Why Society Needs to Value Health Improvements in Dollars

David Bishai

Hopkins Population Center Papers on Population

WP 98-04


Financial support for this study was provided in part by a grant from the Johns Hopkins Population Center (NICHD grant P30-HD06268). The funding agreement ensured the author’s independence in designing the study, interpreting the data, writing and publishing the report.

Correspondence may be addressed to

Dr. David Bishai, MD, MPH, PhD
Department of Population and Family Health Sciences
Johns Hopkins University School of Hygiene and Public Health, Room W 4503
615 N. Wolfe Street
Baltimore, MD 21205
Phone (410) 955-7807 Fax (410) 955-0792
dbishai@jhsph.edu
Abstract

Background. Lacking information about how much a population wishes to spend for improved health, health planners must use subjectivity in setting health objectives.

Objectives. This paper indicates how the absence of information about how much resources a population would trade for better health could lead to allocations that may be efficient, but which could still lower welfare.

Methods. I depict the process of health production and welfare maximization graphically and mathematically to distinguish efficiency from the social optimum.

Results. Not every point of efficiency will improve the welfare of society. A planner knowing only the relative cost-effectiveness of various health interventions could do harm by imposing an efficient but non-optimal regime.

Conclusions Information about society's preferences could assist exercises such as the US's Healthy People 2010, and WHO's Evidence and Information for Policy. Decision-makers using such information should be apprised of current philosophical and empirical inherent limitations.

Acknowledgement

Helpful comments from W. Henry Mosley, Constance Nathanson, Kenneth Hill, and Robert Kelly are gratefully acknowledged. All errors are my own.
In this paper I develop the proposition that quantitative measures of social preferences for health are a vital missing element of many analyses in the cost-effectiveness literature. I will use some basic tools from welfare economics to show how an exclusive focus on achieving health outcomes based only on technical feasibility can actually lower social welfare. I will argue here in favor of a cost benefit approach as opposed to an exclusive cost-effectiveness focus. Both measure the cost of a desired outcome; in cost effectiveness analysis the outcome remains defined in terms of clinical parameters such as QALYS (quality adjusted life years) or DALYS (disability adjusted life years) etc., while in cost benefit analysis there is an attempt to measure the benefit in dollar terms which will permit decision-makers to decide on tradeoffs between health outcomes and non-health outcomes. Analysts are attracted to the cost-effectiveness approach because it appears to avoid the thorny issues of placing a dollar value on health. True, these issues are thorny, but ignoring them does not make them go away (1). When decision-makers outside the health sector must daily make budgetary tradeoffs between health and other sectors it is of no assistance to state that health is a sacred commodity beyond monetary value. Whatever the attraction of such a stance, it has the practical effect of leaving decision-makers in an information vacuum.

Healthy People 2000 is one example of an exercise in priority setting (2). The Global Burden of Disease is another example (3). The World Health Organization has already begun efforts to further prioritize health interventions based on cost-effectiveness with the development of a cluster devoted to evidence and information for policy. With costs attached to some objectives, planners can begin to prioritize on the basis of technical efficiency. It is the hope of fueling such prioritization exercises that drives much of the frenzy to compute the cost-effectiveness of various interventions. The vast accumulation of information on cost-effectiveness that has occurred over the last 10 years will probably have an influence on the health objectives set for the next decade. Planners realize that a list of health objectives that pays no regard to the cost of their attainment is of less use than a list that does. How does one shop from a Sears Wishbook that doesn’t list prices?

The immense amount of effort ongoing to measure cost-effectiveness is testament to the wide recognition that cost information is important for policy. I have no intention of belaboring this point here. Rather, I argue here that even if a list of health objectives incorporated unlimited information on costs per adjusted life year it would still fail to guide planners in the optimal direction. A planner must know about the demand for health. Stretching the
Sears Wishbook analogy further, how does one successfully shop for a friend (with their money) without knowing their preferences (and budget limits)?

