutions. In *Parham v. J.R.*, 442 U.S. 584 (1979), the Supreme Court rejected the need for a judicial hearing in an involuntary civil commitment proceeding for a minor, citing *Eldridge* and the principle of medical deference. *Id.* at 607-09. In *Youngberg v. Romeo*, 457 U.S. 307 (1982), the Court adopted the professional judgment rule, directing courts to defer to physicians' decisions regarding treatment of institutionalized persons unless a doctor's decision was "such a substantial departure from accepted professional . . . standards as to demonstrate that the person responsible actually did not base the decision on such a judgment." *Id.* at 323. The due process right asserted by institutionalized persons has both a substantive component—for example, the right to be free of bodily restraints—and a procedural component, i.e., the right to a set of minimally necessary procedural steps for the review of whether a substantive right has been violated. Washington v. Harper, 494 U.S. 210, 220 (1990). Courts

B. The Hidden Assumptions of Deference

In *Eldridge* the persuasive power of the decision hinges on trust in the proxy for accuracy that the Court has invoked: medical authority. The Court softened the cold edge of a legal standard based on a cost-benefit calculation by relying on perhaps the last bastion of benign paternalism. Yet when one examines closely the basis for expanding deference from a standard of review for past acts into a proxy for fair process in making externally binding, prospective decisions, the rationale does not work.

The assumption that an additional evidentiary hearing would not improve protection against error in a cost-effective way works only if courts accept that medical decisions come with their own intrinsic protections against error, and against exactly the types of error—those involving credibility, veracity and bias—that evidentiary hearings and cross-examination are designed to prevent. But is this really true of medical practice? Medical decisions are certainly subject to the risks common to all decision-making processes against which due process criteria ordinarily provide a shield. Under the basic principle that the degree to which procedural refinements attach is contextual, related to the nature of the decision and its consequences, at least some medical decisions would seem to call for a fuller, rather than a more truncated, set of procedural guarantees. Such decisions have a highly consequential risk of error. Decisions as to the course or scope of treatment (including what is medically necessary) are often the products of informed judgment, not a simple application of scientific fact, and are highly indeterminate.[60] Most medical decisions are made by individuals, rather than by formal institutional processes. And some of those decisions are enforceable by the state. All of these characteristics signal a need for heightened sensitivity to the integrity of the process.

Consider the following four possible risks applicable to medical decisions:

(1) A physician may make an *arbitrary* decision, i.e., a decision which

---

conflated the two, however, under the rubric of the professional judgment standard: "[I]f a plaintiff's substantive due process rights were not violated (because professional judgment was exercised), any procedural due process requirements (which require essentially the same exercise of professional judgment) would undoubtedly have been fulfilled as well." Wells v. Franzen, 777 F.2d 1258, 1261 n.2 (7th Cir. 1985), *quoted in* Susan Stefan, *Leaving Civil Rights to the "Experts": From Deference to Abdication Under the Professional Judgment Standard*, 102 YALE L.J. 639, 651-54, 680 (1992).

60. CARL E. SCHNEIDER, THE PRACTICE OF AUTONOMY: PATIENTS, DOCTORS, AND MEDICAL DECISIONS 130-34 (1998); M. Gregg Bloche, *The Invention of Health Law*, 91 CAL. L. REV. 247, 266-70 (2003); Jay Alexander Gold, *Wiser Than the Laws?: The Legal Accountability of the Medical Profession*, 7 AM. J.L. & MED. 145, 165-71 (1981); Harold S. Luft, *Economic Incentives and Clinical Decisions*, *in* THE NEW HEALTH CARE FOR PROFIT, *supra* note 56, at 103, 105-06.

is the product of an abuse of appropriate process norms. A physician's refusal to consider relevant evidence would be an example of such arbitrariness. The outcomes of such decisions might be good or bad, right or wrong, in both utilitarian and normative terms, but they would be nonetheless tainted by the lack of regularity and predictability.

(2) A physician may make a *biased* decision, i.e., a decision reflecting a pattern of unfairness, which disparages the interests of certain persons or classes of persons. Physicians may adhere to process rules in some cases, but not in others, leading to differential outcomes that are unfair insofar as similar arguments, situations, or conditions are treated differently in different cases.

(3) A physician may make a *corrupted* decision, i.e., a decision driven by the self-interest of the decision-maker. Decisions tainted by conflicts of interest may affect all classes or groups of patients, and may seem arbitrary, but they are in fact motivated by a specific, clear purpose. The purpose which drives the process, however, is illegitimate.

(4) Finally, a physician may make a *careless* decision, i.e., a decision based on hasty, ill-considered, or unsupported beliefs. Insufficiencies in the decision-maker's training, self-discipline, or commitment to improvement can produce decisions that are more negligent than arbitrary, evenly applied and of no personal benefit to the decision-maker.

Protection against all four such risks is necessary for a decision-making process to have integrity as a process.[61] Deference to medical authority makes sense only if it plausibly provides protection against such risks. At least for the first three risks,[62] such protection requires more than a body of expert scientific

61. In an article reviewing the literature on the role of trust in medical relationships, Mark Hall and colleagues conceptualize multiple dimensions of trust in a model that closely approximates my framework for the categories of risk to process integrity. They identify the dimensions of trust as: fidelity ("pursuing a patient's best interests and not taking advantage of his or her vulnerability"); competence (avoiding mistakes and maximizing the optimal outcome); honesty (including avoidance of deception by silence); confidentiality ("proper use of sensitive or private information"); and global trust (the holistic aspect of trust, reflecting the interconnection of the various dimensions). Mark A. Hall et al., *Trust in Physicians and Medical Institutions: What Is It, Can It Be Measured, and Does It Matter?*, 79 MILBANK Q. 613, 620-24 (2001). The authors conclude that professional norms and culture serve the function of preserving trust. *Id.* at 633.

62. As to the fourth concern, hospital-based peer review committees (now called quality improvement committees) provide the chief policing mechanism within medicine. TROYEN A. BRENNAN & DONALD M. BERWICK, NEW RULES: REGULATION, MARKETS, AND THE QUALITY OF

knowledge. What it requires is consistent adherence to the fiduciary principle that a physician will place the interests of her patient above all else. Faith in that principle is central to health process exceptionalism.

The basis for requiring that physicians act as fiduciaries is the recognition that information, power, and susceptibility to harm differ between doctor and patient.[63] Despite high standards within the profession, it is not self-evident that physicians will always properly carry out their fiduciary duties.

The fiduciary principle is broadly conceptualized as an ethical ideal,[64] but much more narrowly enforced as a legal requirement which could be relied on as a substitute for procedural protections in decision-making. As William Sage has noted, whereas corporate law enforces the duty of loyalty by fiduciaries, health law "relegates loyalty to poorly enforced professional disciplinary processes."[65] The fiduciary principle has been most useful in health law for establishing the scope of a physician's duties directly to her patients. As one scholar has noted: "Remarkably, U.S. health law has not yet widely recognized a duty of loyalty, distinct from professional obligations with respect to quality and confidentiality."[66] Even there, while courts and commentators have casually cited it as self-evident,[67] the power of law has been used to enforce the principle mostly in informed consent cases.[68]

The role of payors offers a critical challenge to the fiduciary relationship between patients and physicians. Payors, whether government or private, have significant economic and political power with which to protect themselves from

---

AMERICAN HEALTH CARE 178-85 (1996). Ethics committees also address some of these issues, but they too suffer from process deficiencies. *See* Robin Fretwell Wilson, *Hospital Ethics Committees as the Forum of Last Resort: An Idea Whose Time Has Not Come*, 76 N.C. L. REV. 353 (1998); Susan M. Wolf, *Ethics Committees and Due Process: Nesting Rights in a Community of Caring*, 50 MD. L. REV. 798 (1991).

63. Maxwell J. Mehlman, *Fiduciary Contracting: Limitations on Bargaining Between Patients and Health Care Providers*, 51 U. PITT. L. REV. 365 (1990).

64. *See, e.g.*, Council on Ethical & Judicial Affairs, Am. Med. Ass'n, *Ethical Issues in Managed Care*, 273 JAMA 330 (1995) [hereinafter *Ethical Issues in Managed Care*].

65. Sage, *supra* note 3, at 610; *see also* Mark A. Hall, *Law, Medicine, and Trust*, 55 STAN. L. REV. 463, 503 (2002) (comparing enforcement to a "speakeasy" system).

66. M. Gregg Bloche, *Clinical Loyalties and the Social Purposes of Medicine*, 281 JAMA 268, 273 (1999).

67. Joseph M. Healey, Jr. & Kara L. Downey, *Controlling Conflicts of Interest in the Doctor-Patient Relationship: Lessons from* Moore v. Regents of the University of California, 42 MERCER L. REV. 989, 1001-02 (1991).

68. Hall, *supra* note 65, at 489-90; Marc A. Rodwin, *Strains in the Fiduciary Metaphor: Divided Physician Loyalties and Obligations in a Changing Health Care System*, 21 AM. J.L. & MED. 241, 247 (1995).

decision-making processes that they consider to be unfair or biased against their interests, more so than patients. Moreover, the potential harm of diminished health from inadequate process protections that patients face is less remediable than the risk of lost revenue to payors. Until the 1990s, however, insured patients were largely protected from those consequences by the same flaws in the medical marketplace that drove up costs: the incentives under fee for service to provide every legitimate treatment to every patient. It was that aberration, more than the fiduciary principle itself and certainly more than any enforcement of the fiduciary principle, that mitigated the need for process protections on behalf of patients seeking reimbursement of medical expenses.

In the payment realm the advent of managed care, driven by cost containment imperatives, killed deference, thus ending the economically strange regime of fee-for-service systems. Massive, spiraling health care costs could not be attacked without a rollback of the degree of physician authority that had characterized fee-for-service insurance payment systems. Managed care struck at the heart of unnecessary costs by the combination of a capitated pricing and payment system,[69] financial incentives to doctors and other providers to conserve care, and financial penalties to patients who had to pay "extra" in some form to go outside the system.[70] Managed care companies continue to use withholds, or the sequestration of some part of overall compensation owed to a physician until the company performs an accounting of whether amounts will be deducted because that doctor over-utilized expensive treatments or referrals. Or, they may achieve essentially the same goal by paying bonuses to doctors who meet utilization goals. Even without formal corporate policies of either kind, pressure has been building on physicians since the costs explosion of the 1980s to engage in "bedside rationing."[71]

Beginning in the 1980s health policy analysts recognized the growing potential for financial conflicts of interest between doctor and patient, since doctors operating under the fee-for-service system could multiply their income by ordering unnecessary, if usually not harmful, tests and procedures.[72] The first

---

69. Under a capitation system, providers receive a fixed monthly or yearly amount for each patient enrolled with that provider.

70. See description of financial incentives in David Orentlicher, *Paying Physicians More To Do Less: Financial Incentives to Limit Care*, 30 U. RICH. L. REV. 155, 158-60 (1996).

71. Bloche & Jacobson, *supra* note 12; Einer Elhauge, *Allocating Health Care Morally*, 82 CAL. L. REV. 1449, 1457-65 (1994); Mark A. Hall, *Rationing Health Care at the Bedside*, 69 N.Y.U. L. REV. 693 (1994).

72. AM. MED. ASS'N, REPORT OF THE COUNCIL ON ETHICAL AND JUDICIAL AFFAIRS ON CONFLICTS OF INTEREST (1986); GRAY, *supra* note 31, at 251-55; James S. Todd & Janet K. Horan, *Physician Referral—The AMA View*, 262 JAMA 395 (1989). For a longer historical perspective on the conflicts of interest intrinsic to medical care structured as businesses for providing services, see

judicial decision to focus on financial incentives was *Moore v. Regents of the University of California*,[73] in which the California Supreme Court ruled that a physician who profited by sale of stem cells from a patient who was unaware that his tissue was being used to harvest those materials could be liable for failure to obtain informed consent. Yet what now seems most significant about *Moore* is the extent to which it did not generate any further new law.[74]

The transformation to managed care in the 1990s, in which financial incentives exert pressure for under- rather than over-treatment, exacerbated the risk of tainted decision-making, even producing debates over whether there should be a new version of the professional custom rule, in which physicians undertook fiduciary duties and concomitant responsibilities for the purpose of cost containment.[75] Researchers found that doctors did not propose treatments which they believed would be medically beneficial because of insurance coverage restrictions.[76]

Despite such pressures on ethical aspirations, courts by and large have not expanded the *Moore* principle to address problems of physician conduct related to managed care.[77] The circuits are split on the question of whether to impose a fiduciary obligation under ERISA that would require insurers to notify patients of the financial incentives operating in their plan.[78] The reality, however, is that

---

Veatch, *supra* note 56.

73. 793 P.2d 479 (Cal. 1990).

74. Joan H. Krause, *Reconceptualizing Informed Consent in an Era of Health Care Cost Containment*, 85 IOWA L. REV. 261, 340-41 (1991); Lori Potter, *Failure To Disclose HMO Incentives and the Breach of Fiduciary Duty: Is a New Cause of Action Against Physicians the Best Solution?*, 34 U.S.F. L. REV. 733, 755 (2000).

75. Angell, *supra* note 29; Thomas H. Boyd, *Cost Containment and the Physician's Duty to the Patient*, 39 DEPAUL L. REV. 131, 154-55 (1989); Jerome P. Kassirer, *Managing Care—Should We Adopt a New Ethic?*, 339 NEW ENG. J. MED. 397 (1998); Roger Menendez et al., *A New Ethic for Medicine?*, 339 NEW ENG. J. MED. 1326 (1998); E. Haavi Morreim, *Medicine Meets Resource Limits: Restructuring the Legal Standard of Care*, 59 U. PITT. L. REV. 1, 27-35, 64-72 (1997); Edmund D. Pellegrino, *Interests, Obligations, and Justice: Some Notes Toward an Ethic of Managed Care*, 6 J. CLINICAL ETHICS 312, 314-16 (1995).

76. Sherry Glied & Sarah E. Little, *The Uninsured and the Benefits of Medical Progress*, 22 HEALTH AFF. 210 (2003); Matthew K. Wynia et al., *Do Physicians Not Offer Useful Services Because of Coverage Restrictions?*, 22 HEALTH AFF. 190 (2003) (noting that thirty-one percent of doctors did not discuss useful treatments with patients because of insurance restrictions). Physicians also manipulate insurers' rules in order to secure coverage which otherwise would not be covered. Matthew K. Wynia et al., *Physician Manipulation of Reimbursement Rules for Patients*, 283 JAMA 1858 (2000).

77. PETER D. JACOBSON, STRANGERS IN THE NIGHT: LAW AND MEDICINE IN THE MANAGED CARE ERA 222-49 (2002); Rodwin, *supra* note 68.

78. *See, e.g.*, Horvath v. Keystone Health Plan E., Inc., 333 F.3d 450, 460-62 (3d Cir. 2003)

financial incentives may create conflicts even for treating physicians (and not just for administrators of MCOs).[79] One commentator has suggested that physicians have gone from being patients' agents, to being double agents, to now being free agents.[80] Haavi Morreim observed that the acid test of whether a fiduciary relationship is operating properly is whether the more powerful party is engaged in conduct which she would feel uncomfortable revealing to the other party.[81] In practice, however, fiduciary relationships are seldom tested. Doctors have resisted starting down the road of financial disclosures to patients,[82] expressing concern that such disclosure would undercut trust,[83] and only a handful of courts have compelled such disclosure.[84]

(holding that an HMO had no duty under ERISA to disclose financial incentives plan to patient); Ehlmann v. Kaiser Found. Health Plan of Tex., 198 F.3d 552 (5th Cir. 2000) (same); Shea v. Esensten, 107 F.3d 625, 629 (8th Cir. 1997) (finding that plaintiff stated a viable claim that ERISA imposed such a duty); 42 C.F.R. 422.210(b) (2004) (disclosure of financial incentives required for Medicare plans); *see also* Timothy S. Hall, *Bargaining with Hippocrates: Managed Care and the Doctor-Patient Relationship*, 54 S.C. L. REV. 689, 716-17 (2003); Tracy E. Miller & William M. Sage, *Disclosing Financial Incentives*, 281 JAMA 1424, 1425-26 (1999). The AMA has adopted the distinction between duties of the MCO and duties of the physicians. The House of Delegates voted in 1997 to place "primary responsibility for disclosure on health plans," with physicians having "some obligation" to describe incentives if the plan does not. *Id.* at 1425. In 1995 its Council on ethical issues recommended a requirement of disclosure only by the plan. *Ethical Issues in Managed Care*, *supra* note 64, at 335.

79. Sage, *supra* note 3, at 609.

80. Hoangmai H. Pham et al., *Financial Pressures Spur Physician Entrepreneurialism*, 23 HEALTH AFF. 70, 79 (2004).

81. E. Haavi Morreim, *Economic Disclosure and Economic Advocacy: New Duties in the Medical Standard of Care*, 12 J. LEGAL MED. 275, 314 (1991).

82. According to a report by two health policy researchers, "The AMA's code of ethics requires that physicians discuss with new patients how they are paid by the plan. . . . In the absence of any effective enforcement mechanism, it seems unlikely that many physicians actually convey this information." David Mechanic & Mark Schlesinger, *The Impact of Managed Care on Patients' Trust in Medical Care and Their Physicians*, 275 JAMA 1693, 1694 (1996).

83. Bloche, *supra* note 66, at 273; Miller & Sage, *supra* note 78.

84. *See* Neade v. Portes, 710 N.E.2d 418, 425-27 (Ill. App. Ct. 1999), *aff'd in part and rev'd in part*, 739 N.E.2d 496 (Ill. 2000) (applying fiduciary requirement to physicians). Minnesota requires physician disclosure by statute. MINN. STAT. ANN. § 62J.72(1) (West 2005). Courts rejected claims based on fiduciary principles in *D.A.B. v. Brown*, 570 N.W.2d 168, 172 (Minn. Ct. App. 1997), a decision which may have triggered the legislature to enact the previously-cited statute, and *Spoor v. Serota*, 852 P.2d 1292, 1294-95 (Colo. Ct. App. 1992). *Cf.* Batas v. Prudential Ins. Co., 724 N.Y.S.2d 3, 13 (App. Div. 2001) (holding that no fiduciary relationship exists between a health insurer and an insured party). For arguments that courts should use fiduciary principles to analyze managed-care-related conflicts of interests, see Peter D. Jacobson & Michael T. Cahill, *Applying Fiduciary Responsibilities in the Managed Care Context*, 26 AM. J.L. & MED. 155, 165-66 (2000).

A combination of the status and culture of the medical profession, together with irrational market forces, resulted in deference by payors lasting as long as it did, but the economic irrationality of the old system eventually broke down. Large institutional payors abandoned the custom of deference to physicians, which in effect pulled the financial rug out from under the fiduciary principle.[85] Physician adherence to a fiduciary responsibility has now become far more genuinely altruistic than it formerly was.

In cases at the intersection of medicine and due process, courts used medical authority as a substitute for what otherwise could have been a fuller examination of process questions. The tendency of courts in the procedural due process cases to over-read the fiduciary principle corrupted the process component of health law. With managed care the central process question now is not the extent to which physicians' decisions are binding on third parties, but the extent to which medical caregivers are bound *by* third parties.

C. From Deference to Due Process

1. Truth Gives Way to Justice

The use of deference in tort and contract law functioned as an example of what procedure scholars John Thibaut and Laurens Walker called "truth-seeking" processes: processes designed to resolve a cognitive conflict, to find an answer that would be recognized by all parties as "correct" according to mutually accepted standards.[86] Writing in the late 1970s, when other scholars were analyzing and dissecting the terms of the procedural due process debates epitomized by such cases as *Mathews v. Eldridge*,[87] Thibaut and Walker drew on social psychology research for empirical evidence as to which kinds of procedural devices would most enhance the quality of conflict resolution. One of their threshold conclusions was that conflicts could be divided into cognitive disputes, for which a right answer existed, and distributive disputes, which centered on fairness.

In the traditional cultural view of medicine, physicians have been seen as truth-seekers. Dan Fox has described a medical culture in which authority is asserted "from the top of a hierarchy in which power is derived from knowledge. . . . A principal goal of medical education is to inculcate lifelong

85. John V. Jacobi, *Patients at a Loss: Protecting Health Care Consumers Through Data Driven Quality Assurance*, 45 U. KAN. L. REV. 705, 721 (1997) ("That game has been lost.").

86. John Thibaut & Laurens Walker, *A Theory of Procedure*, 66 CAL. L. REV. 541, 543-44 (1978).

87. *See supra* text accompanying notes 52-59.

habits of deference to superior knowledge and experience."[88] For truth-seeking or cognitive disputes, Thibaut and Walker asserted that a single autocratic decision-maker would be the best mechanism for resolving the conflict.[89] They reasoned that a disinterested third party with control over both process and outcome could best ascertain true facts, better than one constrained to respond to adversarial presentations.[90] The professional custom standard and the broad practice of judicial deference to medical authority fit within this model, as indicated by the Supreme Court's characterizations of doctors in cases like *Eldridge*.

Thibaut and Walker's alternative model for process was "justice-seeking," by which they meant a process needed to resolve conflicts of interest, in which answering factual questions would serve only as a means to the end goal of the fairest division of resources.[91] For this purpose, Thibaut and Walker asserted that it was important to promote the participation of the disputants, to allow them to present their cases as best they could, primarily because the legitimacy of the system as a whole would turn on society's perception of its legitimacy.[92] Subsequent research has found that for participants in a dispute over justice or distribution, having meaningful control over the process for resolution of the conflict, for example by presenting evidence, is critical to perceptions of fairness and legitimacy.[93] However, such control is not a significant factor in perceptions of fairness in conflicts over what is perceived as truth.

One way to understand the backlash of discontent against managed care is as a response to the rupturing of the cultural norm of belief in medical authority and to the conversion of a cognitive dispute into a resource dispute without the consent of those whose disputes were to be adjudicated. As a result, health process is now in transition from the truth-seeking to the justice-seeking model. With the widespread enactment of external review systems for appealing MCO denials of care, the pendulum has swung, so that due process-style guarantees have migrated to medical decision-making, providing insulation from what patients suspect are unreliable medical assessments. External reviews systems are the primary form of the due process model being brought into medical decision-making.

The transition is apparent in the rhetoric of managed care debates. The

---

88. Daniel M. Fox, *Physicians Versus Lawyers: A Conflict of Cultures*, *in* AIDS AND THE LAW: A GUIDE FOR THE PUBLIC 210, 213 (Harlon L. Dalton et al. eds., 1987); *see also* Luft, *supra* note 60, at 105 (describing physician as "seeker of truth").

89. Thibaut & Walker, *supra* note 86, at 544.

90. *Id.* at 547-48.

91. *Id.* at 541-42.

92. *Id.* at 551.

93. TOM R. TYLER ET AL., SOCIAL JUSTICE IN A DIVERSE SOCIETY 92 (1997).

rhetoric used by advocates for external review systems did invoke the cultural power of medical authority and patient trust.[94] But although the older trope of deference to medical authority has not been completely superseded, the source of protection for patients has definitively shifted to procedural protections.[95] Where health law was once a process zone largely divorced from procedural due process norms, it has now become a prime location for contemporary procedural due process concerns to blossom.

In a trio of decisions the Supreme Court has affirmed the transition from deference to due process in health process law.[96] In the first it revealed its own loss of allegiance to the traditional approach to deference evident in *Eldridge*.[97] In the second and third it ruled that state action did not attach to these decisions, but accepted that the actions of privately-run external review systems constitute state action.[98]

2. Pegram v. Herdrich

The Supreme Court's decision in *Pegram v. Herdrich*[99] illustrates that the contemporary image of a physician is considerably less idealized than it once was. There the plaintiff asserted that she had received substandard care because of an eight-day delay in receiving an ultrasound test so that it could be done by a clinic which was affiliated with the HMO with which her employer had contracted for medical services. She attributed her resulting harm to the fact that

---

94. *See, e.g.*, the comments of Senator Edward Kennedy (D-Mass.) in support of a federal Patients Bill of Rights, including an external review provision: "My final point is the underlying commitment of this legislation to make sure that doctors are going to make the decisions. Trained medical personnel and families are going to make these judgments and decisions." 147 CONG. REC. S6905 (daily ed. June 26, 2001).

95. *See, e.g.*, U.S. GEN. ACCOUNTING OFFICE, HMO COMPLAINTS AND APPEALS: MOST KEY PROCEDURES IN PLACE, BUT OTHERS VALUED BY CONSUMERS LARGELY ABSENT (1998) [hereinafter GAO REPORT]; SHARON WILCOX, CONSUMER PROTECTION IN PRIVATE HEALTH INSURANCE: THE ROLE OF CONSUMER COMPLAINTS (2000); OFFICE OF INSPECTOR GEN., DEP'T OF HEALTH AND HUMAN SERVS., PUB. NO. OEI-07-94-00281, MEDICARE HMO APPEAL AND GRIEVANCE PROCESSES: BENEFICIARIES' UNDERSTANDING (1996); OFFICE OF INSPECTOR GEN., DEP'T OF HEALTH & HUMAN SERVS., PUB. NO. OEI-95-00430, BENEFICIARY PERSPECTIVES OF MEDICARE RISK HMOS (1998); KAREN POLLITZ ET AL., THE HENRY J. KAISER FAM. FOUND., ASSESSING STATE EXTERNAL REVIEW PROGRAMS AND THE EFFECTS OF PENDING FEDERAL PATIENTS' RIGHTS LEGISLATION, at v (2002) ("External review is a formal process for resolving disputes between health plans and consumers.").

96. Pegram v. Herdrich, 530 U.S. 211 (2000); Grijalva v. Sullivan, 526 U.S. 1096 (1999); Am. Mfrs. Mut. Ins. Co. v. Sullivan, 526 U.S. 40 (1999).

97. *Pegram*, 530 U.S. 211.

98. *Grijalva*, 526 U.S. 1096; *Am. Mfrs. Mut. Ins. Co.*, 526 U.S. 40.

99. 530 U.S. 211 (2000).

the physician who had examined her was a part owner of the HMO.

On the question presented the Court rejected her argument that the HMO had undertaken fiduciary obligations pursuant to ERISA's provisions governing plan administrators. Additionally, it found her experience to be less than shocking. "Although it is true that the relationship between sparing medical treatment and physician reward is not a subtle one under the [defendant's] scheme, no HMO organization [sic] could survive without some incentive connecting physician reward with treatment rationing."[100]

Nor did the Court nostalgically invoke medical authority in older delivery structures as unproblematic:

> [I]n an HMO structure, a physician's financial interest lies in providing less care, not more. The check on this influence . . . is the professional obligation to provide covered services with a reasonable degree of skill and judgment in the patient's interest. . . . The adequacy of professional obligation to counter financial self-interest has been challenged no matter what the form of medical organization. HMOs became popular because fee-for-service physicians were thought to be providing unnecessary or useless services . . . .[101]

Instead, the Court adopted an explicitly market-centered conceptualization of the physician's role, which is an implicit foundation for the understanding that mechanisms beyond deference would be needed for patient protection.

3. Grijalva v. Shalala *and* American Manufacturers Mutual Insurance Co. v. Sullivan

Standards for medical decision-making based explicitly on due process principles provide a shield against inadequacies in the "professional obligation to counter financial self-interest."[102] The most frequent incorporation of those standards has occurred in the law regulating public systems of health care delivery, especially Medicare.[103] In a series of court decisions[104] alternating with

---

100. *Id.* at 220.

101. *Id.* at 219-20.

102. *Id.* at 220.

103. Bill Sage points out that Medicare has provided a vector for the introduction of process norms in medicine: "Before Medicare, the medical profession frequently viewed fair process as whatever physicians said it was and sometimes trampled procedural rights in order to achieve desirable substantive results. After Medicare, physicians increasingly bought into a model of fairness defined instead by law and lawyers . . . ." William M. Sage, *The Lawyerization of Medicine*, 26 J. HEALTH POL. POL'Y & L. 1179, 1188 (2001).

104. 946 F. Supp. 747 (D. Ariz. 1996), *aff'd*, 152 F.3d 1115 (9th Cir. 1998), *vacated and remanded*, 526 U.S. 1096 (1999).

regulatory[105] and statutory[106] changes, the law governing appeals of treatment denials by Medicare beneficiaries enrolled in managed care organizations has moved toward much more elaborate due process-based protections.

In *Grijalva v. Shalala*, first the district court and then the Ninth Circuit Court of Appeals upheld challenges by beneficiaries based on procedural deficiencies.[107] In so doing, both courts applied the *Mathews v. Eldridge* balancing test, and distinguished that case based on the finding that the plaintiff Medicare enrollees had a stronger interest in their benefits than Eldridge had in his disability payments. The key factor for why an individual's interest in a recoupable medical expense was entitled to more constitutional weight than a disabled person's monthly cash allowance for living expenses was the structure of managed care. The lower courts found that the Medicare HMO case more closely resembled *Goldberg v. Kelly* than it did *Eldridge*, "especially because [plaintiffs] are HMO enrollees," on the ground that "[t]he HMO's initial adverse coverage determination in many cases prevents receipt of medical care," and a reversal "may come too late to rectify the situation."[108] The decision turned not on medical authority or on deference to a treating physician, but on the weight of a claim to needed medical care, to which Medicare beneficiaries had an entitlement. It was the gate-keeping system used by HMOs which effectively created the claim to care.

The Supreme Court's second critical decision for health process law came in a case related to *Grijalva*. In *American Manufacturers Mutual Insurance Co. v. Sullivan*,[109] the Court ruled that private insurers' decisions about whether a treatment met the criteria of medical necessity was not state action. The case involved claimants who sought payment under the workers compensation system for treatments prescribed by their physicians. The employer's insurance company denied coverage on the ground that the treatments in question were not medically necessary, and the state provided for an appeal to a utilization review entity, a private company which provided independent review. The Court ruled that the workers had no statutory entitlement to the benefits unless and until the treating physician's judgment was confirmed on appeal; thus no constitutionally protected deprivation of property had occurred. The insurers' denial was not state action; neither their role in administering the workers compensation program nor

105. Medicare Program: Medicare Appeals of Individual Claims, 42 C.F.R. § 405 (2004); Medicare Program: Improvements to the Medicare+Choice Appeal and Grievance Procedures, 42 C.F.R. § 422 (2004).

106. Balanced Budget Act of 1997 § 4001, Pub. L. No. 105-33, 111 Stat. 251.

107. *See supra* note 104.

108. 946 F. Supp. at 757.

109. 526 U.S. 40 (1999).

the existence of a state statute authorizing withholding of benefits pending utilization review sufficed to establish that their decisions could be attributed to the state.[110]

The *Sullivan* ruling was critical because HMOs which delivered health care to Medicare beneficiaries had been subjected to procedural due process scrutiny in *Grijalva* based on the finding that their treatment decisions constitute state action.[111] The decisions of the Medicare HMOs, like those of the workers compensation insurers in *Sullivan*, were subject to review by an outside organization. The Supreme Court vacated and remanded the Ninth Circuit decision in *Grijalva* in light of its ruling in *Sullivan*.

However, the Court in *Sullivan* also accepted the viability of procedural due process challenges to utilization review decisions themselves. The Court explained that "[utilization review organizations (UROs)] are private organizations consisting of health care providers" licensed in the same specialties as the provider whose decision is under review.[112] Despite the fact that UROs were private entities, the Court declared that "the State's role in creating, supervising, and setting standards for the URO process [do not] differ in any meaningful sense from the creation and administration of any forum for resolving disputes . . . [thus] the decision of a URO, like that of any judicial official, may properly be considered state action."[113] *Sullivan* establishes the basis for constitutional challenges to the procedures used by external review organizations.[114]

Although the Court in *Sullivan* cut at least one of the doctrinal legs out from under *Grijalva* by reiterating the principle that initial denials by insurers were not state action, it suggested that termination of an *ongoing* service by a Medicare HMO could be subject to due process protections.[115] Given the somewhat uncertain implications of the *Sullivan* decision, the parties in *Grijalva* settled after the Ninth Circuit decision was remanded.[116] Despite the shakiness of their claim after *Sullivan*, plaintiffs were nonetheless able to achieve a significant

---

110. *Id.* at 51-58.

111. Grijalva v. Shalala, 946 F. Supp. 747, 751-53 (D. Ariz. 1996), *aff'd*, 152 F.3d 1115 (9th Cir. 1998), *vacated and remanded*, 526 U.S. 1096 (1999).

112. *Am. Mfrs. Mut. Ins. Co.*, 526 U.S. at 46.

113. *Id.*at 54.

114. *See also* Vellios v. IPRO, 765 N.Y.S.2d 222 (N.Y. Sup. Ct. 2003) (allowing a state administrative law challenge to a denial of private insurance coverage by an external review organization).

115. The Court distinguished *Goldberg v. Kelly* and *Mathews v. Eldridge* on that basis. *Sullivan*, 526 U.S. at 61.

116. Settlement Agreement, Grijalva v. Shalala, No. CIV 93-711 TUC ACM (D. Ariz. Aug. 9, 2000).

victory. What emerged from the judicial, administrative, and legislative branches was a detailed set of regulations that addressed such issues as the nature and format of the notice of appeal, the requirement that the reason for denial be stated with sufficient clarity to enable the enrollee to argue an appeal, the type size necessary for readability, rules for hearings, review by independent entities, the enrollee's access to HMO documents and other evidence, and both advance notice and a grace period after notice of a denial during which the terminated services would continue.[117]

Grijalva and *Sullivan* together illustrate that a "due process" discourse has arisen around health care. The assumptions typical of a distributive fairness dispute behind "justice-seeking" models have permeated discussions of external review systems, which are routinely described as centering on patient complaints and appeals.[118] Even though advocates of external review laws stress their value in avoiding lawsuits,[119] external review is framed as a mechanism for resolving disputes over entitlement to care much more frequently than they are described as "truth-seeking" disagreements over the medical, scientific accuracy of coverage decisions.

D. Summation: The Normalization of Health Process

Managed care has profoundly altered health care financing, delivery, and quality. One of its most significant effects has been on the process component of health law, where the transformation to managed care has undercut the old paradigm of deference to medical judgment. It has done so by revealing or creating (or both) sharp tensions between the interests of physicians and patients. The visibility of these tensions has in turn produced a crisis of confidence in the fiduciary principle and a resultant search for some other mode of protection for patients. That search has led health law to the realm of procedural due process. Due process is the closest thing to trust that the law can construct or the market provide.

The American Medical Association (AMA) accepted the principle that professional and ethical legitimacy attaches to both the treating physician's mandate of fidelity to the individual patient and to an exception from that duty for payors and the physicians in their employ who seek to contain costs.[120] This AMA ethical statement marks the junction where process rules determine

117. Medicaid Program: Medicaid Managed Care, 42 C.F.R. §§ 400, 430, 431-34, 435, 438, 440, 447 (2004).

118. POLLITZ ET AL., *supra* note 95, at v; Berman-Sandler, *supra* note 9, at 237-38.

119. *See infra* text accompanying notes 259-260.

120. Council on Ethical & Judicial Affairs, Am. Med. Ass'n, *Ethical Issues in Managed Care*, 273 JAMA 330, 331 (1995).

outcome. The apparent simplicity of procedure law hides its importance. When conflicts involving a treatment decision arise between those two legitimate ethical positions, those of treating physicians and those of managed care medical directors, it is the law of procedure that will determine how the conflict is to be resolved.

Moreover, the trend toward due process norms in health law is broader than the issue of identifying which formal adjudicatory procedures should be used. Scholars are shifting to a process values approach, even if not labeling it as such, for other purposes as well.[121] It is evident in Sara Rosenbaum's proposed standard for assessments of medical necessity;[122] in Jacobson and Cahill's suggestion that a process based on fiduciary duty principles be entrusted with resolving the tension between expense and quality care;[123] and in Greaney and Boozang's suggestion of a "mission primacy" standard for reviewing actions of trustees, such that the burden would rest on trustees to demonstrate why charitable mission should not trump resource maximization.[124] Each of these proposals is grounded in an attempt to instantiate a new set of process criteria.

The old exceptionalism of health process law is dead, along with the fee-for-service model of organization and relatively smaller-sized corporate ownership. Due process values now dominate.

III. EXTERNAL REVIEW ORGANIZATIONS AS ADJUDICATION SYSTEMS

From 1992 to 1998, the percentage of persons in the private workforce who were enrolled in managed care plans for their health insurance coverage increased by more than fifty percent.[125] As the public witnessed "the transformation of health care payers from passive poolers of risk to aggressive managers of cost,"[126] concern erupted over the incentives for under service,[127] fed

121. My thanks to Peter Jacobson for pointing out this thread.

122. Rosenbaum et al., *supra* note 30.

123. Peter D. Jacobson & Michael T. Cahill, *Applying Fiduciary Responsibilities in the Managed Care Context*, 26 AM. J.L. & MED. 155, 165-66 (2000).

