

Introduction: Evaluation in Analytic Theory and Political Practice

Lawrence D. Brown
Columbia University

Michael K. Gusmano
The Hastings Center

Abstract The development of professional policy analysis was driven by a desire to apply “science” to policy decisions, but the vision of apolitical policy analysis is as unattainable today as it was at the inception of the field. While there is powerful evidence that schemes to “get around” politics are futile, they never seem to lose their popularity. The contemporary enthusiasm for health technology assessment and comparative-effectiveness research extends these efforts to find technical, bureaucratic fixes to the problem of health care costs. As the benefits and costs of health care continue to grow, so too will the search for analytic evidence and insights. It is important to recognize that the goal of these efforts should not be to eliminate but rather to enrich political deliberations that govern what societies pay for and get from their health care systems.

Professional policy analysis initially was driven largely by a desire to apply “science” to policy decisions. From its beginning, the policy sciences movement has centered on connecting knowledge, policy making, and power. Harold Laswell, the founding father of the movement, wanted policy making to be informed by knowledge. Leading scholars in public administration likewise hoped that policy analysis might dispel the tension between effective government administration and politics (Long 1962; Mosher 1982; Waldo 1948, 1980), which has often been portrayed as incompatible with scientific management in public administration (Taylor 1967; Wilson 1887). As Woodrow Wilson and others argued, elected officials, legitimate because accountable to the public, have the constitutional responsibility to determine the direction of policy, and administrators have the duty to carry out those policies. Ever since, policy

Journal of Health Politics, Policy and Law, Vol. 38, No. 6, December 2013
DOI 10.1215/03616878-2373139 © 2013 by Duke University Press

analysts have aspired to identify the technically “right” solutions to policy problems that would take politics out of the picture.

Policy analysis has come to be dominated by welfare economics, based on Paretian criteria, and gives preference to a particular definition of efficiency. It is not clear, however, that it is appropriate, desirable, or even possible to impose a universal criterion on all issues and across societies. Economist and philosopher Amartya Sen (1970) shows, for example, that if individuals have “nosey” preferences, liberalism is inconsistent with the Pareto principle. Sen claims that we should solve this so-called liberal paradox by placing greater weight on liberal values than we do on efficiency. But potential conflicts between personal rights and Pareto efficiency raise a crucial constitutional question for every political system: What choices should be left to the individual and which should be made collectively? A society may place a high value on liberty and wish to secure a wide sphere in which liberal values trump meddlesome preferences and the Pareto principle.

While there is powerful evidence that schemes to “get around” politics are futile, they seem never to lose their popularity. This history of health care policy is replete with attempts to find technical solutions to challenging political problems that beset the allocation of health care resources, the evaluation of health care technologies, and the trade-offs between spending on health care and on other things that society values. According to James Morone (1993: 723), this perennial apolitical “technocratic wish” led the United States in the 1980s and 1990s into “a distinctly bureaucratic health care regime.” Rather than scrutinize the practices of other Western health care systems, the United States forged its own path, relying on payment reforms (Medicare’s “diagnosis related groups” and “resource-based relative value scale”) and health services research (spurred partly by the creation of new federal agencies such as the Office of Technology Assessment and the Agency for Health Care Policy and Research [AHCPR]) that would calculate the value of medical technology, cut spending on medical care that was ineffective, and help the government set prices in ways that encouraged hospitals and physicians to operate efficiently. These strategies “appear to operate automatically, scientifically—without visible decisions by politicians or bureaucrats” (730).

The contemporary enthusiasm for health technology assessment and comparative effectiveness research originates in and extends these efforts to find technical, bureaucratic fixes to the problem of health care costs. The articles in this issue show, however, that the “technocratic fix” is no longer a uniquely American pursuit. Writing in the early 1990s, Morone argued

that the quest for technical solutions to the cost crisis arose in the United States partly because we lacked the political will to adopt a national health care budget, on which most countries in the Organisation for Economic Co-operation and Development (OECD) rely to control health care spending. Budget politics can be highly contentious, of course, and using a budget to control health care costs raises the fearsome specter of government rationing. The United States apparently prefers to ration by ability to pay, and so, declining to set national budget limits on health care spending, policy makers continue to experiment with technological solutions that supposedly take politics out of policy judgments that are inescapably political.

