College of Business Administration Northern Arizona University Box 15066 Flagstaff AZ 86011 520-523-3657

Limitations of Cost-Effectiveness Analysis: Defective Data and Significant Digits

William Hildred and James Pinto

To become a practicing physician, one passes through a demanding course of instruction and apprenticeship. The profession guards the reputations of its collective number by administering a licensing program. Economists are available with varying kinds of preparation, though the standard tends to involve intensive study of neoclassical wisdom and completion of an acceptable research endeavor. No licensing beyond conferral of the academic degree is required, and some employment is available without any formal credential in the discipline.

Most citizens cherish their freedom to offer even the most bizarre personal interpretations of economic matters, and some obtain a public forum for the purpose. During the national debate about health care reform in 1995, newspaper readers might not have considered it unusual for a highly regarded local physician to be offered extensive space in a metropolitan newspaper to present his solutions to problems of the health care system. Prestigious medical journals often offer the observations of their editors and subscribers on such issues. Of course, as participants, their views of the operation of the health care system have some value in the public debate, with readers presumably competent to give appropriate weight to self-interest.

However, such performances establish an environment within which professionals certified in medical fields (among others), but with little or no formal training in the discipline of economics, are allowed to publish studies of economic issues, with little concern for rigor and soundness. Their colleagues, including peer reviewers for medical journals, seem to accept such work uncritically, unaware of the degree to which they misuse technical economic concepts and methods.

We follow the eminent Institutionalist, Marc Tool, in believing that "... the only serious grounds on which political economists for centuries have claimed significance for their work have been its alleged relevance to the resolution of real problems facing real people in real political economies" [Tool 1979, 293]. Accordingly, we consider it appropriate to point out the irrelevance, indeed harmfulness, of trends in public policy debate that have the potential to undermine sound decision making about the use of resources in the health care system. The trends in question unite sophisticated computer techniques with training in statistical inference to develop conclusions that

are intended to influence health care policy, all covered by casual acceptance of misused economic theory.

Three Dangers

A danger to health professionals who prize their independence to practice medicine, and to their patients, is that the resources necessary for accomplishing their work are likely to be restricted by stringent financial criteria. Salaries and bonuses, along with ability to prescribe treatments and diagnostic tests, are at stake. These considerations are widely discussed in the industry and among patients, third-party payers, and the general public.

Unfortunately, failure to anticipate distant perils enables some physicians to contribute to trends that may well increase the probability of their occurrence. An eagerness is apparent, on the part of academically based physicians in particular, to adopt the methods, language and mentality of the financial world in evaluating the desirability of medical practices, specifically the use of cost-effectiveness analysis (CEA). Perhaps the technique is attractive to these writers because they believe in its power to resolve difficult issues, and they may enjoy the prestige associated with showing that they are conversant with the method. Alternatively, they may use it defensively, because it is forced upon them by powerful financial professionals who manage insurance firms and health care delivery systems. Other motivations may encourage the behavior, but in every case it should be noted that uncritical acceptance and dissemination of the CEA tool enhances its image and power, when an actually pressing need is to question its meaningfulness. Finally, from the perspective of Institutional economics, a quite serious danger confronts the larger community that is losing its control over democratic processes that determine the level and structure of resource use and the composition of output of the health care industry, especially that part of the industry supported by public funds. This danger is that decisions are increasingly made without significant participation of the public, in contravention of the ideals set forth most cogently by John Dewey.

Dewey's Instrumentalism

Dewey [1927]argued forcefully that a democratic society could not use undemocratic means to find solutions to its problems. Indeed, only through democratic processes that rely on the method of scientific inquiry could a modern society achieve judgments that are adequate for identifying and solving social problems. He offered prescriptions both for identifying democratic truth and for using that insight to improve the lives of the members of the society.

Inasmuch as humans exist as humans only in societies, their well-being is never individual; it is always social. The constant interplay between the growth of individual persons and the improvement of the society within which they live requires personal participation in political life. In that life, the person who contributes through open and honest communication with the community the understandings and insights that are peculiarly his or hers advances both personal and social well-being. Precluding such participation deprives both individual and community of the requisites of improvement. Of course, this conclusion entails the additional recognition that experts do not have either adequate understanding or moral authority to impose their interpretations on others, whether overtly through coercive assertion of their presumed superiority or covertly by cloaking their arguments in a disciplinary language unintelligible to the untrained.

