Cost-of Illness Approach

Cost-of-illness (COI) estimates provide an accounting of the dollars spent on medical expenses and the dollars of employment compensation that are forgone as a result of illnesses, accidents, or premature deaths. Such an accounting provides useful information to economists and policymakers because it indicates the magnitude of the economic flows resulting from government programs that improve public health.

COI estimates frequently have served as a measure of the monetized benefits of government programs that promote health and reduce the number of premature deaths, illnesses, or injuries (the value of program benefits are the costs that are avoided). However, COI estimates do not provide reasonable measures of the social value of program benefits nor do they provide a consistent gauge of the severity of illness. COI measures are influenced by transitory variables such as the distribution of education, employment opportunities, income, the current state of medical technology, and the characteristics of the institutions through which medical services are bought and sold. The influence of these transitory variables and the effects of income and circumstance erode the usefulness of COI estimates as measures of social welfare or disease severity.

The Components of Cost-of-Illness

COI estimates are composed of two types of costs: direct and indirect. Direct costs are expenditures for medical goods and services such as medications, doctor visits, and hospitalization. Indirect or human capital costs are the present value of labor earnings that are forgone as a result of an adverse health outcome.[5] Specifically, indirect costs per person can be expressed as

$$Indirect\ costs = \sum_{t=0}^{\infty} \frac{P_t E_t}{(1+r)^t} \qquad (4)$$

where E_t is the individual's earnings in year t, P_t is the probability of surviving until year t, and r is an interest rate, measuring the opportunity cost of lost earnings. The discount rate, $(1 + r)^{-1}$, converts future losses into today's dollars. Direct and human capital costs are summed to yield a COI estimate.

For a particular illness, the comparative sizes of direct and indirect costs depend on the characteristics of the illness and the technologies associated with the illness. Mushkin (1979) argued that, over time, biomedical research, technological change, and new diagnostics should result in proportionally higher direct costs. She showed that from 1900-1975, direct costs did rise as a proportion of total costs: in 1900, direct costs of illness were 10 percent of total cost while in 1975 they were 25 percent of the total. Mushkin hypothesized that medical advances would serve to equate direct and indirect costs, and in fact, there are many cases where advances have raised direct costs and lowered indirect costs. For example, Calabresi and Bobbitt (1978) note that, prior to the invention of the kidney dialysis machine, kidney failure was quickly fatal. COI estimates from kidney failure prior to invention of the machine would include very low direct costs and high indirect costs. Using the machine, COI estimates would include very high direct costs (especially in the years immediately following its invention) and reduced indirect costs (as patients return to work).[6]

In cases where illness results in extensive morbidity or premature mortality, indirect costs still greatly outweigh direct costs. Experience with COI estimates for foodborne pathogens suggests that, in general, the relative share of total costs due to medical expenses is lower for pathogens that are more likely to cause deaths or disability. Extrapolations from published estimates of foodborne-illness costs (Buzby et al., 1996) indicate that in 1993, direct medical expenditures accounted for between 30 and 50 percent of

[5] The COI approach is sometimes called the earnings-expenditure approach.

[6] Tullock (1995) argues that the current debate over the extent of government involvement in health care was precipitated by increasing ability, at increasing cost, of curing diseases.

total costs of illness for cases of *Salmonella* (non-typhoid), *Campylobacter jejuni* or *coli*, *Staphylococcus aureus, Clostridium perfringens*, and *Listeria monocytogenes*, 12 percent for *E. coli* O157:H7, and only 3 percent for cases of *Toxoplasma gondii* (Golan et al., 1998). The distribution of costs between medical and productivity loss depends on the rate of death and disability for each pathogen.

In the sections that follow, we examine and critique the theoretical basis for using the COI approach to measure costs associated with morbidity and mortality. We first examine the theoretical underpinnings of the indirect or human capital component of the COI measure and then turn to the direct cost component.