It was not accidental that Healthy People 2000 and Global Burden of Disease do not incorporate measurements of the social demand for health. The real problem for the authors of these projects is that other than some attempts to value changes in environmental quality there is almost no basis in the literature for inferences on the demand for most public health goods. And there still is remarkably little known about this issue. A recent review of applications of the WTP technique found only 48 such studies in the literature (4). This stands in contrast to the thousands of published attempts to answer “Is technology T an efficient way to produce health?” using cost-effectiveness analysis. In short, we have had few serious attempts to address empirically the question: How healthy does this society want to be?

Past arguments in favor of the CBA approach have stressed advantages such as the greater transparency of distributional assumptions (5, 6) and the ability to easily adapt to accommodate multiple program outcomes including survival, quality of life, lost productivity and altruistic concerns about fellow human beings (6). Indeed some criticisms of basic tools in CEA center around the subtleties of distribution and adaptability (7). In this paper, I focus on the difficulty a society would face in attaining an allocation of health care resources that is distributionally and technically optimal unless society’s preferences are assayed.

To support my argument that the optimal health objectives depend on a knowledge of social preferences, I will appeal to an economic definition of the social optimum and explain how this is defined. Social preferences can accommodate the value that society places on equity as well as efficiency. I hold that an exercise in defining health objectives ought to show how the objectives are related to the social optimum, and ideally would define them as exactly that. I will show how knowledge of social preferences is fundamental to defining the social optimum. Without this knowledge it is possible to reach a point that might be more technically efficient but could offer lower social welfare than the status quo.

The Optimum: Where Cost = Benefit and Benefit Includes Equity

The standard notion of economic efficiency in the pursuit of an objective is to devote resources to it until “(marginal) cost equals (marginal) benefit”. Applying this concept to the setting of a nation’s health one would
define the social optimum as the point at which the "marginal social cost equals marginal social benefit". Because this concept is fundamental to any exercise in objective setting, the meaning of each term will be defined.

Marginal Cost

Let us suppose that all government funding for public health activity, M, is drawn from total government spending, G where G is fixed by the political process. Spending M leaves G-M=Z to be spent on other valuable public goods such as roads, monuments, and the military. The public health sector is assumed to use M in the most efficient manner to improve the health, H, of the population by lowering age-specific morbidity and mortality rates. (Leave aside temporarily decisions about social preferences for targeting mortality vs. morbidity and targeting the young vs. the old.) If M=0, then health is at a minimum, H0, determined by the personal efforts of individuals with no help from the government. If M=G, then health is at a social maximum, but Z=0 and there is no public spending on anything except health. To depict the tradeoff between H and M, economists use a graph such as the one in Figure 1 called a production possibility frontier.

If the society starts at point A shown in the diagram, it is able to enjoy ZA=G-MA public goods and a level of health H_A units above what it would have if there were no public health spending. It achieves H_A by spending M_A on public health. To achieve a higher increment of health denoted ΔH, a decrement of other goods -ΔZ must be endured to finance the investment of +ΔM. At any given point on the curve, the tradeoff between health and other public goods is the government's marginal cost of producing health. The marginal cost is the inverse of the slope of the curve shown in Figure 1. The word "marginal" is used here to describe the notion that the cost is that of an infinitesimal increment. It is a notion that is best described with the notation of calculus, dM/dH. Note that the marginal cost of health is very low (small change in M associated with big change in H) when health is near the minimum and the cost is very high when health is near the maximum. This conforms to the notion that in prioritizing, the public health sector will do the most cost effective things first and reserve more costly interventions for later. A common mistake is to confuse point X, where H+Z is at a maximum with the socially optimal point. This is untrue. The social optimum depends upon what society holds valuable. Figure 1 only describes the supply side of public health. What remains completely unknown is the demand side of public health.
Marginal Benefit

To depict the notion that both health, H, and other goods, Z, are valued by society, a natural tool would be the social welfare function \( W(H, Z) \). This function has the property of assigning number values to combinations of H and Z. One assumes that as H and Z increase \( W(H, Z) \) increases. One also might assume that as H and Z increase, \( W(H, Z) \) increases somewhat less rapidly for equivalent increments in H or Z as society shows satiation. Equity concerns can be embedded in \( W(H, Z) \)—W may rise faster when health improvements are distributed equally in society or alternatively when they are concentrated in the least well off.