Public policy options should be offered in the manner of scientific hypotheses. These are to be tested with respect to their fruitfulness, and modified as necessary to achieve desired outcomes. As in science, small gains in understanding the nature of problems and the consequences of policy choice add cumulatively as participants in public dialogue verify the acceptability of analysis and policy proposals through disciplined observation and open discussion. "Warranted knowledge" is the guide and the process corrects its own errors. Any argument not based upon or proceeding to develop warranted knowledge is suspect; most obviously, deceit in the form of fabricated data or defective reasoning are condemned as destroyers of democratic progress. No amount of skillful rhetorical flourish is to be accepted as justification for granting an advantage to one party or another in the effort to discover compromises that are socially beneficial.

A Current Problem: Allocating Resources for Health Care

Because identification of desirable health care reforms is so difficult, many people are attracted to financial criteria or what they believe to be economic principles for making choices. A major issue concerns the "standard" of care that will be available, whether care is paid for with involvement of insurance companies, medical management consultants, health maintenance organizations, or from the public purse. For example, some women with breast cancer can receive autologous bone marrow transplantation with their insurers' blessing, but other insurers regard the treatment as "experimental" and deny payment for their clients. Similarly, surgery and external-beam radiation for prostate cancer are considered standard practice by third-party payers, even though there is no decisive evidence that either treatment extends life. Other treatments for this cancer are viewed negatively by many managed-care administrators, who deny payment for the interventions. The latitude available to payers in these kinds of decisions is probably inevitable because medical literature is often inconclusive about the efficacy of treatments.

Two troublesome outcomes are immediately apparent. (1) Patients are presented a menu of interventions for which payment is permitted. Some of the treatments may be effective, or all may be ineffective, but even the well-informed patient must make a choice from the menu. This is not a new circumstance. (2) As managed care firms, especially those with a profit goal, become dominant in the allocation of health care resources, uncertainty about the efficacy of treatments provides an opportunity to employ ill-suited and ill-understood canons of economic efficiency, i.e., Cost-Benefit Analysis and its mutant progeny, Cost-Effectiveness Analysis and Cost-Utility Analysis. Whether medical evidence is ambiguous or quite firm, the financial managers who will control medical practice will almost certainly justify their decisions with appeal to financial variables. This is a new tyranny that must be resisted, along with that of bogus demonstrations of economic efficiency.

The presumed prestige of neoclassical economic analysis is often sought to support or oppose financial arrangements. Concepts of efficiency, often quite incoherent and incorrect (from the conventional-theory perspective) are bandied about, and esoteric exercises are conducted to explore efficiency implications of various policy options. Many such studies are published by academic physicians who display little acquaintance with the fine points (nay, even the rough points) of conventional economic theory. Their economic studies seldom advance public understanding of health care issues. More often, their work detracts from warranted knowledge that might assist the public in its deliberations. Altogether, but especially with respect to taxation and expenditure of public funds, this is a most undesirable situation. The public cannot conduct its business democratically and effectively without honest debate about the general nature and specific details of problems.

This problem has been investigated in at least two other articles. The untenable technique of Cost-Utility Analysis and several examples of misuse of cost data for the related approach of Cost-Effectiveness Analysis have been criticized [Hildred and Beauvais 1995, Hildred and Watkins 1996]. The latter article pointed out the general error of using "charge" data from a single hospital as a measure of opportunity cost for CEA. This paper extends that analysis by demonstrating the more serious errors that occur when highly aggregated data from many hospitals are incorporated into CEA, and manipulated to produce impossibly precise results. It is important that those who use CEA and those who would use its results understand its limitations. If faulty data are used and the degree of precision of the several steps of the process is not carefully controlled, confidence in the results derived with the tool may be unjustified.

Improper Data Sets

Dozens of articles published in medical journals each year purport to show that one or another medical intervention is or is not "cost-effective." The very meaning of the term is notoriously fluid [see the discussion in Hildred and Watkins 1996], making it imprudent to accept at face value the findings presented in the reports. More seriously, many of these reports are probably useless because of their use of faulty data and failure to respect the problem of significant digits.