124. Thomas L. Greaney & Kathleen M. Boozang, *Mission, Margin, and Trust in the Nonprofit Health Care Enterprise*, 5 YALE J. HEALTH POL'Y L. & ETHICS 1, 84 (2005).

125. Laurie McGinley, *HMO Fracas Moves to Who Makes Medical Decisions*, WALL ST. J., Feb. 18, 1999, at A24.

126. M. Gregg Bloche, *Fidelity and Deceit at the Bedside*, 283 JAMA 1881, 1883 (2000).

127. George J. Annas, *Patients' Rights in Managed Care—Exit, Voice, and Choice*, 337 NEW ENG. J. MED. 210 (1997); R. Adams Dudley & Harold S. Luft, *Managed Care in Transition*, 344 NEW ENG. J. MED. 1087 (2001); Rhonda L. Rundle, *HMOs Brace Themselves for 'Avalanche' of New Laws: California Bills on Dispute-Resolution Process Could Be Model for U.S.*, WALL ST. J., Feb. 20, 1998, at B4.

by horror stories of death or disability resulting from care denied or diagnoses missed when MCOs refused to grant pre-authorization for expensive tests or treatments.[128] As one commentator has explained: "[M]ost Americans are discomforted by the idea of having their care rationed and, at some level, they understand that managed care is a mechanism for doing so."[129] The creation of review systems for denial of health coverage and care was a core component of the response to the managed care transformation of service delivery. Because external review organizations (EROs) directly addressed the need to have a check on financial incentives for cost-cutting, provisions creating them became a standard component of "patients' bill of rights" legislation.[130]

By the time of the Supreme Court's decision in *Moran*, external review systems had become the new norm.[131] Not only were these laws enacted in the great majority of states, but a system of external review for denials of authorizations for treatment was also adopted in Medicare, Medicaid, the military health care system, and the system of health care for federal employees.[132] The nature of these new entities—are they adjuncts to the medical system or to the legal system? —became the threshold issue in *Moran*.

A. The Supreme Court's Analysis of EROs in Rush Prudential HMO, Inc. v. Moran

Debra Moran was diagnosed as suffering from a combination of two painful and debilitating conditions in her right shoulder.[133] Her health insurance plan

---

128. Although the allusion to horror stories usually refers to the media, such stories also have surfaced in the text of federal court decisions. *See, e.g.*, Pryzbowski v. U.S. Healthcare, Inc., 245 F.3d 266 (3d Cir. 2001) (MCO delayed in approving back surgery by an out-of-network surgeon, leading to persistent, excruciating pain). For examples of media portrayals, see Michael A. Hiltzik & David R. Olmos, *The Health Care Revolution: Remaking Medicine in California* (pts. 1-7), L.A. TIMES, Aug. 27-Aug. 31, 1995; Robert Kuttner, *The Lethal Side Effects of Managed Care*, BUS. WEEK, Aug. 7, 1995, at 16; Ellyn E. Spragins, *Beware Your HMO*, NEWSWEEK, Oct. 23, 1995, at 54.

129. David Mechanic, *The Managed Care Backlash: Perceptions and Rhetoric in Health Care Policy and the Potential for Health Care Reform*, 79 MILBANK Q. 35, 37 (2001).

130. George J. Annas, *A National Bill of Patients' Rights*, 338 NEW ENG. J. MED. 695, 697 (1998); Tracy E. Miller, *Center Stage on the Patient Protection Agenda: Grievance and Appeal Rights*, 26 J.L. MED. & ETHICS 89 (1998); Rhonda L. Rundle, *External Review of HMO Decisions Becomes Hot Issue*, WALL ST. J., Dec. 3, 1998, at B2.

131. Brief for Respondent State of Illinois at *11-12, Rush Prudential HMO, Inc. v. Moran, 536 U.S. 355 (2002) (No. 00-1021), 2001 WL 1548340 (listing states).

132. Mary C. Morgan, *Independent Review of Managed Care Decisions*, HEALTH L. NEWS, Winter/Spring 2000, at 5, 13-15.

133. The facts in this and the following paragraph are from *Rush Prudential HMO, Inc. v.*

required her to go through an in-network provider prior to receiving any specialty care. The in-network primary care physician initially treated her symptoms with physiotherapy, but the symptoms continued. She sought out a surgeon who specialized in treating her particular condition (brachial plexopathy and thoracic outlet syndrome), who recommended a new form of surgery which the specialist had herself originated. After further consultations with in-network physicians, her primary care doctor recommended that Moran receive the newly-developed surgical procedure out-of-network.

Moran's health insurance coverage derived from her husband's group plan at his place of employment. Under that plan Rush Prudential HMO was both the insurer and the provider of care. The plan contract provided that beneficiaries would receive "medically necessary" care. It also provided that Rush Prudential would retain "the broadest possible discretion" to interpret the meaning of "medically necessary" and to determine which services fell within that category. Rush Prudential denied Moran's request for authorization of the out-of-network surgical procedure on the ground that it was not medically necessary. Moran paid the $94,841 charge for the surgery herself and then contested the denial.

Moran's health insurance plan was governed by ERISA because it was a private workplace benefits plan.[134] Because the employer had purchased the plan, rather than self-insure and bear the risk itself, the plan was also subject to state law regulation.[135] Under the "savings clause" of ERISA, state laws that regulate insurance are exempt from preemption by ERISA;[136] such state laws directly regulate the insurers within the state that sell plans to employers, and thereby set limits on the terms of plans which are available for purchase by employers.[137] Therefore, fully-insured ERISA plans are directly regulated by federal law and indirectly regulated by state law.

Under ERISA Moran could have filed suit asserting a denial of benefits due to her under the plan.[138] However, the standard of review that the court would use to assess whether there had been a wrongful denial of benefits recognized that Rush Prudential was entitled to have exercised discretion in making its

---

Moran, 536 U.S. 355, 360 (2002).

134. The Employee Retirement Income Security Act (ERISA), 29 U.S.C. §§ 1001-1461 (2000) regulates the structure and administration of employee benefits offered by private sector employers. *See* Phyllis C. Borzi, *Distinguishing Between Coverage and Treatment Decisions Under ERISA Health Plans: What's Left of ERISA Preemption?*, 49 BUFF. L. REV. 1219, 1222-24 (2001).

135. Self-insured employer-sponsored health plans are deemed not to be insurance for the purpose of state regulatory laws. Thus, even state insurance laws cannot regulate self-insured plans. 29 U.S.C. § 1144(b)(2)(B) (2000); FMC Corp. v. Holliday, 498 U.S. 52, 61 (1990).

136. 29 U.S.C. § 1144(b)(2)(A) (2000).

137. Metro. Life Ins. Co. v. Massachusetts, 471 U.S. 724, 739 (1985).

138. 29 U.S.C. § 1132(a) (2000).

decision.[139] Only if a court found that there had been an abuse of such discretion would Moran have been entitled under ERISA to recover the costs of her surgery.[140]

Instead, Moran filed suit to enforce an Illinois statute that established an external review system when an HMO denied treatment that had been recommended by the patient's primary care physician on the ground that it was not medically necessary, as was the case here. Under the Illinois system the HMO was compelled to submit the case to an independent physician jointly selected by the patient, her primary care provider, and the HMO. If the independent physician determined that in her judgment the care in question was medically necessary, the HMO was obligated to provide it or, as here, compensate the patient for its cost. A state court ordered Rush Prudential to comply with this statute, and the third-party physician declared that her surgery had been medically necessary.[141]

Rush Prudential nonetheless refused to pay, and defended against Moran's claim for reimbursement on the ground that ERISA preempted the Illinois external review statute. Its primary argument was based on case law which had held that even state laws regulating insurance would be preempted if they altered the enforcement and remedy scheme established by ERISA, reasoning that this "super preemption" was justified by the clear Congressional intent that employers operating group benefits plans should be able to rely on the understanding that they would be liable for only one nationally uniform set of remedies, rather than the differing remedies available under state law.[142]

The federal district court to which Rush Prudential had removed the case concluded that there was preemption and that the ERISA standard, not the results of the external review, governed the question of whether Moran's claim was valid.[143] The court ruled that it was bound to accord the "broadest possible discretion" to the HMO's decision, and found that there had been no abuse of that discretion.[144]

---

139. Firestone Tire & Rubber Co. v. Bruch, 489 U.S. 101, 115 (1989).

140. *See, e.g.*, Jones v. Kodak Med. Assistance Plan, 169 F.3d 1287, 1292 (10th Cir. 1999); Bedrick v. Travelers Ins. Co., 93 F.3d 149, 152 (4th Cir. 1996).

141. Rush Prudential HMO, Inc. v. Moran, 536 U.S. 355, 362 (2002).

142. This interpretation of ERISA originated in *Pilot Life Insurance Co. v. Dedeaux*, 481 U.S. 41 (1987), where the Court preempted a state-common-law-based suit for bad faith breach of contract. Rush Prudential also made a separate preemption argument that the Illinois statute was not a law regulating insurance, and thus not protected by the savings clause, an argument that the Court rejected. *Moran*, 536 U.S. at 365-73.

143. Moran v. Rush Prudential HMO, Inc., No. 98 C 442, 1999 WL 417384, at *2-4 (N.D. Ill. June 15, 1999)

144. *Id.* at *5.

The Seventh Circuit reversed, holding that the Illinois statute "simply adds to the [insurance] contract . . . an additional dispute resolving mechanism."[145] The court acknowledged that the Fifth Circuit had recently reached the opposite conclusion in a case involving a substantially similar Texas external review law, finding that the Texas law was preempted because it directly conflicted with the ERISA enforcement scheme by "establish[ing] a quasi-administrative procedure for the review of [a decision to deny benefits] and [binding] the ERISA plan to the decision of the independent [reviewer]."[146] Four judges on the Seventh Circuit agreed with the Fifth Circuit, characterizing the Illinois statute as a "law establish[ing] a system of appellate review of benefits decisions" in conflict with the ERISA provisions.[147]

Thus, as the case went to the Supreme Court with an inter-circuit conflict, the question regarding deviation from ERISA's enforcement provisions was framed as whether the Illinois law created an alternative enforcement scheme for adjudicating disputes between patients and HMOs about benefits due under ERISA plans, or whether the law merely mandated one of the terms that insurance contracts sold by HMOs in Illinois were required to include. Most of the parties and amici on both sides found comfort in describing the external review system as quasi-arbitration.

The State of Illinois asserted that its law "effectively requires binding arbitration."[148] Also arguing in support of the law, the United States described the system as a "private, non-judicial arbitration-like mechanism to settle disputes."[149] The AMA used the phrase "independent external review or arbitration."[150] The National Association of Insurance Commissioners called it "an arbitration-like procedure."[151] Moran's brief used the phrase "an objective, independent third-party review."[152] These briefs argued that arbitration did not constitute a separate enforcement scheme than ERISA's because the plaintiff was

145. Moran v. Rush Prudential HMO, Inc., 230 F.3d 959, 972 (7th Cir. 2000).

146. *Id.* at 971 (quoting Corporate Health Ins., Inc. v. Tex. Dep't of Ins., 215 F.3d 526, 539 (5th Cir. 2000), *vacated and remanded sub nom.* Montemayor v. Corporate Health Ins., 536 U.S. 935 (2002)).

147. *Id.* at 973 (Posner, J., dissenting from denial of rehearing en banc).

148. Brief for Respondent State of Illinois, *supra* note 131, at *29.

149. Brief for the United States as Amicus Curiae at *23, Rush Prudential HMO, Inc. v. Moran, 536 U.S. 355 (2002) (No. 00-1021), 2001 WL 1480556.

150. Brief of the American Medical Ass'n et al. as Amici Curiae in support of Respondents at *22, *Moran*, 536 U.S. 355 (No. 00-1021), 2001 WL 1480546.

151. Brief of the National Ass'n of Insurance Commissioners as Amicus Curiae in support of Respondents at *18, *Moran*, 536 U.S. 355 (No. 00-1021), 2001 WL 1673395.

152. Brief for Respondent Debra C. Moran at *13, *Moran*, 536 U.S. 355 (No. 00-1021), 2001 WL 1684554.

entitled to no greater remedy (e.g., no punitive damages) in arbitration than was provided under ERISA.[153] All of them, however, framed the question as involving a dispute resolution mechanism.

The opponents of the law also found the arbitration analogy useful. Rush Prudential characterized the Illinois law as a command for "binding de novo arbitration."[154] Such a system provided a different remedy, not a contractual right, it argued. The "only pertinent difference between [this] law and a judicial remedy is that the decision-maker is a physician."[155] The American Association of Health Plans identified "the distinguishing characteristic" of the law as the fact "that strangers to the plan . . . make judgments about how plan assets must be spent that are binding on plan administrators."[156]

The four Justices who agreed that the Illinois external review system operated as essentially an arbitration procedure held that this form of dispute resolution was indeed an alternative enforcement mechanism and thus should be preempted. "[A]s a binding decision on the merits of the controversy[,] the [external] review resembles nothing so closely as arbitration."[157] Because the independent physician's determination was conclusive on the ultimate question of whether benefits were due, the dissenting Justices found that the Illinois statute set up a procedure for decision-making that conflicted with the procedure established in ERISA. "There is no question that arbitration constitutes an alternative remedy to litigation."[158]

By contrast, the linchpin of the majority's decision was its rejection of the arbitration analogy and, with it, its rejection of the framing of the statute as centering on a form of dispute resolution. The Court essentially reached into the medical deference model and reconfigured the external review system as a state-mandated requirement for a third medical opinion, a formulation that had not been advanced by any of the parties or amici.[159] It found that the external review process "does not resemble either contract interpretation or evidentiary litigation before a neutral arbiter, as much as it looks like a practice (having nothing to do

153. *See, e.g.*, Brief for Respondent State of Illinois, *supra* note 131, at *24; Brief for the United States as Amicus Curiae, *supra* note 149, at *21; Brief of the American Medical Ass'n as Amici Curiae, *supra* note 150, at *22; Brief of the National Ass'n of Insurance Commissioners as Amicus Curiae, *supra* note 151, at *18.

154. Brief for Petitioner at *22, *Moran*, 536 U.S. 355 (No. 00-1021), 2001 WL 1090765.

155. Reply Brief for Petitioner at *8, *Moran*, 536 U.S. 355 (No. 00-1021), 2001 WL 1590509.

156. Brief of American Ass'n of Health Plans, Inc. et al. as Amici Curiae in Support of Petitioner at *26, *Moran*, 536 U.S. 355 (No. 00-1021), 2001 WL 1077919.

157. *Moran*, 536 U.S. at 395 (Thomas, J., dissenting).

158. *Id.*

159. The AMA brief came the closest. *See* Brief of the American Medical Ass'n as Amici Curiae, *supra* note 150, at *25.

with arbitration) of obtaining another medical opinion. . . . [It] is far removed from any notion of an enforcement scheme."[160] The Court cautioned that its interpretation could vary, depending on the procedural regime of a particular external review system.[161]

Driven by the harshness of unremediable wrongs, the Court made the right decision for the wrong reason in *Moran*. The Court elected to stuff the complex reality of external review into the overly simple box of a "second opinion," rather than to reconsider the portion of *Pilot Life* which gave ERISA enforcement mechanisms overly broad reach.[162] In 2004 the Court reaffirmed the scope of ERISA preemption, ruling that patients harmed by an MCO's wrongful denial of coverage for medically necessary care could not obtain damages based on state law.[163]

The Court's decision in *Moran* immeasurably increased the importance of external review laws by giving a green light for their continued and increased use.[164] However, one ironic consequence of the Court's finding that external review systems are not a form of remedy could be to insulate them from due process review. The more distant a mechanism is from the basic adjudicatory function, the less likely it is that it will be scrutinized for procedural fairness,[165] the very motivation for external review laws in the first place. Conversely, the presence of a more fully developed set of procedures within an external review system makes ERISA preemption more likely.

The first signal of weakness created by stronger procedural protections for patients came in *Hawaii Management Alliance Ass'n v. Insurance Commissioner*,[166] in which the Hawaii Supreme Court held that ERISA preempted that state's external review law because it "too closely resemble[d] adjudication."[167] In distinguishing *Moran* the Hawaii court found it "fatal" to that state's system that it allowed judicial review of external review decisions, incorporated portions of the state's Administrative Procedure Act, established

160. *Moran*, 536 U.S. at 383.

161. *Id.* at 381.

162. Pilot Life Ins. Co. v. Dedeaux, 481 U.S. 41, 52 (1987) (holding that the civil enforcement provision in ERISA is the exclusive remedy for enforcement of a right to benefits).

163. Aetna Health Inc. v. Davila, 542 U.S. 200 (2004).

164. Since *Moran*, external review laws in Texas and Maryland have been upheld against ERISA challenges. Corporate Health Ins., Inc. v. Tex. Dep't of Ins., 314 F.3d 784 (5th Cir. 2002); Conn. Gen. Life Ins. Co. v. Ins. Comm'r, 810 A.2d 425 (Md. 2002).

165. Richard C. Reuben, *Constitutional Gravity: A Unitary Theory of Alternative Dispute Resolution and Public Civil Justice*, 47 UCLA L. REV. 949 (2000).

166. 100 P.3d 952 (2004), *cert. denied sub. nom.* Baldado v. Haw. Mgmt. Alliance Ass'n, 125 S. Ct. 2524 (2005).

167. *Id.* at 966.

procedural requirements for hearings, and provided for review by a three-member panel.[168] Numerous other state external review schemes include at least some of those same features.[169]

B. External Review Systems—An Overview

In forty-one states and the District of Columbia, legislatures have enacted laws establishing external review systems.[170] Although the Court in *Moran* noted the varied nature of these laws, neither the Justices nor any of the parties or amici examined the most salient features of such laws, beyond the specifics of the Illinois statute. Had their variety been acknowledged, the Court would have had even more difficulty in casting aside the dispute resolution function which animates these laws.

In general, all of the state systems set forth rules for notification to the patient of an adverse determination, for what share of fees for the review is paid by each party, and for the time deadlines for each stage of the review process.[171]

---

168. *Id.*

169. *See infra* text accompanying notes 195-209, 240-241.

170. ALASKA STAT. § 21.07.050 (2004); ARIZ. REV. STAT. ANN. § 20-2537 (West 2002); CAL. HEALTH & SAFETY CODE § 1370.4 (West 2000 & Supp. 2005); CAL. INS. CODE § 10169 (West 2005); COLO. REV. STAT. ANN. § 10-16-113.5 (West 1999); CONN. GEN. STAT. § 38a-478n (2005); DEL. CODE ANN. tit. 16, § 9119 (2003); D.C. CODE ANN. § 44-301.07 (LexisNexis 2001); FLA. STAT. ANN. § 408.7056 (West 2002); GA. CODE ANN. § 33-20A-39 (West 2002); HAW. REV. STAT. § 432E-6 (LexisNexis 2004); 215 ILL. COMP. STAT. ANN. 134/45 (West 2000); IND. CODE ANN. § 27-13-10.1-1 (West 2003); IOWA CODE ANN. §§ 514J.1-514J.15 (West Supp. 2005); KAN. STAT. ANN. § 40-3228 (2000); KY. REV. STAT. ANN. § 304.17A-623 (LexisNexis Supp. 2004); LA. REV. STAT. ANN. § 22:3081 (2004); ME. REV. STAT. ANN. tit. 24-A, § 4312 (West Supp. 2004); MD. CODE ANN., INS. § 15-10A-05 (LexisNexis 2003); MASS. GEN. LAWS ANN. ch. 176O, § 14 (West Supp. 2005); MICH. COMP. LAWS ANN. §§ 550.1901-550.1929 (West 2002); MINN. STAT. ANN. § 62Q.73 (West 2005); MO. ANN. STAT. § 376.1387 (West 2002); MONT. CODE ANN. § 33-31-303 (2003); N.H. REV. STAT. ANN. §§ 420-J:5-a to 420-J:5-e (LexisNexis 2004); N.J. ADMIN. CODE § 8:38-8.7 (Supp. 2000); N.M. STAT. ANN. § 59A-57-4 (Supp. 2004); N.Y. INS. LAW § 4914 (McKinney 2000); N.C. GEN. STAT. § 58-50-62 (2003); OHIO REV. CODE ANN. § 1751.85 (LexisNexis 2001); OKLA. STAT. ANN. tit. 63, §§ 2528.1-2528.10 (West 2004); OR. REV. STAT. §§ 743.857-743.864 (2003); 40 PA. CONS. STAT. ANN. § 991.2162 (West 1999); R.I. GEN. LAWS § 23-17.12-10 (2001); S.C. CODE ANN. §§ 38-71-1910 to 38-71-2060 (2004); TENN. CODE ANN. § 56-32-227 (2000); TEX. INS. CODE ANN. § 21.58C (Vernon 2004); UTAH CODE ANN. § 31A-22-629 (2001); VT. STAT. ANN. tit. 8, § 4089f (2001); VA. CODE ANN. § 38.2-5901 (Michie 1999 & Supp. 2002); WASH. REV. CODE ANN. § 48.43.535 (West Supp. 2005); W. VA. CODE ANN. § 33-25C-6 (West 2002); WIS. STAT. ANN. § 632.835 (West 2004).

171. POLLITZ ET AL., *supra* note 95, at vii. Pollitz provides the best comparative analysis of the external review laws, but the report does not cite to specific statutory provisions.

In nearly all states, a final external review decision that a benefit is covered by the policy is binding on the MCO,[172] although in two states there is a mechanism by which a health plan can opt out of participation in the external review system.[173] Beyond those broad principles, however, there are many variations.[174] The external review laws have produced finely-calibrated adjudication systems, as one sees by using traditional categories of procedural jurisprudence to describe them.

As in any adjudication system, jurisdictional authority is a threshold question. In all states an external review complaint can be brought against managed care plans, subject to the exemption for self-insured plans under ERISA; the majority of external review laws cover all insured health plans.[175] Treatment decisions based on a determination that a particular service was not medically necessary are subject to review in all states; in addition, some state systems also review denials based on the assertion that a treatment was experimental or investigational, and thus not covered by the contract of insurance.[176]

There is significant variation among the states in the standards of review for determining what is medically necessary. This is the most important point of law at issue in the external review systems, and the one as to which a body of external review decisions may ultimately have the greatest impact on health policy. The critical question is the extent to which external review mechanisms will force health plans to change how the necessity standard is applied to particular treatments.

At one end of the continuum, the MCOs receive considerable deference. North Carolina law specifically authorizes insurers to "compar[e] the cost-effectiveness of alternative services or supplies."[177] Several states explicitly limit

---

172. Neither party is bound by the external review decision in the District of Columbia. D.C. CODE § 44-301.07(p) (2001).

173. In Oregon, the health plan can opt out of the external review system; if so, it must inform enrollees of that fact. OR. REV. STAT. §§ 743.859, .863, .864 (2003). West Virginia allows a managed care plan to get an exemption from the system if it has "an established external review procedure in place." W. VA. CODE R. § 114-58-12.1 (2002).

174. For state-by-state summaries of the laws, see Joyce Krutick Craig, *Managed Care Grievance Procedures: The Dilemma and The Cure*, 21 J. NAT'L ASS'N ADMIN. L. JUDGES 336 (2001).

175. POLLITZ ET AL., *supra* note 95, at 8-9.

176. *Id.* States differ as to how and by whom cases are screened for whether they present issues which are eligible for external review. Regulatory agencies perform this function in most states, but health plans do the screening in eight states. *Id.* at 16-17.

177. N.C. GEN. STAT. § 58-3-200(b) (2003). However, North Carolina law also guarantees that the external review will be conducted on a de novo basis. N.C. GEN. STAT. § 58-50-80(j) (2003).

the scope of review by mandating that reviewers will be bound by the health plan's definition of medically necessary,[178] an escape valve that could be used to render the review process a facade if plans succeeded in marketing contracts that simply excluded a large number of recurringly problematic disputes. In these states, the external review can be essentially limited to a check on whether the health plan complied with its own protocols. However, such language can be combined with language providing more latitude, as in West Virginia, where the ERO is limited to applying the terms of the plan, but regulations also define medical necessity to include "[t]he most appropriate available supply or level of service."[179] Moreover, regulators in some states may informally exercise discretion even when they are bound by the plan's definition.[180]

Other state laws have placed thumbs on the opposite end of the scale. Some states have created a rebuttable presumption in favor of the care recommended by the treating physician. Iowa, for example, requires that such treatment be upheld unless it is found to be neither medically necessary nor consistent with clinical standards of medical practice.[181] Although contained in regulation rather than statute, District of Columbia law achieves the same end by defining medically necessary in terms of the treating physician's recommendation.[182]

These standards may at first appear to be pro-patient, but that reaction is based upon the assumption that external reviews arise when a treating physician's recommendation is vetoed by the managed care entity. While that situation is not uncommon, the dispute may also arise between a treating physician, possibly receiving a financial incentive to underutilize care, and an outside specialist secured by the patient, as was the case in *Moran*,[183] in which case the impact of the rebuttable presumption in favor of the treating physician could favor the insurer.

Redefinition of medical necessity definitely occurs. Washington grants external review entities specific statutory authority to override the health plan's medical necessity standards if they are determined "to be unreasonable or inconsistent with sound, evidence-based medical practice."[184] It is unclear whether such a finding could be utilized in a subsequent claim for malpractice

178. ALASKA STAT. § 21.07.050(d)(1) (2004); ARIZ. REV. STAT. ANN. § 20-2537(E) (2002); KAN. STAT. ANN. § 40-22a14(b) (2000); OR. REV. STAT. § 743.862(2) (2003) (unless plan's definition is "unreasonable"); TENN. CODE ANN. § 56-32-227(b)(6) (2000); WIS. STAT. ANN. § 632.835(3m) (West 2004).

179. W. VA. CODE R. § 114-58-5.5 (2002).

180. Berman-Sandler, *supra* note 9, at 268-69.

181. IOWA CODE ANN. § 514J.12 (West Supp. 2005).

182. D.C. MUN. REGS. tit. 22, § 6099.1 (2005).

183. *See supra* text accompanying note 133.

184. WASH. REV. CODE ANN. § 48.43.535(5) (West Supp. 2005).

against an MCO, not merely a physician, based on corporate negligence in standard-setting. In practice Michigan allows an external review decision to deviate from the terms of the contract if the medically necessary standard is being applied in an obsolete or unworkable fashion.[185]

A number of states do not address the degree of deference to be shown to a health plan's definition of medically necessary, but instead invoke general concepts of consistency with accepted standards of medical practice and professionally established protocols, such as Minnesota's definition of medically necessary care as "health care services appropriate, in terms of type, frequency, level, setting and duration, to the enrollee's diagnosis or condition, and diagnostic testing and preventive services."[186] Some states look to the "most appropriate available supply" relevant for the particular treatment,[187] or require compliance with "scientific evidence" as to the usefulness of disputed services.[188] Regulations in Vermont specify that the ERO is "not bound by the insurer's clinical protocols or practice guidelines."[189] About half of the states with external review systems allow not only for reversal or affirmation of the health plan's denial but also for modification, perhaps to allow some portion of the treatment being sought.[190]

The Georgia statute exemplifies the final option used by states, which is to craft an exceedingly detailed definition of medical necessity solely for external review purposes. Georgia defines medical necessity as

> care based upon generally accepted medical practices in light of conditions at the time of treatment which is: (A) Appropriate and consistent with the diagnosis and the omission of which could adversely affect or fail to improve the eligible enrollee's condition; (B) Compatible with the standards of acceptable medical practice in the United States; (C) Provided in a safe and appropriate setting given the nature of the diagnosis and the severity of the symptoms; (D) Not provided solely for the convenience of the eligible enrollee or the convenience of the provider or hospital; and (E) Not primarily custodial care, unless custodial care is a covered service or benefit under the eligible enrollee's coverage.[191]

---

185. Berman-Sandler, *supra* note 9, at 268.

186. MINN. R. 4685.0100(9)(b) (2005).

187. MASS. REGS. CODE tit. 105, § 128.020 (2005).

188. *Id.*

189. 21-040-012 VT. CODE R. § 7(B) (2004).

190. POLLITZ ET AL., *supra* note 95, at 18-19.

191. GA. CODE ANN. § 33-20A-31(5) (2000 & Supp. 2003). Maine also has a highly specific, although less complex, statutory definition. ME. REV. STAT. ANN. tit. 24-A, § 4301-A(10-A) (Supp. 2003).

A number of states specify that the external review shall be conducted as a de novo review.[192] The critical question of which party has the burden of proof is left unanswered in every state except Maryland,[193] which assigns it to the managed care entity.

What follows is a detailed description of the procedural workings of external review systems. Here, however, the devil is *not* in the details. Although the particularities are necessary to an understanding of the broader themes, the greatest significance of external review laws is that they illuminate what are the currently dominant norms and values related to adjudication. External review laws tell us that faith in adjudication has not only failed but collapsed.[194] Imagine a game of "Legal Jeopardy" with this problem: If external review is the answer, what is the question? The correct response: If legislators starting from scratch invented an adjudication system today, what would it look like?

1. Evidence and Hearings

Some statutes, especially those which require consistency with accepted standards of practice, specify what evidence can or must be considered in the ERO's decision-making process.[195] In some states external reviewers are required

---

192. ALASKA STAT. § 21.07.050(d)(1) (LexisNexis 2000); MASS. GEN. LAWS ANN. ch. 176O, § 14(a) (West Supp. 2005); MICH. COMP. LAWS ANN. § 550.1911(8) (West 2002); MINN. STAT. ANN. § 62Q.73(7) (West 2004); N.H. REV. STAT. ANN. § 420-J:5-b(IX) (2004); N.C. GEN. STAT. § 58-50-80(j) (2003); OHIO REV. CODE ANN. § 1751.84(E) (Anderson 2004); S.C. CODE ANN. § 38-71-1970(D)(2) (2002); 14 VA. ADMIN. CODE § 5-215-70(b) (2005).

193. MD. CODE ANN., INS. § 15-10A-05 (LexisNexis 2002).

194. The "failure" reference is to Judith Resnik, *Failing Faith: Adjudicatory Procedure in Decline*, 53 U. CHI. L. REV. 494 (1986).

195. A typical provision states:

> [A]n independent review entity . . . shall take into account all of the following:
>
> (a) Information submitted by [a party] including:
>
> 1. The covered person's medical records;
> 2. The standards, criteria, and clinical rationale used by the insurer to make its decision; and
> 3. The insurer's health benefit plan.
>
> (b) Findings, studies, research, and other relevant documents of government agencies and nationally recognized organizations [giving examples]; and
>
> (c) Relevant findings in peer-reviewed medical or scientific literature, published opinions of nationally recognized medical specialists, and clinical guidelines adopted by relevant medical societies.

KY. REV. STAT. ANN. § 304.17A-625(1) (LexisNexis 1996 & Supp. 2004).

to consider evidence in each listed category;[196] in other states they are merely authorized to do so.[197] Other states, like Alaska, specify both, listing certain categories of evidence which must be considered and others that are discretionary.[198]

Several states have established procedures for information-sharing between the parties that parallel discovery. Health plans may have an affirmative duty to send the patient a copy of any documents that it submits to the external reviewer, including clinical review criteria, unless the documents contain proprietary information.[199] Other states provide patients a right to see all evidence in the case, but do not impose a duty on the health plan to furnish it to them;[200] in New York and Oklahoma, patients may see the evidence upon request.[201] EROs in West Virginia have subpoena power.[202]

In general, external review systems conduct exclusively paper reviews, but there are, nonetheless, a wide variety of provisions regarding in-person hearings before which the patient or her representative can appeal. In some states the external review proceeding is a function of a state agency, so that the final decision is an agency decision, even if it consists only of a commissioner's approval of the decision of a private external review organization. In those states the possibility of a hearing may be governed by the state's body of administrative law, or there may be an explicit provision for the external review process.[203] In Missouri external reviews of treatment denials are specifically excluded from the provision of state administrative law that would require a hearing.[204] A hearing "may" be held in Maryland and New Hampshire.[205]

---

196. *See, e.g., id.*; N.H. REV. STAT. ANN. § 420-J:5-b(IX) (2004); OHIO REV. CODE ANN. § 1751.84(D)(8) (LexisNexis 2004).

197. *See, e.g.*, LA. REV. STAT. ANN. § 22:3082(A) (West 2004).

198. ALASKA STAT. § 21.07.050(d)(3), (4) (2004).

199. *See, e.g.*, GA. CODE ANN. § 33-20A-36(3) (2000); 3 COLO. CODE REGS. § 4-2-21(8)(C)(2) (2005).

200. *See, e.g.*, MASS. REGS. CODE tit. 105, § 128.403(B) (2005); N.H. REV. STAT. ANN. § 420-J:5-b(VII) (LexisNexis 2004); N.C. GEN. STAT. § 58-50-80(d)(3) (2003); TENN. CODE ANN. 56-32-227(b)(3)(B) (2000).

201. N.Y. INS. LAW § 4912(g) (McKinney 2000); OKLA. STAT. ANN. tit. 63, § 2528.7(B) (West 2004).

202. W. VA. CODE ANN. § 33-25C-6(m) (LexisNexis 2003).

203. Patients in Florida and New Mexico have a right to an administrative hearing. *See* FLA. STAT. ANN. § 408.7056(3) (West 2002); N.M. CODE R. § 13.10.17.31(A) (Weil 2005).

204. MO. ANN. STAT. § 376.1387(1) (West 2002), exempting external review from MO. ANN. STAT. § 536.010 (West 2000).

205. MD. CODE ANN., INS. § 15-10A-04(a)(3) (2002); MD. CODE REGS. 31.10.18.15(B); N.H. REV. STAT. ANN. § 420-J:5-b(IV)(a) (LexisNexis 2004).

Among the majority of states where the final decision emerges from the external review organization without adoption by a state agency, only the District of Columbia and Maine provide for the patient to appear at a hearing and present evidence.[206] Vermont law provides for a teleconference if the patient requests one.[207] Massachusetts leaves the question of whether to hold an in-person hearing to the discretion of the reviewer.[208] Wisconsin explicitly forbids appearances before the ERO by a party, a party's representative, or a witness.[209]

In the four states in which she interviewed regulatory officials, Berman-Sandler found a pattern of mostly paper-only review, similar in form to alternative dispute resolution, but with the "plaintiff" retaining the right to sue.[210] Extensive contacts have developed between state agencies and the independent review companies, with staff of the former reporting increasingly higher quality levels for the work done by the latter.[211] Attorneys are seldom involved; their presence tends to occur, predictably, only in big-dollar cases.[212] On the whole she found that external review "looks like a [state agency's] consumer protection and complaint system . . . assess[ing] whether a health plan has followed prescribed regulatory conduct. The review of conduct includes elements of procedural due process . . . ."[213]

2. Selection and Structure of the EROs

Independence and impartiality on the part of adjudicators is a cardinal principle in American law.[214] Most states include provisions prohibiting conflicts of interest between EROs and their staff and the MCOs which are the source of the claims denials being reviewed, as well as any conflicts that might arise as to particular individual patients or their primary care physicians.[215] Most states have some procedure for designating or certifying which entities may operate as EROs for appeals within that state.[216] In a number of instances, review entities that have

206. D.C. CODE ANN. § 44-301.07(k) (LexisNexis 2001); ME. REV. STAT. ANN. tit. 24-A, § 4312(5)(B) (West Supp. 2004).

207. 21-040-012 VT. CODE R. § 7(C) (2004).

208. MASS. ANN. LAWS ch. 176O, § 14 (West Supp. 2005).

209. WIS. STAT. ANN. 632.835(3)(d) (West 2004).

210. Berman-Sandler, *supra* note 9, at 256-57.

211. *Id.* at 258.

212. *Id.* at 261-62.

213. *Id.* at 257.

214. The mechanisms described in this paragraph are discussed in Craig, *supra* note 174; POLLITZ ET AL., *supra* note 95, at 16-18.

215. POLLITZ ET AL., *supra* note 95, at 17-18.

216. Berman-Sandler, *supra* note 9, at 239.

been accredited by a private accreditation agency are deemed to be certified.[217]

There is greater variance in the statutory requirements as to who selects which ERO will review which claim and in the degree of oversight of the EROs. The majority of states provide for an ERO to be assigned to each appeal by a state regulatory office. However, fourteen states allow the health plan to exercise some degree of control over the selection of the ERO.[218] In ten states the health plan simply gets to choose which ERO will hear the case.[219] In Ohio the plan chooses one of two EROs recommended by the state.[220] In Oregon the plan can reject the state's selection of an ERO once in any proceeding.[221] Illinois and Montana both provide that the patient and the plan will jointly select the ERO,[222] although it is difficult to imagine that the patient, unless represented by counsel with substantial external review experience, would be able to make an informed choice. The same problem would be true also in Wisconsin, where the patient

---

217. *See, e.g.*, COLO. REV. STAT. ANN. § 10-16-211(11)(F) (West 1999).