This argument suggests that countries operating within a budget constraint would have little to gain from these analytic techniques, but, as sometimes happens, innovations designed in the United States are implemented more aggressively elsewhere. Soon after the United States eliminated the Office of Technology Assessment and changed both the mandate and the name of the Agency for Health Care Policy and Research (while slashing its budget), England launched the National Institute for Health and Clinical Excellence (NICE), which has established itself as one of the leading agencies for economic evaluation of health technology in the world. Since then, Canada, Australia, and several European countries have started using formal economic evaluations to inform resource allocation decisions and clinical practice guidelines (Oliver and Sorenson 2009). The aim of economic evaluation of health care services in England, and many other countries, is not to control overall health expenditures but to ensure that the money spent on health care generates “value for money,” a concept that is gaining ground in the United States as well (Gusmano and Callahan 2011).

But how has economic evaluation actually worked outside the United States? What are the strengths and limits of this approach? Does economic evaluation in practice fulfill the “technocratic wish” described by Morone? The articles in this volume go a long way toward answering these questions.

The article by Michael Drummond offers a succinct, sharp review of economic evaluation that dissects its strengths and limits absent illusions that politics can or should be expunged from the process. In this context his discussion of the criteria adopted for assessing medical technology is especially important. In the quest to get it “right,” especially in public bodies, one might expect a steady expansion of evaluative criteria justified in the name of accountability, responsiveness, and reasonableness. This is understandable and perhaps desirable: Economic evaluation is not supposed to impose values on society, but rather allow for a more systematic evaluation of choices that reflect societal values. At some point, however,

one may accumulate enough criteria to justify just about anything. How should evaluators tackle criterion “creep” that, by injecting subjective considerations into the analysis, arguably undermines the hope for apolitical analysis?

Some of the most important, but controversial, applications of economic evaluation are assessments of technologies that may extend the life of a patient with a life-threatening illness. For years, analyses of health care expenditures have reminded us that health care spending is concentrated in the last six months of life (Wennberg et al. 2004). How much should societies spend on expensive health care technologies that may extend life by only a few months? Is ever-increasing health care spending at the end of life a sensible and humane investment, or is it evidence that medicine has lost its way and is creating an “economic quagmire” (Callahan and Nuland 2011)? Richard Cookson’s article explores NICE’s “end-of-life premium” and asserts that health technology appraisal committees should “give special additional weight to health gains from life-extending end-of-life treatments.” His thorough critique of eleven ethical arguments for such a premium shows how, notwithstanding the appeals of rational criteria for resource allocation in health services, one soon bumps up against their limitations. Cookson’s exploration of decisions about end-of-life care calls into question how airtight these calculi can or should be. After reviewing and rebutting many possible justifications for NICE’s policy, one may reach the (non)conclusion that the main ethical justification for an end-of-life premium is simply that death is different. If so, NICE’s remit would seem to end at the point that death impends. But then one risks descending the slippery slope to US practices in which any life-prolonging service is considered “appropriate” and the sky is the limit. If the ethical approach falls somewhere between the rigor of NICE and the extravagance of the United States, should society rely on institutions like NICE to say what makes analytic sense and then call in elected officials to retreat from rationality and indulge in what Klein and Maybin (2012: 13) call the “emotional resonance” of end-of-life care? Then again, labeling politics as the realm of nonrational (or irrational) may damage the public interest by tarnishing the legitimacy of the political sector. Is it more helpful simply to accept that there are various types of rationality and that NICE practices one and politicians another? These questions admit no ready answers, and Cookson’s article does a great service by illuminating them.

Along with the recognition that criteria on which economic evaluations rest involve normative questions from which political considerations cannot be expelled, these essays remind us that efforts to act on economic

evaluation will always invoke interest group responses. When the Congressional Budget Office concluded in 2008 that “some medical services could be used more selectively without a substantial loss in clinical value” (2), it implied that health care spending could decline without compromising quality. But as Uwe Reinhardt (2012: 41) reminds us, “every dollar of health care spending is a dollar of someone’s health care income.” While economic evaluation is sometimes sold as a “painless prescription” that will eliminate only medical care technologies from which there is little or no benefit, these technologies unquestionably benefit the manufacturers that produce them and the medical professionals who use them.

In the United States in the mid-1990s, the AHCPR suffered politically damaging attacks by stakeholders who were offended that it had questioned the value of certain health care technologies. In 2009 Congress responded to stakeholder concerns about the Preventive Services Task Force’s call for limiting the use of screening mammograms among women aged forty to forty-nine years without a family history of breast cancer (US Preventive Services Task Force 2009) by inserting in the Patient Protection and Affordable Care Act (PPACA) an amendment that required the federal government to ignore the task force’s recommendations on mammography (Gusmano and Gray 2010). Even in the United States, successful stakeholder attacks on public agencies that evaluate health technology are rare, but Mara Airoidi’s article suggests that it is important to anticipate them and to consider strategies to overcome this resistance.