A recent publication [Lieu et al. 1994] is reviewed next to demonstrate the nature of these problems and to provide a basis for evaluating the strength of the conclusions in such studies. The study selected for this analysis is neither the best nor the worst of the genre; it is simply representative.

The Lieu study compared the efficacy of a varicella (chicken pox) vaccination program for children with the likely outcomes if the program were not carried out. The concluding argument of the article involved a comparison of several cost-effectiveness calculations and cost-benefit ratios of the hypothetical varicella vaccination program with other preventive-medicine programs, and concluded that the varicella vaccination program was superior to the others because its cost-per-life-year-saved was lower than that of the other programs. Proponents of the vaccination program would probably find this comparison advantageous, but people making a dispassionate decision about using public resources for medical care might not be persuaded. Indeed, further examination of the derivation of the study's cost-effectiveness findings, beginning with the financial data underlying the conclusion, suggests the wisdom of skepticism.

Not in dispute are the epidemiological data, such as rates of infection, complications, death, disability, or vaccine efficacy. Lieu et al. cite what are probably well-accepted data bases and published studies as the foundation for their projections. However, the cost data that they incorporate in their CEA do not carry the same authority.

Use of "Charges" to Represent "Costs"

The general problem: accounting cost vs. opportunity cost. Hildred and Watkins [1996] demonstrated that most medical CEAs are highly suspect because of their reliance on *charges* to represent costs of medical interventions. Neoclassical economic theory identifies the socially optimum price of a commodity as one that equals the "opportunity cost" of the resources used in its production. Opportunity cost is the amount that must be paid by the producer to attract and hold these resources away from competitive uses in their highest-valued alternative employment. This is also shown to equal the "marginal cost" of the firm's use of the resource, defined as the addition to total cost associated with production of the last unit of the good. Hildred and Watkins [1996] showed that derived accounting costs are greater than marginal cost, and in the typical hospital environment, charges usually exceed accounting costs (therefore necessarily exceeding marginal cost as well). Therefore,

charge data are not appropriate for CEA unless modified through the technique of "shadow pricing" (not "adjusting" by a "cost to charge ratio") to reflect marginal cost and opportunity cost.

The Lieu article contains many errors of this kind; we illustrate the ramifications of the problem by directing attention to the authors' specification of \$7,482 as the cost of treating 5- to 9-year old patients hospitalized for varicella complications. First, since the \$7,482 datum is charge-based, it is not acceptable as a measure of the "cost" of hospitalization for complications from varicella.

Specific defect: use of a mean. More seriously, \$7,482 is a *mean* of treatment charges and as such is a measure of charges for confinements that differ in unknown ways. The total amount charged for treating the 5-9 age cohort undoubtedly includes a distribution of expensive, less-expensive, moderately-expensive, and inexpensive cases, and the mean is influenced by the relative number of cases in each portion of the distribution, especially at the extremes. A measure of variance would indicate the importance of this factor, and readers have no context for interpreting the significance of the mean without information about dispersion.

Additionally, in the hospital environment, charges are *administered* prices, not an outcome of the impersonal competitive market forces of neoclassical theory but determined by the relative bargaining power of health care providers and those who pay for medical care, whether as patients (with little bargaining power) or as insurers (with varying degrees of bargaining power). Consequently, otherwise identical treatment provided to different patients generates different prices, both at a moment in time and at different times.

This mean, therefore, is not a basis for measuring "the" cost of hospitalizing children for pneumonia and encephalitis resulting from chickenpox.

Specific defect: aggregated data base. \$7,482 is the mean of treatment calculated from a *statewide data base* developed from many hospitals [377, 378]. The problem inherent in using a mean, as noted above, is compounded in an indeterminate manner by the use of aggregated data. As the mean from any single hospital masks the variety of charges and conveys no reliable information about opportunity cost, so the mean of data from multiple hospitals further buries any possibility of identifying opportunity cost. There is no reason to believe that all the variations among treatment intensities, charges, and myriad hospitals are eliminated by the mathematical process of calculating a mean, mystically leaving only the pure indicator of opportunity cost.