Figure 2 depicts how \( W(H, Z) \) which is a three-dimensional surface, increases in H and Z. The dark bands are called indifference curves. Indifference curves are collections of points depicting combinations of H and Z that offer the same amount of welfare, W. The indifference curves furthest to the NorthEast offer the highest welfare level. The (negative) slope of any indifference curve can be called the marginal benefit of substituting H for Z. The socially optimal allocation is where the amount of Z and H is deployed in the most efficient manner by the supply side to permit the enjoyment of welfare on the highest indifference curve attainable. This point, S.O., is where the slope of the production possibility curve and the slope of the indifference curve are equal. It is where the marginal cost of substituting Z for H equals the marginal benefit. Note well that point X while technically very attractive actually offers a lower level of social welfare than point S.O. This is due to the social preference structure inherent in \( W(H, Z) \).

**CE Ratios Are Not Enough**

Since there is little known about efficient public health production and less known about the preferences of the public in trading public health against other public goods, it is quite likely that the present allocation is some distance away from both a point of efficiency and the social optimum. Billions of dollars are probably also being wasted because of obstacles to the definition, dissemination and implementation of technically efficient practices in public health. Billions more dollars are probably being wasted annually through over or under provision of public health. Technical inefficiency means that society is operating on a point inside the production possibility curve, such as point Y in Figure 2. But the argument that there is plenty of work to be done in defining technically cost effective practices is no reason to defer efforts to define social preferences. Simply attaining any point on the production possibility curve may leave society worse off than it is at point Y. Note in Figure 2 that a movement from point Y to an “efficient” point such as X would represent a decrease in welfare! A policy maker at point Y needs to know
which direction to aim in, in order to reach the social optimum—only knowledge of social preferences can provide that.

In theory, democratic governments take account of the public’s preferences through the electoral system. If public health priorities are greatly out of step with the preferences of the electorate, presumably elected officials will become aware of this and signal public health planners accordingly. For example, Washington DC abounds with health interest groups using the political system to signal planners about public health priorities. The transmission of public preferences for public health priorities exclusively through the electoral system has the disadvantage of delivering such a weak, potentially distorted, and intermittent signal regarding the shape of the true social welfare function that it behooves planners acting in the public interest to find techniques to augment their insights into the nature of the public interest. W.H. Frost writing in 1925 presaged this need as follows:

As various lines of activity are suggested to him (the health officer) by special propagandists it is his duty to consider what returns are to be expected from each one and to decide how much of his capital he will put into each one. Since his capital comes entirely from the public, it is reasonable to expect that he will be prepared to explain to the public his reasons for making each investment, and to give them some estimate of the returns which he expects. Nor can he consider it unreasonable if the public should wish to have an accounting from time to time, to know what returns are actually being received and how they check with the advance estimates which he has given them. Certainly any fiscal agent would expect to have his judgement thus checked and to gain or lose his clients’ confidence in proportion as his estimates were verified or not (8).

The Economic Model of Health Production

Note that in Figure 2, H does not represent the total stock of the nation’s health, but merely that portion of the nation’s health that can be influenced by public spending on public health. The following adaptation of the Mokyr health production function can make this concept more precise(9). This mathematical depiction breaks health production into 3 contributions: from the environment, from the public health sector, and from individual activities (which may be partially financed by government). The formula is given as:

\[ h_{ij} = E + G \left[ (B_j - \phi_j)M \right] + F \left[ (A_i - e_{ij})X_{ij} \right] \]