Human Capital Costs and the Wealth of Nations

The human capital approach is based on the assertion that the cost to society of adverse health outcomes is the impact that such outcomes have on national income. Robinson (1986) traces the philosophical underpinnings of the human capital approach to the economic doctrine dominant from the beginning of the 19th through the middle of the 20th centuries. According to this doctrine, the best government policy is the one that most effectively furthers the "wealth of nations," as measured by national income. The human capital approach to valuing life is consistent with this doctrine. With the human capital approach, the value of a life is measured in terms of its contribution to national income, i.e., to the wealth of the nation. The human capital approach is based on the assertion that social welfare is diminished by illness, disability, and premature death to the extent that these outcomes diminish national income.

The use of forgone earnings to measure the value of health and life therefore hinges on two assertions. First, that changes in health status are reflected in changes in earnings and national income and, second, that national income is a valid measure of social welfare. Both assertions must hold for the COI approach to provide a valid measure of change in societal well-being. The first assertion is often, but not always, true, and the second is usually false. Both are examined below.

Early proponents of the human capital approach argued that investments in health contribute to economic growth (particularly Mushkin, 1962), and this notion is reflected in many modern debates concerning investment in health and safety. It is argued that though many investments in human health and safety might appear to restrain production and national income (through, for example, restrictions on unsafe but productive production procedures), these investments ultimately augment human capital and lead to increases in both the number and quality of people in the workforce, thereby increasing national income and social welfare.

It is reasonable to assume that a healthy labor force is more productive than an unhealthy one, and empirical work has established a connection between health and ability to earn income (for example, Mullahy and Sindelar, 1995). However, health and income need not move lock-step for everyone. Harberger (1971) presents the example of a coal miner with silicosis who voluntarily quits a $7-an-hour job in the mine to take a $2-an-hour job clerking in a grocery store. Though the miner's health would improve, his earnings and his productivity as measured by the value of production would go down. Clearly, increases in health do not necessarily lead to increases in national income.

The second assertion, that national income is an accurate indicator of societal well-being, is even more problematic. As Mishan (1975) observes, there are many ways to increase output without necessarily increasing a society's welfare:

> . . . although financial journalists manage to convey the contrary impression, maximizing GNP is not an acceptable goal of economic policy. If it were, the simplest way of promoting it would be to adopt a policy of virtually unlimited immigration—accepting immigrants up to the point at which the value of their marginal product is zero. (p. 301)

Mishan continues by noting that simply tabulating income and the number of productive bodies in a society is not an accurate gauge of social welfare. If it were, then the death of someone with a negative net present value earnings stream would result in a net benefit to society. Indeed, strict adherence to a national income theory of well-being leads to the

uncomfortable conclusion that unproductive members of society detract from social welfare. Mishan rejects this "net output method" criterion because it does not incorporate the welfare of the potential victims, restricting itself to society *ex post* and ignoring society *ex ante*. Other economists have rejected the moral implications of a criterion that provides assistance only to those whose contribution to net output is positive (Devons, 1961; Ridker, 1967).

Other criticisms of national income as an indicator of societal well-being are a bit more subtle. Samuelson (1950) was one of the first to criticize national income as a welfare measure with his observation that when there are two or more individuals in society, maximizing aggregate income yields ambiguous welfare implications. However, not only does national income gloss over distributional issues in calculating welfare (the aggregation problem), it also fails to account for "non-material utilities." National income accounts only for goods and services that are bought and sold, meaning that as long as society places any value on non-market goods, services, or intangibles, social welfare measures derived with national income measures will diverge from true social welfare. For public health issues, the obvious failing of COI is its inability to account for the value of pain and suffering avoided. For example, the observation that the ex-coal miner breathes more easily after changing professions is not reflected by an increase in national income.