Specific defect: aggregate adjustment factor. Finally, the \$7,482 datum was obtained by multiplying "charges" by .53 [377], the mean ratio of cost-to-charge for

all confinements in all hospitals in California for some unspecified year. This number is unacceptable as a multiplier, first because the required adjustment must be one that gets at marginal cost, not a derived accounting cost, and the administered pricing problem is ineluctable. Secondly, a statewide ratio (presumably the calculated mean of all hospitals' cost/charge ratios for all services provided by those hospitals) conceals more than it reveals.

For example, within each hospital, a stated cost/charge ratio for coronary bypass surgery, would be the mean of charges collected from patients without insurance, patients with private insurance (whose company has negotiated a reimbursement schedule for the service), and patients covered by Medicare or Medicaid (each with a negotiated reimbursement schedule, possibly quite different from each other and certainly different from the rate laid upon the uninsured person). As a consequence, the calculated cost/charge ratio for bypass operations for any time period will depend on the mix of patients and their payment sources. Moreover, the cost/charge ratio for each hospital's provision of, e.g., maternity care, will be subject to the same variation, as will the cost/charge ratio for every service provided by the hospital. That variation should be expected to differ with each contract with third-party payers. Some services may be reimbursed more or less generously with each contract, as bargaining power shifts one way or another. The upshot of the situation is that the mean cost/charge ratio for any single hospital is misleading; for all hospitals in a state, the mean of the ratios for each discrete service is meaningless, and the mean of the ratios of all hospitals for all services is totally vacuous.

Questionable Representativeness of Data

Other sources of data are problematic. The research team reported prices of emergency room visits for varicella complications at \$163, outpatient visits at \$42, and \$3 for telephone advice, all the result of "consultations" with staff of the Kaiser health care group. Since HMOs are touted as "different" from other kinds of providers, chiefly with respect to their cost structures (and, accordingly, their accounting systems), there is no compelling reason to accept these data as representative of all health care providers.

It is truly astounding that the cost of the vaccine was not known by the investigators. "Informal discussion" is not an acceptable scientific foundation for a decision to specify a price of \$35 for each dose. What range of prices is found? What accounts for the differences in price? What proportion of buyers obtains the vaccine at the various prices? All of these questions bear heavily on the believability of the \$35 price. It is especially significant that the dosage price was the only dollar-value parameter selected for sensitivity analysis: along with vaccine administration strategies, it was "varied over plausible ranges" [Lieu et al. 1994, 378]. A reader of the article could not

determine from the information provided whether the sensitivity test variations were able to produce anything other than "noise" -- as Kamarck warns about (see below). This "datum" is so important that inability to identify it precisely is a fatal weakness.

Finally, the choice of \$10,000 as the annual cost of long term care for people disabled by the disease appears to be without any foundation. The authors noted that there are no data on this phenomenon, and simply chose \$10,000 for the calculation.

Turning now to the problem of estimating error, we believe that attention to significant digits would help achieve recognition of the range of error that may be encountered in carrying out this component of a CEA. We follow the lead of Andrew Kamarck, whose work rests firmly on the nature of measurement; its fundamental postulates are quite sound, regardless of other extensions of measurement theory that may be found in methodological literature.

The Significant Digits Problem

Kamarck has recommended that economists take a page from their colleagues in the physical sciences and adopt the practice of specifying the margin of error of their results by reference to significant digits. Noting the inordinate commitment to refining data sets in complete disregard of their accuracy [Kamarck, 119-120], he advised researchers to remember the difference of meaning when a number is used in a theoretical manner, as distinct from real-world measurement [135].

To illustrate with an example, the number 220 has three significant digits, with the number of hundreds and tens known to be two, but the number of units has resulted from rounding of a real world measurement. If the units were known exactly, the number would be written as 220.0 and have four significant digits.

The rules of significant digits include as significant all the digits to the right of the decimal point and non-zero digits to the left of the decimal point. Continuing the example, the number 200 indicates a true value between 150 and 250, while 220 denotes a true value between 215 and 225; 222 means that the true value is between 221.5 and 222.5.