218. ALASKA STAT. § 21.07.050(c) (LexisNexis 2004); IND. CODE ANN. § 27-13-10.1-2 (LexisNexis 1999); IOWA CODE ANN. § 514J.7(1)(a) (West Supp. 2002); KY. REV. STAT. ANN. §§ 304.17A-623(7), -627 (LexisNexis 2004); 806 KY. ADMIN. REGS. 17:290 (2005); LA. REV. STAT. ANN. § 22:3082(A) (2004); LA. ADMIN. CODE tit. 37, pt. XIII, § 6227(A) (2005); OKLA. STAT. ANN. tit. 63, § 2528.6(B) (West 2004); S.C. CODE ANN. § 38-71-1970(B)(1)(a) (2002); TENN. CODE ANN. § 56-32-227(a) (2000); UTAH CODE ANN. § 31A.22-629(1)(b)(iii) (2003); UTAH ADMIN. CODE R590-203-6(3) (2005); WASH. REV. CODE ANN. § 48.43.535(3) (West Supp. 2005). In Iowa and Oklahoma, the patient can object to the state regulatory agency if she believes the ERO to be biased. In Tennessee patients may file appeals either with a state agency, in which case agency staff conduct the review, or with their health plan, in which case the plan selects the reviewer. OHIO REV. CODE ANN. § 3901.80(c) (Anderson 2002). OR. REV. STAT. § 743.858 (2003). 215 ILL. COMP. STAT. ANN. 134/45(f)(3)(A) (West 2000); MONT. CODE ANN. § 33-37-103 (2003); MONT. ADMIN. R. 37.108.305(1) (2005). In Montana the state agency will select an ERO if the parties cannot agree.

219. ALASKA STAT. § 21.07.050(c) (LexisNexis 2004); IND. CODE ANN. § 27-13-10.1-2 (LexisNexis 1999); IOWA CODE ANN. § 514J.7(1)(a) (West Supp. 2002); KY. REV. STAT. ANN. §§ 304.17A-623(7), 304.17A-627 (LexisNexis 2004); 806 KY. ADMIN. REGS. 17:290; LA. REV. STAT. ANN. § 22:3082(A) (2004); LA. ADMIN. CODE tit. 37, pt. XIII, § 6227(A) (2005); OKLA. STAT. ANN. tit. 63, § 2528.6(B) (West 2004); S.C. CODE ANN. § 38-71-1970(B)(1)(a) (2002); TENN. CODE ANN. § 56-32-227(a) (2000); UTAH CODE ANN. § 31A.22-629(1)(b)(iii) (2003); UTAH ADMIN. CODE R590-203-6(3) (2005); WASH. REV. CODE ANN. § 48.43.535(3) (West Supp. 2005). In Iowa and Oklahoma, the patient can object to the state regulatory agency if she believes the ERO to be biased. In Tennessee patients may file appeals either with a state agency, in which case agency staff conduct the review, or with their health plan, in which case the plan selects the reviewer.

220. OHIO REV. CODE ANN. § 3901.80(c) (Anderson 2002).

221. OR. REV. STAT. § 743.858 (2003).

222. 215 ILL. COMP. STAT. ANN. 134/45(f)(3)(A) (West 2000); MONT. CODE ANN. § 33-37-103 (2003); MONT. ADMIN. R. 37.108.305(1) (2005). In Montana the state agency will select an ERO if the parties cannot agree.

alone makes the selection.[223]

The issue of control over selection of the decision-maker is all the more critical in external review because of the open-ended nature of most of the standards that will be used. There are obviously serious problems with a system in which one party, with the advantages of repeat-player status, could develop unspoken relationships and understandings with private judges before whom the plan will appear again and again.

3. Remedies, Precedent, and Further Proceedings

a. Remedies and Precedent

The basic remedies from external review are simple. A patient who appeals a denial of treatment is entitled, in most states, to a written decision. The successful patient is entitled to an order directing the health plan to provide (or pay for, if it is a retrospective case) the treatment which was initially denied. No monetary damages are awarded in any external review system. Interim relief is available in Massachusetts and Washington state, where EROs can order the plan to continue providing a service pending the external review decision.[224]

Several statutes have language designed to insure that the decision is written in clear language, understandable to the layperson.[225] Michigan requires a "plain English" explanation of the basis for the decision.[226] In several states, the decision must list the evidence which was considered by the panel and relied on as the basis for the decision.[227]

Perhaps the single most remarkable characteristic of the external review systems, from a due process point of view, is the absence of precedent. With three exceptions, there is no readily accessible body of prior written decisions. Public availability is required by law only in Oregon, which posts redacted versions of decisions on a web site.[228] California and Michigan have elected to do

223. WIS. STAT. ANN. § 632.835(3)(a) (West 2004).

224. MASS. ANN. LAWS ch. 176O, § 14(b) (LexisNexis 2002) (requiring showing of substantial harm); WASH. REV. CODE ANN. § 48.43.535(8) (West Supp. 2005) (stating that if the health plan wins the appeal, it can seek recoupment from the patient of the expense of the service).

225. *See, e.g.*, KY. REV. STAT. ANN. § 304.17A-625(6) (LexisNexis 2004); 105 MASS. CODE REGS. § 128.415(B) (2005); S.C. CODE ANN. § 38-71-1970(H)(3) (West 2002).

226. MICH. COMP. LAWS ANN. § 550.1911(16) (West 2002).

227. *See, e.g.*, MONT. ADMIN. R. 37.108.305(3) (2005); N.H. REV. STAT. ANN. § 420-J:5b(X) (LexisNexis 2004).

228. OR. REV. STAT. § 743.862(5) (2001); OR. ADMIN. R. 836-053-1355 (2005).

the same.[229] Decisions which go through an administrative agency may be accessible through other means.[230] Kentucky's law addresses the question of precedent in a different way, by specifying that the effect of each decision will be limited to that one case only.[231]

Thus, precedent can build up or affect future or other cases only on an informal basis. This situation creates an enormous imbalance between the parties: repeat players, most often the MCOs, can create and maintain their own compilations of decisions and use them as guidance. Only individuals who pursue appeals with the assistance of similarly-organized representatives would have a comparable information base. The absence of publicly-accessible statements of reasons for external review decisions also fundamentally disables systemic oversight of health plans.

b. Further Review and Litigation

One major mitigating factor for all of the due process deficiencies in external review systems is that, as a general matter, patients can subsequently bring suit against the health plan for injuries caused by the denial of treatment.[232] However, the law shapes the effect that an external review decision could have on a later lawsuit in a number of ways, some of which would bar litigation.

In West Virginia, the plaintiff must have gone through and prevailed in the external review process in order to bring suit.[233] In seven other states the plaintiff must have first pursued external review before she can litigate.[234] In yet other states the outcome of the external review creates a rebuttable presumption in a

229. *See* Cal. Dep't of Managed Health Care, Independent Medical Review Decisions, http://wp.dmhc.ca.gov/imr (last visited Nov. 1, 2005); Mich. Dep't of Labor & Econ. Growth, Labor & Economic Growth, http://www.michigan.gov/cis/0,1607,7-154-10555_20594_20596--,00.html (last visited Nov. 1, 2005).

230. In Arizona external review determinations become "final administrative decisions" which can be obtained in the same way as other administrative decisions. ARIZ. REV. STAT. ANN. § 20-2537(H) (2002). A newspaper was able to obtain redacted copies of the decisions in Maryland through a freedom of information request to the Insurance Administration. Bill Brubaker, *Health Insurance Consumers Wield the Power of Appeal*, WASH. POST, July 3, 2004, at E1.

231. KY. REV. STAT. ANN. § 304.17A-625(3), (9) (LexisNexis Supp. 2004).

232. As a practical matter, however, this option may not offer much relief. Persons insured through a private sector workplace plan can bring suit only under ERISA, under which they can recover only actual expenses and attorneys fees. 29 U.S.C. § 1132(a) (2000).

233. W. VA. CODE ANN. § 33-25C-7(e) (LexisNexis 2002).

234. CAL. CIVIL CODE § 3428(k)(1) (West 2004); GA. CODE ANN. § 51-1-49 (2000); ME. REV. STAT. ANN. tit. 24-A, § 4313(2)(B) (2003); OKLA. STAT. ANN. tit. 36, § 6594(A) (West 2004); 14-000-016 R.I. CODE R. § 6.1.12 (Weil 2005); TEX. CIV. PRAC. & REM. § 88.003(a)(2)(B) (Vernon 2005); WASH. REV. CODE ANN. § 48.43.545(7)(a)(ii) (West Supp. 2005).

later lawsuit.[235] Some states specify that an external review decision is admissible in subsequent proceedings;[236] Kentucky courts must accept the decision as "scientifically valid and accurate."[237] At the other extreme South Carolina bars its use as evidence,[238] and the District of Columbia specifies that the decision "shall not affect any other legal cause of action."[239]

Most state laws do not provide for further appeal of the external review decision itself. Exceptions are Alaska, Delaware, Michigan, Minnesota, New Mexico, Pennsylvania, and Rhode Island.[240] Some states limit judicial review only to "arbitrary and capricious" decisions by an ERO.[241]

4. Utilization and Reversal Rates

Differences in jurisdiction, definitions, and reporting make it difficult to assess the extent to which patients actually utilize external review systems. The Kaiser Family Foundation and American Association of Health Plans found a national average of 0.7 reviews per 10,000 enrollees, but that average was derived from an enormous range: from 0.2 to 1.7 per 10,000.[242] There are no national data on the number of external review appeals filed. In the four states that she examined in detail, Berman-Sandler found a pattern of rapid growth in the early years of external review, followed by a leveling off of the number of appeals. From 1999 to 2002 in those states, 5129 cases entered the external review systems.[243] There are large variations among the states. In the four largest states, in the most recent year for which data were available in 2004, California

235. COLO. REV. STAT. ANN. § 10-16-113.5(11) (West 1999); GA. CODE ANN. § 33-20A-37(b) (2000) (presumption available only to a prevailing MCO); MO. ANN. STAT. § 376.1387(1) (West 2002); 40 PA. STAT. ANN. § 991.2162(c)(5) (West 1999).

236. ARIZ. REV. STAT. ANN. § 20-2537(J) (2002); KY. REV. STAT. ANN. § 304.17A-625(8) (LexisNexis 2004); N.C. GEN. STAT. ANN. § 90-21.55(a) (West 2003); OHIO REV. CODE ANN. § 1751.88 (West 2004).

237. KY. REV. STAT. ANN. § 304.17A-625(8) (LexisNexis 2004).

238. S.C. CODE ANN. § 38-71-1990(B) (2002).

239. D.C. CODE ANN. § 44-301.07(p) (LexisNexis 2001).

240. ALASKA STAT. § 21.07.050(d)(8) (Michie 2004); DEL. CODE ANN. tit. 18, § 332 (1999); MICH. COMP. LAWS ANN. § 550.1915(1) (West 2002); MINN. STAT. ANN. § 62Q.73(8) (West 2005); N.M. STAT. ANN. § 59A-4-20 (Michie 2002); 40 PA. CONS. STAT. ANN. § 991.2162(c)(5) (West 2005); R.I. GEN. LAWS § 23.17.12-10(b)(6) (2001).

241. *See, e.g.*, MINN. STAT. ANN. § 62Q.73(8) (West 2005).

242. Berman-Sandler, *supra* note 9, at 249.

243. Berman-Sandler, *supra* note 9, at 301-02 app. B, ch. II. This total includes 528 cases decided in the Pennsylvania system, which did not report a number for "cases accepted." These numbers can serve as only a rough gauge because the start date for external review varied in each of the states.

processed 586 cases; New York adjudicated 796 external reviews; Texas decided 231 cases; and Florida reported only 29.[244] In Maryland 280 cases went into external review in fiscal year 2003,[245] while in the two much larger states of Michigan and Ohio, the most recent annual numbers were 129 and 176, respectively.[246]

Observers have consistently found serious problems in access to external review, growing primarily from the requirement that the plan's system of internal reviews be exhausted and from failures to effectively communicate to patients how external review works or even that it exists.[247] The Kaiser Family Foundation report criticized this exhaustion of remedies requirement, together with weak notice provisions, for resulting in many fewer appeals than would otherwise be the case, because patients may be required to go through one or even two levels of internal review, and sometimes are not told that external review exists until they have completed internal review.[248]

A study of the California system, the most detailed examination of the workings of an external review system, found an enormous drop-off in the number of complaints regarding denials of care and the number that ultimately went to independent review: from 6127 to 299. [249] Even in the same state, there was a huge difference among the six plans' studies: Rates ranged from three to twenty-five percent of complaints proceeding to external review.[250] Two-thirds of

---

244. CAL. DEP'T OF MANAGED HEALTH CARE, INDEPENDENT MEDICAL REVIEW DETERMINATIONS, *available at* http://www.hmohelp.ca.gov/imr/stats.pdf; N.Y. STATE DEP'T OF INS., NEW YORK STATE EXTERNAL APPEAL PROGRAM 2002, at 27, *available at* http://www.ins.state.ny.us/acrobat/extapp02.pdf; TEX. OFFICE OF PUB. INS. COUNSEL, COMPARING TEXAS HMOS 2003, at 109, *available at* http://www.opic.state.tx.us/docs/238_printout2003hmo.pdf; FLA. AGENCY FOR HEALTH CARE ADMIN., STATEWIDE PROVIDER AND HEALTH PLAN CLAIM DISPUTE RESOLUTION PROGRAM, ANNUAL REPORT 2004, at 3, *available at* http://www.fdhc.state.fl.us/MCHQ/Managed_Health_Care/SPHPClaimDRP/Annual_Report_2004.pdf. These figures include only medical necessity cases in states where experimental treatment decisions are also reviewed.

245. Brubaker, *supra* note 230.

246. Mich. Office of Financial & Ins. Servs., HMO Complaint Information, http://www.michigan.gov/documents/cis_ofis_compinfo_28032_7.html (last visited Oct. 29, 2005); Jane DuBose, Healthleaders, Ohio's External Review Experience, (July 2, 2004), http://www.healthleaders.com/news/feature56164.html.

247. POLLITZ ET AL., *supra* note 95, at vii, 8-14; Berman-Sandler, *supra* note 9, at 247-53.

248. POLLITZ ET AL., *supra* note 95, at 5-7, 10, 12; *see also* Berman-Sandler, *supra* note 9, at 247-53.

249. JILL K. SILVERMAN ET AL., INST. FOR MED. QUALITY, INDEPENDENT MEDICAL REVIEW EXPERIENCES IN CALIFORNIA, PHASE II: CASES INCLUDING MEDICAL NECESSITY 36-37 (2003) [hereinafter IMR 2003].

250. *Id.*

patients were unaware of the system until their individual cases arose,[251] and a majority of them were not informed of the deadlines that applied.[252] These data do not tell us whether further appeals were not pursued because patients were satisfied with the outcome or because of problems with access. However, in the Medicare program, where appeals are automatic rather than contingent on patient filings, the rate of utilization is dramatically higher than in state external review programs.[253]

When patients make it as far as the external review system in a state, they have a significant win rate. Although it is impossible to evaluate what any reversal rate means, or whether it means anything more than survival of the strongest cases, patients who appeal do have a meaningful chance of success. In the largest states, external review panels reversed denials of care approximately forty percent of the time.[254]

C. Non-Judicial Conferences: The Framers of External Review

The structure of the state external review laws, with their components modeled on the components of a dispute resolution process, make clear that these systems are not merely commands to secure an additional physician's opinion, as the Supreme Court's majority described them in *Moran*. Although external review systems can fairly be described in part as attempts to restore some deference to medical authority, the states accomplished the restoration by means of an adjudication system. The primary voices in the debates over enacting these laws were not those whom we think of as proceduralists, such as judges or law professors, but those of the interest groups directly in conflict. Because of that, the tradition that the process of making process[255] should be a function of neutral experts did not carry over into the process of making external review systems.[256]

---

251. *Id.* at 12.

252. *Id.* at 14.

253. KAREN POLLITZ ET AL., INST. FOR HEALTH CARE RES. & POL'Y, EXTERNAL REVIEW OF HEALTH PLAN DECISIONS: AN OVERVIEW OF KEY PROGRAM FEATURES IN THE STATES AND MEDICARE, at iv-v, viii, 17-18 (1998).

254. Berman-Sandler, *supra* note 9, at 256.

255. The phrase is Robert Bone's. Robert G. Bone, *The Process of Making Process: Court Rulemaking, Democratic Legitimacy, and Procedural Efficacy*, 87 GEO. L.J. 887 (1999).

256. Beginning with the Rules Enabling Act in 1934 and the adoption of the Rules in 1938, Congress has delegated primary responsibility for the adoption and revision of the Rules of Civil Procedure to the Supreme Court. The prevailing understanding was that in vesting this power in the Court rather than in Congress or the Justice Department, "Congress placed rulemaking under the institution it perceived to be least responsive to interest group politics." Paul D. Carrington, *Making Rules To Dispose of Manifestly Unfounded Assertions: An Exorcism of the Bogy of Non-Trans-*

Instead, the enactment of external review laws illustrates the emergence of openly adversarial proceduralism.

The central players in the legislative battles fell into five general camps: insurance and other business interests who sought to limit the scope and extent of review; consumer groups who sought to maximize it; medical groups who sought to shift, or at least share, potential malpractice liability with managed care organizations; governmental regulatory agencies; and private organizations that accredit health care organizations.[257] Stakeholders engaged in debates over external review based on bottom-line interests.[258] Consumer advocates proposed

---

Substantive Rules of Civil Procedure, 137 U. PA. L. REV. 2067, 2075 (1989). In turn, the Court shared its authority with the Judicial Conference and subordinate committees, a structure thought to be "substantially immunized from the possibility of influence resulting from direct interest or coercive pressures brought to bear by organized groups." *Id.* at 2077.

In recent years, the mask of neutrality has fallen away from the court rule-making model. Openly interest-based advocacy began in earnest in the 1980s, and has continued, especially in hard-fought contests over Rule 11 and the rules governing discovery. *See* Bone, *supra* note 255, at 903-04; Richard L. Marcus, *Of Babies and Bathwater: The Prospects for Procedural Progress*, 59 BROOK. L. REV. 761, 805-12, 819 (1993); Linda S. Mullenix, *Hope over Experience: Mandatory Informal Discovery and the Politics of Rulemaking*, 69 N.C. L. REV. 795, 851-55 (1991); Kent Sinclair, *Service of Process: Rethinking the Theory and Procedure of Serving Process Under Federal Rule 4(c)*, 73 VA. L. REV. 1183, 1197-212 (1987); Jeffrey L. Stempel, *Politics and Sociology in Federal Civil Rulemaking: Errors of Scope*, 52 ALA. L. REV. 529 (2001); Jeffrey L. Stempel, *New Paradigm, Normal Science, or Crumbling Construct? Trends in Adjudicatory Procedure and Litigation Reform*, 59 BROOK. L. REV. 659 (1993).

257. The Senate Committee on Health, Education, Labor and Pensions heard testimony in 1998 on "health care quality: grievance procedures" and in 1999 on "key patients' protections." A total of fifteen witnesses testified at the two hearings. *Health Care Quality: Grievance Procedures: Hearing on S. 326 Before the S. Comm. on Health, Education, Labor, & Pensions*, 105th Cong. (1998) (witness list); *Key Patients' Protections: Lessons from the Field: Hearing on S. 326 Before the S. Comm. on Health, Education, Labor, & Pensions*, 106th Cong. (1999) (witness list). They included six government officials, five industry representatives, three consumer representatives, and one medical spokesperson. There was no testimony from a bar association or other legal representative or from anyone representing the standpoint of procedural expertise. *Id.* In its evaluation of HMO complaints and appeals procedures, the GAO identified one industry group, one consumer group, the association of state insurance regulators, and two private accrediting organizations as the most influential advocates involved in the issue. GAO REPORT, *supra* note 95.

258. Shirley Eiko Sanematsu, *Taking a Broader View of Treatment Disputes Beyond Managed Care: Are Recent Legislative Efforts the Cure?*, 48 UCLA L. REV. 1245, 1263 (2001) ("The reason most often cited for the [California] legislation was to curb managed care's profit-making incentive to deny care."); Nathaniel S. Shapo, *In the Eye of the Storm: A Regulator's Perspective on Managed Care Organization Liability*, 30 J. LEGAL STUD. 669, 683 (2001) ("view[ing] the MCO liability debate [in Illinois] through the prism of interest group politics, where the clash between MCOs and doctors' groups is extraordinar[ily] hostile"); Lisa Strauss, *Managed Healthcare Plans*,

external review laws as vehicles for inserting a patients' rights orientation into the administration of managed care to balance cost containment incentives, with procedural design oriented toward that goal; and physicians sought reinstatement of at least some of their power vis-à-vis payors. The response of the insurance industry was mixed: initially negative, then increasingly supportive of external review as a relatively inexpensive way to curb litigation and limit liability.[259] When the idea of external review reached Congress, debate centered on its usefulness for decreasing litigation[260] and the impact of lawsuits on the managed care industry.[261]

Although counter to the norm of rule neutrality, this transparently pluralist negotiation of an adjudication mechanism has advantages. Direct interest group negotiation leads to a balancing of the real interests in play, rather than hiding them behind a screen of what is sometimes false objectivity.

Openly pluralist negotiation of procedure also has its costs, however. Even though judicial authors do not guarantee impartial rules of procedure, the court rule-making model reflected in the Federal Rules provides certain institutional constraints. Judges and others more detached from interest group advocacy

---

16 GA. ST. U. L. REV. 151, 152 (1999) (describing Georgia external review law as "a political compromise between managed care, business lobbyists, and the Governor"); Louise G. Trubek, *Public Interest Lawyers and New Governance: Advocating for Health Care*, 2002 WIS. L. REV. 575, 590-92, 596 (describing negotiations over bill between consumer-physician alliance and MCOs in Wisconsin and positions on external review).

259. Scott Thornton, Comment, *The Texas Health Care Liability Act: Managed Care Organizations Can Say Goodbye to Their Extensive Immunity from Lawsuits—Or at Least That Was How It Was Supposed To Work*, 30 TEX. TECH. L. REV. 1227, 1271 (1999). As part of class action settlements, several insurers, including Aetna, CIGNA, Health Net, and Anthem/Wellpoint, have agreed to establish separate external review boards for appeals by physicians. *See* Conn. State Med. Soc'y, Aetna Settlement Summary, http://csms.org/content/showpage.asp?page=news17 (summarizing the terms of settlement reached in multi-district litigation in the Southern District of Florida); Conn. State Med. Soc'y, CIGNA Settlement Summary, http://csms.org/content/showpage.asp?page=news18 (same); Conn. State Med. Soc'y, Health Net Settlement Summary, http://csms.org/content/showpage.asp?page=news19 (same); Conn. State Med. Soc'y, Anthem/Wellpoint Settlement Summary, http://csms.org/content/showpage.asp?page=news20 (same). *See generally* Ceci Connolly, *Insurer Agrees To Pay Doctors $198 Million*, WASH. POST, July 12, 2005, at A2.

260. The dispute centered on whether other restraints on litigation should be imposed as well. STEPHANIE LEWIS, THE HENRY J. KAISER FAMILY FOUND., A GUIDE TO THE FEDERAL PATIENTS' BILL OF RIGHTS DEBATE 6-9 (2001).

261. Perhaps the most colorful articulation of the defense against such lawsuits came from Senator Phil Gramm (R-Tex.), who told the Senate, "I have seen people healed in hospitals, doctors' offices, clinics. I have even seen people healed in tent revivals. But I have never, ever seen anybody healed in a courthouse." 147 CONG. REC. S6567 (daily ed. June 21, 2001).

presumably bring with them some degree of allegiance to process integrity as a primary value. The fashioning of external review systems proceeded with little or no input from proceduralists; the legal profession qua legal profession was not involved.

IV. RAMIFICATIONS FOR PROCEDURAL THEORY

The creation of external review systems represents not only the attempt to substitute due process principles for the erosion of medical authority, but also the attempt to require accountability from powerful private sector entities. External review laws use process itself as a structure of accountability. Increasing accountability to the public from the market is the most self-evident purpose animating the creation of external review systems. One finds this theme repeatedly emphasized in the legislative and political histories of these laws; it speaks directly to the popular outcry against increased reliance on profit as the driving force behind health care delivery systems.

This Part traces the evolution of procedural due process theory. A conceptualization of "process values" arose in the context of disputes between individuals and the government, with an emphasis on the right to an evidentiary hearing prior to the denial or termination of certain benefits. In response to the ascendancy of a degree of corporate power that rivals state power, greater importance attaches to the potential for process mechanisms to foster accountability in the private sector and to promote dialogic exchange among multiple sectors engaged in decision-making. In this new economic environment, attention to accountability and deliberativeness may equal the old paradigm's focus on individual dignitarian concerns. This section concludes with an analysis of how procedural rules and devices perform regulatory functions.

A. Assumptions of the Old Paradigm

Most procedural theory scholarship has focused on cases like *Goldberg* and *Eldridge*, where the question is what process must be accorded the individual who faces adverse action by the state.[262] In these cases the primary process

262. JERRY L. MASHAW, DUE PROCESS IN THE ADMINISTRATIVE STATE (1985); Frank H. Easterbrook, *Substance and Due Process*, 1982 SUP. CT. REV. 85, 116-17; Owen M. Fiss, *Reason in All Its Splendor*, 56 BROOK. L. REV. 789, 792-94 (1990); Friendly, *supra* note 43, at 1268-70; Jerry L. Mashaw, *Administrative Due Process: The Quest for a Dignitary Theory*, 61 B.U. L. REV. 885 (1981) [hereinafter Mashaw, *Administrative Due Process*]; Mashaw, *supra* note 55; Rand E. Rosenblatt, *Health Care Reform and Administrative Law: A Structural Approach*, 88 YALE L.J. 243 (1978); Richard B. Saphire, *Specifying Due Process Values: Toward a More Responsive Approach to Procedural Protection*, 127 U. PA. L. REV. 111, 113 (1978). Scholars who rejected *Goldberg* as

concern is restraint of governmental abusiveness, as that abuse manifests itself in the procedures by which government deprives citizens of liberty or property. Grandly put, "the fundamentally important idea towards which the entire constitutional phrase [procedural due process] is reaching . . . is to safeguard the individual from government power that strikes arbitrarily and unfairly."[263]

Because public sector cases involve an opposition between citizen and state, they are the source of the greatest concern about the quality of interaction that occurs in the process by which state actors determine whether an individual has a legitimate claim or entitlement. Dignitarian values are most at risk when the state is acting against one of its citizens. It is in this context that disrespectful treatment expresses as well as imposes second-class status.[264]

Similarly, the concept of process as values-affirming, independently of substantive law, has been associated primarily with public law issues and large-scale litigation, rather than with private, bipolar disputes.[265] Also, the claims at issue when individuals litigate on theories of recovery derived from public law are most likely to fall within the category that we think of as "rights."[266] It is in

---

"wrongly decided," Richard Epstein, *No New Property*, 56 BROOK. L. REV. 747, 748 (1990), or who endorsed the *Eldridge* criteria, Richard J. Pierce, *The Due Process Counterrevolution in the 1990s*, 96 COLUM. L. REV. 1973, 1999 (1996), have not been widely followed by other scholars.

263. Cynthia R. Farina, *Conceiving Due Process*, 3 YALE J.L. & FEMINISM 189, 213 (1991).

264. Abram Chayes, *The Role of the Judge in Public Law Litigation*, 89 HARV. L. REV. 1281 (1976); MASHAW, DUE PROCESS IN THE ADMINISTRATIVE STATE, *supra* note 262, at 23-24, 45-49; Frank I. Michelman, *The Supreme Court and Litigation Access Fees: The Right To Protect One's Rights*, 1973 DUKE L.J. 1153, 1173-74. For a discussion of "dignitary theories," see MASHAW, DUE PROCESS IN THE ADMINISTRATIVE STATE, *supra* note 262, at 162, 172-82; Mashaw, *Administrative Due Process*, *supra* note 255, at 886; Saphire, *supra* note 262, at 117-25.

265. Owen M. Fiss, *The Supreme Court, 1978 Term—Foreword: The Forms of Justice*, 93 HARV. L. REV. 1, 30-31 (1979). Ironically, it was on the basis of this distinction that the father of dignitarian process values jurisprudence himself—Justice Brennan—opened the door to the huge increase in the number of cases going into arbitration. In *Moses H. Cone Memorial Hospital. v. Mercury Construction Corp.*, 460 U.S. 1 (1983), involving a contract dispute, Justice Brennan's opinion for the Court proclaimed that Congress intended by the Federal Arbitration Act "to move the parties to an arbitrable dispute out of court and into arbitration as quickly and easily as possible." *Id.* at 22. The Court later broadened the category of what it considered arbitrable disputes, with Brennan in dissent, but continued to rely on a public policy of facilitating enforcement of arbitration clauses. Rodriguez de Quijas v. Shearson/Am. Express, Inc., 490 U.S. 477, 480-86 (1989).

266. Harry T. Edwards, *Alternative Dispute Resolution: Panacea or Anathema?*, 99 HARV. L. REV. 668, 671-72 (1986); Harry T. Edwards, *Where Are We Heading with Mandatory Arbitration of Statutory Claims in Employment?*, 16 GA. ST. U. L. REV. 293, 294, 297 (1999). When government action impinges on rights recognized as carrying special weight, courts are particularly sensitive to procedural norms. *See, e.g.*, M.L.B. v. S.L.J., 519 U.S. 102 (1996) (holding that the

the interplay between a powerful government and rights-bearing individuals that a decision-making process most significantly generates and reinforces norms.

These traditional procedural due process concerns generated a body of scholarship that largely overlooked the issue of how procedure functions to assure the accountability of private entities. As we have seen, the central theoretical insights of the leading procedure scholars concerned the connection between procedure and individual liberty and dignity, not issues of systemic oversight. For this reason, the procedural due process canon as a whole lacks appreciation of the importance of an analysis that focuses on the accountability of private sector enterprises.

B. The New Primacy of Accountability and Deliberativeness

Although a new paradigm for procedural theory has not yet crystallized, one can see the emergence of two primary values slighted by the old canon: accountability and deliberativeness. Health process law in general, and external review systems in particular, offer well-defined examples of each.

1. Accountability

In health law most of the coverage decisions about whether treatment is medically necessary arise in the private sector. As advocates of patients' bills of rights saw, medical necessity disputes highlight the potential for the use of process in public law efforts to achieve accountability of powerful private entities. This context for analysis of process values resonates with a different line of procedural due process cases, which arose in consumer protection law.

The consumer due process cases typically have involved a dispute in which one of the parties was able to secure a remedy, such as a lien, which was made available by the state without providing adequate procedural safeguards for the other party. The rules of procedure then extant allowed one category of private litigants access to a powerful remedy, thereby systematically depriving another category of litigants of procedural safeguards. The successor to *Goldberg v. Kelly* in this line of cases is not *Eldridge* but *Fuentes v. Shevin*,[267] in which the Court ruled that common law replevin statutes could not authorize the seizure of chattel goods without prior notice and opportunity for a hearing.

The *Fuentes* line of cases grew out of efforts to reform predatory consumer

---

state must waive filing fee for indigent appellant when considering the termination of parental rights); Santosky v. Kramer, 455 U.S. 745 (1982) (holding that in parental termination proceeding, state must meet the clear and convincing evidence standard).

267. 407 U.S. 67 (1972). The most recent significant decision in this line of cases is *Connecticut v. Doehr*, 501 U.S. 1 (1991).

credit practices.[268] By analyzing the process aspects of the state's enforcement of judgments against borrowers, the Court addressed the question of the optimal balance between allowing free space for market forces and requiring accountability of powerful private parties. The procedures of adjudication functioned as a structure of accountability.[269]

Grijalva provides an example of health process law operating at the intersection of the progeny of both *Goldberg-Eldridge* and *Fuentes*. Under the reasoning of the lower federal courts, the defendant Medicare MCOs were private sector entities which functioned as state actors. After finding that state action existed, the courts' analyses focused on an accountability theme deriving from the interaction between patient and MCO. The *Grijalva* decisions are silent on *Goldberg*-like intimations of dignity and participation. They are also not based on *Eldridge* conclusions that the heightened procedural protections were needed to avert a high risk of erroneous decisions. Plaintiffs won in the lower courts because multiple judges concluded that the private MCO system lacked sufficient accountability. Although this rationale was expressed in individual rights terms, it was fundamentally a systemic argument, not an individual one.

The systemic concern is more oriented toward a focus on the product of procedure, the outcome, than on procedure itself qua procedure. In both *Fuentes* and *Grijalva*, plaintiffs challenged procedural rules that exacerbated the imbalance of power between the parties. In both, the existing procedures allowed the stronger party to use law as a tool to diminish the contractual rights held by the weaker party. The procedural due process claim operated to impose accountability and restraint on the stronger party's actions, not solely because of dignitarian harms, but because of what was at stake. It would be naïve to believe that the courts found the stakes to be high purely from the perspective of an individual plaintiff; at issue was the operation of a system.

---

268. A field research project on predatory practices during the period prior to *Fuentes* observed that "[p]erhaps the most objectionable practice in consumer credit is that of requiring the consumer who buys on credit to sign an authorization of entry of confession of judgment at the time he makes his purchase." Note, *Translating Sympathy for Deceived Consumers into Effective Programs for Protection*, 114 U. PA. L. REV. 395, 418 (1966); *see also* George Brunn, *Wage Garnishment in California: A Study and Recommendations*, 53 CAL. L. REV. 1214, 1248-49 (1965) (recommending an end to garnishments arising from installment credit contracts and to all pre-judgment garnishments).

269. In subsequent decisions involving garnishment and sequestration, the Court backed away from *Fuentes*, not unlike the way the Court retreated from *Goldberg*, by allowing procedural protections other than an evidentiary hearing to suffice prior to seizure, so long as a prompt post-seizure hearing was afforded. *See, e.g.*, Mitchell v. W.T. Grant Co., 416 U.S. 600 (1974). However, *Fuentes* remains good law. N. Ga. Finishing, Inc. v. Di-Chem, Inc., 419 U.S. 601, 608 (1975) (Stewart, J., concurring).

When the defendant is a government agency, courts can assume that other public systems such as administrative law or legislative oversight will provide accountability. The background assumption of the presence of these other systems makes the accountability function less visible, and leaves courts free to focus on individual dignity and norms of participatory democracy. By contrast, the private sector procedural due process cases highlight the accountability function. And, in today's world of massive concentration of power in private entities, coupled with the privatization trend in which public functions are increasingly being performed in the private sector, the accountability function of procedure will likely become more significant.

Although accountability emerges as the dominant process value in the private party health process cases, dignitarian concerns are not absent. However, dignity norms in private cases differ from those in citizen-government disputes. The right to a hearing is not the only, nor necessarily the most significant, aspect of dignitarian rights. The related right to be told why—to be accorded the respect of a full explanation for the decision—is at least equally compelling. I would call that the deliberative, rather than the dignitarian, aspect of public value process norms. In *Grijalva* both the district and appeals court treated the right to a full explanation of the reasons for a denial as essential.[270]

2. Deliberativeness

Deliberativeness is the second process value which has grown in importance with the decline of the old paradigm. In today's world deliberativeness in decision-making is more important, more of the time, than the right to an evidentiary hearing, especially in a technologically sophisticated field such as health. A well-reasoned explication of the decision in even a relatively simple medical case will necessitate careful review of scientific evidence. In medical necessity reviews, a mandate for explanation is likely to produce a decision of higher quality than would be produced by adversarialism per se.

The value of deliberativeness also synchronizes well with the increasing variety in types of procedural mechanisms, because of its intrinsic flexibility and adaptability. Academic debates about *Eldridge* treated the values underlying procedural due process as an unvarying package, applicable to any context. Most scholars invoked the right to an evidentiary hearing prior to a deprivation as the

270. Grijalva v. Shalala, 152 F.3d 1115, 1123 (9th Cir. 1998); Grijalva v. Shalala, 946 F. Supp. 747, 757-59 (D. Ariz. 1996). Similarly, reflecting on a large number of similar cases, Judge Weinstein described the pre-*Grijalva* Medicare system as providing "'gobbledegook' notices" without any real or understandable reason given for the denial. Weinstein, *Adjudicative Justice*, *supra* note 17, at 401.

benchmark for adequacy, and paid little separate attention to deliberativeness.[271] I would argue, however, that in adjudicating complex issues, participation is a secondary value.

Mandating deliberativeness incentivizes decision-makers to obtain full input from the parties, so that they can prepare a fully reasoned and explicated decision. The rationale behind the requirement for a full explanation in every case also creates the strongest foundation for maximum application of a right to a hearing. Using an incentives approach, rather than the mandate for an evidentiary hearing adopted by the Court in *Goldberg*, would tend to produce hearings in situations where they would actually add the greatest value, and would avoid some of the systemic congestion caused by the *Goldberg* approach.