In her case study of an evaluation of treatments for eating disorders, Airoidi argues that although economic evaluations of health technology often meet with stakeholders’ resistance, particularly when the results call for disinvestment in existing technologies (as distinct from funding decisions about new technologies not yet in use), policy makers may successfully implement disinvestments if the process they adopt generates consensus among relevant interests. In her case study of an economic evaluation that was used to withdraw funding for residential care services designed to address eating disorders, stakeholder resistance was assuaged by deliberations that brought together a wide range of relevant parties, including provider organizations and clinicians, government agencies, and patient representatives. The process adopted reduced mistrust among stakeholders and encouraged participants to reach agreement by framing the debate not as one pitting providers against purchasers but as one about how best to identify a broad strategy to benefit the patients they served. The deliberations did not hinge on what Airoidi calls “rigorous” cost-effectiveness analysis focused on specific interventions. Instead, the exercise

succeeded in good part because the evidence considered in the stakeholders' deliberations kept the whole care pathway fully and continually in view instead of focusing narrowly on particular services and treatments, thereby averting the "stakeholder entrenchment" that often marks efforts at disinvestment.

The case, and Airoidi's intriguing analysis of it, raises important questions. The public exercise clearly gathered and shared copious information so that deliberations could be well grounded. But was adequate information available on all the topics the deliberations explored? From what sources did this information come? How much of it was solid, and how much included guesstimates? If, as Airoidi remarks, the literature on the effectiveness of interventions was inconclusive because the studies' sample sizes were too small, how firm was the base of evidence on which the participants relied? And why did participants not challenge premises, methods, and conclusions as unsound if they feared losing money or authority? The articles in this issue by Drummond and Alan Maynard combine with the characterization of evidence used in the process Airoidi describes to raise concerns about the adequacy of the evidence base for high-quality public deliberations about the value of health care technologies.

One wonders too whether Airoidi's case does not depict a presumably atypical situation in which there was a preexisting consensus that community services ought to substitute for overused hospital care. Did the deliberations lead participants all into the light or instead offer up "scientific" and collective confirmation of an item of conventional wisdom? Cases in which some sort of prior consensus has not taken shape may not be amenable to the deliberative designs that Airoidi describes. As she observes, although the exercise successfully steered disinvestment at the planning stage, it did not involve "rationing," that is, "the particular choice to provide a specific intervention to a patient." This case may therefore illustrate both the power *and* the limits of what a deliberative process can achieve.

Maynard's article offers additional rich insights into the power and limits of economic evaluation. One central issue is whether, and how, economic evaluation can keep up with the speed of advancements in medical technology. Do considerations of cost, rigor, and timeliness consign economic evaluations to stay one or more chapters behind the technological class? Maynard also highlights critical questions that confront widely cited small-area analyses of health care spending and use. Although John Wennberg and colleagues at Dartmouth College suggest that Medicare expenditures could be reduced by 30 to 40 percent—a claim echoed by the Obama administration during the health care reform debate of 2009–2010

(Gusmano 2011)—others question whether the variations in spending noted in the Dartmouth Atlas of Health (and the impressive potential savings these reductions supposedly entail) reflect wasteful clinical behavior or poor measurements of variations in the genuine clinical conditions and needs of patients. Be that as it may, such variations have persisted for decades; as Maynard remarks, and as Joe White (2011) has argued, even if these variations in health care spending reflect unnecessary volume, we “know very little” about how to fix the problem.

The articles in this issue show that economic evaluations can impressively improve the quality of public debates and decisions about the evaluation of health care technologies, but also that the development and deployment of the evidence that legitimates these analytic exercises should be held to rigorous standards—standards that are by no means easily met. Randomized control trials, often viewed as the gold standard, are expensive, lengthy, problematic in their end points, increasingly left incomplete, and less than golden in the narrowness of their samples and exclusion of affected populations (Epstein 2007). Less-exalted evaluative methods pose troubling questions of their own. As “evaluation” spreads to motley assortments of assessors in multiple sites in myriad political contexts, how does one ensure the rigor and quality of what gets done in its name? How can we tell when a little knowledge is proving to be a dangerous thing? And might evaluation be spreading along a rather thin knowledge base that is explored and exploited by people with rather modest skills? Do good studies tend to cost more than policy makers are prepared to tolerate? When it comes to quality of analysis, how often do they get what they pay for?

Today these questions remain wide open because we know strikingly little about how evaluation is, in fact, deployed in the field. Surely it would be helpful to invite sociologists and experts in organizational theory to investigate how evaluation gets implemented in practice. How are methods, criteria, guidelines greeted and applied? Because we know rather little about whether and how far localities follow the “rules,” it is hard to say, for example, whether variations in postcode prescribing, which NICE was created to address, have been reduced.