Important as it may be to identify the precision of a measurement in this way, it is more crucial to be aware that the results of familiar mathematical procedures depend for their aggregate precision on the precision with which every component of the process is measured. Kamarck reminds us that when mathematical manipulations are carried out on a set of numbers, the resulting answer is "... accurate to a number of significant digits no greater and usually less than the number of significant digits in *any one* of the original numbers" [136, emphasis in original]. Continuing our

example, if the number 220 (known to be significant to two digits) is divided by 0.5 (significant to a single digit), the resulting quotient of 440 is significant to the least significant digit used in the division. In this case, the one significant digit found in 0.5 makes the quotient of 440 significant to the hundreds digit only. All that can be said is that the true number is somewhere between 350 and 450. Thus, in a calculation involving many different numbers with both large and small numbers of significant digits, the more likely it ". . . is useless and misleading to manipulate to a large number of places when one item is known only very roughly" [136].

As to the possibility of rescue through the application of sensitivity analysis, Kamarck is equally cautionary.

• Very often the results of the usual type of sensitivity analysis are completely meaningless because they are swamped by the 'noise' in the numbers concerned. For example, the parameter being examined is taken as, say, '3167' when there is in reality one significant digit: the true number should have been written as 3000. A sensitivity analysis that tests a 10 per cent variation is in fact a useless exercise since the margin of error, +500, in the number is greater than the +317 of the variation tested [140].

The first table of the varicella article [377] presents nonfinancial probability data, extending many places to the right of the decimal (in the case of the probability of death as a complication, the probability presented is 0.0000082). In addition to death, probabilities are listed for the population rate of infection (for 5-9 year-olds, serving as an example: each age group has a different probability), rates per case of major complications requiring hospitalization (and of these, specific rates for pneumonia and encephalitis), and a rate of long-term disability from encephalitis. The table also lists outcome probabilities of several medical interventions, vaccine efficacy variation, and work loss. These data are followed by another table with estimates of costs for the medical interventions, medication and vaccine, long-term care for victims disabled by the disease, work-loss by parents, adult sufferers, and the disabled and dead victims.

The key table [378] presents summative data on the findings of the simulation concerning reduction of: number of cases, major complications (including pneumonia and encephalitis), disability and death, and costs due to treatment and work loss.

These data underlie the calculation of the "cost-effectiveness" of the proposed vaccination program. Five effectiveness outcomes are presented, with and without discounting 30 years into the future. Thus, we learn that it would cost \$4.20 to prevent a case of chickenpox, \$1,650 to prevent a major complication, \$832,000 to prevent each instance of long-term disability, \$294,000 to prevent a death, and \$16,000 to achieve each "life-year saved" [379].

The team of authors conceded that *life-saving results are seldom an issue in this disease*, but because the ostensibly neutral common denominator for comparisons among medical interventions is cost per life-year saved, concluded their argument with such a comparison. Thus, the varicella program is rated against others[380]:

• Varicella vaccination, children \$16,000

Hypertension screening, mid-age men 21,300

Hepatitis B vaccination, infants 41,000

Colorectal cancer screening, elderly 45,000

Cholesterol reduction 132,000

On this reading, the varicella prevention program should be preferred to the others. However, the data problems set forth above should give pause about the cogency of the comparison, and turning to the significant digits issue, we recall Kamarck's stricture regarding the element of the chain with the fewest significant digits. In this study, it seems that the culprit has at best two significant digits, constituting the weakest link in the analysis. Beginning with the probabilities relating to effects of the disease, and putting all in the same scientific notation, we observe that there would be [378]:

• Without Program With Program

New cases $3.953 (10^6) .24 (10^6)$

Major sequellae .009930 (10⁶) .0000610 (10⁶)

Disability .000020 (10⁶) .0000001 (10⁶)

Death $.000056 (10^6)$. $.0000004 (10^6)$

In this array, the number of new cases with the program (240,000) is presented with two significant digits; it is the foundation from which the frequency of the other events associated with the vaccination program is calculated. However, following Kamarck, none of the succeeding estimates can contribute meaningfully to the calculation of effects, despite their derivation from probabilities with as many as seven significant digits

The cost data are similarly constrained: in their third table [378], the authors state the program cost at \$88 (10^6) and the reduction of costs of the disease as \$10 (10^6). Though these may be the result of summing various numbers with more significant digits, there are only two significant digits here. Thus, on both the cost side and the effectiveness side of the consideration, two significant digits limit the calculations, and most of the profuse number-manipulation yield mostly "noise."

Finally, the authors' decision to state their case for their program in terms of the cost-per-life-year-saved criterion seems completely self-defeating, given their explicit admission of the unimportance of saving lives by intervening in this disease. There are probably some persuasive considerations to support a varicella immunization program, but they are obscured by CEA, and especially by mention of cost-per-life-year-saved.