This is essentially the argument behind Melvin Eisenberg's deconstruction of the participation norm in adjudication. Building on Lon Fuller's analysis of adjudication as a form of social ordering,[272] Eisenberg argues that Fuller's "Participation Thesis"—the right of a party in adjudication to participate in the decisional process—breaks down into three norms: (1) "that the adjudicator should attend to what the parties have to say;" (2) that the adjudicator should explain her decision in a manner that responds to the parties; and (3) that the decision "should proceed from and be congruent with" the parties' evidence and arguments.[273] Eisenberg also argues, however, that some matters are appropriate for the weak responsiveness of a consultative process, in which the decision-maker is free to draw on sources other than the parties' submissions.[274] A consultative process incentivizes the decision-maker to treat the parties' input seriously, but also leaves her free to use a considerable amount of managerial discretion.

Eisenberg's approach would not satisfy the strongest dignitarians. His willingness to accept a consultative process in *Eldridge*, which he uses as an

---

271. The only major exception is Morgan, *supra* note 44. Frank Michelman notes that "[t]he individual may have various reasons for wanting to be told why, even if he makes no claim to legal protection, and even if no further participation is allowed him." Frank I. Michelman, *Formal and Associational Aims in Procedural Due Process*, *in* NOMOS XVIII: DUE PROCESS 126, 127 (J. Rowland Pennock & John W. Chapman, eds. 1977). The remainder of his essay, however, focused on the right to a hearing. Deliberativeness was also discussed briefly in Mashaw, *Administrative Due Process*, *supra* note 262, at 913-915; Robert S. Summers, *Evaluating and Improving Legal Processes: A Plea for "Process Values,"* 60 CORNELL L. REV. 1, 26 (1974).

272. Lon L. Fuller, *The Forms and Limits of Adjudication*, 92 HARV. L. REV. 353, 365-72 (1978).

273. Melvin Aron Eisenberg, *Participation, Responsiveness, and the Consultative Process: An Essay for Lon Fuller*, 92 HARV. L. REV. 410, 411-12 (1978).

274. *Id.* at 414-17.

example of its appropriate use,[275] makes clear that the consultation concern would not necessarily sustain the right to a hearing. Edmund Pincoffs explained the logic behind prizing a right to hearing over the right to a responsive decisional rationale in the hierarchization of values as "whatever reasoning justifies participation (the opportunity to contest the official's reasoning) will justify revelation to the person affected of the official's reasons."[276] The reverse is not necessarily true.

I would argue, however, that deliberativeness is more important than the right to participate in a hearing, even though that contention skewers a sacred cow of progressive procedural theory. The importance of deliberativeness derives from the fact that its impact matters not just at the level of the individual case, but also at a system level.

Philosopher Jurgen Habermas argued that democracy worked best in an "ideal communication situation," a context in which social consensus as to norms guiding behavior grew out of reasoned exchange, without constraints on either the content of or access to the exchange.[277] The insight that dialogic exchange rather than individual rights could be the core requirement for fair process applies to external review systems. Habermas's approach reinforces the importance of substantive engagement as a basis for the legitimacy of legal rules.

The value behind the distinct function of deliberativeness derives from the value placed on cultural interconnectiveness, reflected in an insistence on reasoned application of law's coercive power. As with accountability, the concern behind this process value is less with valuation of individuals than with restraint of institutional arrogance and unconstrained economic, as well as governmental, power.

C. Procedure as Regulation

We have analyzed accountability as a process value oriented toward the intrinsic fairness of a decision-making procedure, as a kind of policing mechanism to guard against the abuse of procedure itself. Now we examine procedure as a structure of distributive accountability, with a focus more on

---

275. *Id.* at 421.

276. Edmund L. Pincoffs, *Due Process, Fraternity, and a Kantian Injunction*, *in* NOMOS XVIII: DUE PROCESS, *supra* note 271 at 172, 173.

277. JURGEN HABERMAS, BETWEEN FACTS AND NORMS: CONTRIBUTIONS TO A DISCOURSE THEORY OF LAW AND DEMOCRACY 165-68, 321-28, 360-63 (William Rehg trans., 1996). Lawrence Solum has linked Habermas's theory to issues of procedural justice, noting that the litigation system has many devices (e.g., discovery) which seek to create the conditions that would allow for at least an approximation of the ideal communication situation. Lawrence B. Solum, *Procedural Justice*, 78 S. CAL. L. REV. 181, 268-271 (2004).

outcomes of procedure rather than on procedural devices themselves. Assuming a regulatory purpose,[278] the issue here is how well external review systems operate to bring public norms and values to bear on the allocational choices of what is otherwise a private rationing system.

One can envision the matrix of process values along two dimensions. One dimension is defined by one's concept of the function of adjudication, i.e., whether one prioritizes the role of adjudicatory systems as dispute resolution mechanisms or as expressions of public values.[279] The second is determined by whether one evaluates such a system by the quality of its outcome or by the quality of its procedures qua procedures. The following chart illustrates the possible array of values:

EVALUATION BASED ON:	*FUNCTION OF ADJUDICATION SYSTEM*	
	Dispute Resolution	*Serving Social Goals*
Quality of Outcome	Accuracy	Regulatory Free Market
Quality of Process	Responsiveness Ethicality (how parties are treated) Efficiency	Participation Dignity Cost-Effectiveness

On this understanding, procedure and regulation are not two utterly distinct categories as much as they are reflections of different emphases on function and evaluative norms. Consigning individual coverage decisions to external review panels operated by private companies creates a perfect merger between adjudication and the fundamental paradox of American health care: that both are public goods treated as private commodities, delivered primarily within private

278. Reviewing the literature of pro- versus anti-regulatory arguments is beyond the scope of this Article. The most extensive discussions of decision-making procedures as a mechanism of regulating managed care are in NORMAN DANIELS & JAMES E. SABIN, SETTING LIMITS FAIRLY: CAN WE LEARN TO SHARE MEDICAL RESOURCES? 25-66 (2002); and Norman Daniels & James Sabin, *Limits to Health Care: Fair Procedures, Democratic Deliberation, and the Legitimacy Problem for Insurers*, 26 PHIL. & PUB. AFF. 303 (1997) [hereinafter Daniels & Sabin, *Limits to Health Care*]. *See also* Sharona Hoffman, *Unmanaged Care: Towards Moral Fairness in Health Care Coverage*, 78 IND. L. REV. 659, 693, 711-12 (2003). For a summary of the contentions that regulatory mechanisms for managed care, including external review, will not work, see David A. Hyman, *Regulating Managed Care: What's Wrong with a Patient Bill of Rights*, 73 S. CAL. L. REV. 221 (2000).

279. David Luban, *Settlements and the Erosion of the Public Realm*, 83 GEO L.J. 2619, 2622-2640 (1995).

systems.[280] The accretion of coverage decisions operates as a micro-rationing device. Thus, promoting accountability for outcomes is fundamentally a regulatory function.[281] Because the decisions of external review panels essentially constitute direct allocations of resources, they provide especially clear examples of the blend of adjudication and regulation.

External review systems are related to, but not the same as, privatization, the contracting out of what have typically been government functions or services to private sector profit or non-profit entities. In the realm of procedure, privatization is most visible in the rapid increase in the use of arbitration and other private dispute resolution mechanisms outside of the public courts,[282] a development which external review laws obviously resemble. In public services privatization generally has led to "new blends of public and private power at all levels of government," and, more fundamentally, a "redefinition of what is public and what is private."[283]

Yet external review systems could be labeled equally well either as private sector administrative tribunals or as public law arbitration panels.[284] They are hybrid mixtures of procedural due process norms and precepts being operationalized and administered by privately owned and operated adjudicators. External review laws have a "publicizing," as well as a privatizing, impact because they extend public law values into the marketplace, farther than due process doctrine itself could reach.[285] Indeed, in several states the procedures

---

280. *See id.* (discussing adjudication as a public good).

281. Martha Minow, *Public and Private Partnerships: Accounting for the New Religion*, 116 HARV. L. REV. 1229, 1259 (2003); *see also* Jody Freeman, *Private Parties, Public Functions, and the New Administrative Law*, *in* RECRAFTING THE RULE OF LAW: THE LIMITS OF LEGAL ORDER 331, 336 (David Dyzenhaus ed., 1999).

282. Judith Resnik, *Migrating, Morphing,* and *Vanishing: The Empirical and Normative Puzzles of Declining Trial Rates in Courts*, 1 J. EMPIRICAL LEGAL STUD. 783, 819-22 (2004); Judith Resnik, *For Owen M. Fiss: Some Reflections on the Triumph and the Death of Adjudication*, 58 U. MIAMI L. REV. 173, 188-89 (2003); Stephen J. Ware, *Default Rules from Mandatory Rules: Privatizing Law through Arbitration*, 83 MINN. L. REV. 703 (1999); Jack B. Weinstein, *Some Benefits and Risks of Privatization of Justice Through ADR*, 11 OHIO ST. J. ON DISP. RESOL. 241 (1996).

283. Alfred C. Aman, Jr., *Globalization, Democracy, and the Need for a New Administrative Law*, 49 UCLA L. REV. 1687, 1688-89 (2002). Aman argues that the scope of this change calls for a new conceptualization of administrative law, which he calls the "neo-corporatism theory of administrative law." *Id.* at 1703-04.

284. Note that an administrative law judge wrote one of the first comprehensive surveys of these laws. *See* Craig, *supra* note 174.

285. Jody Freeman, *Extending Public Law Norms Through Privatization*, 116 HARV. L. REV. 1285 (2003). Freeman argues that external review laws are examples of a successful use of publicization. *Id.* at 1330.

accompanying external review are incorporated into the body of state administrative law.[286]

However, although external review entities are more heavily "publicized" than most other private companies, external review laws substantially privatize and deregulate the process for adjudicating complaints that needed care is being denied, if one uses the litigation system or the administrative law model as the benchmark. The trade-off for the ease of streamlined procedures is that external review systems have many fewer protections for the weaker party than public law rules of procedure, as we have seen. On the other hand, they are subject to more regulation than private arbitration generated by contracts, which has become the primary alternative to litigation.

Moreover, in another sense, the degree to which collective norms are being extended by external review laws past their prior reach is more limited than for other such hybrid entities. Before such laws were enacted, patients relied on deference to medical judgment and the fiduciary principle to protect their interests (however naïve the high level of trust may have been). The pre-ERO world was also a pre-MCO world.

It is difficult to assess the extent to which external review systems serve this regulatory or collective accountability function. The device that most effectively renders litigation a regulatory institution is class actions. Yet at least some state regulatory officials describe external review as working so well that it amounts to precisely that: "an ongoing class action."[287] Of course, the fact that regulators are "overwhelmingly satisfied" with external review[288] does not necessarily tell us much about its quality as regulation.

The regulated industry reports that it responds to external review decisions in a systemic fashion. Although the data are limited to California, all but one of the plans studied there in 2003 reported that external review decisions influenced their coverage policies.[289] The impact of industry responsiveness would be heightened by mechanisms for input by patient or consumer organizations.[290] If meaningful oversight existed on a national scale, external review really might amount to an ongoing class action.

Norman Daniels and James Sabin bring the most optimistic approach to external review. Although not analyzing external review systems specifically, they see the best solution to the problems of distributive justice in managed care

---

286. *See supra* text accompanying notes 203, 230.

287. Berman-Sandler, *supra* note 9, at 270.

288. *Id.* at 275.

289. IMR 2003, *supra* note 249, at 4, 28-29.

290. Ongoing consumer advocacy has been particularly strong in Wisconsin. *See* Trubek, *supra* note 258.

decision-making as the adoption of regulation which is "*process focused.*"[291] Daniels and Sabin argue for a combination of the components of accountability discussed in this Article—stated rationales, publicly-accessible decisions, mechanisms for appeal or reconsideration, and some form of regulation—together with a master norm to guide outcomes: that the rationale must be based on "a reasonable construal of how to meet the medical needs of a covered population under acceptable resource constraints."[292] By emphasizing process over any substantive choices to guide allocational decisions, they hope to avoid the "tragic choices" problem, i.e., the resistance to accepting the need to make such decisions that leads to their being made in disguised ways, beyond public control.[293] For Daniels and Sabin external review or similar systems can extend deliberative democratic principles to the private institutions in which, for better or worse, much health policy is made.[294]

Using Daniels and Sabin as a benchmark illustrates how far short of that accountability paradigm external review systems currently operate. Again, the fundamental deficiency is the lack of access to the decisions in almost all the states. Without that public resource there cannot be the kind of "institutional reflective equilibrium"[295] which they seek, nor any basis for self-conscious acknowledgment of resource constraints, outside of the terms of a contract. Yet since essentially all contracts operate under the same term of "medically necessary," Daniels and Sabin are correct to insist that the contract subject to interpretation in all these cases is, at least to some extent, a social one.

A different kind of accountability relates to quality of care. Although physicians were among the stakeholders who designed external review systems, their primary goals were defensive: to avoid being left holding the bag for both liability[296] and cost[297] when managed care organizations denied authorization for

---

291. Daniels & Sabin, *Limits to Health Care*, *supra* note 278, at 348.

292. *Id.* at 307.

293. *Id.* at 318-21; *see also* David Orentlicher, *The Rise and Fall of Managed Care: A Predictable "Tragic Choices" Phenomenon*, 47 ST. LOUIS U. L.J. 412 (2003) (arguing that managed care failed to sustain its strongest cost containment mechanisms because they could not survive public scrutiny as a rationing device, as predicted in GUIDO CALABRESI & PHILIP BOBBITT, TRAGIC CHOICES (1978)).

294. Daniels & Sabin, *Limits to Health Care*, *supra* note 278, at 323-24.

295. *Id.* at 328.

296. A payor's erroneous refusal to authorize reimbursement for medically necessary care does not absolve the treating physician of professional responsibilities to provide such care. *See, e.g.*, Wickline v. State, 239 Cal. Rptr. 810, 819 (Ct. App. 1986). The availability of external review makes that outcome less likely.

297. "[O]ne can view external review requirements as a means to promote physician payment as much as patient care." William M. Sage, *Physicians as Advocates*, 35 HOUS. L. REV. 1529, 1541

reimbursement of expenses for treatment. Very little attention has been paid, either at the time of enactment or since, to the impact of external review on quality of medical practice, another question which merits further study.

To some extent accountability may be in tension with therapeutic goals. Based on how external review systems are structured, Bill Sage is skeptical that they will improve the quality of care or the degree of trust between doctor and patient.[298] Sage recommends that review systems be designed based on therapeutic, rather than adversarial, principles, recognizing that while "legal fairness demands due process, often through adversarial advocacy . . . medical fairness . . . demands beneficence and respect for persons."[299] He favors a mediation approach, especially for patients who have chronic conditions and are thus likely to be repeat users of a review system.[300]

Finally, in some situations involving medical necessity decisions, the value of compassion may justifiably trump all others, including accountability, quality of care, and accuracy. Patients with terminal conditions may seek "last-chance therapies," treatments with little if any proven efficacy which insurance companies can properly deny as experimental. In those instances Kathy Cerminara argues that the protections of external review are largely beside the point.[301] She advocates the development of personalized therapeutic counseling approaches, using a multi-disciplinary team of professionals.[302]

A bioethics or therapeutic model is perfectly consistent with an emphasis on deliberativeness: Good faith dialogic exchange is central to both, even if accountability plays a lesser role. Yet while it may be impractical or even callous to expect patients in greatly weakened physical condition to navigate appellate or arbitration-style systems, it would be equally naïve to assume that payors will adopt more resource-intensive, and thus more expensive, ways of saying no without some degree of formal or informal pressure. Ultimately, then, the prerequisite for a therapeutic approach is a foundation of accountability.

(1999).

298. Sage, *supra* note 3, at 632.

299. *Id.* at 635.

300. *Id.* at 645-46; *see also* Kathy L. Cerminara, *Contextualizing ADR in Managed Care: A Proposal Aimed at Easing Tensions and Resolving Conflict*, 33 LOY. U. CHI. L.J. 547 (2002) (evaluating non-litigation options as a way to resolve health care disputes without increasing patient stress). There have been promising results with mediation in the context of malpractice litigation. Edward A. Dauer & Leonard J. Marcus, *Adapting Mediation To Link Resolution of Medical Malpractice Disputes with Health Care Quality Improvement*, 60 LAW & CONTEMP. PROBS. 185 (1997).

301. Kathy L. Cerminara, *Dealing with Dying: How Insurers Can Help Patients Seeking Last-Chance Therapies (Even When the Answer Is "No")*, 15 HEALTH MATRIX 285 (2005).

302. *Id.*

V. DO THE MCOS COME OUT AHEAD?

A. Applying Process Values to External Review Systems

Thirty years ago Marc Galanter developed a structural framework for analysis of adjudication that divided parties into repeat players and one-shotters, examined how lawyers affected the power dynamics of litigation, and described a number of ways in which the passivity of legal institutions enhanced the power of repeat players.[303] Galanter's now classic article described a system in which certain classes of parties enjoyed structural advantages. His conclusion about why the "haves" systematically come out ahead emphasized the intrinsic weakness of law as a mechanism for seeking greater social justice, but it can also be read as an indictment of how the litigation system distorts all of the process values that we have discussed, from the utilitarian/efficiency cluster, to the old and new dignitarian norms, to accountability.

External review laws were enacted as part of a backlash against a system in which MCOs were perceived to always come out ahead. The external review systems were designed to mitigate an imbalance between patients and payors (as well as between physicians and payors), and to rectify the operations of a payment process in which decisions were effectively unreviewable because the practical impact of the process was to deny treatment for which reimbursement had not been authorized. When one applies the range of process values to the external review statutes, how well do they measure up?

In this Section I evaluate external review systems by applying utilitarian, dignitarian, and accountability critiques. Although external review falls short in many respects, it retains significant appeal, and may provide a model for forms of independent private review in contexts other than denial of medical claims. At the end of the section, we return to the prospect of health process exceptionalism, but in a new form: to ask not whether the degree of deference should be reinstated, but whether the level of process protections should be elevated.

1. The Utilitarian Critique

The greatest appeal of the external review concept is its claim to promote greater efficiency than exists in litigation. The entire procedure is streamlined. In some states appeals can be filed on a web site.[304] Some states operate consumer

303. Marc Galanter, *Why the "Haves" Come Out Ahead: Speculations on the Limits of Legal Change*, 9 L. & SOC'Y REV. 95 (1974).

304. Arkansas, Illinois, Indiana, Iowa, Kansas, Kentucky, Missouri, North Carolina, Ohio, Oklahoma, Oregon, Pennsylvania, Texas, and Utah.

assistance offices,[305] and lawyers are not required.[306] The sources of evidence are often specified in statute or regulation. There are deadlines for decisions and shorter deadlines for decisions in cases involving exigent need for care. Medical professionals make the final decisions, as in specialized courts or well-run arbitration systems. The relatively low utilization rates keep down operational costs.

It is difficult to evaluate external review systems on the acid test of how great the risk is of erroneous decisions or, more precisely, how likely external review organizations are to correctly identify wrong decisions by MCOs. Reversal rates alone shed no light on this.[307] The system lacks sufficient mechanisms either to correct for possible bias or to enhance quality by review of the most difficult, and indeterminate, medical judgments.

In most states there is very little oversight of the external review decisions themselves. In Michigan, where external review decisions must be approved by a state official before they become final, the state agency accepts ninety-eight percent of the ERO decisions.[308] In California a clinical advisory panel meets quarterly to assess quality of the decision-making.[309] Most states rely on private accreditation mechanisms to assure quality;[310] in addition, a national association of independent reviewers has developed.[311]

Several commentators have worried that external review laws, because of the political climate that produced them, are susceptible to bias in favor of consumers. They argue that the goal should be to enforce the actual contract language, as medical necessity is interpreted by the health plan.[312] Especially problematic from this perspective is the extent to which some states have legislated a definition of medical necessity and/or provided for de novo review of

---

305. POLLITZ ET AL., *supra* note 95, at 11-12.

306. However, lawyers tend to get involved when significant amounts are at stake. Berman-Sandler, *supra* note 9, at 261-62. Additionally, private companies have begun to market such services. Peter Landers & Amy Dockser Marcus, *You Can Make Them Pay: New Ways To Appeal Make It Easier To Take on Health Insurers and Win*, WALL ST. J., Sept. 17, 2002, at D1; Barbara Martinez, *Aetna Insiders Now Advocate for Patients*, WALL ST. J., Jan. 28, 2003, at B1.

307. *See supra* text accompanying note 254.

308. Berman-Sandler, *supra* note 9, at 259.

309. CAL. HEALTH & SAFETY CODE § 1347.1(03) (West 2000).

310. Berman-Sandler, *supra* note 9, at 239.

311. *See* National Association of Independent Review Organizations, http://www.nairo.org (last visited Dec. 4, 2005).

312. Hyman, *supra* note 278, at 250; Aaron Seth Kesselheim, Comment, *What's the Appeal? Trying To Control Managed Care Medical Necessity Decision-Making Through a System of External Appeals*, 149 U. PA. L. REV. 873, 915-16, 918-19 (2001); Sloan & Hall, *supra* note 4, at 193.

such determinations.[313] Some support for their concern is evident in the request by health plans in California for an agency that will review the independent reviewers.[314]

Market asymmetries, especially in access to information, suggest that the question is more complicated than the possibility of bias in one direction or another. Should consumers/patients be held to the terms of a bargain, or as Wendy Mariner argued, should the "health plan's obligations . . . be those that reasonable managed care organizations, and reasonable patients, with equal bargaining power and good information, would expect as fair and reasonable for the stated price"?[315] If the latter is accurate—and it probably is more consistent with what enrollees in a group plan, who had no role in negotiating the contract, believe they are getting—then standards for reasonableness imposed through publicly accountable institutions such as the legislature should be acceptable as methods of perfecting the market.

In either case, it may nonetheless be true that one could due process to death a useful reform by adding more of the indicia of a litigation system—such as more extensive discovery or hearings—to the external review systems. Such reforms could lead to nothing more than procedural formalism without functional improvement. As utilitarians would point out, humanistic process values can be lost in systemic sclerosis.

2. The Dignity and Deliberativeness Critique

As effectuators of traditional dignity values, external review systems have serious deficiencies. Few states provide opportunities for patients to present their case, other than by the submission of appeals forms or paper records.[316] Dignitarians would argue that even though hearings would delay decisions, patients could decide whether to make that trade-off or not, depending on their most significant concerns.

The traditional dignitarian critique is less applicable to external review than to other contexts, however. External review boards are usually assessing the competing recommendations of two (or more) physicians. Thus the voice for the weaker party that would be most significant in the great majority of cases would be that of the physician recommending the treatment that the health plan had denied. This gives a different valence to norms of individual dignity and the

313. *See supra* text accompanying notes 174-182.

314. IMR 2003, *supra* note 249, at 29.

315. Wendy K. Mariner, *Standards of Care and Standard Form Contracts: Distinguishing Patient Rights and Consumer Rights in Managed Care*, 15 J. CONTEMP. HEALTH L. & POL'Y 1, 27, 43 (1998).

316. POLLITZ ET AL., *supra* note 95, at 20.

collective ideal of participation. Some form of consultation between the external reviewers and the physician(s) who recommended a particular treatment could well enhance accuracy, however, even if there was little or no gain to dignity.

If one separately examines the deliberativeness critique, external review systems receive at best weak reviews. Presumably all EROs state reasons for their decisions, but it is difficult to know with how much clarity or whether the reasons are communicated in terms that the average layperson would understand. Data from California are not encouraging. The 2003 audit of the system found that patients, their physicians, and the health plans whose denials had been overturned all wanted more information and explanation of the rationales of the external reviewers.[317] The fact that such problems plague the system in California, which has more transparency than in any other state, does not augur well for the quality of deliberativeness nationwide.

3. The Accountability Critique

As mechanisms of accountability designed to prevent abusive use of procedures by more powerful parties, external review systems are also problematic. If the utilitarian critique evinces concern with bias in favor of consumers, an accountability model focuses attention on the risk of bias against consumers. External review laws create advantages for repeat-player MCOs in multiple ways. As a threshold matter, there are wide disparities among the states on three important indicia of accountability: (1) independence of the external reviewer;[318] (2) practical access to the system;[319] and (3) public availability of the decisions being rendered.[320] All of these factors operate to exclude or disable the one-shot patient party, even if she is represented by counsel.

One additional problematic characteristic of all external review systems is their individuation and disjoinder effects. All cases are individual cases, and can only be litigated as such. There is no possibility that a series of similar denials of care, for example, can be joined either in the remedy of seeking injunctive relief or in the form of constituting a class action. Because of the way that decisions about what is medically necessary function as a micro-rationing system, this individuation and disjoinder is an especially significant process cost, more so than in more routine consumer cases such as *Fuentes*.

---

317. IMR 2003, *supra* note 249, at 3-5. Of the unsuccessful appellants, none said they understood the reasoning of the IRO. *Id.* at 32.

318. *See supra* text accompanying notes 218-223.

319. *See supra* text accompanying notes 247-252.

320. *See supra* text accompanying notes 228-231. Legitimate concerns about the confidentiality of medical information can be accommodated by the redaction of personal identifiers, as occurs in the three states that make decisions public.

Possible remedies obtainable from external review are quite limited, but that trade-off is central to the function of the system. If successful complainants could recover monetary damages, for example, the speed and informality would have to be sacrificed. Moreover, patients can usually elect to sue without going through external review or after they lose an appeal.[321] But external review systems would be better equipped to prevent abusive tactics by the more powerful party if clear penalties existed for conduct such as excessive delay or refusal to produce documents. In some states the law authorizes what is in essence a judgment by default in such situations.[322] Most state laws, however, lack these self-policing mechanisms.

To the extent that one can judge fairness by the perception of fairness, the few data which exist are not reassuring. In a representative sample study, only about half of patients who went through external review in California expressed confidence that their individual medical status had been considered, that the review had been fair and impartial, or that the reviewers had thoroughly considered the available scientific information.[323] Their assessment was almost totally contingent on outcome; those who had prevailed found the process fair, those who had lost found it unfair.[324] Since the hallmark of legitimacy for a process system is that those who do not prevail nonetheless accept the procedure as fair,[325] there appears to be a serious problem in at least the public's belief that external review achieves the basic accountability function that produced these laws in the first place: protecting individual patients against MCO abuses.[326]

Too often, external review systems represent missed opportunities for accountability. Even if they provide a reasonable degree of protection against improper denials of care or reimbursement, they do not even seriously attempt to enhance overall quality of care by sharing information with physicians conscientiously attempting to solve difficult medical problems. Nor do they offer a venue in which distributional concerns might be opened up for greater public deliberativeness. In these respects, they are strongly aligned with the goals of individualized dispute resolution, rather than with any broader functions.

4. External Review as Creation of a New Model

A final consideration in applying the metric of process values to external

321. *See supra* text accompanying note 232.

322. *See, e.g.*, W. VA. CODE ANN. § 33-25C-6(j) (LexisNexis 2003).

323. IMR 2003, *supra* note 249, at 18.

324. *Id.* at 32.

325. TYLER, *supra* note 93, at 79-80, 166-67.

326. To some extent, the dissatisfaction may stem from the lack of a hearing or other opportunity for patient participation. *See supra* text accompanying notes 92-93.

review systems is the question of whether they will themselves become a new marker for minimally acceptable procedural due process, not only in the private marketplace but also, in a kind of doctrinal blowback, in public law as well.[327] Litigating in the shadow of arbitration came to influence many aspects of court procedure,[328] and reactions to due process shortcomings in private arbitration are now forcing it to include more procedural protections.[329] Private external review companies will seek larger markets, and their efficiency and low operating costs could make them a model for resolving non-managed care disputes, setting off a competitive dynamic similar to that between litigation and arbitration. Establishment of an independent review process specifically for physicians has become one of the terms of settlement in class action suits against large insurance companies, in which doctors alleged that the insurers repeatedly underpaid them.[330]

In the end, with so little empirical data on how external reviews are actually operating, we do not yet know whether they will provide appropriate and sufficient relief for patients or whether the MCOs will always come out ahead. Alternatives to litigation certainly are not insulated from the same problems that Galanter identified in the court system.[331] The promise of external review systems is that they will be streamlined but careful, efficient but fair, providing accountability without being cumbersome. If they live up to that standard, they may emerge as the FedEx of adjudication systems, leading the older public sector institutions (i.e., courts) to look to them for high quality customer service in areas

---

327. *See* Robert Pear, *Bush Pushes Plan To Curb Medicare Appeals*, N.Y. TIMES, Mar. 16, 2003, at A1 (discussing proposal to substitute arbitrators for administrative law judges to hear second-level appeals of denials of claims for services or payment and to add requirement that deference be given to agency policies, as well as to statutes and regulations).

328. For discussions of the impact on litigation of arbitration and other alternatives, see Kenneth S. Abraham & J.W. Montgomery III, *The Lawlessness of Arbitration* (Univ. of Va. School of Law, Public Law Research Paper No. 02-09, 2002), *available at* http://ssrn.com/abstract=353340; and Bryant Garth, *From Civil Litigation to Private Justice: Legal Practice at War with the Profession and Its Values*, 59 BROOK. L. REV. 931, 948 (1993).

329. *See, e.g.*, Hooters of Am., Inc. v. Phillips, 39 F. Supp. 2d 582, 618-20 (D.S.C. 1998), *aff'd and remanded*, 173 F.3d 933, 938-40 (4th Cir. 1999); Armendariz v. Found. Health Psychcare Servs., Inc., 6 P.3d 669, 682-89 (Cal. 2000). The American Bar Association Section on Dispute Resolution has created a model due process protocol for mediation and arbitration in health care disputes. Margaret M. Harding, *The Limits of the Due Process Protocols*, 19 OHIO ST. J. ON DISP. RESOL. 369, 405-08 (2004).

330. Ceci Connolly, *Insurer Agrees To Pay Doctors $198 Million*, WASH. POST, July 12, 2005, at A2.

331. *See* Carrie Menkel-Meadow, *Do the "Haves" Come out Ahead in Alternative Judicial Systems?: Repeat Players in ADR*, 15 OHIO ST. J. ON DISP. RESOL. 19, 32-53 (1999).

involving scientific or technical expertise. Or they may never amount to more than "low-end justice for the rank and file."[332] Only time and more data will tell us the answer.

B. Health Process Exceptionalism Redux

We began with the story of health process exceptionalism: how deference dominated due process. External review laws bring us full circle to the questions behind exceptionalism. Should the exigencies of managed care practice lead us to conclude that health process should be different after all, by becoming more medicalized? Bill Sage's argument for models of therapeutic mediation would lead to structures even less tied to procedural norms than before. Or, with medical deference now largely abandoned, should there be *greater* due process protections than the current norm, in order to enhance the accountability function of external review?

One marker of the friction between the exceptionalist or medical model and the legal model is the divergence between the professions in how they conceptualize the role of prior case-specific decisions. The professional culture of medicine emphasizes the importance of change geared to improving the quality of care as much as the goal of adherence to norms based on current knowledge. Dan Fox describes medical culture as granting authority to "physicians, armed with the latest knowledge, and not [bound by] precedent."[333] A core value of the legal system, by contrast, is "[t]o stand by things decided, and not to disturb settled points."[334] Not only does precedent as such lack value in the methodology for production of medical knowledge, but payors may be legitimately apprehensive that decisions in highly unusual cases could create mandates.

For these reasons precedent could not operate in external review systems in the way that it does in other adjudication systems. Adopting a concept of binding precedent would disadvantage patients and society as well as flout the norms of medicine. But the misfit between medical and legal concepts of precedent does not justify blocking the gains that would ensue from the greater transparency built into making decisions accessible to the public. Knowledge of outcomes in

---

332. Bryant G. Garth, *Tilting the Justice System: From ADR as Idealistic Movement to a Segmented Market in Dispute Resolution*, 18 GA. ST. U. L. REV. 927, 932 (2002).

333. Fox, *supra* note 88, at 213. Bill Sage concurs, arguing that gaps in medical knowledge cannot be filled by "the mortar of . . . interpretation," as is done with case law that builds upon case law, but must await moments of new scientific discovery to establish the validity of evolving clinical practices. Sage, *supra* note 297, at 1604.

334. BLACK'S LAW DICTIONARY 1414 (7th ed. 1999) (definition of *stare decisis et non quieta movere*).

prior similar cases provides a baseline, offering guidance to patients, protection against the potential flaws of medical decision-making,[335] and fulfillment of deliberativeness as a dignitarian process value.

The availability of the body of decisions is fundamental to medical transparency, even if precedent as lawyers understand it may prove counterproductive. The approach to evaluating external review as a regulatory mechanism should focus on aggregate accountability that includes a range of measures, from procedural audits to data-driven analysis. Oversight, too, can be innovative.[336] Thoroughly public or completely transparent systems are not necessary for legitimacy and fairness in these decisions, but accountability *to* the public is.

CONCLUSION

In a rare example of legislative convergence, forty-two jurisdictions in the United States have adopted external review laws, almost all in less than a decade. These laws represent both a new paradigm in health law and a new generation of adjudication mechanisms. Where the two fields intersect, external review furnishes the latest chapter in a long-running saga of tension between deference to physicians and the norms of procedural due process.

External review systems combine private judging companies and standards set by public law. They illustrate both how due process norms are manifest in privatized dispute resolution and how hybrid public-private networks function as a key mechanism for adjudication. External review represents a new model for resolving disputes about whether care is medically necessary, and may foreshadow similar models in other fields.

In external review systems, the dominant process norms are accountability and deliberativeness, rather than a model of rights associated with entitlement to a hearing. The purpose of eliciting dialogic engagement combines with the use of procedure itself as regulation. However, those goals cannot be attained or even confidently measured without greater transparency than now exists in these new systems.

335. *See supra* text accompanying note 61.

336. See Louise Trubek's call for "stepping outside the regulatory box." Trubek, *supra* note 258, at 583-84.

NOTE

Does "Reparative" Therapy Really Constitute Child Abuse?: A Closer Look

Sean Young*

INTRODUCTION

The political rhetoric surrounding children and homosexuals often overlooks the interests of children who are themselves homosexual. Yet under the political radar, legal scholars are becoming increasingly cognizant of the need to discuss the rights of sexual minority children vis-à-vis their parents. As one author notes, "Psychological abuse from family members affects queer youth more than any other group of adolescents,"[1] and sixty percent of gay-related violence suffered by these children takes place in the home.[2] Queer youth who disclose their sexual identity to unaccepting parents may suffer emotional and physical abuse. Parents "may banish the child from the house and shirk their legal duty to provide financial support because they want to disown the sexuality of their child."[3] Within this web of physical and emotional abuse lie parental efforts to change their child's sexual orientation, otherwise known as "reparative therapy."[4]

---

* J.D. candidate, Yale Law School.

1. Sonia Renee Martin, *A Child's Right To Be Gay: Addressing the Emotional Maltreatment of Queer Youth*, 48 HASTINGS L.J. 167, 169 (1996).

2. Valerie Lehr, Parental Rights as if Queer Youth Mattered 10 n.59 (Aug. 29, 2002) (unpublished manuscript prepared for the Annual Meeting of the American Political Science Association), *available at* http://it.stlawu.edu/~vleh/APSA%202002%20—final.pdf.

3. Miriam Aviva Friedland, *Too Close to the Edge: Lesbian, Gay, Bisexual and Transgender Youth in the Child Welfare System*, 3 GEO. J. GENDER & L. 777, 792-93 (2002).

4. Also known as "conversion therapy" or "reorientation therapy." In 1996, Sonia Martin noted that parents' reactions to their child's coming out is often accompanied by "seeking psychiatric therapy or institutionalization for their queer children." Martin, *supra* note 1, at 174.

In 1999, Karolyn Ann Hicks proposed that subjecting one's child to reparative therapy can and should constitute child abuse.[5] Using New York state child abuse law as a framework,[6] Hicks claimed that abuse depends on the "reasonably prudent parent standard."[7] A reasonably prudent parent researching reparative therapies would discover that they are potentially dangerous and not accepted in the mainstream medical community. Further, taking existing societal homophobia into account, a parent would not subject her child to such therapy.[8] Therefore, a parent who subjects her child to reparative therapy should be guilty of child abuse. Hicks concluded with guarded optimism, however, by stating that since the Supreme Court's 1986 decision in *Bowers v. Hardwick* held that a belief in homosexuality's immorality was a valid public policy consideration,[9] it would be difficult for states to prosecute parents under this theory.

Hicks's theory has recently taken on new relevance. In 2004, what Hicks claimed as the primary obstacle to her theory's implementation was removed: *Bowers* was overruled by *Lawrence v. Texas*,[10] which held that a belief in homosexuality's immorality could not serve as a rational basis for any state law.[11] Regardless of whether *Lawrence* really means that Hicks's theory may now be implemented,[12] In re *E.L.M.C.*[13] demonstrates that such a prospect is not

---

5. Karolyn Ann Hicks, *"Reparative" Therapy: Whether Parental Attempts To Change a Child's Sexual Orientation Can Legally Constitute Child Abuse*, 49 AM. U. L. REV. 505 (1999). Hicks does not explicitly distinguish between the concepts of child abuse and child neglect. This is understandable, as both concepts generally depend on similar standards. "Neglect has also been broadly defined as the disregard of one's duty, owing to indifference or willfulness." 3 AM. JUR. 2D *Proof of Facts* § 1 (2005); *see, e.g.*, *In re* C. Children, 583 N.Y.S.2d 499 (App. Div. 1992) (using same evidence to determine child abuse and child neglect). For the sake of simplicity, the differences between child abuse and neglect are beyond the focus of this Note, which focuses on child abuse; this is because subjecting a child to reparative therapy involves an affirmative act that more properly falls under the category of abuse. In any case, the distinction between child abuse and neglect should not be overemphasized.