The vision of apolitical policy is as unattainable today as it was at the inception of policy analysis. Precisely because the authors of the articles in this special issue have made extraordinary contributions to the practice as well as to the theory of evaluative techniques in health policy, their articles illuminate not only the boundaries of these techniques but also their interfusion with political considerations that evaluators cannot, and should not seek to, quarantine. As the benefits and costs of health care continue to

grow steadily, so too will the search for analytic evidence and insights that do not eliminate but rather enrich the political deliberations that govern what societies pay for and get from their health care systems.

References

- Callahan, Daniel, and Sherwin B. Nuland. 2011. "The Quagmire." *New Republic*, May 19. www.newrepublic.com/article/economy/magazine/88631/american-medicine-health-care-costs#.
- Congressional Budget Office. 2008. *Technological Change and the Growth of Health Care Spending*. (CBO Publication No. 2764). Washington, DC: US Government Printing Office. Retrieved from www.cbo.gov/sites/default/files/cbofiles/ftpdocs/89xx/doc8947/01-31-techhealth.pdf.
- Epstein, Steven. 2007. *Inclusion: The Politics of Difference in Medical Research*. Chicago: University of Chicago Press. doi:10.7208/chicago/9780226213118.001.0001.
- Gusmano, Michael K. 2011. "Do We Really Want to Control Health Care Spending?" *Journal of Health Politics, Policy and Law* 36, no. 3: 495–500. doi:10.1215/03616878-1271153.
- Gusmano, Michael K., and Daniel Callahan. 2011. "Value for Money: Use with Care." *Annals of Internal Medicine* 154, no. 3: 207–8. doi:10.7326/0003-4819-154-3-201102010-00012.
- Gusmano, Michael K., and Bradford H. Gray. 2010. "Evidence and Fear: Navigating the Politics of Evidence Based Medicine." *AcademyHealth Reports* 38, June 1. www.academyhealth.org/files/publications/academyhealthreports/evidenceandfear.pdf.
- Klein, Rudolf, and Jo Maybin. 2012. *Thinking about Rationing*. London: King's Fund. www.kingsfund.org.uk/sites/files/kf/field/field_publication_file/Thinking-about-rationing-the-kings-fund-may-2012.pdf.
- Long, N. E. 1962. *The Polity*. Chicago: Rand McNally.
- Morone, James A. 1993. "The Health Care Bureaucracy: Small Changes, Big Consequences." *Journal of Health Politics, Policy and Law* 18, no. 3: 723–39. doi:10.1215/03616878-18-3-723.
- Mosher, F. C. 1982. *Democracy and the Public Service*. New York: Oxford University Press.
- Oliver, Adam, and Corinna Sorenson. 2009. "The Limitations and Challenges to the Economic Evaluation of Health Technologies." In *The Economics of New Health Technologies: Incentives, Organization, and Financing*, edited by Joan Costa-Font, Christophe Courbage, and Alistair McGuire, 205–16. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199550685.003.0013.
- Reinhardt, Uwe. 2012. "Divide et Impera: Protecting the Growth of Health Care Incomes (Costs)." *Health Economics* 21, no. 1: 41–54. doi:10.1002/hec.1813.

- Sen, Amartya. 1970. "The Impossibility of a Paretian Liberal." *Journal of Political Economy* 78, no. 1: 152–57. doi:10.1086/259614.
- Taylor, F. W. 1967. *The Principles of Scientific Management*. New York: Norton.
- US Preventive Services Task Force. 2009. "Screening for Breast Cancer: U.S. Preventive Services Task Force Recommendation Statement." *Annals of Internal Medicine* 151, no. 10: 716–26. doi:10.7326/0003-4819-151-10-200911170-00008.
- Waldo, D. 1948. *The Administrative State: A Study of the Political Theory of American Public Administration*. New York: Ronald Press.
- Waldo, D. 1980. *The Enterprise of Public Administration*. Novato, CA: Chandler and Sharp.
- Wennberg, John E., et al. 2004. "Use of Hospitals, Physician Visits, and Hospice Care during Last Six Months of Life among Cohorts Loyal to Highly Respected Hospitals in the United States." *British Medical Journal* 328, no. 7440: 607–10. doi:10.1136/bmj.328.7440.607.
- White, Joseph. 2011. "Prices, Volume, and the Perverse Effects of the Variations Crusade." *Journal of Health Politics, Policy and Law* 36, no. 4: 775–90. doi:10.1215/03616878-1334731.
- Wilson, W. 1887. "The Study of Administration." *Political Science Quarterly* 2, no. 2: 197–222. doi:10.2307/2139277.