- \( h_{ij} \) is the health of the i-th household in area J
- E is environment
- G is Government health production
- \( B_j \) is Best available Public Health Technology in area J
- \( \phi_j \) is the gap between Best and Actual in area J
- M is goods and services purchased by govt. e.g. wastewater treatment plants, safety patrols, and communicable disease control departments
- F is a household health production function for household i.
A_i is Best available household technology in community i
\( \varepsilon_{ij} \) is the gap between best and actual by ith house in community j
X_{ij} is goods and services purchased by household i. e.g. medications, surgeries, medical counsel

The advantage of the Mokyr production function is that it offers a division of responsibility between households investing resources X_{ij} and governments investing resources M. Households demand X_{ij} and finance it over the lifecycle by making payments of money (and service) to providers, to private insurers and to government-sponsored social insurance programs. Thus in this framework much of government involvement in the health sector (Medicare, Medicaid, VA, Champus) is selectively facilitating the trading that permits households pursuit of personal health through F[X_{ij}]. In the U.S. system (and even in a centrally funded health system) individual health choices, X_{ij} still dominate the decision to seek and comply with medical treatment.

Many of the objectives listed by the Public Health Service in Healthy People 2000 are clearly mediated by household F[X_{ij}] decisions, for example exercise objectives, dietary intake, and substance use. The government can seldom choose to expend X_{ij} on a subject's behalf. A government can devote a portion of M to cajole, subsidize, entice, or legislatively coerce consumption of X_{ij}. Finally, a small amount of government spending is allocated to health promoting activities which are on a scale that households cannot pursue. Sanitation, public safety, and the control of communicable diseases are mainstays of the G[M_j] function. Over time a government can use M to fund research which will raise the public and private levels of technology, A and B.

**Social Welfare Functions**
As discussed above, I am keeping all of the choices regarding equity, redistribution, coverage, and funding of government sponsored health insurance functions out of the current analysis. Policy choices in this arena can affect public health by affecting the prices of X, but they operate through F(X_{ij}). After policy sets the health insurance and access environment, the choices between medical payments/premiums and other spending are the household’s.

Leaving taxation, redistribution, and entitlements aside, the social planner must choose Z and M to give society the highest social welfare (maximum of W(H,Z) ) provided that health is produced according to equation [1] and that the sum of spending on public health and everything else not exceed the available budget (Z+M=G). The point depicted by S.O. in Figure 2 will have the following mathematical property:
Social Benefit from Improved Health \[ \text{Social Benefit from Other Govt. Spending} \] \[
\frac{dW}{dH} \bigg|_{P_H^*} = \frac{dW}{dZ} \bigg|_{P_Z}
\]
which says in English

Social Benefit from Improved Health \[ \text{Social Benefit from Other Govt. Spending} \]
\[
\frac{\text{Effective Price of Health}}{P_H^*} = \frac{\text{Effective Price of Other Govt. Spending}}{P_Z}
\]

Here \( P_H^* = \frac{P_M}{\frac{dH}{dM}} \) is renamed the "effective" price of H. It states that the optimal allocation between Z and M is found where the marginal benefit to social welfare of having a little more H and a little less Z is equal to the effective marginal cost of reallocating spending to M from Z. Marginal social benefit equals marginal social cost.

**Philosophical Problems**

Intensely difficult problems in social policy and social choice theory abound in this framework. Foremost are the ethical problems of defining aggregate health, \( H_i \), and aggregate social welfare. The decision about how to base a concept of social good on measures of individual good is an ethical one. For instance, the simplest aggregation strategy makes profound moral claims. Suppose I state that public health is an aggregate of individual health as follows:

\[ H_j = \sum_{i=1}^{N} \alpha_i h_i \]

or that social welfare is an aggregate of individual welfares (utilities) as follows:

\[ W(H,M) = \sum_{i=1}^{N} \gamma_i U(h_i, s_i, M_j) \]

where \( h_i \) is the health status of the ith household measured as the stock of expected quality adjusted life years remaining, \( s_i \) is the consumption of all other goods, and \( M_j \) is the countrywide availability of all public goods. \( U \) is a utility function.