6. Hicks, *supra* note 5, at 520.

7. *Id.* at 523 (citing Enright v. Busy Bee Playschool, 625 N.Y.S.2d 453, 454 (App. Div. 1995); *In re* Robert "YY," 605 N.Y.S.2d 418, 420 (App. Div. 1993)).

8. Hicks, *supra* note 5, at 524-25.

9. 478 U.S. 186 (1986); Hicks, *supra* note 5, at 546.

10. 539 U.S. 558 (2003).

11. *Id.* at 577 ("'[T]he fact that the governing majority in a State has traditionally viewed a particular practice as immoral is not a sufficient reason for upholding a law prohibiting the practice.'") (quoting *Bowers*, 478 U.S. at 216 (Stevens, J., dissenting)).

12. There are likely other obstacles to Hicks's theory being implemented. There are unique barriers to prosecuting child abuse in situations in which the abuse is emotional. The harmful effects of emotional abuse are hidden to third parties, who would ordinarily be able to report it. Adolescents themselves construe such abuse as "normal." There is little awareness of intervention

implausible. That 2004 case involved a child custody dispute between a same-sex couple, in which one parent began to believe homosexuality was immoral after the separation. The trial court below awarded the parties joint parental responsibility but required the parent who believed homosexuality was immoral to "make sure that there is nothing in the religious upbringing or teaching that the minor child is exposed to that can be considered homophobic."[14] The appellate court remanded the case to determine whether homophobic teachings would "significantly impair her emotional development."[15] Although this was a child custody case, the evidence used to determine custody is often equally applicable in child abuse cases.[16] And although this case did not involve reparative therapy, the motivating factor behind the parent's potentially homophobic teachings and reparative therapy was the same: the belief, usually religious, that homosexuality is immoral. Therefore, this case shows that a court may not consider that belief to be sacrosanct territory immune from judicial intervention.

This Note argues that, for evidentiary reasons, contrary to what Hicks may suggest, reparative therapy cannot be considered child abuse under current law.[17] There is no reliable evidence that reparative therapy works, and there is also no empirical evidence that reparative therapy is harmful. Professional psychological associations' codes of ethics continue to remain silent on the practice, and the courts do not require psychological treatments to empirically demonstrate their effectiveness in order to justify their appropriateness. This Note discusses these evidentiary factors in detail.

In responding to Hicks's piece, this Note also presents a nuanced, up-to-date, and practical legal framework through which to analyze reparative therapy. This is necessary to the extent legal scholars continue to theorize causes of action against and protections from reparative therapy. For instance, Laura Gans

---

options. And states require children to be represented by an adult. John Alan Cohan, *Parental Duties and the Right of Homosexual Minors To Refuse "Reparative" Therapy*, 11 BUFF. WOMEN'S L.J. 67, 78-79 (2002-2003); *see also* Sana Loue, *Redefining the Emotional and Psychological Abuse and Maltreatment of Children*, 26 J. LEGAL MED. 311, 336 (2005) (discussing reasons for underreporting of emotional child abuse as compared to physical child abuse and listing recommendations to encourage the prosecution of emotional child abuse).

13. 100 P.3d 546 (Colo. Ct. App. 2004).

14. *Id.* at 563.

15. *Id.*

16. *See generally* James Lockhart, Annotation, *Cause of Action for Modification of Child Custody or Visitation Arrangement Based on Abuse of Child*, 6 CAUSES OF ACTION 2D 287 (2004) (discussing how child abuse determinations influence custody determinations).

17. This is not to ignore the fact that political reasons play a role in whether such prosecutions could happen. It is simply more useful to legal scholarship to analyze the evidentiary barriers rather than to attribute an undue amount of influence to politics.

proposes an "intentional infliction of emotional harm" cause of action against reparative therapists;[18] David Cruz advocates a less deferential standard of informed consent in reparative therapy;[19] John Alan Cohan asserts that adolescents should have the constitutional right to refuse reparative therapy;[20] and James Gilliam highlights the problem of placing homosexual children in the foster care of those who employ reparative therapy.[21] In addition to these theories, this Note's framework will also affect the standards for determining whether the institutionalization of homosexual juveniles for reparative therapy purposes is constitutional,[22] as well as determining how child custody is awarded.

Part I critiques Hicks's definition of reparative therapy and proposes a useful alternative. Part II shows that Hicks's analysis turns on the admissibility of psychological testimony as evidence. It discusses the evidentiary standards set out in *Daubert v. Merrell Dow Pharmaceuticals, Inc.*[23] and *Frye v. United States*,[24] surveys the way the standards have been applied to psychological testimony generally, and then applies the standards to the reparative therapy scenario. Part III distinguishes among and analyzes the legal significance of the data presented by Hicks and other reparative therapy opponents and shows how this evidence would be ineffective in demonstrating that reparative therapy constitutes child abuse. Part IV explores the ambivalent relationship between professional codes of ethics and position statements of psychology associations and explains how only the codes of ethics are relevant to whether reparative therapy constitutes child abuse. Part V discusses the "empirically validated treatment" (EVT) controversy that is brewing in the practice of psychology. Part VI discusses the implications of these analyses for whether reparative therapy constitutes child abuse, discusses the implications for reparative therapy in other legal contexts, and proposes recommendations for dealing with reparative

18. Laura A. Gans, *Inverts, Perverts, and Converts: Sexual Orientation Conversion Therapy and Liability*, 8 B.U. PUB. INT. L.J. 219, 245 (1999).

19. David B. Cruz, *Controlling Desires: Sexual Orientation Conversion and the Limits of Knowledge and Law*, 72 S. CAL. L. REV. 1297, 1361 (1999) (discussing informed consent in reparative therapy context).

20. Cohan specifically advocates for adolescents to possess the right to refuse reparative therapy, just as they currently have the constitutional right to contraceptives, testing for sexually transmitted diseases, and abortions. Cohan, *supra* note 12, at 75.

21. James W. Gilliam, Jr., *Toward Providing a Welcoming Home for All: Enacting a New Approach To Address the Longstanding Problems Lesbian, Gay, Bisexual, and Transgender Youth Face in the Foster Care System*, 37 LOY. L.A. L. REV. 1037, 1039 (2004).

22. *See infra* Section I.B. (discussing *Parham* standard of due process for committing youths to mental institutions).

23. 509 U.S. 579 (1993).

24. 293 F. 1013 (D.C. 1923).

therapy in the legal system generally.

I. DEFINING REPARATIVE THERAPY

Hicks provides the following definition of reparative therapy: "Reparative therapy, a program of psychotherapy, attempts to 'cure' homosexuals by turning them into heterosexuals."[25] These therapies may include:

> [B]ehavioral therapy, electrical shock therapy, chemical aversive therapy, drug and hormone therapy, surgery, and psychotherapy. Other accounts are similar and include homophobic counseling, religious propaganda, isolation, unnecessary medication (including hormone treatment), subliminal therapies designed to inculcate 'feminine' or 'masculine' behavior and 'covert desensitization' therapies that teach a young person to associate homosexual feelings with disgusting images.[26]

This definition is too broad to be analytically or practically useful. This Part explains why this is the case and proposes a revised definition for purposes of legal analysis.

A. From Electroshock Therapy to Psychotherapy: A Critical Distinction

Hicks's account of reparative therapy is misleading, because it does not recognize the decline in the more physically invasive methods of reparative therapy and the rise of "purely" psychotherapeutic reparative therapies. A child "praying to be saved"[27] is quite different from a child undergoing electrical shock therapy.[28]

Other scholars, on the other hand, have made this critical distinction. For instance, Gans notes, "The earliest forms of conversion therapy included injecting patients with substances, such as testosterone, estrogen, animal organ extracts, and cocaine, performing 'castration, hysterectomy, and vasectomy,' and surgically removing the ovaries and clitoris."[29] Then she distinguishes psychotherapy, explaining that "psychiatrists also employed [other] techniques specifically targeting the mind. . . . Psychotherapy appears to be one of the more popular therapeutic formats through which to carry out conversion attempts."[30]

---

25. Hicks, *supra* note 5, at 513.

26. *Id.* at 515.

27. *Id.* at 521.

28. *Id.* at 515.

29. Gans, *supra* note 18, at 223 (quoting JONATHAN N. KATZ, GAY AMERICAN HISTORY: LESBIANS AND GAY MEN IN THE U.S.A. 130 (ed. rev. 1992)).

30. *Id.* at 223-24.

Kenji Yoshino specifically notes that reparative therapy today is primarily focused on psychotherapy. He remarks that "[e]ven mental health professionals who currently advocate psychoanalytic therapy for homosexuals deride such physical interventions as 'quackeries,'" and that "[v]irtually every sexual orientation therapy ever formulated has typically passed into history along with its originators, . . . [but that] [p]sychoanalysis has proved one exception to this rule of obsolescence."[31] As early as 1996, commentators noted that reparative therapy was most often "conducted through 'conventional' therapy, i.e., psychotherapy."[32] Cruz also recognizes that "perhaps the most enduring psychic approaches [to reparative therapy] besides aversive techniques have been psychoanalytic in nature,"[33] and that "[t]he psychic, primarily verbal conversion techniques in current circulation avoid the appearance of outright torture that marked many past practices."[34]

State statutes themselves already generally recognize the unconscionability of more physically invasive practices by permitting minors to refuse such treatments. "The laws in many states[35] support a minor's right to refuse extreme treatments such as electroconvulsive therapy, psychosurgery, and behavior modification programs utilizing deprivation or aversive techniques."[36] On the

---

31. Kenji Yoshino, *Covering*, 111 YALE L.J. 769, 789 (2002) (quoting TIMOTHY F. MURPHY, GAY SCIENCE: THE ETHICS OF SEXUAL ORIENTATION RESEARCH 82-83 (1997)).

32. Gans, *supra* note 18, at 221 n.12 (citing Lili Wright, *The Straight Truth: No One Knows if Gays Can Change*, SALT LAKE TRIB., May 12, 1996, at A1).

33. Cruz, *supra* note 19, at 1307.

34. *Id.* at 1309-10.

35. *See, e.g.*, ALASKA STAT. § 47.30.825(g) (2004) ("In no event may treatment include psychosurgery, lobotomy, or other comparable form of treatment without specific informed consent of the patient, including a minor . . . ."); CAL. WELF. & INST. CODE § 5326.6(d) (West 2005) ("Under no circumstances shall psychosurgery be performed on a minor."); LA. CHILD. CODE ANN. art. 1409.O (2005) ("Prefrontal lobotomy shall be prohibited as a treatment solely for medical or emotional illness of a minor patient."). *But see* KAN. STAT. ANN. § 59-2978(a)(6) (2004) ("Every patient being treated in any treatment facility, in addition to all other rights preserved by the provisions of this act, shall have the following rights: . . . not to be subject to such procedures as psychosurgery, electroshock therapy, experimental medication, aversion therapy or hazardous treatment procedures without the written consent of the patient or the written consent of a parent or legal guardian, if such patient is a minor or has a legal guardian provided that the guardian has obtained authority to consent to such from the court which has venue over the guardianship following a hearing held for that purpose . . . .").

36. Cohan, *supra* note 12, at 82 (referring to "[a]versive techniques" as the use of physical restraints or seclusion to negatively condition undesired behaviors); *see also, e.g.*, Heller v. Doe *ex rel.* Doe, 509 U.S. 312, 345 (1993) (listing examples of such techniques used on mentally retarded); Natrona County Sch. Dist. No. 1 v. McKnight, 764 P.2d 1039, 1044 (Wyo. 1988) (including the following definition of "aversive" techniques: "ignoring, [saying] no, token fine, water spray, vapor

other hand, no state appears to give minors the right to refuse psychotherapy.[37]

B. Institutionalization of Adolescents

Hicks also highlights another unique method used by parents who wish to change their child's sexual orientation: the institutionalization of adolescents in psychiatric centers against their will.[38] In some cases, children are "kidnapped, taken to an in-patient center, and drugged for most of their teenage years . . . ."[39]

However, this attempt to change a minor's sexual orientation does not fall under child abuse law, because legal doctrine attributes the act of institutionalization to the state, not the parent.[40] As both Hicks and Cohan

---

spray, vision-occluding, and sound-masking helmet, ammonia, taste aversive, cool shower, muscle squeeze, spank, pinch, time-out helmet with safety tube and optional automatic vapor spray, contingent physical exercise, and hand squeeze").

37. One Florida statute may give minors a right to refuse psychotherapy. It provides that: "When any minor age 13 years or older experiences an emotional crisis to such degree that he or she perceives the need for professional assistance, he or she shall have the right to request, consent to, and receive outpatient crisis intervention services including individual psychotherapy, group therapy, counseling, or other forms of verbal therapy provided by a licensed mental health professional . . . ." FLA. STAT. ANN. § 394.4784 (West 2005). Although the statute appears to give the minor a right to "consent to . . . psychotherapy," the context of the statute is meant to give minors a right *to* psychotherapy. Consent is mentioned, because typically consent by the parent is required; but in a situation in which a minor has an emotional crisis, the minor receives the right to consent for himself instead of relying on the parent. This concept is better illustrated by an analogous Illinois statute: "Any minor 12 years of age or older may request and receive counseling services or psychotherapy on an outpatient basis. The consent of his parent, guardian or person in loco parentis shall not be necessary to authorize outpatient counseling or psychotherapy." 405 ILL. COMP. STAT. ANN. 5/3-501 (West 2005). A New Mexico statute contains a similar provision and is immediately followed by its prohibition on the more invasive treatments being performed on a minor, which emphasizes the fact that no such protections exist for psychotherapy: "(A) Any child shall have the right, with or without parental consent, to consent to and receive individual psychotherapy, group psychotherapy, guidance, counseling or other forms of verbal therapy that do not include any aversive stimuli or substantial deprivations. (B) No psychosurgery or convulsive treatment shall be performed on a child . . . ." N.M. STAT. ANN. § 32A-6-14 (Michie 2005).

38. Hicks, *supra* note 5, at 521; *see also* Ruthann Robson, *Our Children: Kids of Queer Parents & Kids Who Are Queer: Looking at Sexual Minority Rights from a Different Perspective*, 64 ALB. L. REV. 915, 934 n.77 (2001) (citing literature).

39. Hicks, *supra* note 5, at 521. It is worth noting that, legally, a parent cannot "kidnap" her own children absent a court order denying custody.

40. This is presumably because upon the parent's request, the state is the actor that ultimately decides whether a child may be institutionalized. *See* Parham v. J.R., 442 U.S. 584, 598-99 (1978) ("In an earlier day, the problems inherent in coping with children afflicted with mental or emotional abnormalities were dealt with largely within the family. . . . As medical knowledge about the

recognize, the institutionalization of children generally is reviewed under the Due Process Clause of the Federal Constitution, which governs actions by the *state*. In *Parham v. J.R.*,[41] the Supreme Court held that before a parent may commit her child, a neutral fact-finder, relying upon psychiatric standards, must determine whether the child can be committed.[42] Because juvenile institutionalization does not properly fall under child abuse doctrine,[43] this Note's definition of reparative therapy will not include it.[44]

C. A Revised Definition of Reparative Therapy

This Note defines reparative therapy as the attempt, *through psychotherapy*, to 'cure' homosexuals by turning them into heterosexuals. It adopts the view that "[p]sychotherapy is distinct from therapy that employs 'medical treatments directed primarily at the patient's body or treatment involving the use of chemical or mechanical means.' Instead, it is a 'treatment of mental and emotional problems by psychological methods.'"[45]

There is no "standard" method of performing reparative therapy.[46] A vast number of books and guides articulate various methodological and theoretical

---

mentally ill and public concern for their condition expanded, the states, aided substantially by federal grants, have sought to ameliorate the human tragedies of seriously disturbed children. Ironically, as most states have expanded their efforts to assist the mentally ill, their actions have been subjected to increasing litigation and heightened constitutional scrutiny.").

41. *Id.*

42. *Id.* at 606-09 (describing procedures required by due process).

43. There do not appear to be any cases in which the act of a parent committing a child to an institution was alleged to be child abuse. The Supreme Court has recognized that the institutionalization of a minor is a separate issue from the existence of child abuse: "In defining the respective rights and prerogatives of the child and parent in the voluntary commitment setting, we conclude that our precedents permit the parents to retain a substantial, if not the dominant, role in the decision, absent a finding of neglect or abuse . . . ." *Id.* at 604.

44. Nonetheless, because the constitutionality of juvenile institutionalization ultimately rests upon the psychiatric standards on which the neutral fact-finder bases her decision, and because psychiatric standards are a prominent factor in determining whether reparative therapy is child abuse, this Note's analysis also indirectly affects *Parham* cases. *See infra* Section VI.A.

45. Gans, *supra* note 18, at 224.

46. The best description of the reparative therapy method from a non-reparative therapist is Douglas C. Haldeman, *Gay Rights, Patient Rights: The Implications of Sexual Orientation Conversion Therapy*, 33 PROF. PSYCHOL.: RES. & PRAC. 260, 260 (2002) (noting that "[p]sychoanalytic theories, still promoted by some advocates of conversion therapy, suggest that homosexuality constitutes a form of arrested psychosexual development. According to this notion, lesbians and gay men suffer from an incomplete bond and resultant identification with the same-sex parent, which is then symbolically repaired in psychotherapy").

approaches to conversion therapy.[47] Therapies are usually complemented by a formal or informal support network.[48] Therapy also attempts to undo psychological or emotional patterns of thinking, such as rejection, shame, and animosity toward the same-sex parent. The patient is also encouraged to develop non-sexual bonds with members of the same sex.[49] Religion, through prayer or Bible study, is often integrated into the practice.[50]

This Note focuses on psychotherapy for three reasons. First, the conscience-shocking physical interventions listed above, as opposed to psychoanalysis, are themselves derided by some reparative therapists.[51] Second, although such physical interventions undoubtedly still occur today, the pervasiveness of non-physical psychoanalytical reparative therapy provides an easy moral and legal defense for parents facing critics, who usually base their criticism on such physical interventions. Third, limiting the definition to psychoanalysis is more analytically useful. If the definition includes the invasive physical interventions, an analysis of whether reparative therapy constitutes child abuse would be confused with an analysis of whether physically invasive therapy in general constitutes child abuse. By limiting the definition to psychoanalysis, the inquiry is limited to whether the act of attempting to change a child's sexual orientation without physically invasive techniques is itself child abuse.

Limiting the definition of reparative therapy to psychotherapy should not be read to imply that psychotherapy is somehow more benign or harmless by nature. As Judge Stephen Hjelt notes:

> Psychotherapy is the principle product, good, or service of the mental health

---

47. *See generally* Regeneration Books, http://www.regenbooks.org. Regeneration Books is a Christian "ex-gay" ministry that serves as a clearinghouse for religious reparative therapy literature.

48. Rob G., New Direction for Life, Getting Out: Some Things You Should Know About the Journey Out of Homosexuality, http://www.freetobeme.com/r_primer.htm (last visited Jan. 25, 2005).

49. These broad generalizations of reparative therapy are drawn from several examples of positive and negative testimonials regarding experiences with conversion therapy. *See generally* Ben Newman [pseud.], Nat'l Ass'n for Research & Therapy, Change of Heart: "My Two Years in Reparative Therapy," http://www.narth.com/docs/ben.html (last visited Nov. 11, 2005); HUMAN RIGHTS CAMPAIGN FOUNDATION, FINALLY FREE: PERSONAL STORIES: HOW LOVE AND ACCEPTANCE SAVED US FROM "EX-GAY" MINISTRIES 7 (2000), *available at* http://www.hrc.org/Content/ContentGroups/Publications1/Finally_Free/FinallyFREE.pdf.

50. Gans, *supra* note 18, at 226 ("One 'discipleship program' combines 'four meetings a week of Bible study, church worship, and group therapy to examine behavior patterns, lifestyle changes, and the underlying psychological causes of homosexuality.' Some of the other techniques reportedly used include a '14-step recovery program' and the playing of team sports, such as baseball and basketball.").

51. *See supra* notes 31-32.

> profession. It is a treatment that can do great good or great harm. It is a functional analog to a drug or medical device. It can relieve symptoms and resolve conditions. It can cure. It can kill. It can also cause adverse reactions. Like any drug or medical device, psychotherapy has contraindications as well as dose-specific impacts.[52]

Therefore, focusing on psychotherapeutic forms of reparative therapy does not lessen the critical stakes involved—the health and well-being of homosexual children.

II. THE EVIDENCE PROBLEM: THE ADMISSIBILITY OF PSYCHOLOGICAL TESTIMONY

Hicks's thesis that reparative therapy ought to constitute child abuse is based upon child abuse law in the state of New York.[53] Generally, a parent in New York who does not treat her child as would a "reasonably prudent parent" is subject to child abuse prosecution. Hicks argues that when a "reasonably prudent parent" considers whether to send her child to a reparative therapist, she would discover that the practice is harmful and consequently refrain from subjecting her child to reparative therapy.[54] Thus, if that parent still sends her child to reparative therapy, thereby violating the "reasonably prudent parent" standard, she will have abused her child.

Hicks fails to recognize the evidentiary problems with her analysis by failing to acknowledge that the judicial determination of child abuse in the reparative therapy context rests largely on expert testimony. In order to show why such a case depends largely on expert testimony, it is necessary to briefly examine what legally constitutes child abuse and the evidentiary standards required to prove it.

A. Why the Determination of Whether Reparative Therapy Constitutes Child Abuse Depends on Expert Testimony

The definition of child abuse varies by state statute, but the minimum standard is defined by a federal statute, the Child Abuse Prevention and Treatment Act, which conditions federal funding on state adoption of this definition of child abuse: "Any recent act or failure to act on the part of a parent

---

52. Stephen Hjelt, *Informed Consent and Psychotherapy: Apples and Oranges in the Garden of Doctrine*, 22 J. NAT'L ASSOC. ADMIN. L. JUDGES 1, 2 (2002).

53. *See* Hicks, *supra* note 5, at 520 ("The State of New York is a sizable state with thorough statutory law and sufficient case law on child abuse and neglect. This Comment, therefore, selected the laws of the State of New York as a framework for analyzing whether 'reparative' therapy constitutes child abuse and neglect.").

54. Hicks, *supra* note 5, at 523-25.

or caretaker, which results in death, serious physical or emotional harm, sexual abuse or exploitation, or an act or failure to act which presents an imminent risk of serious harm."[55]

Since this Note's definition of reparative therapy excludes physically invasive techniques,[56] the analysis of whether reparative therapy constitutes child abuse focuses on the emotional or psychological form of child abuse. According to the National Clearinghouse on Child Abuse and Neglect Information, emotional abuse is now widely considered to be a form of child abuse:

> All States and territories except Georgia and Washington include emotional maltreatment as part of their definitions of abuse or neglect. Typical language used in these definitions is 'injury to the psychological capacity or emotional stability of the child as evidenced by an observable or substantial change in behavior, emotional response, or cognition,' or as evidenced by 'anxiety, depression, withdrawal, or aggressive behavior.'[57]

As the above quotation suggests, and as J. Robert Shull explains in his detailed survey of state statutory treatment of emotional child abuse,[58] despite the fact that popular and professional psychological conceptions of "emotional child abuse" focus on the actions of the parent, the legal system focuses on the existence of actual emotional harm in the child, in addition to whether the parent's actions caused the harm.[59]

---

55. 42 U.S.C.A. § 5106g(2) (West 2003).

56. *See supra* Section I.C.

57. Nat'l Clearinghouse on Child Abuse & Neglect Info., Definitions of Child Abuse and Neglect, http://nccanch.acf.hhs.gov/general/legal/statutes/define.cfm#bfn6 (last visited Oct. 10, 2005); *see, e.g.*, CAL. WELF. & INST. CODE § 18951(e)(4) (West 2001) ("'Child abuse' as used in this chapter means a situation in which a child suffers from any one or more of the following: . . . Willful mental injury."); N.Y. FAM. CT. ACT § 1012(e-i) (McKinney 2005) ("protracted impairment of physical or emotional health"); WIS. STAT. § 948.04(1) (West 2005) ("Whoever is exercising temporary or permanent control of a child and causes mental harm to that child by conduct which demonstrates substantial disregard for the mental well-being of the child is guilty of a Class F felony."); *see also* Cohan, *supra* note 12, at 77 ("Today most states recognize that mental and emotional abuse fall under child abuse and neglect statutes.").

58. J. Robert Shull, Note, *Emotional and Psychological Child Abuse: Notes on Discourse, History, and Change*, 51 STAN. L. REV. 1665 (1999).

59. *Id.* at 1672 (analyzing Alaska statute as an example, noting that "[t]he emphasis is not on the actions of the parent—verbal castigation, close confinement, whatever—but instead on the results, the measurable ('observable') and severe ('substantial') effects on the child's development ('impairment in the child's ability to function'). Damage, apparently, must be actual and not inferred"); *see also* Loue, *supra* note 12, at 317 (discussing professional psychological definitions of emotional child abuse, and noting that "[t]here exists a tension between those definitions

Whether actual emotional harm exists and whether the parent caused such emotional harm are factual inquiries,[60] and the most effective and widely used method for convincing the fact-finder that abuse exists is to use expert opinion testimony.[61] In fact, some statutes explicitly require such testimony to establish the existence of emotional abuse.[62]

Therefore, in order to successfully prosecute a parent who subjects her child to reparative therapy, expert testimony is almost always required. Presumably, expert testimony conveying the arguments and data presented by Hicks's piece would be sufficient to establish the link between the reparative therapy and the resulting emotional harm. But would such testimony even be admissible in court? This Part argues that it is doubtful.

---

focusing primarily on the resulting harm to the child, such as causing impairment, and those that focus on the behavior of the abuser, such as exposing the child to verbal insults. Still other definitions focus on the child's response to the abuser's behavior").

60. Jenna Mella, Annotation, *Termination of Parental Rights Based on Abuse or Neglect*, 9 CAUSES OF ACTION 2D 483 § 6 (1997) ("A finding of abuse is usually based, on the testimony or evidence given by an expert witness, usually a doctor, who is qualified to make a judgment on whether the child's injuries were the result of abuse, or accidental in nature based on a thorough physical examination.").

61. *See, e.g.*, Smith v. Smith, No. 2004 CU 2168, 2005 WL 2374721, *8 (La. Ct. App. Sept. 28, 2005) ("After Dr. Pellegrin reviewed the tape, she opined that the child was clearly being subjected to severe emotional abuse by Michaelle Duncan, in that Michaelle Duncan was clearly alienating the child from her father, encouraging the child to spy on her father and family, and asking her to perform poorly in school. This testimony was not contradicted by Michaelle Duncan or by any other evidence . . . ."); *In re* K.A.W., 133 S.W.3d 1, 14 (Mo. Super. Ct. 2004) (reversing trial court determination of existence of emotional child abuse by evaluating competing testimonies by experts); Skye W. v. Jennifer W., 704 N.W.2d 1 (Neb. Ct. App. 2005) (holding that state failed to establish existence of emotional abuse to seek termination of parental rights because state provided no expert testimony).

62. *E.g.*, ALASKA STAT. § 47.17.290(9) (2004) ("'[M]ental injury' means a serious injury to the child as evidenced by an observable and substantial impairment in the child's ability to function in a developmentally appropriate manner and the existence of that impairment is supported by the opinion of a qualified expert witness . . . ."); S.C. CODE ANN. § 20-7-490(5) (2004) ("'Mental injury' means an injury to the intellectual, emotional, or psychological capacity or functioning of a child as evidenced by a discernible and substantial impairment of the child's ability to function when the existence of that impairment is supported by the opinion of a mental health professional or medical professional."); TENN. CODE ANN. § 37-1-102(21)(B) (West 2001) ("'Severe child abuse' means . . . [s]pecific brutality, abuse or neglect towards a child which in the opinion of qualified experts has caused or will reasonably be expected to produce severe psychosis, severe neurotic disorder, severe depression, severe developmental delay or retardation, or severe impairment of the child's ability to function adequately in the child's environment, and the knowing failure to protect a child from such conduct.").

The discussion deserves its own Part for three reasons. First, the evidentiary rules concerning this important factor are complex and vary by state. Second, as this Part shows, psychological testimony is fundamentally different from traditional scientific testimony, introducing additional wrinkles that will be relevant to the reparative therapy scenario. Third, there is an enormous amount of literature that appears to make conflicting assertions concerning how these rules apply specifically to psychological testimony, and these conflicting assertions deserve analysis.

There are two primary evidentiary standards. *Frye* generally governed federal and state courts for most of the twentieth century,[63] but was replaced in federal courts by *Daubert* in 1993. Since then, twenty-four states[64] have incorporated the *Daubert* test into their jurisprudence, while sixteen states and the District of Columbia still adhere to the *Frye* test.[65] Six states apply a hybrid of the tests[66] and four states have developed their own tests.[67]

B. The "Conservative" Frye Standard

The *Frye* standard states that novel scientific knowledge must have gained "general acceptance" in the scientific community in order to be admissible.[68] In 1923, the District of Columbia Appellate Court set the standard that would eventually be adopted by the rest of the United States. In deciding that a newly developed lie-detecting test was not well-established enough to be admitted into evidence, the court stated:

> Just when a scientific principle or discovery crosses the line between the experimental and demonstrable stages is difficult to define. Somewhere in this twilight zone the evidential force of the principle must be recognized, and while courts will go a long way in admitting expert testimony deduced from a well-recognized scientific principle or discovery, the thing from which the deduction is made must be sufficiently established to have gained general

63. *See* Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579, 585 (1993) ("In the 70 years since its formulation in the *Frye* case, the 'general acceptance' test has been the dominant standard for determining the admissibility of novel scientific evidence at trial.").

64. *See* Alice B. Lustre, Annotation, *Post-*Daubert *Standards for Admissibility of Scientific and Other Expert Evidence in State Courts*, 90 A.L.R.5TH 453 (2005).

65. *Id.*

66. *Id.* (Alabama, Hawaii, Massachusetts, Nevada, New Hampshire, and New Jersey).

67. *Id.* (Georgia, Utah, Virginia, and Wisconsin). Discussing each of these unique jurisdictions is, unfortunately, outside the scope of this Note.

68. *Daubert*, 509 U.S. at 585.

acceptance in the particular field in which it belongs.[69]

This has been considered a stringent test because it essentially shuts out new scientific theories which have not yet gained acceptance in the general scientific community, even when those theories are based on sound methodologies. This test was not revisited until 1993.

C. The "Liberal" Daubert Standard

The *Daubert* standard intended to liberalize the rule by adding other factors courts should consider when faced with novel scientific theories. Once it is established that the witness is qualified to serve as an expert, the judge must ensure that the scientific testimony is both relevant and reliable.[70] In doing so, it must consider the following non-exhaustive factors: (1) whether the theory "can be (and has been) tested";[71] (2) whether the theory "has been subjected to peer review and publication";[72] (3) the "known or potential rate of error";[73] and (4) the "general acceptance"[74] factor from *Frye*.[75]

Daubert was meant to be a *liberalizing* rule. The Supreme Court notes that the Federal Rules of Evidence contain a "liberal thrust" and a "general approach of relaxing the traditional barriers to 'opinion' testimony."[76] The liberal nature of *Daubert* is often captured by the distinction between admissibility and weight. Admitting scientific evidence liberally does not mean a "free-for-all in which befuddled juries are confounded by absurd and irrational pseudoscientific assertions."[77] On the contrary, while such evidence may be *admitted* relatively easily, its *weight* can readily be "attacked by cross-examination and refutation."[78]

---

69. Frye v. United States, 293 F. 1013, 1014 (App. D.C. 1923).

70. *Daubert*, 509 U.S. at 589.

71. *Id.* at 593.

72. *Id.*

73. *Id.* at 594.

74. *Id.*

75. Federal Rule of Evidence 702 states that expert testimony in the form of "scientific, technical, or other specialized knowledge" may be admitted if it is probative, and *Daubert* only applies to "scientific" knowledge. But the question of whether psychology is "scientific," "technical," or "other specialized" knowledge became moot when, in 1999, *Kumho Tire Co. v. Carmichael*, 526 U.S. 137 (1999), extended the *Daubert* standard to apply to "technical" and "other specialized" knowledge as well.

76. *Daubert*, 509 U.S. at 588 (quoting Beech Aircraft Corp. v. Rainey, 488 U.S. 153, 169 (1988)).

77. *Id.* at 596.

78. Barmeyer v. Mont. Power Co., 657 P.2d 594 (Mont. 1983).

D. Whether Psychological Testimony Should Be Treated as Scientific Testimony

Doctrinally, psychological testimony should be treated in the same way as other scientific testimony. However, the legal scholarship is split as to whether this should be the case as a normative matter.

On the one hand, psychological testimony should be treated in the same way as other scientific testimony, because psychology is the "science of mind and behavior"[79] and employs rigorous standards for determining the validity and reliability of psychological studies.[80] Admitting psychological testimony does not force the trier of fact to accept it, and psychological evidence is not so ungraspable by the lay person that it will "elicit unquestioning acceptance by the trier of fact."[81] On the other hand, a psychologist often relies on subjective observations and experience; many psychology studies rely on retrospective observation rather than controlled experimentation;[82] and psychological testimony does not "utilize machines or formulas that result in calculated and tangible findings."[83] Furthermore, triers of facts, faced with an "aura of certainty, glossed by a psychiatric diploma and the façade of superior knowledge,"[84] can be led to overlook the testimony's "merely conjectural nature,"[85] erroneously crediting psychological testimony with as much weight as other scientific testimony.

As a predictive matter, many scholars felt that instead of liberalizing evidentiary standards, *Daubert* would make such standards more rigid as applied specifically to psychological testimony, since it explicitly introduced traditionally "scientific" factors such as error rate and testability, which do not apply well to the psychological field.[86] For instance, immediately following *Daubert*, one

---

79. Henry F. Fradella et al., *The Impact of* Daubert *on the Admissibility of Behavioral Science Testimony*, 30 PEPP. L. REV. 403, 412 (2003).

80. *See id.* ("These standards include, but are not limited to, (1) replicability, (2) logic, (3) adherence to recognized methodologies, (4) construct validity (i.e., how well data analysis 'fits' into preexisting theory), (5) adherence to proper statistical sampling and statistical procedures for data analysis, (6) avoidance of bias, and (7) qualifications of the researcher.").

81. Cheri L. Wood, *The Parental Alienation Syndrome: A Dangerous Aura of Reliability*, 27 LOY. L.A. L. REV. 1367, 1402 (1994) (citing David McCord, *Syndromes, Profiles and Other Mental Exotica: A New Approach to the Admissibility of Nontraditional Psychological Evidence in Criminal Cases*, 66 OR. L. REV. 19, 85-86 (1987)).

82. Fradella, *supra* note 79, at 412.

83. Wood, *supra* note 81, at 1401.

84. *Id.* at 1402 (citing *Evidence of "Acute Grief Syndrome" Cannot Be Used as Expert Testimony*, 207 N.Y. L.J. 26, 27 (1992)).

85. *Id.*

86. *See* Veronica B. Dahir et al., *Judicial Application of* Daubert *to Psychological Syndrome*

scholar noted that "the testability or falsifiability and potential error rate factors for appraising [social science evidence] will rarely be sufficiently present to meet the *Daubert* standard."[87] Another stated, "the Court's opinion read literally would dictate the end of the receipt of psychiatric and psychological testimony in federal courts."[88] Even as late as 2004, scholars noted that "syndromes can never satisfy *Daubert* because they are not testable. Courts that choose to rely on evidence stemming from a theory about this or that syndrome embrace what is in reality not a scientific theory at all."[89]

E. The Increasing Importance of the "Generally Accepted" Factor

Any veracity in these predictions and descriptions of *Daubert*'s stringency has been undermined by the fact that judges in *Daubert* jurisdictions apparently treat psychological testimony the same way *Frye* jurisdictions and pre-*Daubert* courts generally have treated them, in part because both standards share the same "general acceptance" factor. In 1996, one scholar noted that *Daubert* courts "have sometimes sought refuge in *Frye*-like principles in order to bring some order out of what appears to be a chaotic situation of there appearing to be a new syndrome developed every month."[90] Based on the results of a survey filled out by 325 state judges before 1999, one scholar concluded that "*Daubert's* impact on the admissibility of psychological syndrome and profile evidence [is]

---

and Profile Evidence, 11 PSYCHOL. PUB. POL'Y & L. 62, 65 (2005) ("[T]here are particular difficulties with psychological syndromes and profiles meeting the falsifiability and error rate guidelines. For example, some syndromes may not be testable at all because of their etiological uncertainty . . . . The related error rate guideline would be difficult for some psychological syndromes and profiles to meet as well, because if something cannot be tested, there is no basis on which to assess known or potential rates of error.").