Conclusion

As a prudent reader must entertain serious doubt about the relative economic attractiveness of the varicella vaccination program, so should he or she recognize the danger in uncritical acceptance of the CEA-determined value of any medical intervention. It is untenable to concatenate treatment outcome measures, some elements of which may be accurate to many significant digits and others much less accurate, in highly aggregated summary indicators of "effectiveness" of an intervention. Some credence might be granted to a competently performed analysis of the costs of treatments, taken as a separate inquiry, but joining the unavoidable and probably large inaccuracies of cost data with considerably more accurate "effectiveness" measures only generates "noise," not scientific insight.

Users of faulty data and practitioners of faulty analysis may be expected to react negatively to criticism of their work, insisting that the critic set forth a better way, directed to the development of procedures to improve the measures and analysis. However, it is not better measurements that are required for the elevation of public discussion of health care resources. Indeed, the drive to quantify might well distract from the substance of a public dialogue that could engender understanding and selection of desired policies, especially if the results of the experts' "measurements" are deceptive. Nor is improvement in computerized manipulation of those data a solution to the problem discussed here. Even in the unlikely event that the opportunity cost data were routinely recorded, and ever more complex algorithms developed, the results of such projects must be debated intelligently by those affected by the choices to be made, i.e., the citizens of a democratic society, an activity not encouraged by obscurantist analyses undertaken to confer advantage on a particular course of action.

The "better way" is the method of social inquiry proposed by Dewey [1927]. There is no neoclassical "natural law" solution awaiting discovery; there are political solutions awaiting creation. These arise from inquiry into the sources and causes of social problems, being based on warranted knowledge, that is, conclusions reached through dialogue about verifiable evidence. Those who wish the community to acknowledge great merit in a medical procedure should do so in an atmosphere in which they must prove that merit, not least by establishing the soundness of the data and calculations offered in its defense. The community must develop a more honest and responsible discussion, in which advocacy of a position cannot take the form of the following, a comment [Anonymous, 1996] made by one who had just read the foregoing article:

I take my hat off to Lieu et al. They are pushing a good thing. . . . If they can get this vaccine in use, it will save lives. If they can cook up some calculation that improves the chances of approval, more power to them.

Bibliography

Anonymous, personal communication, August 27, 1996.

Dewey, J. The Public and Its Problems. New York: Swallow Press, 1927.

Hildred, W. and F. Beauvais. "An Instrumentalist Critique of 'Cost-Utility Analysis." *Journal of Economic Issues*, December 1995, pp. 1083-1096.

Hildred, W. and L. Watkins. "The Nearly Good, The Bad, and The Ugly in Cost Effectiveness Analysis of Health Care." *Journal of Economic Issues*, September 1996, pp. 755-775.

Kamarck, A. *Economics and The Real World*. Philadelphia: University of Pennsylvania Press, 1983.

Lieu, T. A., S. L. Cochi, S. B. Black, M. E. Halloran, H. R. Shinefield, S. J. Holmes, M. Wharton, and A. E. Washington. "Cost-effectiveness of a Routine Varicella Vaccination Program for US Children." *Journal of the American Medical Association*, February 2, 1994, 375-381.

Sox, Jr., H..D., M. A. Blatt, M. C. Higgins, and K. I. Marton. *Medical Decision Making*. Boston: Butterworth Publishers, 1988.

Tool, M. The Discretionary Economy. Santa Monica, CA: Goodyear, 1979

College of Business Administration Northern Arizona University Box 15066 Flagstaff AZ 86011 520-523-3657

Limitations of Cost-Effectiveness Analysis: Defective Data and Significant Digits

Working Paper Series 97-01 — April 1996

William Hildred

Professor of Economics
College of Business Administration
Northern Arizona University
Box 15066
Flagstaff AZ 86011
Bill.Hildred@nau.edu

James Pinto

Professor of Economics and Statistics
College of Business Administration
Northern Arizona University
Box 15066
Flagstaff AZ 86011
James.Pinto@nau.edu

A version of this paper was read at the Association for Institutional Thought section,
Western Social Science Association annual meeting in Reno, Nevada, April 19,

1996