The usual utilitarian stipulation that \( \alpha_i = 1 \) and that \( \gamma_i = 1 \) for all \( i \) is not trivial. This would be a moral claim that the health of each individual is of equal importance. A Rawlsian might claim that the least well off ought to weigh more heavily, e.g. \( \alpha_i > 1 \) if \( y_i = y_{\min} \). Myriad variations abound. Which one is fair? How could one even
know which one is fair? It is not “obvious” that voting to decide issues of fairness is fair? Moral judgements about interpersonal tradeoffs will have to precede application of the willingness to pay method (11).

One approach to addressing equity concerns might be to directly assess the public’s preferences for equity (12, 13). An interim approach might be to begin to present WTP results based on permutations of weighting schemes for $\alpha_i$ and $\gamma_i$ (14). For example various estimates of cost-benefit could be presented based on egalitarian weights, Rawlsian weights, Benthamite weights, etc. Two practical results might flow from an accumulated stock of results which demonstrated how various moral assumptions would translate into different resource allocations in the health sector. First, these efforts could speed the evolution of ethical thought by offering “policy consequences” to those whose moral judgements are swayed by consequences. Second these results could also speed progress by underscoring a desperate need for guidance in an area where policy decisions are made daily with unexamined assumptions and rules of thumb (15).

**Arrow’s Theorem**

Not only is it difficult to assign weights, but there are special properties that one would expect a social welfare function to have that are impossible to achieve. One would want a social welfare function such as $W(H,M)$ to 1) be able to reach decisions across the entire range of options; 2) to represent the welfare of more than one person; 3) to increase even when only one household achieved higher utility leaving all others unchanged. Arrow (16) was able to prove that any social welfare function that obeyed these 3 conditions would fail to have a fourth desirable property requiring that choices between $H$ and $Z$ should be unchanged by considerations of how $H$ is preferred to some unrelated choice. There can be no social welfare function that obeys all of the conditions that pure logic demands. Given this result one can choose a nihilistic view that since there can be no “logical” social welfare function, random guessing in policy decisions is as good as reason. In contrast to nihilism, Nobel Laureate Amartya Sen (17) has given a thoughtful consideration of which of the 4 Arrow criteria are most dispensable in different settings and offers grounds for compromising the so-called “purity” of a social welfare function in order to take advantage of the guidance it can offer.

The bulk of policy decisions are guided with minimal data on effectiveness and with the decision-maker’s “gut instinct” about what is best for social welfare. Given the Arrow theorem, would policy decisions guided by effectiveness data and a social welfare function be any better for society than those guided by instincts? Would a
social welfare function that must fail to meet at least one of the four Arrow conditions be better than instincts which are not obligated to meet any of the four?

These tremendous philosophical quandaries are daunting. Yet each day vital decisions are being made under a status quo of immense uncertainty. I only propose that by injecting a dose of improved information about social preferences into the system, one can do better than the status quo while adhering to a policy of primum non nocere. One need not aspire to rewrite the constitution nor re-organize the government. Leaving the institutional structure as it is, I encourage more researchers to begin to offer decision-makers quantitative information about social preferences which can supplement gut level instinct about what best serves the public interest. Such information needs to be presented with the fullest possible disclosure of its limitations and assumptions.

**Measuring Social Welfare Functions**

Each method of measuring the social demand for health is fraught with potential errors and requires various assumptions. Information produced by a single method may be unreliable. For this reason, I suggest using several methods jointly to reconnoiter the profile of the social welfare surface shown in Figure 2.

The most obvious method for measuring social preferences is known as the contingent valuation method. Marketed goods can be easily assigned a value by observing the price that people actually pay for them. For non-marketed goods such as public spending (M) on health, a hypothetical or “contingent” market can be devised and a set of representative subjects can be asked how they would evaluate items in such a market. Thus for a health objective such as the reduction of exposure to air pollutants, one could confront the subject with a hypothetical market where living conditions at various levels of air pollutants are described at length with photographs and vignettes and then ask subjects how much they would have to be paid to accept each of the various air pollutant levels. Conversely they could be asked how much they would wish to pay to obtain an improvement in the air pollutant levels. Results such as these have been accumulating in the environmental economics literature(18-20), and the technique has been extended to other public health goods such as poisoning risks (21), automobile safety (22), and occupational safety (23).