87. Christopher Slobogin, *Pragmatic Forensic Psychology*, 9 PSYCHOL. PUB. POL'Y & L. 275, 287 n.77 (2003) (citing Michael H. Graham, Daubert v. Merrell Dow Pharmaceuticals, Inc.: *No* Frye, *Now What?*, 30 CRIM. L. BULL. 153, 162 (1994)). *But see* Wood, *supra* note 81, at 1412-13 (fearing in 1994 that *Daubert*'s leniency would make it easier to admit Parental Alienation Syndrome testimonies).

88. Michael H. Gottesman, *Admissibility of Expert Testimony After* Daubert: *The "Prestige" Factor*, 43 EMORY L.J. 867, 875-76 (1994).

89. J. Eric Smithburn, *The Trial Court's Gatekeeper Role Under* Frye, Daubert, *and* Kumho: *A Special Look at Children's Cases*, 4 WHITTIER J. CHILD & FAM. ADVOC. 3, 16 (2004).

90. *Id.* at 15 n.99 (citing James T. Richardson, *Dramatic Changes in American Expert Evidence Law*, 2 JUD. REV. 13, 23 (1996)); *see also* William M. Grove & R. Christopher Barden, *Protecting the Integrity of the Legal System: The Admissibility of Testimony from Mental Health Experts Under* Daubert/Kumho *Analyses*, 5 PSYCHOL. PUB. POL'Y & L. 224, 238 (1999) ("Following *Daubert/Kumho*, federal judges are now on notice by the U.S. Supreme Court that they bear an affirmative duty to actively exclude junk science testimony.").

negligible"[91] and suggested that:

> [J]udges [may not be] concerned with the two more technical factors of falsifiability and error rate because (a) they assume general acceptance and peer review and publication are proxies for scientific reliability and are sufficient factors for determining scientific validity, and (b) because they do not understand how to apply them.[92]

A more recent study notes that *Daubert* courts end up relying heavily on *Daubert*'s "general acceptance" prong.[93] This also appears to be the case in medical cases generally, as one legal scholar noted in 2005: "[T]he admissibility of expert medical testimony in civil and criminal litigation after *Daubert* looks much like that before *Daubert*."[94]

Certain psychological testimonies are regularly admitted regardless of jurisdiction. Psychological testimony with regard to the existence of false confessions; competency to stand trial; an individual's mens rea; personality disorders; and the causes, types, and problems of emotional distress are generally accepted.[95] Testimonies on Post-Traumatic Stress Disorder (PTSD),[96] Battered Child Syndrome,[97] and a child's tendency to delay reporting sexual abuse[98] are

---

91. Dahir, *supra* note 86, at 78.

92. *Id.* at 77.

93. *See* Fradella, *supra* note 79, at 443-44 ("Finally, although courts pay lip service to *Daubert*, it appears the *Frye* test is alive and well. Cases in which methods and/or conclusions were being offered that conformed to those that are 'generally accepted in the relevant scientific community' are the ones in which testimony is deemed admissible. In contrast, when an expert varies from that which is generally accepted, courts are quick to exclude the testimony citing the very same factors that were relevant under *Frye*.").

94. Stephen Chris Pappas, *Curing the* Daubert *Disappointment: Evidence-Based Medicine and Expert Medical Testimony*, 46 S. TEX. L. REV. 595, 615 (2005).

95. Fradella, *supra* note 79, at 441-42.

96. *See* Isely v. Capuchin Province, 877 F. Supp. 1055, 1066 (E.D. Mich. 1995) ("[T]he Court finds that Dr. Hartman is qualified to testify as an expert and give opinion testimony regarding PTSD and repressed memory in this case."); Toro v. State, 642 So. 2d 78, 82-83 (Fla. Dist. Ct. App. 1994) (using *Frye* to admit PTSD testimony, albeit noting that standards may change in future); State v. Alberico, 861 P.2d 192, 208 (N.M. 1993) ("We hold that PTSD testimony is grounded in valid scientific principle."); State v. Martens, 629 N.E.2d 462, 466-68 (Ohio Ct. App. 1993) (adopting *Daubert* test and accepting PTSD testimony).

97. United States v. Boise, 916 F.2d 497, 503-04 (9th Cir. 1990) (applying *Frye* and recognizing that several circuit and state courts have recognized that Battered Child Syndrome is an accepted medical diagnosis); State v. Heath, 957 P.2d 449, 464 (Kan. 1998) (applying *Frye* in admitting testimony concerning general acceptance of Battered Child Syndrome).

98. State v. J.Q., 599 A.2d 172, 174 (N.J. Super. Ct. App. Div. 1991) ("We hold that CSAAS evidence is generally reliable to explain secrecy, belated disclosure and recantation by a child sex

also accepted. Parental Alienation Syndrome (PAS), "the systematic denigration by one parent by the other with the intent of alienating the child against the other parent,"[99] has gained acceptance over the past decade.[100] One scholar in 2003 expressed fears that while *Frye* jurisdictions would and do accept testimonies on Battered Woman Syndrome,[101] that *Daubert* jurisdictions would reject them due to their strict scientific standards.[102] However, this has not been the case, as *Daubert* jurisdictions have also accepted them.[103]

While other psychological testimonies are more controversial, the controversy over whether to accept a certain testimony does not split along *Daubert* and *Frye* lines. For instance, prior to 2000, psychological testimonies on the profiles of sex offenders were rejected in both *Daubert* and *Frye* jurisdictions.[104] In 2000, a new assessment tool called the Abel Assessment for

---

abuse victim . . . ."); State v. Marrington, 73 P.3d 911, 917 (Or. 2003) (remanding case to admit testimony in accordance with proper standards).

99. Stan Hayward, UK Men and Father's Rights, A Guide to the Parental Alienation Syndrome, http://www.coeffic.demon.co.uk/pas.htm (last visited Oct. 29, 2005).

100. *See* Carol S. Bruch, *Parental Alienation Syndrome and Parental Alienation: Getting It Wrong in Child Custody Cases*, 35 FAM. L.Q. 527, 537 (2001) ("An electronic search for all reported U.S. cases between 1985 and February 2001 employing the term 'parental alienation syndrome' revealed numerous mental health professionals in addition to Gardner who have testified that PAS was present."); Alayne Katz, *Junk Science v. Novel Scientific Evidence: Parental Alienation Syndrome, Getting It Wrong in Custody Cases*, 24 PACE L. REV. 239, 239-40 (2003) (noting the use of PAS in child custody cases); Wood, *supra* note 81, at 1401 (noting and denouncing growing acceptance of PAS); *see also In re* Marriage of Bates, 819 N.E.2d 714, 730 (Ill. 2004) (admitting PAS testimony, noting recognition by American Psychological Association). *But see* People v. Loomis, 172 Misc. 2d 265, 267-68 (N.Y. County Ct. 1997) (rejecting PAS testimony in New York).

101. *See, e.g.*, People v. Humphrey, 921 P.2d 1, 2 (Cal. 1996); State v. Smullen, 844 A.2d 429, 449 (Md. 2004); State v. Reese, 692 N.W.2d 736, 741 (Minn. 2005) (affirming line of cases recognizing Battered Woman Syndrome since 1989).

102. *See* Jay B. Rosman, *The Battered Woman Syndrome in Florida: Junk Science or Admissible Evidence?*, 15 ST. THOMAS L. REV. 807, 858 (2003) ("Under a strict *Daubert* test requiring proof of all four 'prongs,' the admissibility of the battered spouse syndrome would most likely fail.").

103. *See, e.g.*, United States v. Young, 316 F.3d 649, 657 (7th Cir. 2002); Harris v. State, 84 P.3d 731, 747 n.13 (Okla. Crim. App. 2004); State v. Weaver, 648 N.W.2d 355, 363-64 (S.D. 2002).

104. *See, e.g.*, United States v. Powers, 59 F.3d 1460, 1471 (4th Cir. 1995) (applying *Daubert* to reject use of penile plethysmograph test for sex offender profiling); State v. Floray, 715 A.2d 855, 857-61 (Del. 1997) (applying *Daubert* standard to reject use of pedophile profiling); Flanagan v. State, 625 So. 2d 827, 828 (Fla. 1993) (applying *Frye*); State v. Parkinson, 909 P.2d 647, 652-53 (Idaho 1996) (applying *Daubert* to reject MMPI sex offender profiling method); State v. Cavaliere,

Sexual Interest had started to gain steam in the psychological community. Since then, the profiling tool has met a mixed response independent of *Daubert* or *Frye*.[105] Similarly, testimonies based on repressed memories have been accepted[106] and rejected[107] in both *Daubert* and *Frye* jurisdictions.[108] Lastly, Child Sexual Abuse Accommodation Syndrome (CSAAS), where a child covers for the parent who is sexually abusing her, has also met with a mixed reaction independent of jurisdiction.[109]

---

663 A.2d 96, 97-100 (N.H. 1995) (applying *Daubert* standard).

105. United States v. Robinson, 94 F. Supp. 2d 751, 753 (W.D. La. 2000) (applying *Daubert* to accept Abel Assessment); *In re* Ready, 824 N.E.2d 474, 476-78 (Mass. App. Ct. 2005) (applying *Daubert* to reject Abel Assessment); People v. Franks, 195 Misc. 2d 698, 702 (N.Y. County Ct. 2003) (rejecting Abel Assessment because it relied on polygraph); State v. Bieck, No. 2003-CA-66, 2004 WL 1490129, at *3 (Ohio Ct. App. 2004) (admitting Abel Assessment).

106. Some *Daubert*-jurisdiction cases accept testimony. *E.g.*, Hoult v. Hoult, 57 F.3d 1, 3-4 (1st Cir. 1995) (affirming trial court's application of *Daubert*); Isely v. Capuchin Province, 877 F. Supp. 1055, 1066 (E.D. Mich. 1995) ("In this case, Dr. Hartman knowledgeably testified about several studies which have validated the theory of repressed memory."); *see also* Fradella, *supra* note 79, at 442 ("[Repressed memory retrieval] has been generally accepted by the scientific community [as of April 2003] and is, therefore, generally accepted under *Daubert*."). There is at least one *Frye*-jurisdiction case which accepts testimony. Wilson v. Phillips, 86 Cal. Rptr. 2d 204, 208 (Cal. Ct. App. 1999) (holding that repressed memory theory is not "scientific" and therefore no *Frye* hearing necessary).

107. Some *Daubert*-jurisdiction cases reject testimony. *E.g.*, Gier v. Educ. Serv. Unit No. 16, 845 F. Supp. 1342, 1353 (D. Neb. 1994) ("Plaintiffs have failed to demonstrate by a preponderance of the evidence that their experts' methodologies for evaluating the plaintiffs in this particular case are reliable."); State v. Cressey, 628 A.2d 696, 699 (N.H. 1993) ("[T]he evaluation of a [sexually abused] child is partly a science and partly an art form."); People v. Murphy, 654 N.Y.S.2d 187, 190 (App. Div. 1997) (accepting repressed memory testimony only if independently corroborated); State v. Quattrocchi III, No. P92-3759, 1999 WL 284882, at *10 (R.I. Apr. 26, 1999) ("As the testimony at the preliminary hearing indicates, the experts are deeply divided on the reliability or accuracy of recovered memories."); Hunter v. Brown, No. 03A01-9504-CV-00127, 1996 WL 57944, at *6 (Tenn. Ct. App. Feb. 13, 1996) (Franks, J., concurring) ("*Daubert* requires the trial judge to 'ensure that any and all scientific testimony or evidence admitted is not only relevant, but reliable.' I believe it is well documented that the current state of scientific knowledge about repressed memory is too contradictory and inconclusive to be a reliable basis for expert testimony at this stage."). At least one *Frye*-jurisdiction case rejects testimony. *E.g.*, Hearndon v. Graham, 767 So. 2d 1179, 1182 (Fla. 2000) ("[W]e recognize that the acceptance of theories supporting memory loss of childhood sexual abuse is a disputed area of psychological study.").

108. *See generally* Joseph A. Spadaro, *An Elusive Search for the Truth: The Admissibility of Repressed and Recovered Memories in Light of* Daubert v. Merrell Dow Pharmaceuticals, Inc., 30 CONN. L. REV. 1147 (1998).

109. Some *Daubert*-jurisdiction cases reject CSAAS theory. *E.g.*, Steward v. State, 652 N.E.2d 490, 493-94 (Ind. 1995) (limiting CSAAS testimony to impeachment of child's testimony);

F. Other Factors for Rejecting Testimonies

Whether a psychological testimony is accepted does not appear dependent upon whether the jurisdiction is a *Daubert* or *Frye* jurisdiction. Nonetheless, there are some consistent reasons why psychological testimonies are rejected. As stated before, lack of "general acceptance" is a common reason for rejection.

Psychological testimonies concerning issues common to the experience of mankind or easily understood by the lay juror have consistently been rejected. For instance, in *Commonwealth v. Francis*,[110] a psychologist testified that "memories fade over time, that people under severe stress do not acquire information as well as alert persons not under stress, and that people tend unconsciously to resolve apparent inconsistencies between their memories and after-acquired facts."[111] Because the jurors had a "general understanding" of such principles, the expert testimony was excluded.[112] And in *State v. Roquemore*,[113] a

---

Newkirk v. Commonwealth, 937 S.W.2d 690, 693 (Ky. 1996) ("The foregoing authorities demonstrate unmistakably that this Court has not accepted the view that the CSAAS or any of its components has attained general acceptance in the scientific community justifying its admission into evidence to prove sexual abuse or the identity of the perpetrator."); State v. Foret, 628 So. 2d 1116, 1125 (La. 1993) ("[H]is use of CSAAS is seen as having highly dubious value by many members of the psychological treatment community."). Some *Frye*-jurisdiction cases reject CSAAS theory. *E.g.*, Hadden v. State, 690 So. 2d 573, 577 (Fla. 1997) ("We likewise agree with Judge Ervin's conclusions that syndrome testimony in child abuse prosecutions must be subjected to a *Frye* test and that such evidence has not to date been found to be generally accepted in the relevant scientific community."); Commonwealth v. Evans, 603 A.2d 608, 611-12 (Pa. Super. Ct. 1992). At least one *Daubert*-jurisdiction case accepts CSAAS theory. *E.g.*, State v. Edelman, 593 N.W.2d 419, 423 (S.D. 1999) ("The trial court did not abuse its discretion in allowing CSAAS testimony."). Some *Frye*-jurisdiction cases accept CSAAS theory. *E.g.*, State v. Curry, 931 P.2d 1133, 1139 (Ariz. Ct. App. 1996); People v. Peterson, 537 N.W.2d 857, 867 (Mich. 1995) ("Qualified experts on child sexual abuse may, therefore, use evidence of CSAAS characteristics of sexually abused children *for the sole purpose* of explaining a victim's specific behavior which might be incorrectly construed as inconsistent with an abuse victim *or* to rebut an attack on the victim's credibility."); State v. Doan, 498 N.W.2d 804, 809 (Neb. Ct. App. 1993) (adopting New Jersey standard that "CSAAS evidence is generally reliable to explain secrecy, belated disclosure and recantation by a child sex abuse victim; [and] that syndrome evidence including CSAAS is not reliable to prove that sex abuse, in fact, occurred"); People v. Higgins, 784 N.Y.S.2d 232, 235 (App. Div. 2004) ("It is well settled that such expert testimony 'may be admitted to explain behavior of a victim that might appear unusual or that jurors may not be expected to understand,' such as a child's failure to promptly complain of abuse.") (citation omitted).

110. 453 N.E.2d 1204 (Mass. 1983); *see also* State v. McClendon, 730 A.2d 1107, 1116 (Conn. 1999) (citing *Francis* with approval).

111. *Francis*, 453 N.E.2d at 1210.

112. *Id.*

rape case required jurors to determine whether an element of rape existed: whether the defendant's actions were violent. The state called upon an expert to testify that the crime scene photos, police reports, and pathological reports fit known patterns of violent behavior. Noting that if "the subject of the testimony is within the understanding of the jury, the expert testimony is inadmissible,"[114] the court rejected the testimony, because the "jury is perfectly capable of making the analysis and factual determinations without opinion testimony."[115]

When proffered evidence is more prejudicial than probative, trial courts must bar admission, an evidentiary rule commonly used to exclude psychological testimony.[116] For instance, in *Pennell v. State*,[117] psychological testimony concerning the profile or general characteristics of a serial murderer as compared with the defendant was rejected. "Such evidence is of little probative value and extremely prejudicial to the defendant since he is, in a sense, being accused by a witness who was not present at any of the crimes."[118] CSAAS testimonies are also sometimes rejected on this basis.[119]

Other reasons for rejecting expert testimony include an inappropriate reliance upon a review of the literature and not on the expert's own experiences or expertise,[120] lack of expert qualification, lack of relevance or applicability,[121] or inappropriate interference with witness credibility.[122] However, these factors will not be as salient to reparative therapy. First, this Note assumes that the testifying psychologist is not merely reviewing and summarizing the literature of

113. 620 N.E.2d 110 (Ohio Ct. App. 1993).

114. *Id.* at 114.

115. *Id.*

116. FED. R. EVID. 403.

117. 602 A.2d 48 (Del. 1991).

118. *Id.* at 55; *see also* State v. Floray, 715 A.2d 855, 859 (Del. Super. Ct. 1997) (rejecting testimony regarding the profile of a sex offender because it was more prejudicial than probative).

119. Rosemary L. Flint, *Child Sexual Abuse Accommodation Syndrome: Admissibility Requirements*, 23 AM. J. CRIM. L. 171, 173 (1995) ("However, some courts prohibit testimony concerning CSAAS, arguing that its probative value is outweighed by its prejudicial impact and that it is unreliable.").

120. United States v. Paul, 175 F.3d 906, 912 (11th Cir. 1999) (rejecting testimony by law professor concerning handwriting analysis, because "[h]is skill, experience, training and education as a lawyer did not make him any more qualified to testify as an expert on handwriting analysis than a lay person who read the same articles"); Carroll v. Litton Sys., Inc., No. B-C-88-253, 1990 WL 312969, at *50 (W.D.N.C. Oct. 29, 1990) ("The mere recitation of a list of studies is not a magical incantation paving the way to the witness stand unless it is accompanied by reasoned and scientifically accepted analysis.").

121. Dahir, *supra* note 86, at 71.

122. Smithburn, *supra* note 89, at 22 ("Some courts reject traits evidence for its improper bolstering of the credibility of the alleged victim and thereby invading the province of the jury.").

reparative therapy. Second, this Note also assumes that psychological testimony is given by a qualified expert. Third, the testimony of psychologists regarding the psychological harm of reparative therapy is especially relevant and applicable, since child abuse cases commonly rely upon expert testimony and reparative therapy only consists of psychoanalysis with no "physical" effects. Finally, testimony rejected due to undue interference with witness credibility mostly involves psychological theories concerning a certain individual's propensity to lie under various circumstances, so this factor is irrelevant here.

Expert testimony that children suffer psychological harm from reparative therapy is required to show that the therapy constitutes emotional child abuse. This Part has demonstrated that whether the jurisdiction is under *Daubert* or *Frye*, such testimony's admissibility depends primarily on the following factors: general acceptance, whether the testimony concerns a subject common to the experience of mankind, and whether the testimony will be more prejudicial than probative.

III. SORTING OUT THE EVIDENCE ON REPARATIVE THERAPY

Although Hicks marshals a substantial amount of data, consisting primarily of surveys and anecdotes, to support her claim that a reasonably prudent parent would know of reparative therapy's harmfulness, much of the data would either be inadmissible in court or ineffective even if admitted. Gans's and Yoshino's critiques of reparative therapy, as well as critiques of reparative therapy generally, also contain data that would face similar evidentiary problems if proffered as evidence.

This Part discusses several categories of data that have been presented by legal scholars and shows how each category of data is either inadmissible in court or ineffective in showing that reparative therapy constitutes child abuse. The categories of data include: sexual orientation's link to nature as opposed to nurture, the influence of homophobia, preexisting psychological problems in homosexual youth, anecdotal evidence, the inability of reparative therapy to change sexual orientation, and homosexuality's status as a mental illness.

After considering each of these categories of data relied upon by legal scholars, this Part discusses the most critical piece of data that these scholars have failed to produce that would actually be admissible and effective in showing that reparative therapy constitutes child abuse: evidence of reparative therapy's harmfulness.

A fine distinction should be made, however. In a child abuse case involving reparative therapy, while the state may not be able to find evidence that reparative therapy is per se harmful, it is entirely plausible for the state to find an expert witness who would testify that reparative therapy in the individual child's

case was harmful to that child. Such testimony would probably be countered by the parent's expert testifying that the alleged harm to the child was not a result of the reparative therapy, but other factors;[123] a jury would then have to decide whom to believe.[124] It is generally difficult for either side to prove their case under this scenario, because in many emotional harm child abuse cases, "[d]epression, antisocial behavior, and other behaviors may be evidence of emotional maltreatment, but they could just as will [sic] be attributed to other causes."[125] In any case, this Note does not speculate on the existence or effectiveness of such testimony, since that fully depends on the facts of each individual case. Rather, this Note focuses on Hicks's assertion that reparative therapy should constitute per se child abuse.

A. Sexual Orientation Is a Product of Nature

Some have rightly noted that pro-gay activists and intellectuals often focus unnecessarily and unhelpfully on whether sexual orientation is a product of nature or nurture.[126] Fortunately, Hicks does not make the mistake of relying on a nature/nurture argument to show how reparative therapy would be child abuse.

---

123. As will be discussed *infra*, many proponents of reparative therapy believe that if someone with homosexual attractions is suffering from psychological malaise such as depression, the depression is a result of his homosexuality, not any form of reparative therapy.

124. Some states may permit a psychologist to testify as to her *opinion* about whether the child in question has suffered harm from the reparative therapy, based on her specialized experience; this is known as "pure opinion testimony." *See* Hadden v. State, 690 So. 2d 573, 579-80 (Fla. 1997) ("While an expert's pure opinion testimony comes cloaked with the expert's credibility, the jury can evaluate this testimony in the same way that it evaluates other opinion or factual testimony.").

125. Cohan, *supra* note 12, at 77. For examples of narratives framing the ex-gay process generally as psychologically beneficial, see Amy E. Peebles, *It's Not Coming Out, So Then What Is It? Sexual Identity and the Ex-Gay Narrative*, 47 TEX. LINGUISTIC F. 155 (2003), *available at* http://studentorgs.utexas.edu/salsa/salsaproceedings/salsa11/SALSA11papers/peebles.pdf.

126. *See generally* Janet E. Halley, *Sexual Orientation and the Politics of Biology: A Critique of the Argument from Immutability*, 46 STAN. L. REV. 503 (1994); *see also* Devon W. Carbado, *Straight Out of the Closet*, 15 BERKELEY WOMEN'S L.J. 76, 109 n.205 (2000) ("[T]he treatment of homosexuality in antiracist discourse should not hinge on whether it is attributable to 'nature' or 'nurture,' nor should it hinge on the question of 'choice.' What is almost always true about efforts to locate the 'cause' of homosexuality is that such efforts are buttressed by the idea that homosexuality is deviant."); Nancy J. Knauer, *Law and Sexuality: A Review of Lesbian, Gay, Bisexual and Transgender Legal Issues*, 12 L. & SEXUALITY 1, 5 (2003) ("By premising their rights claims and related appeals to equality principles on assertions of immutable status, pro-gay advocates have entrusted the success of a major social and political movement to the reliability of a few inconclusive studies concerning, inter alia, the size of the hypothalamus in the cadavers of gay men and the inner ears of lesbians.").

However, the subject is still being discussed enough in reparative therapy debate to warrant a brief mention here.

While scientific evidence can be marshaled to support both the argument that sexual orientation is a result of nature and the argument that sexual orientation is a result of nurture,[127] reparative therapy proponents make the claim that sexual orientation is a product of nurture and therefore changeable.[128] In response to this charge, some gay rights advocates, perhaps recognizing that the nature versus nurture debate is both futile and unhelpful to their cause, do not rely on the proposition that sexual orientation is biological, but instead attack reparative therapy itself.[129] In the midst of this debate, no professional association asserts that sexual orientation is an innate trait, suggesting a growing recognition of the distinction's moral irrelevance.

Hicks rightly notes that whether sexual orientation is a product of nature or nurture is not relevant to whether reparative therapy is harmful.[130] Therefore, psychological testimony that sexual orientation is innate would be rejected due to its clear lack of general acceptance, and psychological testimony that sexual orientation is changeable by the environment is irrelevant as to whether reparative therapy is harmful. Hicks properly frames the issue in terms of whether reparative therapy is harmful when she says that "regardless of the 'correct' answer, 'reparative' therapy is psychologically damaging and should not be administered on gay, lesbian, bisexual and transgender people, and

---

127. Hicks, *supra* note 5, at 511-13 (reviewing scientific data); *see also* A. Dean Byrd, Nat'l Ass'n for Research & Therapy, The Innate-Immutable Argument Finds No Basis in Science, http://www.narth.com/docs/innate.html (last visited Apr. 1, 2005) ("[A]lthough the issue is enormously complex and simply cannot be reduced to a matter of nature vs. nature [sic]—the answer to that debate is probably 'yes'—it is likely that homosexual attraction, like many other strong attractions, includes both biological and environmental influences."); Cohan, *supra* note 12, at 73 ("The question over whether gay, lesbian, bisexual, and transgender people are born into their sexual orientation remains inconclusive. I believe that sexual orientation is at least partially determined genetically.").

128. *See, e.g.*, Traditional Values Coalition, Sexual Orientation: Fixed or Changeable?, http://traditionalvalues.org/urban/seven.php (last visited Apr. 1, 2005). A. Dean Byrd & Stony Olsen, *Homosexuality: Innate and Immutable?*, 14 REGENT U. L. REV. 383 (2001-2002), is a good example of a recent attempt by reparative therapy proponents to try to focus the debate on nature versus nurture.

129. *See, e.g.*, Press Release, Human Rights Campaign, American Psychiatric Association Bolsters Condemnation of Reparative Therapy To Change Gays (Dec. 11, 1998), http://www.hrc.org/Content/ContentGroups/News_Releases/19981/AMERICAN_PSYCHIATRIC_ASSOCIATION_BOLSTERS_CONDEMNATION_OF_REPARATIVE_THERAPY__TO_CHANGE__GAYS.htm (responding to charge that homosexuality is changeable by highlighting abuses of reparative therapy).

130. Hicks, *supra* note 5, at 512-13.

especially not on children."[131] As the remainder of this Part shows, Hicks's data may support her assertion in the court of public opinion, but it would not be able to support her assertion in the courtroom.

B. The Strengthening of Internalized Homophobia

Another attack mounted against reparative therapy is that it strengthens a child's internalized homophobia. The idea is that a child naturally adopts homophobic attitudes from the society around her (societal homophobia), and when the child becomes aware of her own homosexual attractions, the child then turns her preexisting homophobic attitudes toward herself, resulting in shame, guilt, and self-hatred.[132] These resulting feelings are known as "internalized homophobia." Given this backdrop, the introduction of reparative therapy into the child's life further reinforces the child's internalized homophobia. Therefore, opponents of reparative therapy argue, a parent who subjects her child to reparative therapy subjects her child to increased feelings of shame, guilt, and self-hatred, resulting in emotional child abuse. Hicks alludes to this process when she states, "Parents of gay, lesbian, bisexual, and transgender children should also be aware of hardships that their children face as a result of societal homophobia."[133]

This argument initially appears to be admissible, because the idea of internalized homophobia as a psychologically harmful phenomenon would likely be considered "generally accepted" among licensed psychologists. One study endorsed by the American Psychological Association (APA) showed that "gay men scoring high on a measure of internalized homophobia were significantly more likely than less homophobic gay men to experience sexual dysfunction and relationship instability, and to blame themselves for anti-gay victimization."[134] Another professional argues that clinicians "who offer or even consider

131. Hicks, *supra* note 5, at 513.

132. *See* Cohan, *supra* note 12, at 72-73 ("Many gay and lesbian individuals who are raised in a society like ours that disapproves of homosexuality will internalize those negative attitudes and values. Every time such a person feels sexual desire for someone of the same sex, he will experience shame, guilt and self-hatred without necessarily understanding why, because these feelings often operate on an unconscious level.").

133. Hicks, *supra* note 5, at 524-25; *see also id.* at 524 n.100.

134. AM. PSYCHOLOGICAL ASS'N, GUIDELINES FOR PSYCHOTHERAPY WITH LESBIAN, GAY & BISEXUAL CLIENTS (2000), *available at* http://www.apa.org/pi/lgbc/guidelines.html (citing Ilan Meyer & Laura Dean, *Internalized Homophobia, Intimacy, and Sexual Behavior Among Gay and Bisexual Men*, *in* 4 PSYCHOLOGICAL PERSPECTIVES ON LESBIAN AND GAY ISSUES: STIGMA AND SEXUAL ORIENTATION: UNDERSTANDING PREJUDICE AGAINST LESBIANS, GAY MEN, AND BISEXUALS 160 (Gregory Herek ed., 1998)).

conversion therapy for their clients are ignoring the sociopolitical context that perpetuates both external and internal homophobia."[135] Although it is somewhat problematic that there is no "official" definition of homophobia,[136] there is also no "official" definition of racism, which is generally accepted to be psychologically harmful.[137] Moreover, even if reparative therapists do not agree with the conceptual validity of homophobia, "general acceptance" does not require unanimity.[138]

However, there are two interrelated problems with the admissibility of these data. First, the testimony would have to be extremely careful not to link homophobia with reparative therapy in a way that may be more prejudicial than probative, especially given the fact that the context behind reparative therapy is usually a type of conservative Christianity. The obvious religious context might make a judge more sensitive as to whether testimony concerning homophobia is more prejudicial than probative. In *Valentin v. New York City*,[139] a case involving a sexual harassment claim against a police department, the plaintiff sought to submit a qualified expert's testimony concerning "institutionalized sexism and homophobia" within "police culture."[140] The court rejected such testimony. Because the expert had no knowledge of the plaintiff's particular work environment and was "not present when the specific incidents allegedly occurred," such evidence would have been more prejudicial than probative.[141] Similarly, in *State v. Haynes*,[142] expert testimony concerning whether the

135. Kathleen J. Bieschke et al., *Programmatic Research on the Treatment of Lesbian, Gay, and Bisexual Clients: The Past, the Present, and the Cause for the Future, in* HANDBOOK OF COUNSELING AND PSYCHOTHERAPY WITH LESBIAN, GAY, AND BISEXUAL CLIENTS 309, 313 (Ruperto M. Perez et al. eds., 2000).

136. Lester W. Wright, Jr. et al., *Development and Validation of the Homophobia Scale*, 21 J. PSYCHOPATHOLOGY & BEHAV. ASSESSMENT 337, 338-46 (1999), *available at* http://www.springerlink.com/media/f83ebu4qmh1jvl9twc2l/contributions/t/7/3/1/t73103hp41744507.pdf (noting the discrepancies between different studies of homophobia and that "there is no universally accepted definition of homophobia").

137. *See* Am. Psychological Ass'n, Resolution Against Racism and in Support of the Goals of the 2001 UN World Conference Against Racism, Racial Discrimination, Xenophobia, and Related Intolerance (June 10, 2001), *available at* http://www.apa.org/pi/racismresolution.html.

138. Barmeyer et al. v. Mont. Power Co., 657 P.2d 594, 598 (Mont. 1983) ("Absolute certainty of result or unanimity of scientific opinion is not required for admissibility.") (citing United States v. Baller, 519 F.2d 463, 466 (4th Cir. 1975)); *see also* Kaelbel Wholesale, Inc. v. Soderstrom, 785 So. 2d 539, 546 (Fla. Dist. Ct. App. 2001) (finding that clear majority sufficient for general acceptance).

139. No. 94 CV 3911(CLP), 1997 WL 33323099 (E.D.N.Y. Sept. 9, 1997).

140. *Id.* at *19.

141. *Id.*

142. No. 4310, 1988 WL 99189 (Ohio Ct. App. Sept. 21, 1988).

defendant's killing was motivated by homophobia was rejected partly because "the prejudicial impact outweighed probative value, as it tended to 'sensationalize' the facts and issues."[143]

The second problem is that while homosexual attraction may not be common to the experience of mankind, homophobia arguably is. In seeking to bar admission, the parent may object that just as a professional psychologist would not be required to testify about the existence and harmful effects of racism due to widespread knowledge and experience of the phenomenon, a professional psychologist would not be required to testify about the existence and harmful effects of homophobia. For instance, in *Haynes*, the expert testimony sought to analyze whether the defendant's act of killing was a "homophobic murder," defined by the expert as a murder resulting from "a panic that ensued after an unwanted homosexual encounter."[144] This expert testimony was rejected, because it was "well within the understanding of the average juror."[145] Similarly, if the prosecution wished to advance its basic argument that when a reparative therapist tells a child that the child should overcome deep seated attractions, the child will become emotionally harmed to a degree constituting child abuse, it would not require a psychologist to explain the term "internalized homophobia" to further advance its argument. Bereft of the backing of expert testimony, the strength of that argument must then depend on other evidence, the possibilities of which will be explored *infra*. This section simply shows that the prosecution cannot rely on the "internalized homophobia" argument that Hicks employs.[146]

C. Preexisting Psychological Problems in Homosexual Youth

One of the more scientific arguments against reparative therapy is an extension of the homophobia argument. It is undisputed that homosexual, bisexual and transgender youths suffer from depression at significantly higher rates than heterosexual youths. For example, one study cited by Hicks noted that 28.1% of gay males and only 4.2% of heterosexual males attempt suicide.[147] She also notes that homosexual youths are more likely to be victims of assault, which

---

143. *Id.* at *4.

144. *Id.*

145. *Id.*

146. However, there is nothing preventing the prosecution from employing the phrase "internalized homophobia" in its opening or closing argument, as long as it does not claim expert backing.

147. Hicks, *supra* note 5, at 518 n.57 (citing Gary Remafedi, *The Relationship Between Suicide Risk and Sexual Orientation: Results of a Population-Based Study*, 88 AM. J. PUB. HEALTH 57, 57-60 (1998)).

has an undeniable psychological impact.[148] The APA also recognizes that "[r]esearch has shown that gay men are at risk for mental health problems and emotional distress."[149]

These facts nonetheless have no bearing on whether reparative therapy is harmful, because preexisting depression only speaks to the state of the youth *prior* to the reparative therapy. Even reparative therapy proponents do not dispute these facts. In fact, they often rely on these facts to argue *for* reparative therapy,[150] claiming that reparative therapy is required to bring such youths out of depression. The debate between proponents and opponents of reparative therapy does not center around whether homosexuals are in fact more depressed but on the *causes of the depression*. Therefore, psychological testimony regarding the increased likelihood of depression in homosexual youths would be rejected. Whether reparative therapy *actually* increases or decreases depression in homosexual youths is a separate inquiry that Section G of this Part discusses.

D. Anecdotal Evidence

Anecdotal evidence is a powerful tool in the marketplace of ideas, and as such it is heavily relied upon by both sides of the reparative therapy debate.[151] For example, Yoshino introduces his discussion of reparative therapy with a powerful and horrific anecdote from an individual undergoing electroshock therapy.[152] More significant to the child abuse analysis, the vast majority of Hicks's evidence for the harmfulness of reparative therapy includes both comical

---

148. *Id.* at 518 n.56 ("Fifty-nine percent of gay men and twenty-one percent of lesbians report victimization in high school, and fifty percent and twelve percent, respectively, report victimization in junior high school.") (quoting Anthony R. D'Augelli, *Lesbian, Gay, and Bisexual Development During Adolescence and Young Adulthood*, *in* TEXTBOOK OF HOMOSEXUALITY AND MENTAL HEALTH 279 (1996)); *see also id.* at 518 n.58 ("[L]esbian, gay, and bisexual youths are at risk for psychological problems.") (quoting Anthony R. D'Augelli & Scott Hershberger, *Lesbian, Gay, and Bisexual Youth in Community Settings: Personal Challenges and Mental Health Problems*, 21 AM. J. COMMUNITY PSYCHOL., 421, 443-44 (1993)).

149. AM. PSYCHOLOGICAL ASS'N, *supra* note 134.

150. Dale O'Leary, Nat'l Ass'n for Research & Therapy, Recent Studies on Homosexuality and Mental Health, http://www.narth.com/docs/recent.html (last visited Nov. 18, 2005) ("[S]ome social conservatives will attribute the findings [that homosexuals are more likely to be depressed] to the inevitable consequences of the choice of a homosexual lifestyle.").

151. For testimonials from people with self-reported negative experiences with conversion therapy, see HUMAN RIGHTS CAMPAIGN, *supra* note 49, at 7. For testimonials from people with self-reported positive experiences with conversion therapy, see Nat'l Ass'n for Research & Therapy, Interviews/Testimonials, http://www.narth.com/menus/interviews.html (last visited Nov. 18, 2005).

152. Yoshino, *supra* note 31, at 784-85; *see also* Cruz, *supra* note 19, at 1352-53 (citing anecdotes).

and horrific anecdotes,[153] such as exorcisms,[154] physical restraints,[155] and kidnappings.[156]

The use of anecdotes is not only practically persuasive, but also plays a significant role in the field of psychology. Because psychological theories do not have the same kind of validity as other scientific theories,[157] psychologists "can at best offer only 'anecdata': information obtained through experience in dealing with psychological problems, reading about case studies, and extrapolation from the theoretical speculations of others."[158]

The 1999 Supreme Court case of *Kumho Tire Co. v. Carmichael*[159] suggests that anecdotal testimony is permissible, provided it is based on the psychologist's own experiences: "An expert, whether basing testimony upon professional studies or personal experience, employs in the courtroom the same level of intellectual rigor that characterizes the practice of an expert in the relevant field."[160] Therefore, a psychologist might testify to her experiences with her clients who have undergone reparative therapy and are seeking to recover from its harmful effects. For instance, one psychologist documents his own experiences with post-reparative therapy clients, noting "chronic depression, low self-esteem, difficulty sustaining relationships, and sexual dysfunction."[161] On the other hand, testimony recounting the stories of people a psychologist has never met who have undergone reparative therapy will not be admitted, as it would constitute a mere review of the literature that any lay person could perform.

Sometimes, a state court may permit personal experience testimony from a psychologist by considering such testimony not "scientific," thereby circumventing the *Daubert* or *Frye* test. "Testimony that is based solely on the expert's own clinical observation and experience is not subject to *Frye*."[162]

---

153. Discussing all of Hicks's anecdotal evidence would be impractical. Anecdotes are found at Hicks, *supra* note 5, at 515-20 nn.39-53 & 60, 524-25 nn.100-01, 103-04, 106, & 108.

154. *Id.* at 525 n.106.

155. *Id.* at 515 n.40, 516 n.46.

156. *Id.* at 516 n.45; *see also* Cohan, *supra* note 12, at 82 n.66.

157. Christopher Slobogin, *Doubts About* Daubert*: Psychiatric Anecdata as a Case Study*, 57 WASH. & LEE L. REV. 919, 921 (2000) ("[T]he validity of psychiatric opinion is hard to gauge.").

158. *Id.* at 922.

159. 526 U.S. 137 (1999).

160. *Id.* at 152.

161. Haldeman, *supra* note 46, at 261. Haldeman also notes that the severity of the psychological harms resulting from reparative therapy depends on the therapy involved and the person's natural resilience.

162. James P. Flannery et al., Frye, Daubert, Donaldson, *and Junk Science: The Admissibility of Novel Scientific Evidence in Illinois*, 18 CBA REC. 30, 36 (2004).

However, this is only the case in states that have not imported the *Kumho* holding that eliminates the need to distinguish between "scientific" and "technical" or "other specialized" knowledge. And such experience-based testimony may ultimately be subject to the same *Daubert*-esque factors in order to test for reliability.[163]

While anecdotal testimony may be permitted if the anecdotes derive from the psychologist's own experiences, it is important to remember that emotionally powerful anecdotes can come from both sides. In fact, it is precisely because of this result, with one side pitting its poster children against the other's, that a judge may bar such anecdotal testimonies altogether for their inflammatory, and therefore prejudicial, effect. Even if the judge does not bar such testimony, anecdotal evidence should not be the primary legal weapon in the arsenal of reparative therapy opponents because it is easily countered.

E. Reparative Therapy Does Not Work

A major locus of the debate over reparative therapy is whether it can truly change homosexuals into heterosexuals, and this argument often complements the argument that reparative therapy is a manifestation of societal homophobia. This has taken on increased significance as the debate has shifted away from whether sexual orientation is a product of nature or nurture. Citing the APA, Hicks notes that "scientific evidence does not show that 'reparative' or conversion therapy works."[164] Quoting the American Psychiatric Association, Gans also notes that "[t]here is no evidence that any treatment can change a homosexual person's deep-seated sexual feelings for others of the same sex."[165] There is copious evidence demonstrating the low success rate of reparative therapy,[166] and evidence of "success" is often discredited, legitimately or not, by

163. Logerquist v. McVey, 1 P.3d 113, 131 (Ariz. 2000) ("[S]ome of *Daubert*'s questions can help to evaluate the reliability even of experience-based testimony. In certain cases, it will be appropriate for the trial judge to ask, for example, how often an engineering expert's experience-based methodology has produced erroneous results, or whether such a method is generally accepted in the relevant engineering community. Likewise, it will at times be useful to ask even of a witness whose expertise is based purely on experience, say, a perfume tester able to distinguish among 140 odors at a sniff, whether his preparation is of a kind that others in the field would recognize as acceptable.") (citing Kumho Tire Co. v. Carmichael, 526 U.S. 137, 151 (1999)).

164. Hicks, *supra* note 5, at 513-14, 514 n.36.

165. Gans, *supra* note 18, at 227; *see also id.* at 227 nn.52-55.

166. Hicks, *supra* note 5, at 518 nn.59-60 (citing studies indicating three percent success rate and ten percent success rate). Hicks also suggests that the leaders of the reparative therapy movement are predominantly heterosexual. *Id.*; *see also* Cruz, *supra* note 19, at 1378-81 (citing evidence of failures).

claiming that "'success' stories are really stories of bisexuals who are responding to heteroerotic inclinations that were already present."[167] Proponents of reparative therapy, in response, tend to rely on the flexible nature of sexual orientation or the *possibility* of change,[168] presumably because it is easier to hedge by arguing that change is "possible" rather than to declare outright that reparative therapy "works." Arguing that change is "possible" needs only be supported by one "success" story, whereas arguing that reparative therapy "works" requires a greater burden that proponents may know they cannot meet. To support the claim that change is possible, proponents often cite case studies of self-reporting individuals who have "changed."[169]

The latest manifestation of this debate revolved around a study published in 2003 by Dr. Robert Spitzer, who was instrumental in eliminating homosexuality from the American Psychiatric Association's recognized list of mental diseases in 1973. Dr. Spitzer studied 200 cases of individuals who had undergone reparative therapy and reported some measure of change.[170] In response, professional associations remained relatively silent while individual psychologists have challenged the methodology and reliability of the study.[171] Some legal scholars

---

167. Mark A. Yarhouse & Warren Throckmorton, *Ethical Issues in Attempts To Ban Reorientation Therapies*, 39 PSYCHOTHERAPY: THEORY/RES./PRAC./TRAINING 66, 69 (2002).

168. *See* Benjamin Kaufman, *Why NARTH? The American Psychiatric Association's Destructive and Blind Pursuit of Political Correctness*, 14 REGENT U. L. REV. 423, 430 n.35 (2001-2002) (citing studies of reparative therapy's "success"). It is subtle but significant that the title of Kaufman's section documenting such studies is "Change is Possible." *Id.*

169. *See* New Direction for Life Ministries, Summary of Evidence Found by the Homosexuality and the Possibility of Change Project, http://www.newdirection.ca/research/evidence.htm (last visited Nov. 3, 2005) (containing links to studies showing a total of 86 persons with a change in sexual orientation behavior, 287 persons with partial sexual orientation shift, and 45 persons with total sexual orientation shift). *But see* Yarhouse & Throckmorton, *supra* note 167, at 73 ("Critics are right to point out that many studies cited to support the effectiveness of professional change therapies and religion-based ministries suffer from poor methodologies, including small sample sizes, lack of clear definitions and consistency in measures of change or success, and use of therapist report and self-report of change. However, poor methodologies do not disprove success; what is needed are prospective, longitudinal studies of those entering such change programs and greater consistency as to what constitutes 'success.'").

170. Robert L. Spitzer, *Can Some Gay Men and Lesbians Change Their Sexual Orientation? 200 Participants Reporting a Change from Homosexual to Heterosexual Orientation*, 32 ARCHIVES SEXUAL BEHAV. 403 (2003).

171. *See* B.A. Robinson, Ont. Consultants on Religious Tolerance, Analysis of Dr. Spitzer's Study of Reparative Therapy, http://www.religioustolerance.org/hom_spit.htm (Feb. 16, 2002). The APA cancelled a scheduled debate on the topic. *See id.* ("The doctors who were to debate on the topic decided there was not enough scientific information to have a proper debate. They felt that any debate would turn into a political debate and not a true scientific debate. While there is

rightly saw this as another round in the irrelevant nature/nurture debate.[172] In any case, today, it is "generally accepted" in established professional associations that reparative therapy does not work. Even reparative therapy proponents do not make the bald assertion that reparative therapy indeed does "work." Instead they merely assert that change is *possible*.[173]

Nonetheless, psychological testimony regarding the ineffectiveness of reparative therapy in changing homosexuals into heterosexuals would be rejected on the basis of irrelevance, because arguing that something does not work does not mean such a thing is *harmful*—sugar pills may not cure a patient of the common cold, but they do not *harm* the patient.[174] Similarly, whether reparative therapy works or not is irrelevant to whether reparative therapy is *harmful*. For instance, many therapies in general have questionable "success" rates when it comes to curing or alleviating widely recognized mental ailments. As Christopher Slobogin notes, "Even many symptoms—such as whether a person is 'depressed,' 'anxious,' or suffering from 'low self-esteem'—are unverifiable in the same way a physical fact is because the terms themselves are so amorphous and subjective."[175] The questionable success rates of such non-controversial therapies do not therefore show that such therapies for alleviating depression and other disorders are harmful.[176]

---

information on reorientation therapy, there have been no controlled research studies.").

172. Knauer, *supra* note 126, at 2-10 (discussing theoretical and political implications of Spitzer's study).

173. Nat'l Ass'n for Research & Therapy of Homosexuality, Gay-to-Straight Research Published in APA Journal, http://www.narth.com/docs/throckarticle.html (last visted Nov. 18, 2005) (interviewing Warren Throckmorton, who stated, "[S]exual orientation, once thought to be an unchanging trait, is actually quite flexible for some people"). This assertion is very different from the claim that reparative therapy works.

174. The analogy is not perfect, since unlike the common cold, homosexuality is not considered an illness. A sugar pill may harm the patient by precluding more effective cures, but reparative therapy would not harm the patient because there is no illness, so it is not precluding any other effective treatments.

175. Slobogin, *supra* note 157, at 921.

176. It is this author's opinion that the debate over whether reparative therapy "works" is as irrelevant as the debate over whether sexual orientation arises from nature or nurture, because the inquiry is just as complex. The inquiry itself depends on one's definition of sexual orientation, relies on an unsound distinction between feeling and behavior, and requires breaking into the "black box" that is an individual's sexuality, which is so fluid and socially constructed as to render the categories of sexual orientation meaningless. *See* Warren Throckmorton, *Initial Empirical and Clinical Findings Concerning the Change Process for Ex-Gays*, 33 PROF. PSYCHOL.: RES. & PRAC. 242, 243 (2002) (describing said conceptual problems underlying the attempt to answer the question, "Do ex-gay ministries help people change sexual orientation?").

F. Homosexuality Is Not an Illness

A common refrain among opponents of reparative therapy is the fact that the American Psychiatric Association removed homosexuality from its Diagnostic and Statistical Manual of Mental Disorders (DSM) in 1973, followed by the renunciation of homosexuality as a mental illness by the APA.[177]

The psychological community itself, however, does not grant the DSM much reverence. William Grove, arguing that a listing in the DSM should not constitute "general acceptance," notes the process of creating the DSM: "[S]pecialty subcommittee members were assigned to review aspects of the literature relating to certain categories, using subjective methods . . . . Analyses of existing data sets were sometimes undertaken . . . but [the procedure] is not completely explicit, repeatable, or tied to polling representative samples of scientists."[178] He also adds that the listings in the DSM are "labels, not theories. This is a critical distinction—some experts have misled courts to believe falsely that the existence of a diagnostic label in DSM-IV somehow proves general acceptance of the existence of the described disorder."[179] Slobogin also notes the lack of agreement among psychologists on DSM listings: "[F]ield research indicates that mental health professionals involved in everyday practice may disagree more than half the time even on major diagnostic categories such as schizophrenia and organic brain syndrome."[180] Of the reparative therapy legal scholars, Cruz is the only one who acknowledges the inherent ambiguity in the nature of "mental illness" and who questioned reliance on the DSM.[181]

Nonetheless, the DSM appears to be held in high regard in the courts. When psychological testimonies concern mental diseases that are listed in the DSM, they generally pass the *Daubert* or *Frye* test, despite the fact that many of the DSM listings are falsifiable and should therefore meet a higher standard in *Daubert* jurisdictions, which are supposed to consider falsifiability.[182] Courts that

---

177. Kaufman, *supra* note 168, at 433 ("Those who insist that homosexuality is not an illness point to the APA's 1973 decision to remove homosexuality from its DSM."); *see, e.g.*, Gans, *supra* note 18, at 221-22; Hicks, *supra* note 5, at 518 n.60; Yoshino, *supra* note 31, at 798-99. Tam Tran provides a good account of the history of the DSM. Tam B. Tran, *Using DSM-IV To Diagnose Mental Illness in Asian Americans*, 10 J. CONTEMP. LEGAL ISSUES 335, 336-38 (1999). Her review specifically focuses on the DSM-IV, where "IV" signifies that the edition is the fourth edition.

178. Grove, *supra* note 90, at 230.

179. *Id.*

180. Slobogin, *supra* note 157, at 920.

181. *See* Cruz, *supra* note 19, at 1313-34 (discussing the DSM and the sociological context of "mental illness").

182. *See* Smithburn, *supra* note 89, at 15 n.101 (noting that *Daubert*'s falsifiability requirement may call the DSM into question) (citing James T. Richardson, *Dramatic Changes in American*

permit psychological testimony regarding "compulsive gambling disorder," a DSM-IV listed disease, do not permit testimony on the "pathological gambling lifestyle" because it strays beyond the boundaries of the DSM-IV definition.[183] In criminal defense cases, "whenever a mental disorder is raised as a defense, if it is not listed in the DSM, it is not given much credence."[184] Holding PTSD to be a legitimate disorder, *State v. Alberico* states, "The existence of DSM III-R and its general acceptance in psychology indicate that PTSD has been exposed to objective scientific scrutiny and empirical verification."[185]

Nonetheless, although the practice of reparative therapy certainly traces its roots to the view of homosexuality as a mental illness,[186] it is unclear to what

---

Expert Evidence Law, 2 JUD. REV. 13, 25 (1996)). Just because a court may consider DSM-listed diseases to be bona fide mental diseases, however, does not mean that such testimony may be used for any purpose conceivably related to mental disease. *See, e.g.*, United States v. DiDomenico, 985 F.2d 1159, 1161 (2d Cir. 1993) ("The fact that certain personality traits or conduct may be identified, categorized or characterized by the psychiatric profession by, for example, inclusion in the Diagnostic and Statistical Manual of Mental Disorders (3rd Edition 1980) ('DSM-III') promulgated by the American Psychiatric Association, does not necessarily make the traits or conduct a 'mental disease' or 'mental disorder' that can be the basis of the defense of insanity."); Harris v. Pulley, 885 F.2d 1354, 1383 (9th Cir. 1988) ("[W]hat is classified as a mental disorder by the American Psychiatric Association is not necessarily a condition that a state is constitutionally required to take into account in assessing punishment.").

183. Fradella, *supra* note 79, at 432 (citing United States v. Scholl, 166 F.3d 964 (9th Cir. 1999)); *see also* Comer v. Stewart, 230 F. Supp. 2d 1016, 1052 (D. Ariz. 2002) (excluding testimony based on *Daubert*, because it did not properly apply the standards of diagnosing PTSD in accordance with the definition of DSM-IV); United States v. Harris, No. S192 Cr.455(CSH), 1994 WL 683429, at *4 (S.D.N.Y. Dec. 6, 1994) ("*The New York Times*, in its April 19, 1994 issue, characterized The Diagnostic and Statistical Manual of Mental Disorders hereinafter ("DSM") as 'the psychiatric profession's diagnostic Bible.'"); Cassell v. Lancaster Mennonite Conference, 834 A.2d 1185, 1190 (Pa. Super. Ct. 2003) (reversing trial court decision to exclude expert testimony because expert testimony's reliance on DSM meant it was "generally accepted" and passes *Frye* test). *But see* Mancuso v. Consol. Edison Co. of N.Y., 967 F. Supp. 1437, 1456 (S.D.N.Y. 1997) ("ConEd has cited no cases in which a qualified psychologist was excluded from testifying because she did not follow the DSM-IV.").

184. David Barton, *A Death-Struggle Between Two Civilizations*, 13 REGENT U. L. REV. 297, 342 (2000-2001).

185. 861 P.2d 192, 208 (N.M. 1993). Additional language from the opinion indicates the court's deference toward the DSM. *See id.* ("We hold that PTSD testimony is grounded in valid scientific principle. DSM III-R is specialized literature that specifically catalogues the symptoms of mental disorders and prescribes the method by which the psychological evaluation should take place. DSM III-R, according to the State's experts, is widely used in courtrooms, not only for issues of sex abuse, but for issues concerning sanity and competency as well. PTSD is generally accepted by psychologists and psychiatrists as a valid technique for evaluating patients with mental disorders.").

186. Gans, *supra* note 18, at 223 ("The current practice of conversion therapy attests to the

extent reparative therapy as practiced today does in fact rely on that view. While Hicks asserts that "'reparative' therapists and 'ex-gay' ministries continue to claim that homosexuality is a mental illness that can, and should, be changed,"[187] Yoshino notes that organizations supporting reparative therapy are "relatively insulated from the depathologization of homosexuality, as they are less reliant on a literal disease model to justify their conversion practices."[188] Mark Yarhouse and Warren Throckmorton, proponents of reparative therapy, cite studies showing that not all reparative therapies rely on a pathology-based model of treatment.[189] Another reparative therapist states that "those who embrace reparative therapy as an option would not necessarily need to believe that those who call themselves homosexuals demonstrate more pathology than those who are heterosexuals."[190]

While it is relatively clear that for cases in which the existence of a mental illness is relevant, expert testimony relying on the DSM would probably be admitted, it is unclear whether expert testimony concerning homosexuality's *absence* from the DSM would be admitted in the reparative therapy scenario. The few cases considering the fact that homosexuality is not in the DSM include cases concerning sentencing that improperly factors the defendant's homosexuality as a mental illness,[191] an employment discrimination case,[192] and a case involving the right to be a foster parent.[193] Nonetheless, it is likely that the exclusion of homosexuality from the DSM would not be relevant to whether reparative therapy is harmful, because cases that rely on the DSM do so in order

---

antiquated belief that homosexuality is a disease.").

187. Hicks, *supra* note 5, at 518.

188. Yoshino, *supra* note 31, at 801.

189. Yarhouse & Throckmorton, *supra* note 167, at 67. Another example is stress inoculation training for coping with stressors—this is a "well-established treatment directed at a population other than a DSM diagnostic category." William C. Sanderson, *The Importance of Empirically Supported Psychological Interventions in the New Healthcare Environment, in* 15 INNOVATIONS IN CLINICAL PRACTICE: A SOURCE BOOK 387, 396 (Leon VandeCreek et al. eds., 1997).

190. Charlotte Rosenak, Nat'l Ass'n for Research & Therapy of Homosexuality, Some Psychologists Say Reparative Therapy Is Unethical Yet Modern Methods Are Healing and Client-Centered (last visited Nov. 11, 2005), http://www.leaderu.com/orgs/narth/unethical.html.

191. *See* United States v. Donaghe, 50 F.3d 608, 613 (9th Cir. 1994) ("[T]he district court must not base its departure on Donaghe's motive for committing the crime [upon] Donaghe's diagnosis as a homosexual."); Commonwealth v. Bey, 841 A.2d 562, 565 (Pa. Super. Ct. 2004) (rejecting defendant's homosexuality as basis for sentencing).

192. Rowland v. Mad River Local Sch. Dist., 730 F.2d 444, 454-55 (6th Cir. 1984) (Edwards, J., dissenting).

193. Howard v. Child Welfare Agency Review Bd., No. CV-1999-9881, 2004 WL 3154530, at *6 (Ark. Cir. Ct. Dec. 29, 2004).

to determine whether something is in fact a mental illness. Because reparative therapy does not depend on whether homosexuality is a mental illness, such evidence would be irrelevant.

On the other hand, if a psychologist testifies that reparative therapy causes or exacerbates disorders *that are* listed in the DSM-IV, such as "generalized anxiety disorder" (GAD) (Code 300.02)[194] or "major depressive disorder" (Code 296)[195] (commonly known as and hereinafter depression), such testimony would certainly be relevant to the child abuse claim. Whether such testimony would be reliable is a different question and will be addressed next.

G. Reparative Therapy Causes or Exacerbates GAD or Depression

Herein lies the crux of the evidentiary problem. It is generally accepted that reparative therapy does not work, that homosexuality is not a mental illness, and that homosexuals are much more prone to have mental illnesses than heterosexuals, yet there appears to be scant evidence, apart from anecdotes,[196] that reparative therapy is harmful. One psychologist in 1994 noted the scattered nature of the reparative therapy debate and its resulting lack of focus on this specific harm issue: "The research question, 'What is being accomplished by conversion treatments?' may well be replaced by, 'What harm has been done in the name of sexual reorientation?' At present, no data are extant."[197] Apparently, this lack of data persisted for another decade. In 1997, the APA stated, "Data that *conclusively* indicate harmfulness of conversion therapy do not exist."[198] Cruz notes in 1999 that despite the vast number of accounts of people who have suffered real harms through reparative therapy, there was a "lack of systematic

194. AllPsych Online, Generalized Anxiety Disorder (GAD), http://allpsych.com/disorders/anxiety/generalizedanxiety.html (last visited Apr. 2, 2005) ("GAD is evidenced by general feelings of anxiety such as mild heart palpitations, dizziness, and excessive worry. The symptoms are difficult to control for the individual and are not related to a specific event (such as in PSTD).").

195. AllPsych Online, Major Depressive Disorder (Unipolar Depression), http://allpsych.com/disorders/mood/majordepression.html (last visited Apr. 2, 2005) ("Symptoms of depression include the following: depressed mood (such as feelings of sadness or emptiness), reduced interest in activities that used to be enjoyed, sleep disturbances (either not being able to sleep well or sleeping too much), loss of energy or a significant reduction in energy level, difficulty concentrating, holding a conversation, paying attention, or making decisions that used to be made fairly easily, [and] suicidal thoughts or intentions.").

196. Haldeman notes that "subjects who have undergone failed attempts at conversion therapy often report increased guilt, anxiety, and low self-esteem," but his statement is based on anecdotal evidence. Yarhouse & Throckmorton, *supra* note 167, at 70.

197. *Id.* (citing Douglas C. Haldeman, *The Practice and Ethics of Sexual Orientation Conversion Therapy*, 62 J. CONSULTING & CLINICAL PSYCHOL. 221, 221-27 (1994)).

198. Cruz, *supra* note 19, at 1351.

information about the frequency of these harms." In 2002, Yarhouse and Throckmorton assert that "[a]ccording to the empirical evidence to date, the objective of sexual reorientation has not been demonstrated to be *intrinsically* harmful."[199] Douglas Haldeman, who is not a proponent of reparative therapy, notes the methodological unreliability of anecdotal evidence: "The reports of harm done by conversion treatments, however, are subject to the same methodological limitations as those affecting studies purporting to show a positive treatment outcome."[200] As recently as February of 2003, B.A. Robinson, an opponent of reparative therapy, expressed his frustration at this lack of evidence.[201]

The only published study close to empirically linking reparative therapy with psychological harm was completed by Ariel Shidlo and Michael Schroeder in 2002.[202] Similar to Spitzer's study, they interviewed 202 people who had undergone reparative therapy. Although Cruz looked forward to this study in 1999 as the first scientific basis for the real psychological harms of reparative therapy,[203] the primary focus of the study was, unfortunately, yet again on the efficacy of reparative therapy.[204] The psychologists did ask the participants for whom reparative therapy failed what types of psychological harms they suffered, and many reported increased distress, depression, or spiritual harms.[205] However,

---

199. Yarhouse & Throckmorton, *supra* note 167, at 71.

200. Haldeman, *supra* note 46, at 261.

201. Robinson, *supra* note 171 ("I am personally enraged at the irresponsibility of the large professional mental health organizations. . . . Today, there are at least 1,000 therapists conducting reparative therapy. . . . Yet no statistically valid, peer-reviewed study in this field has ever been attempted. Mental health professionals know that these forms of therapy are dangerous and can induce suicide attempts. Yet the therapy's safety and efficacy can only be guessed at.").

202. *See* Ariel Shidlo & Michael Schroeder, *Changing Sexual Orientation: A Consumers' Report*, 33 PROF. PSYCHOL.: RES. & PRAC. 249 (2002); *see also* Warren Throckmorton, May I Ask Your Evaluation of the Shidlo and Schroeder Study, Which Appears To Have Arrived at Conclusions So Directly Opposed to Spitzer's Study?, http://www.drthrockmorton.com/article.asp?id=11 (last visited Dec. 2, 2005) ("Shidlo and Schroeder's study is the only peer reviewed study I know about that systematically sought to examine those who say they tried reorientation counseling but were not happy with the results.").

203. Cruz, *supra* note 19, at 1354 ("All of these harms are real and serious. However, there is a significant lack of information about their incidence. Psychologists Ariel Shidlo and Michael Schroeder are currently conducting research with people who have attempted sexual orientation conversion, but have yet to publish any results.").

204. *See* Shidlo & Schroeder, *supra* note 202, at 251 (noting that the title of the project had to be changed from "Homophobic Therapies: Documenting the Damage" to "Changing Sexual Orientation: Does Counseling Work?").

205. *See* Shidlo & Schroeder, *supra* note 202, at 254 (noting that their review of psychological harms were qualitative rather than quantitative); *id.* at 254-56 (recounting anecdotes of harm); *see*

some also reported psychological benefits such as an increased sense of hope, coping strategies, and social skill building.[206]

Unsurprisingly, reparative therapy opponents tout the Shidlo/Schroeder study and denigrate the methodology of the Spitzer study,[207] while reparative therapy proponents do the exact opposite.[208] In any case, both studies rely on non-random samples and self-reporting.[209] As Haldeman notes in 2002, "It is nearly impossible to obtain a random sample of research participants who have been treated for their sexual orientation, and it is equally as difficult to assess outcomes in a way that does not contaminate the scientific process with social bias."[210] To overcome this barrier and produce more meaningful studies, Haldeman encourages a "systematic study of motivations of those who seek to change sexual orientation . . . ."[211]

In sum, the idea that reparative therapy is psychologically harmful is not generally accepted in the psychological community, and psychologists are only starting to construct scientific studies to capture this principle. Although it is speculative at this stage, what will likely emerge are studies that attempt to catalogue various forms of reparative therapy, such as aversive treatments versus

---

also New Direction for Life Ministries, Changing Sexual Orientation: A Consumers' Report, http://www.newdirection.ca/research/shidlo.htm (last visited Apr. 5, 2005) ("The authors discussed the various types of harm that the participants reported perceiving, including psychological harm (ie., depression, self-esteem issues, intrusive imagery), social and interpersonal harm (ie., relationships with family of origin, alienation, loss of social supports) and spiritual harm."); Attempts To Change Sexual Orientation, http://psychology.ucdavis.edu/rainbow/html/facts_changing.html (last visited Apr. 5, 2005) ("Drs. Shidlo and Schroeder also reported that many respondents were harmed by the attempt to change.").

206. *See* Shidlo & Schroeder, *supra* note 202, at 256-57; *see also* New Direction for Life Ministries, *supra* note 205.

207. *See, e.g.*, *Attempts To Change Sexual Orientation*, *supra* note 205 (devoting half of the webpage to criticizing Spitzer's study while spending one paragraph discussing the Shidlo/Schroeder study); Barbara Dozetos, Gay.com, Researchers Clash Over 'Ex-Gays,' May 9, 2001, http://www.gay.com/content/tools/print.html?coll=news_articles&sernum=2001/05/09/2&navpath=channels/news (quoting numerous organizations and individuals criticizing Spitzer's study, while noting that "[t]he methodology behind Shidlo and Schroeder's study differed significantly from Spitzer's").

208. The most "balanced" comparison of the studies I could find was at a pro-reparative therapy site, New Direction for Life Ministries. *See* New Direction for Life Ministries, Comparison of the Research of Robert L. Spitzer with the Research of Ariel Shidlo and Michael Schroeder, http://www.newdirection.ca/research/compare.htm (last visited Apr. 10, 2005).

209. *Id.*

210. Haldeman, *supra* note 46, at 261.

211. *Id.* at 262.

psychotherapy,[212] religion-integrated treatments versus secular ones, and treatments that focus on celibacy versus those that focus on increasing attractions for the opposite sex.[213] Such studies will probably also examine such treatments' effects on clients depending on the clients' own motivations, such as whether the client is a minor pressured by her parents[214] or whether the client is religiously motivated.[215]

IV. POSITION STATEMENTS, RESOLUTIONS, AND CODES, OH MY: CONFUSION OVER GUIDELINES IN PSYCHOLOGICAL PRACTICE

In addition to the aforementioned data and arguments raised by opponents of reparative therapy, much of the legal reparative therapy literature also cites various position statements and resolutions (hereinafter written guidelines) of the numerous professional psychological organizations. Many of these written guidelines contain explicit condemnations of reparative therapy. In Hicks's theory, since most professional psychological associations denounce reparative therapy, a "reasonably prudent parent" "would discover that 'reparative' therapy is not accepted in the mainstream medical community."[216] Therefore, a parent who subjects her child to a psychological practice that is denounced by the psychology profession, indicating its harmfulness, would be guilty of child abuse.

Although Hicks's theory, based on New York law, may be read to imply that a parent must *know* that an action is harmful in order to be subject to prosecution for child abuse, it is important to emphasize that this mens rea element is not required for emotional child abuse cases in many states, which simply require the establishment of emotional harm and its causation.[217]

---

212. *See id.* at 261 (noting that in Haldeman's experience, clients who have gone through aversive procedures suffer significantly more than clients who have undergone less intrusive forms of therapy).

213. *See* Martin Koretzky, Nat'l Ass'n for Research & Therapy of Homosexuality, APA Symposium Seeks Common Ground, http://www.narth.com/docs/commonground.html (last visited Apr. 5, 2005) (reporting that Dr. Mark Yarhouse identified a "continuum of service options for clients who experience same-sex attractions, including reorientation/reparative therapy, chastity/celibacy, sexual-identity management, and gay-affirmative therapy").

214. This would be most useful for the type of reparative therapy discussed in Hicks's article.

215. *See* Cruz, *supra* note 19, at 1345-48 (noting difficulty of ascertaining whether a client's proffered religious reason is "truly" individual or derives from desire to conform).

216. Hicks, *supra* note 5, at 524.

217. For a listing of all state statutes' definitions of emotional child abuse as of January 2005, see NAT'L CLEARINGHOUSE ON CHILD ABUSE & NEGLECT INFO., DEFINITIONS OF CHILD ABUSE AND NEGLECT: SUMMARY OF STATE LAWS (2005), http://nccanch.acf.hhs.gov/general/legal/statutes/

The existence of emotional harm requires expert testimony, and the previous Part has demonstrated that several forms of testimony face evidentiary problems. However, if an expert could truthfully testify that it is generally accepted by the psychology profession that reparative therapy is harmful, such testimony would by definition pass the "general acceptance" test of *Daubert* or *Frye*, and it may therefore not even be necessary to have empirical evidence of reparative therapy's harmfulness. Not only would such testimony be admissible, it would be relatively dispositive in proving that reparative therapy is per se emotionally harmful, since the existence of emotional harm is itself determined by the psychology profession. The profession's condemnation of a practice is rooted in its belief that such a practice is harmful to the patient.[218] Having established the emotional harm element of emotional child abuse, the parent would be guilty or liable for child abuse, regardless of whether the parent even knew that such a practice was harmful.

However, the reason why the argument outlined cannot work is because the written guidelines that are commonly cited by legal scholars are not even considered probative of whether a psychological practice is harmful in court. Instead, courts have consistently looked to psychological associations' codes of ethics to determine whether a psychologist has violated a standard of care, thereby exposing her to liability for harming her patient.[219] While many written guidelines have "condemned" reparative therapy, codes of ethics, the only legally relevant guidelines, have been silent.

This Part reviews the legal literature's reliance upon written guidelines, asserts that the only relevant guideline for legal purposes is a code of ethics, and demonstrates that legal scholars and some psychologists fail to acknowledge this.

---

defineall.pdf.

218. This Part shows that it is codes of ethics that are relevant in determining whether a profession has condemned a practice, deeming it violative of the standard of care. Such practices are presumed to be harmful to the patient. It is true that some psychologists feel that codes of ethics have shifted focus from preventing harm to the patient to shielding psychologists from liability. *See infra* note 239. Still, the existence of liability implies the existence of harm, and the explicit purpose of codes of ethics is still to protect the patient from harm. *See, e.g.*, AM. PSYCHOLOGICAL ASS'N, ETHICAL PRINCIPLES OF PSYCHOLOGISTS AND CODE OF CONDUCT (2002), http://www.apa.org/ethics/code2002.pdf ("This Ethics Code is intended to provide specific standards to cover most situations encountered by psychologists. It has as its goals the welfare and protection of the individuals and groups with whom psychologists work and the education of members, students, and the public regarding ethical standards of the discipline.").

219. *See infra* notes 240-254 and accompanying text.

A. Legal Scholars' Reliance on Written Guidelines of Psychological Organizations

Hicks, Yoshino, and Cohan place an uncritical reliance on the written guidelines of professional psychological organizations, while Gilliam[220] and Cruz ignore them.

Hicks relies on these written guidelines to bolster her argument that a reasonably prudent parent should know that reparative therapy is harmful. First, she cites a unanimous board decision made by the American Psychiatric Association in 1998 to oppose reparative therapy (hereinafter 1998 American Psychiatric Association Board Decision).[221] Second, she notes:

> At the 105th annual meeting in Chicago on August 14, 1997, the [APA] announced: '[T]he APA opposes all portrayals of lesbian, gay and bisexual people as mentally ill and in need of treatment due to their sexual orientation and supports the dissemination of accurate information about sexual orientation, and mental health, and appropriate interventions in order to counteract bias that is based in ignorance and unfounded beliefs about sexual orientation.[222]

Third, she cites a policy statement from the American Academy of Pediatrics, published in 1993, stating that therapy "directed at specifically changing sexual orientation . . . can provoke guilt and anxiety while having little or no potential for achieving changes in orientation" (hereinafter 1993 Pediatrics Policy Statement).[223] Finally, she notes that the president of the APA's Society for the Study of Lesbian, Gay, and Bisexual Issues accused reparative therapy of having an inappropriate bias.[224]

Yoshino, carefully cataloguing the history of reparative therapy, states that "the mental health profession has generally marginalized the practice. None of the major mental health associations . . . currently endorses conversion therapy."[225] To back up his claim, he cites a position statement passed by the American Medical Association, a position statement passed by the Board of Trustees of the American Psychiatric Association supplementing the 1998

---

220. Gilliam's article was a legislative advocacy piece, and the main focus of his piece was not on reparative therapy, so his disregard of such written guidelines is understandable.

221. Hicks, *supra* note 5, at 513 n.34.

222. *Id.* at 513 & n.35 (citing Am. Psychological Ass'n Council of Representatives, Resolution on Appropriate Therapeutic Responses to Sexual Orientation (1997) [hereinafter 1997 APA Resolution]).

223. *Id.* at 514.

224. *Id.*

225. Yoshino, *supra* note 31, at 800.

American Psychiatric Association Board Decision,[226] the 1997 APA Resolution, and a position statement passed by the National Association of Social Workers in 2000.[227]

Cohan, arguing for the minor's right to refuse reparative therapy, cites the 1998 American Psychiatric Association Board Decision,[228] the 1993 Pediatrics Policy Statement,[229] and a condemnation of reparative therapy by the American Psychoanalytic Association.[230]

B. The Missing Link: Codes of Ethics

Based on Hicks, Yoshino, and Cohan's reliance on these written guidelines, it is tempting to assume that such universal condemnation of reparative therapy surely qualifies as relevant and reliable evidence in a reparative therapy child abuse case. However, this assumption is rightly questioned in Gans's article, which provides the most nuanced analysis of the legal impact of these written guidelines. She makes this careful distinction: "The fact that neither the American Psychiatric Association nor the [APA] has condemned the use of conversion therapy on the ground that it is unethical also supports its continued practice."[231] Citing a "Fact Sheet" published by the American Psychiatric Association in 1994[232] and the 1997 APA Resolution,[233] she specifically notes that none of those statements condemn reparative therapy outright. For that reason, she surmises that a negligent malpractice action against reparative therapy would be difficult.[234] Cruz also briefly mentions this distinction.[235]

---

226. *Id.* at 800 n.147.