Contingent valuation methods have well-known limitations (24, 25). Subjects might inflate their willingness to pay if they think that others will be asked to subsidize goods on their behalf, they might deflate their answers if they think that the results will be used to selectively increase their share of the tax burden. Subjects also might not understand the hypothetical goods they are being asked to evaluate (26).
Another method that could complement contingent valuation is the property value approach (27-30). Property values are observable, and should in principle reflect some of the value placed by the market on the public goods one can enjoy at any particular location. Controlling for the quality of the physical structure, and location relative to socioeconomic amenities, variation in property values should correlate with variation in air quality, sanitation, safety, etc. Studies that employ this method must be very careful to control for potential omitted variables or confounders. Furthermore not all public health changes can be localized to a piece of property, e.g. information dissemination.

Other limitations of existing social preference elicitation methods are beyond the scope of this paper. The difficulties have led some to abandon efforts to measure social preferences, but the great importance of this information for decision-making has continued to attract a gradually increasing number of investigators (26). Perhaps the best approach to countering the technical limitations is to check the validity of results obtained by one method against results obtained by another (31-33). Summing up a committee charged to review these methods Kenneth Arrow opined, "If you are finding, by two conceptually quite different methods, numbers that are the same or similar, then--while you can't be quite sure that the reality that you are reaching is the reality that you want--at least you are reassured that you are likely to be measuring something real." (32).

Conclusions

Other authors have offered answers to the question "Why should society measure health improvements in dollars?" The answer offered here is that the comparability achieved by having both costs and benefits expressed in the same units is not just a theoretical luxury, it is a necessity for rational resource allocation. Having the wrong dollar limit on total health spending can be a costly mistake for a society. Without clearer signals about the value of both health improvements and the distribution of health improvements mistakenly low or high health budgets will be unavoidable. Spending the wrong amount of money perfectly efficiently can be just as wasteful to an economy and as harmful to people as squandering large fractions of the right amount of health spending. Although health planners often suppose that the dollar limit on public health spending is set by forces beyond their control, this assumption should be questioned. Why do we continue to force legislators who are told the economic returns of building a dam or a bridge to compare these dollar returns to 100,000 QALYS? If researchers in the public health community do not assess the dollar value of health improvements, nobody else will.
I hope that the fair warning offered here about the tremendous obstacles on the road to understanding the dollar value of health is not overly discouraging. Biomedical scientists possessed of vision have traveled more daunting pathways to bring us miracles we now take for granted. It is time for social scientists to take up this challenge because the destination is one we literally cannot afford to fall short of. Evidence is emerging that the research community has the fortitude.

The vision that has propelled government and multilateral agencies to rationalize the allocation of scarce health resources has been a powerful force for institutional transformation. Projects in priority setting that were once the stuff of dreams have already been implemented. Hopefully this vision will soon translate into redoubled efforts for explicitly deferring to the measured preferences of individuals and communities who must forego the resources used to attain health objectives.
Figure 1: Production Possibility Curve: $H=$ Health, $M=$Public goods used to produce health. At $M=0$, nothing is spent on health production and health is at a minimum. At $M=M_A$, the economy produces $H=H_A$. If public goods are not used to produce health they can be used for other valuable purposes and denoted as $Z$. Total government budget, $G=M+Z$. In the figure, point X maximizes the arithmetic sum of $H+Z$, but may not turn out to be the socially optimal point.
Figure 2. Production Possibility Curve (Solid Line) and Indifference Curves (Dotted Lines) with $W_5 > W_4 > W_3$ etc. $H =$ Health $Z =$ All other public goods net of expenditure of M. The Socially Optimal Point is denoted S.O.
REFERENCES