227. *Id.*; Nat'l Ass'n of Soc. Workers, Position Statement: National Committee on Lesbian, Gay & Bisexual Issues: "Reparative" and "Conversion" Therapies for Lesbians and Gay Men (Jan. 21, 2000), *available at* http://www.safeschoolscoalition.org/RG-PositionStatement-nasw.html.

228. Cohan, *supra* note 12, at 75 n.36.

229. *Id.* at 76 n.37.

230. Cohan cited this condemnation but did not provide a citation. *See id.* at 75 (claiming that the "American Psychoanalytic Association and th[e] American Psychiatric Association have expressed their opposition to 'reparative' therapy" but citing only the American Psychiatric Association condemnation in the subsequent footnote).

231. Gans, *supra* note 18, at 227.

232. *Id.* at 228 n.60.

233. *Id.* at 228.

234. *See id.* at 241 (noting that the fact that neither the APA nor the American Psychiatric Association have banned the use of reparative therapy serves as a defense for the therapist).

235. Cruz, *supra* note 19, at 1301 ("For more than twenty-five years some have argued that reorientation efforts are unethical and harmful and should not be countenanced. Both the [APA] and the American Psychiatric Association . . . have considered resolutions to such effect, although neither adopted the resolutions at issue.").

Gans was almost right. The problem is not that these associations have not *condemned* reparative therapy, but that they have not deemed reparative therapy to be *unethical*.[236] Notably absent from the litany of written guidelines cited by other authors are these associations' codes of ethics. Unlike the standard of care used in medical malpractice actions, which "vary substantially in authorship, form, dissemination, and purpose,"[237] psychological standards for legal purposes actually seem to rely greatly on the code of ethics. As some research psychologists note, "In this age of increasing accountability, the formal means by which we psychologists hold ourselves accountable for our behavior is the APA Ethics Committee."[238] One psychologist laments the fact that codes of ethics had become a code of legal liability rather than a code to protect the patient's best interest.[239]

The discussions by psychologists over legal issues usually center on the code of ethics, since they realize that the code is often the bedrock of litigation. For instance, as a result of disagreeing codes of ethics concerning whether it is ethical for psychologists to have sexual relations with ex-clients,[240] one legal scholar recognizes that it would result in ambiguity over liability.[241] Another notes that the APA potentially misled psychologists when it published a set of guidelines for professionals in making child custody evaluations and states that the guidelines were not intended to be "mandatory or exhaustive."[242] This was misleading, because the APA was "quick to point out that the new guidelines build upon the APA Ethical Principles of Psychologists and Code of Conduct."[243] If such guidelines were based upon the code of ethics, the scholar notes, then the

---

236. Gans, *supra* note 18, at 227.

237. Jodi M. Finder, *The Future of Practice Guidelines: Should They Constitute Conclusive Evidence of the Standard of Care?*, 10 HEALTH MATRIX 67, 72 (2000).

238. Kenneth S. Pope et al., *Ethics of Practice: The Beliefs and Behaviors of Psychologists as Therapists*, 42 AM. PSYCHOLOGIST 993, 1004 (1987).

239. Carolyn R. Payton, *Implications of the 1992 Ethics Code for Diverse Groups*, 25 PROF. PSYCHOL.: RES. & PRAC. 317, 320 (1994) ("Previous codes seemed to have been formulated from the perspective of offering protection to the consumers of our discipline . . . . The 1992 revision appears to be more concerned with offering protection to psychologists . . . . It reads as though the final draft was edited by lawyers in the employment of the APA.").

240. Ronald J. Maurer, *Ohio Psychotherapist Civil Liability for Sexual Relations with Former Patients*, 26 U. TOL. L. REV. 547, 548 n.4 (1995) (citing codes of ethics of various major psychological associations).

241. *Id.* at 548 (proposing solution to the ambiguity).

242. Marc J. Ackerman, *APA Guidelines for Child Custody Evaluation in Divorce Proceedings: Overview, Interpretation, Elaboration (Part III of III), Conducting an Evaluation*, FAIR$HARE, Mar. 1996, at 6.

243. *Id.*

following statement would be more accurate: "As guidelines they are not intended to be either mandatory, *unless already embodied in the APA Code of Ethics*, or exhaustive."[244] Other scholars state that the APA code of ethics explicitly separates its aspirational section from the section that "creates enforceable standards of conduct, which are punishable by sanctions if not followed."[245] Legal scholars who have analyzed the ethical conflicts that may arise between lawyers and psychologists working with each other note the centrality of the code of ethics in the conflict.[246] Another psychologist highlights the important legal role of the code of ethics when noting that, because more psychologists believe in the possibility of rational suicide for the terminally ill, the "current APA code of ethics needs to be examined to see if it will allow for participation in assisted suicide."[247]

Many states explicitly set their legal standards of care in accordance with the APA ethical guidelines.[248] And even if statutes do not codify the state's dependence on private associations, state courts often rely upon them to define the standards for malpractice suits anyway. For example, one Georgia psychotherapy malpractice case admitted expert testimony that explicitly relies on the APA ethical standards as evidence for malpractice.[249] A North Carolina

---

244. *Id.*

245. Nola Nouryan & Martha S. Weisel, *When Ethics Collide: Psychologists, Attorneys and Disclosure*, 36 CAL. W. L. REV. 125 (1999).

246. *Id.* at 131-32 (recommending that psychologists notify lawyers of their obligations to the code of ethics).

247. Phillip M. Kleespies et al., *Suicide in the Medically and Terminally Ill: Psychological and Ethical Considerations*, 56 J. CLINICAL PSYCHOL. 1153, 1168 (2000).

248. *See, e.g.*, IDAHO CODE ANN. § 54-2305 (2003) (including "current, and as future amended, ethical standards for psychologists of the American Psychological Association"); N.C. GEN. STAT. § 90-270.15(a)(10) (2003) (authorizing the Board to discipline licensees whose conduct violates either the statutorily-defined Code of Conduct, or the "then-current code of ethics of the American Psychological Association"); OKLA. STAT. ANN. tit. 59, § 1361 (West Supp. 2005) ("The State Board of Examiners of Psychologists shall publish a code of ethics. . . . In developing and revising this code, the Board . . . may take into account the Ethical Principles of Psychologists and Code of Conduct promulgated by the American Psychological Association and the Code of Conduct promulgated by the Association of State and Provincial Psychology Boards."); WIS. STAT. ANN. § 455.08 (West 1998) ("The [Wisconsin] examining board shall adopt such rules as are necessary under this chapter and shall, by rule, establish a reasonable code of ethics governing the professional conduct of psychologists, using as its model the 'Ethical Standards of Psychologists,' established by the American Psychological Association.").

249. Bala v. Powers Ferry Psychological Assoc., 491 S.E.2d 380 (Ga. Ct. App. 1997) (admitting expert testimony, stating that "it is my opinion that Dr. Abby Friedman's disclosure was a deviation from the *standard of care* of a psychologist as set forth by the Ethical Principles and Code of the American Psychological Association").

court affirmed the use of the APA standards as long as the sections relied upon are not so vague that a "reasonably intelligent member of the profession" would not understand what practices are forbidden.[250] Lawsuits have been filed challenging the reliance on private associations as an unconstitutional grant of legislative authority, but courts have not been receptive to this challenge.[251]

In the child abuse context, the touchstone of a court's inquiry is the harm resulting to the child, but among the indicia of this inquiry is the standard of care.[252] Because codes of ethics weigh more heavily, if not exclusively, in court determinations of standard of care for psychotherapy, whether a code of ethics condemns reparative therapy is far more relevant and reliable than whether other

---

250. White v. N.C. State Bd. of Exam'rs of Practicing Psychologists, 388 S.E.2d 148, 149 (N.C. Ct. App. 1990).

251. Farber v. N.C. Psychology Bd., 569 S.E.2d 287, 300 (N.C. Ct. App. 2002) ("We do not conclude that discretionary reference to the ethical code of the American Psychology Association for purposes of determining improper behavior by a licensee to be a delegation of legislative authority to the APA."); *see also* Lucas v. Me. Comm'n of Pharmacy, 472 A.2d 904, 909 (Me. 1984) ("'[S]tatutes whose operation depends upon private action which is taken for purposes which are independent of the statute' usually pass constitutional muster.") (quoting KENNETH C. DAVIS, ADMINISTRATIVE LAW TREATISE § 3:12 (2d ed. 1978)); Bd. of Trs. v. City of Baltimore, 562 A.2d 720, 731 (Md. Ct. Spec. App. 1989) (noting that "courts have sometimes upheld legislative adoption of private organizations' standards which are periodically subject to revision, in limited circumstances such as where the standards are issued by a well-recognized, independent authority, and provide guidance on technical and complex matters within the entity's area of expertise. These cases usually involve accreditation or similar programs by established professional organizations").

252. *See, e.g.*, *In re* Phillip B., 156 Cal. Rptr. 48, 51 (Ct. App. 1979) ("Several relevant factors must be taken into consideration before a state insists upon medical treatment rejected by the parents. The state should examine the seriousness of the harm the child is suffering or the substantial likelihood that he will suffer serious harm; the evaluation for the treatment by the medical profession; the risks involved in medically treating the child; and the expressed preferences of the child. Of course, the underlying consideration is the child's welfare and whether his best interests will be served by the medical treatment."). In the analogous Christian Scientist context, where parents refuse to submit their ill children to medical treatment, courts will take standard of care into account in addition to evidence of harm. *See* Newmark v. Williams, 588 A.2d 1108, 1117 (Del. 1991) ("There are two basic inquiries when a dispute involves chemotherapy treatment over parents' religious objections. The court must first consider the effectiveness of the treatment and determine the child's chances of survival with and without medical care."); *id.* at 1117-18 (surveying different states' approaches); Custody of a Minor, 379 N.E.2d 1053, 1064 (Mass. 1978) (removing parents' legal custody because they refused to use chemotherapy when chemotherapy "has come to be viewed as the ordinary, medically indicated treatment for acute lymphocytic leukemia in children. Moreover, according to the undisputed medical evidence, chemotherapy is the only existing form of treatment which can claim such status"; and where parents used dietary program when "uncontradicted expert testimony revealed that the dietary program suggested by the parents had no value in the treatment of acute lymphocytic leukemia").

written guidelines do the same.[253]

C. Is Reparative Therapy Ethical?

The fact that greater legal weight is placed with codes of ethics raises the question why private psychological associations, which have not had any compunctions condemning reparative therapy via resolutions, policy statements, and board decisions, have nonetheless refrained from labeling the practice unethical. Some speculative reasons may be the aforementioned lack of concrete evidence of its direct harms or the fear of a religious backlash. Perhaps opponents of reparative therapy do not wish to use the force of law to stop the practices of their reparative therapist colleagues. In any case, votes to amend codes of ethics to make the practice unethical failed in 1994 and 1995.[254] The proposal also failed in 2003, apparently due to last-minute maneuvering by reparative therapists.[255]

This schizophrenia between condemning reparative therapy on one hand and not condemning it on the other has led to confusion and careless usage of the word "unethical" in legal scholarship describing reparative therapy. For instance, William Eskridge notes that "[t]he professional organizations in psychiatry and psychology have disavowed reparative therapy as . . . unethical in its asserted manipulation of patients."[256] Hicks cites a news article apparently reporting that

253. Written guidelines are indeed used in some cases (usually equal protection civil rights cases). *See, e.g.*, High Tech Gays v. Defense Indus. Sec. Clearance Office, 895 F.2d 563, 578 (9th Cir. 1990) (holding that same APA resolution not enough to overcome DOD's proffered reason for employing expanded security clearance for homosexual applicants); Jantz v. Muci, 759 F. Supp. 1543, 1548 (D. Kan. 1991) (relying partly on APA resolution stating that homosexual orientation has no effect on job performance in an equal protection employment discrimination case to show that summary judgment is precluded); Baker v. Wade, 106 F.R.D. 526, 536 (N.D. Tex. 1985) (referring to the use of APA, American Psychiatric Association, and AMA resolutions in case challenging sodomy statute). But they are not used in cases requiring a party to establish that some treatment is harmful according to science or standard of care, as noted previously. *See supra* Section IV.B.

254. In 1994, a resolution condemning conversion therapy as unethical was rejected by the American Psychiatric Association. A similar resolution was proposed but rejected by the APA in 1995. B.A. Robinson, Ont. Consultants on Religious Tolerance, Statements by Professional Associations and Their Leaders, http://www.religioustolerance.org/hom_expr.htm (last visited Nov. 18, 2003); *see also* Kaufman, *supra* note 168, at 424 (documenting American Psychiatric Association discussions to label reparative therapy as unethical in 1993).

255. *See* Interview by Joseph Nicolosi with E. Mark Stern, Member of the Nat'l Ass'n for Research & Therapy of Homosexuality Scientific Advisory Comm., *transcript available at* http://www.narth.com/docs/battleapa.html (last visited Nov. 18, 2004).

256. William N. Eskridge, Jr., *No Promo Homo: The Sedimentation of Antigay Discourse and*

"both the [APA] and the American Psychiatric Association have denounced 'reparative' therapy as unethical."[257] Sherry Colb asserts that "the entire enterprise of 'reparative therapy' is ethically questionable and independently troubling,"[258] citing Hicks's article and conflating reparative therapy with the apparently unethical practice of core gender identity conversion.[259] Harris Miller also suggests that reparative therapy is unethical.[260] Although many individual psychologists certainly believe reparative therapy to be unethical, it is dangerous to conflate that with a practice being "unethical" from a legal perspective.

This confusion does not solely exist in the legal realm, but even among some psychologists themselves. In a study conducted in 1987, 55.7% of 456 psychologists believed that considering homosexuality to be per se pathological was unethical.[261] In another study conducted in 1992, despite the fact that no association had declared reparative therapy unethical, one psychologist presumably practicing reparative therapy noted, "My professional association, the APA, has said that my religious beliefs (e.g., that homosexual acts are wrong) are unethical. Therefore, should I quit the APA or my religion?"[262] In an audio documentary sponsored by National Public Radio, Alix Spiegel, the granddaughter of one of the psychologists who was instrumental in removing homosexuality from the DSM, stated in 2002, "It's now considered unethical to treat homosexuality, and any psychiatrist who attempts to change the sexual orientation of his patient can face professional censure."[263]

In sum, given the legal deference granted to codes of ethics to determine the standard of care, it is unlikely that courts will admit testimony on the psychology

---

the Channeling Effect of Judicial Review, 75 N.Y.U. L. REV. 1327, 1367-68 (2000) (citing Cruz article, *supra* note 19, documenting psychological associations' rejection of reparative therapy, though Cruz does not use the word "unethical").

257. Hicks, *supra* note 5, at 513 n.34.

258. Sherry F. Colb, *Oil and Water: Why Retribution and Repentance Do Not Mix*, 22 QUINNIPIAC L. REV. 59, 78 (2003).

259. *Id.* at 78 n.32.

260. Harris M. Miller II, *An Argument for the Application of Equal Protection Heightened Scrutiny to Classifications Based on Homosexuality*, 57 S. CAL. L. REV. 797, 820 n.148 (1984).

261. Pope et al., *supra* note 238. Although the question on ethics was meant to obtain the individual psychologist's view of whether an activity was ethical, it is unclear whether the psychologist responded in terms of whether the practice was objectively ethical, i.e., permitted by the code, or subjectively ethical.

262. Kenneth S. Pope & Valerie A. Vetter, *Ethical Dilemmas Encountered by Members of the American Psychological Association: A Survey*, 47 AM. PSYCHOLOGIST 397, 406 (1992), *available at* http://kspope.com/ethics/ethics2.php.

263. Alix Spiegel, *This American Life: 81 Words* (Chi. Pub. Radio broadcast Jan. 18, 2002), *audio file available at* http://www.thisamericanlife.org/pages/descriptions/02/204.html (53:40).

profession's rejection of reparative therapy as long as codes of ethics are silent on the issue. Consequently, prosecutors cannot hold parents who subject their children to reparative therapy liable for child abuse under this theory.

V. THE EMPIRICALLY VALIDATED TREATMENTS MOVEMENT

A recent movement in the field of psychology, known as the empirically validated treatments ("EVT") movement, could potentially pave the way for Hicks's theory even if no evidence of reparative therapy's harm were to surface and even if professional associations do not explicitly condemn reparative therapy in their codes of ethics.

A. What Is the EVT Movement?

Traditionally, the effectiveness of psychotherapy was not tested with empirical studies. As long as the treatment was working for the patient, that was good enough. Because of the seemingly subjective nature of psychotherapy and the fact that "[c]hoices of treatment depend heavily on the philosophical [as opposed to scientific] basis of the particular form of psychotherapy at issue,"[264] empirical data supporting psychotherapeutic methods were largely lacking.

However, over the past decade there has been a movement within the psychology profession to push for an increased reliance (and for some, an exclusive reliance) on empirically validated treatments. "Generally, [EVTs] are defined as therapies that have been found to be successful in treating psychological disorders in controlled research studies with delineated populations."[265] This movement is largely seen as having been caused by the rise of managed care organizations, whose interests in cost-cutting are threatened by seemingly endless psychotherapy with highly uncertain and subjective results.[266] Ultimately, "managed care organizations [are] interested in clinicians providing the optimal intervention: the least extensive, intensive, intrusive and costly intervention capable of successfully addressing the presenting problem."[267] As Geoffrey Marcyk and Ellen Wertheimer note, "For the first time, those who practice psychotherapy need to show results, and not just to the patient. The

---

264. Geoffrey R. Marcyzk & Ellen Wertheimer, *The Bitter Pill of Empiricism: Health Maintenance Organizations, Informed Consent, and the Reasonable Psychotherapist Standard of Care*, 46 VILL. L. REV. 33, 36 (2001).

265. *Id.* at 78.

266. Sanderson, *supra* note 189, at 388 ("However, in response to the increased costs of psychotherapy, and in particular to the perceived 'endless' nature of psychotherapy, managed care organizations are pressuring clinicians . . . .").

267. *Id.* (citations and quotations omitted).

therapist-patient relationship has acquired a third member: the Health Maintenance Organization (HMO)."[268] Even some psychologists who initially may be loath to support the EVT movement do so out of recognition that if they do not, HMOs will dictate the standards of care for their practices.[269] As the current president of the APA warned in 2005, "If we do not take on this task [of focusing on empirical support], the challenge will not magically disappear. Rather, someone else will dictate what treatments are acceptable and what types of evidence are privileged."[270]

While the EVT movement appears to be sparked by HMO pressure, it has taken on a life of its own, having direct implications for the standard of care.[271] The pressure on psychology to support its treatments with empirical data for the sake of satisfying HMOs gives rise to the implication that if some treatments have more empirical support than others, it may be "negligent for a therapist to fail to offer that treatment . . . . This empirically supported treatment may thus become the standard of care, with the therapist negligent for offering anything else."[272] Although codes of ethics do not explicitly adopt the position that *only* EVTs are ethical,[273] the influence of the EVT movement is increasingly felt in private associations' codes of ethics.[274] Moreover, the EVT movement is seen as a practical defense against critics of psychotherapy who believe that psychotherapy is simply not on par with conventional medicine.[275] Although

---

268. Marcyzk & Wertheimer, *supra* note 264, at 33.

269. Sanderson, *supra* note 189, at 394 ("The question is not whether or not we should develop treatment guidelines whenever possible, but instead, have we already missed the boat?").

270. Ronald F. Levant, *President's Column: Evidence-Based Practice in Psychology*, MONITOR ON PSYCHOL., Feb. 2005, at 5, *available at* http://www.apa.org/monitor/feb05/pc.html.

271. *Id.* ("Some APA members have asked me why I have chosen to sponsor an APA Presidential Initiative on Evidence-Based Practice (EBP) in Psychology, expressing fears that the results might be used against psychologists by . . . malpractice lawyers.").

272. Marcyzk & Wertheimer, *supra* note 264, at 36-37.

273. Instead, the APA ethics code states that "'[p]sychologists work to develop a valid and reliable body of scientific knowledge based on research.' Similarly, the APA Ethical Guidelines note that psychologists 'maintain knowledge of relevant scientific and professional information related to the services they render . . . and make appropriate use of scientific . . . resources.'" *Id.* at 83-84.

274. *See id.* at 77-82 (describing the APA's history with EVTs); Sanderson, *supra* note 189, at 389 (citing a conference in 1991 determining the treatment consensus statement for panic disorders on the basis of the most convincing empirical data).

275. *See, e.g.*, Hjelt, *supra* note 52, at 39 ("Psychotherapy, as a treatment for an emotional condition, is analogous to a drug or medication prescribed to treat a medical condition."); Sanderson, *supra* note 189, at 390 ("[P]sychological interventions seem to be taking a backseat to pharmacological approaches . . . . Unlike pharmaceutical companies that spend a significant amount of money promoting the use of their treatments to consumers and providers, no such profit-

courts appear unaware of the EVT movement,[276] and opposition to it in the psychological community remains sizeable,[277] the EVT movement shows no signs of slowing.[278]

B. Evidence That Reparative Therapy Does Not Work Would Now Be Admissible

If the EVT movement were to take hold in the practice of psychology and then in courts, Hicks's theory would become more plausible. Testimony presenting already existing evidence that reparative therapy does not work, previously inadmissible due to its irrelevance,[279] could then be admitted. The central tenet of the EVT movement is that all treatments that do not work carry too much risk of harm to the patient. Therefore, the fact that reparative therapy does not work means that it is "generally accepted" that reparative therapy carries too much risk of harm to the patient. Such testimony would be probative of reparative therapy's emotionally harmful effects, strengthening the case for characterizing reparative therapy as child abuse.

It is important to note that the EVT movement could have the side effect of making the *Daubert* factors—testability (or falsifiability), rate of error, and peer review—more relevant to psychological testimony, in contrast to current court practice, which leans heavily on the "general acceptance" prong held over from the *Frye* test. As Stephen Pappas noted in 2005:

> Parallel to the evolution of the judicial approach to expert testimony in the decade following *Daubert*, a more critical and objective evidence-based assessment of medical science has evolved. . . . Evaluating the admissibility of expert medical testimony within the objective framework of evidence-based medicine and its focus on reliability and relevance will move the evaluation

---

motivated organization exists for psychotherapeutic interventions.").

276. A search for "empirically validated treatment" in Westlaw of all federal and state cases revealed no results. This is not to suggest that courts do not consider the role of empiricism in admitting psychological evidence, only that they appear unaware of this significant controversy.

277. *See generally* Marcyzk & Wertheimer, *supra* note 264 (opposing the EVT movement). Sanderson notes that EVT opponents argue that empirical studies fail to accurately mirror what occurs in actual practice, fail to take into account therapist and patient variability, and fail to recognize that all psychotherapies are equally effective. Furthermore, empiricism places undue reliance on the DSM. Sanderson, *supra* note 189, at 396; *see also id.* at 393 (quoting EVT opponent warning of "potential disastrous consequences . . . from such arrant foolishness").

278. *See, e.g.*, *supra* note 270 (describing the reasons behind the creation of a 2005 Presidential Task Force on Evidence-Based Practice).

279. *See supra* Section III.E.

closer to the reliability and relevancy standards envisaged by *Daubert*.[280]

Although Pappas refers to an analogous EVT movement in medicine, it applies equally to the psychological context.

However, even if the EVT movement causes *Daubert* courts to actually apply the other factors to psychological testimony, a prosecutor could still admit the empirical evidence that reparative therapy does not work. Such evidence has already been peer reviewed, and even proponents of reparative therapy have not been able to falsify these studies, nor do they even seriously contest them.[281] Therefore, the *Daubert* hurdles would not prevent such testimony from being admitted, and the EVT movement will have deemed such testimony to be probative of a treatment's undue risk of harm to the child.

VI. IMPLICATIONS AND RECOMMENDATIONS

Surveying the data and trends evolving concurrently in the psychology profession and the courts prompts several implications and recommendations. This Part first summarizes how this Note's analysis affects Hicks's theory of child abuse prosecution, then discusses how the analysis affects reparative therapy in the legal contexts that have been introduced by other legal scholars, and finally lists recommendations for how the legal and psychology professions should deal with reparative therapy as well as each other on a general level.

A. Implications for Child Abuse Prosecution

In sum, Hicks's assertion that reparative therapy legally constitutes child abuse requires one of the following three conditions: (1) empirical evidence that reparative therapy results in psychological harm; (2) condemnation of reparative therapy in codes of ethics, thereby affirmatively establishing reparative therapy as a legal violation of the standard of care; or (3) a substantial increase in the influence of the EVT movement in psychology, thereby requiring reparative therapists to affirmatively and empirically show that reparative therapy works rather than rely defensively on the lack of evidence of its harms.

First, if empirical evidence that reparative therapy results in psychological harm were to surface, a child abuse prosecution would be plausible. Such robust evidence would quickly gain the "general acceptance" of the professional community, and expert testimony concerning the generally accepted notion that

280. Pappas, *supra* note 94, at 597.

281. Proponents instead assert the mere *possibility* of change. *See supra* note 168 and accompanying text.

reparative therapy is harmful would be admitted. This testimony would then establish actual emotional harm. Having established emotional harm by expert testimony and the fact that the parent subjected her child to such emotional harm by sending the child to reparative therapy, emotional child abuse could then be shown.

Second, condemnation of reparative therapy in codes of ethics would also permit prosecution for child abuse. Currently, when courts evaluate psychological malpractice liability, they look to the profession's code of ethics. If the code of ethics condemn a practice, then such a practice violates the standard of care, which means that such a practice is harmful to the patient. If the code of ethics condemns reparative therapy, an expert could testify that it is generally accepted that the profession considers reparative therapy to be emotionally harmful. Having established actual emotional harm by expert testimony, emotional child abuse could then be shown.

Third, a substantial increase in the influence of the EVT movement in psychology would also permit prosecution for child abuse. The EVT movement's central tenet is that psychotherapists should only employ practices that have been empirically validated because only such practices are effective, and all other practices are therefore unjustified and carry too much risk of harm to patients. If this movement caught hold, all psychotherapeutic practices that could not empirically demonstrate their effectiveness would be deemed unsafe and unethical. Assuming that reparative therapy continues to lack empirical support that it works, reparative therapy would then be among such practices that the EVT movement would reject. An expert could then testify that since reparative therapy does not work, it presumptively carries too high of a risk of harm. Having established actual emotional harm by expert testimony, emotional child abuse could then be shown.

B. Implications for Reparative Therapy in Other Legal Contexts

Applying these three conditions can also augment the analyses of the other legal articles already addressing reparative therapy. For instance, Gans's proposed intentional infliction of emotional harm cause of action against reparative therapists depends on the existence of "extreme and outrageous conduct."[282] To show that the conduct is extreme and outrageous, Gans relies on the fact that "studies have proven the harmful effects conversion therapy can have on patients."[283] However, this phrase is misleading, as it is quite easy to prove that something *can* happen; one instance of harm proves that fact.

282. Gans, *supra* note 18, at 246.

283. *Id.* at 247.

Moreover, in any case, her assertion relies on claims by psychologists that reparative therapy perpetuates homophobia,[284] testimony that faces problematic evidentiary hurdles, as discussed previously.[285] She also relies on the fact that reparative therapists "are undoubtedly aware of the volatile controversy surrounding their actions,"[286] but an action's controversial character hardly gives rise to its being extreme or outrageous. However, if any of the three aforementioned conditions were satisfied, the "extreme and outrageous" claim would be substantially strengthened and Gans's intentional infliction of emotional harm theory would likewise be more plausible.

The same analysis applies to Cruz's proposal, which advocates for less deference to a client's claim that she really wants reparative therapy. Although a full discussion of the complex doctrine of informed consent is beyond the scope of this Note,[287] it is sufficient to assert that each condition would have a significant impact on the informed consent analysis. If evidence of the harms of reparative therapy surfaced, a reparative therapist would be required to inform the patient of these harms. If any of the latter two conditions occurred, thereby relegating reparative therapy to the position of being violative of the standard of care, the informed consent defense would drop out of the inquiry altogether, as the reparative therapist would simply not be allowed to perform such a practice.

Gilliam's advocacy for protecting homosexual children from reparative therapy in foster care would gain significant credibility if any of the conditions were met. Homosexual juveniles would no longer be institutionalized by their caretakers for reparative therapy purposes, because *Parham* requires that a neutral fact-finder must make findings based on psychiatric standards that institutionalization is appropriate. If any of the conditions were satisfied, a neutral fact-finder would be obligated to find that institutionalized reparative therapy would be inappropriate.[288] Lastly, child custody cases involving homosexual children would certainly be resolved in favor of the parent who chooses not to subject the child to reparative therapy, even if the pro-reparative

---

284. *Id.* at 248 n.185 (citing sources "discussing the reinforcement of homophobia through the use of conversion therapy").

285. *See supra* Section III.B.

286. Gans, *supra* note 18, at 248.

287. *See* Hjelt, *supra* note 52, at 1 ("The amount written about the doctrine of informed consent in the last forty years truly threatens more than one old growth forest.").

288. Arguably, it is already unconstitutional for a parent to commit her child to a mental institution for reparative therapy purposes. The fact that homosexuality is not an illness would remove any basis on which a neutral fact-finder would agree to commit a child. One observer stated in 2002 that it is "technically no longer . . . possible to have a child institutionalized for being homosexual." Lehr, *supra* note 2, at 8. Regardless of whether this observation is true, the conditions would still strengthen the case that the institutionalization is unconstitutional.

therapy parent sought the therapy for religious reasons.[289]

The only issue for which these conditions are probably irrelevant is the proposal by Cohan that adolescents should have the substantive due process right to refuse reparative therapy. *DeShaney v. Winnebago County Department of Social Services*[290] holds that "nothing in the language of the Due Process Clause itself requires the State to protect the life, liberty, and property of its citizens against invasion by private actors."[291] Therefore, whether reparative therapy is harmful or unethical has no bearing on the due process analysis, since the parent, not the state, would be causing the harm to the child.[292] It is worth noting, however, that some legal scholars are pushing for an adolescent's right to make health care decisions,[293] and the conditions certainly bear on that debate.

C. Recommendations

Although this Note has focused on the issue of whether reparative therapy constitutes child abuse, the aforementioned implications in other legal contexts concerning reparative therapy warrant several broader recommendations. These recommendations generally address practical ways through which the law and psychology can address the controversial issue of reparative therapy.

First, the legal scholarship needs to take a more nuanced approach to the practice of reparative therapy, employing the science carefully and accurately. This is in contrast to framing the conflict in terms of right-wing fundamentalists versus gay rights activists, playing fast and loose with scientific data in order to shore up support for one side or the other[294] or carelessly conflating several studies (i.e., data that reparative therapy does not "work" and data on the

289. *In re* E.L.M.C., 100 P.3d 546, 563 (Colo. Ct. App. 2004) ("While courts are precluded by the free exercise of religion clause from weighing the comparative merits of the religious tenets of the various faiths or basing their custody decisions solely on religious considerations, the family is not beyond regulation in the public interest as against a claim of religious liberty, and neither the rights of religious nor rights of parenthood are beyond limitation. Thus, evidence of beliefs or practices which are reasonably likely to cause present or future harm to the child is admissible in a custody proceeding.") (internal citation omitted).

290. 489 U.S. 189 (1989).

291. *Id.* at 195.

292. This raises the question of whether psychologists licensed by the state are state actors. However, that question may be sidestepped; if reparative therapy is deemed unethical, then reparative therapists will not be able to obtain licenses. Therefore, the child would still be harmed by another private actor—the non-licensed reparative therapist.

293. *See generally* Jennifer L. Rosato, *Let's Get Real: Quilting a Principled Approach to Adolescent Empowerment in Health Care Decision-Making*, 51 DEPAUL L. REV. 769 (2002).

294. *See, e.g.*, *supra* notes 207-208 and accompanying text (discussing reporting on the Spitzer and the Shidlo/Schroeder studies by reparative therapy opponents and proponents).

vulnerability of queer youths to depression). As Robert Bayer, a public health historian at Columbia University, states, "Both sides wrap themselves in the mantle of science and both sides charge that the other side is being unscientific."[295] There is no excuse for legal scholarship to succumb to the same intellectual pitfalls of political activists.

Second, both the legal and psychological communities should join the chorus of those who have pressed for more reliable studies concerning the possible harms of reparative therapy. Relying solely on the rhetoric of resolutions, policy statements, or board decisions only highlights the awkward and embarrassing fact that no private association has condemned reparative therapy in its code of ethics—where it really matters.

Third, the controversy over EVT among psychologists should alert courts to the fact that, doctrinally, psychologists' ongoing debate about the role of empirical data differentiates psychology from other scientific disciplines. Some judges and scholars may certainly recognize this fact, but the *Daubert* and *Frye* doctrines are not clear as to how judges ought to consider psychological testimony. Psychological testimony should not face the same barriers as other scientific testimony for admissibility, as long as judges do not allow psychologists to conflate a variety of scientific assertions and make it clear to jurors that psychological testimony should not be considered as authoritative as other scientific testimony. However, if psychology eventually adopts the EVT approach to validating its practices, then courts may properly treat psychology in the same rigorous way it treats other traditional "sciences."

Fourth, private psychological associations need to develop a more structured mechanism or policy regarding their written guidelines. The ways in which legal scholars haphazardly and sometimes inaccurately rely on the plethora of written guidelines only highlights the present confusion. Furthermore, given the significant extent of private psychological associations' political participation, these associations should be even more careful to explain which positions are rooted in science, to what degree, and whether a position means that a majority of its members agree with that position.[296] With regard to codes of ethics, opponents of reparative therapy may not wish to sentence their proponent colleagues to jail

---

295. Spiegel, *supra* note 263, at 51:40. For more information on the latest battle over scientific data, see Robinson, *supra* note 171. *See also* Kaufman, *supra* note 168, at 426-33 (providing extensive overview of scientific evidence mounted by both sides as well as corresponding responses); Robinson, *supra* note 201; Yarhouse & Throckmorton, *supra* note 167, at 70-73 (providing relatively balanced but pro-reparative-therapy-leaning overview of scientific data).

296. *See* Am. Psychological Ass'n, Pub. Policy Office, Advocacy Issues, http://www.apa.org/ppo/issues/homepage.html (last visited Nov. 3, 2005) (listing its public policy advocacy resolutions).

or subject them to civil liability, but there may be a middle ground between rhetorical denunciations on the one hand and a rigidly draconian code of ethics on the other. Associations should consider such possibilities.

Fifth, private psychological associations should specifically condemn non-psychotherapeutic, physically invasive techniques of reparative therapy in their codes of ethics. Little suggests that reparative therapy proponents would put up much of a fight, since many of them likewise denounce physically invasive techniques. There is no reason that the practice of physically invasive reparative therapy should continue, and its elimination will better focus the debate on the practice of psychotherapy itself. Furthermore, opponents of reparative therapy have an incentive to push for this measure, since the conflation of physically invasive methods and psychotherapy only weakens their arguments.[297]

CONCLUSION

Hicks never claims that reparative therapy would be held by all courts to be child abuse, astutely and implicitly recognizing that the legislatures, politics, and cultural attitudes of each state are significant factors in determining whether a court would adopt her theory.[298] However, she neglects to identify the critical evidentiary obstacles that must be surmounted before her theory could be realized.

One's zeal against homophobia may be justified, especially given the sheer number of anecdotes attesting to the evils of reparative therapy, but zeal must be channeled in an effective manner. Courts that hear such child abuse cases are impervious to scattershot approaches to getting psychological testimony against reparative therapy admitted. Unless scientifically sound evidence as to reparative therapy's harms emerges, private associations actually denounce reparative therapy in their codes of ethics, or the EVT movement takes significant hold in the psychological community, the testimony of psychologists on behalf of the child will neither be admitted nor effective.

This controversy also reveals significant holes in the way that courts and the practice of psychology interact. Both institutions need greater clarity and coherence with regard to how they view psychology, whether as a science, an art, or a mixture of both. Their policies should correspond accordingly. Opponents of reparative therapy have to face the reality that while a great deal of evidence and

297. *See supra* Section I.C. (discussing reasons for focusing on psychotherapeutic forms of reparative therapy).

298. *See* Hicks, *supra* note 5, at 543 ("Whether a court or legislature would ever extend protection to juveniles subjected to 'reparative' therapy may turn in part on the jurisdiction and community in which the juvenile resides.").

popular professional opinion appears to be on their side, it does not excuse overreaching beyond what the evidence has established. Proponents of reparative therapy also need to concede that reparative therapy's efficacy simply does not have scientific backing and move away from the tired and irrelevant nature/nurture debate.

Amidst this political furor, legal scholarship needs to develop a scientifically honest, context-sensitive framework for handling cases involving this controversial practice. The political environment is simply too volatile to risk staking out an overly rigid or principled position against reparative therapy as a whole—only a carefully nuanced system will best meet the concerns of all stakeholders and be insulated from shifting political winds as much as possible.