The list of intangibles and non-marketed goods that add to the well-being of a society is quite long, and, as a result, the correspondence between income and social well-being is not a reliable one. Frankel (1952) discusses three general circumstances in which income and well-being may diverge, and his general observations are echoed in many of the modern critiques of national income as a measure of social welfare (Usher, 1994, reproduces this list). Frankel's first observation is that income is only part of welfare and that increases in income may not lead to increases in welfare if another aspect of welfare is affected adversely in the process (examples include economic growth that results in severe environmental degradation or increases in income spurred by abusive use of child labor). Second, Frankel notes that some social problems are perceived only when a degree of prosperity is attained. He observes that the fault-lines of society, such as an inequitable distribution of access to income or unequal civil rights, might become more pronounced as income grows. Frankel's third observation is that social well-being gives meaning to economic welfare and not the other way around. Frankel argues that the nature and composition of income and economic welfare are not found outside society, but are formed and determined by the institutions, laws, customs, and beliefs of each society.

Empirical evidence supports the argument that national income is not a good gauge of well-being. Standard measures of income and wealth, including GDP, have long diverged from a wide class of measures of well-being. Miringoff, Miringoff, and Opdycke (1996) calculate an index of social health composed of indicators of infant mortality, child abuse, children in poverty, drug abuse, unemployment, homicide, and poverty among those over 65. They find that while GDP grew at a rate of 3.2 percent per year between 1970 and 1979, the social health index declined 2.6 percent per year. Though GDP and the Dow Jones have clearly exhibited long-term increases, many measures of well-being, especially for the poor, have declined.

The legitimacy of the human capital measure as an indicator of changes in welfare resulting from changes in health status hinges on the twin assertions that changes in health status are reflected in changes in national income and that national income is a valid measure of well-being. As illustrated above, earnings and national income do not always mirror health status, and national income is not a reliable gauge of social well-being. The human capital measure of the cost of illness does not measure changes in social welfare and these measures are therefore not appropriate for use in cost-benefit analysis.

Direct Costs of Illness Measure Individual Costs

The direct costs of illness, i.e., expenditures on medicines, health services, and other defensive goods and services, provide an indication of *individual* welfare loss. The welfare cost of these direct expenditures to the individual is the forgone utility resulting from the shift in expenditure patterns. To pay for the medical expenses from the illness, the individual must take money out of savings or reduce other consumption

activities thereby losing the utility of these consumption and savings activities.

However, though the amount of money spent on medical care entails an equal drop in consumption or savings for the *individual*, the same is not true at the *societal* level. Direct expenditures do not correspond to a drop in income or consumption for the economy as a whole, they simply constitute a redirection of economic activity, with some sectors of the economy actually benefiting from increased economic activity. In fact, like all defensive expenditures, medical expenditures are registered as positive additions to national income. For example, all of the defensive expenditures resulting from an outbreak of *E. coli*, such as emergency room care and kidney machines, would lead to increased economic activity in the medical services and equipment sectors of the economy. Increased output in these sectors could actually have a positive impact on national product. For example, Golan et al. (1998) found that in the case of foodborne pathogens, diverting expenditure from general goods and services to expenditure on medical goods and services, including residential care facilities, had a positive net impact on economic activity and income.

At a societal level, direct expenditures for medical care stimulate economic activity in some sectors of the economy, producing welfare gains in those sectors, and stifling economic activity and welfare in other sectors. There are gainers and losers from direct expenditures; these numbers do not correspond to a simple drop in social welfare. So, though the direct costs of illness measure individual costs, simply summing these costs does not result in an accurate measure of societal costs.

The fact that human capital costs strive to estimate societal costs while direct costs measure individual costs results in an uneasy marriage when the two are combined in COI estimates. This internal inconsistency further undermines the usefulness of COI as a measure of either societal or individual welfare.

COI as a Measure of Disease Severity

Practitioners of the COI approach tend to concede its limitations as a measure of individual welfare changes resulting from illness or premature death but defend the approach as a straightforward measure of the economic impact of disease. However, the straightforwardness of the approach is misleading. Practitioners of COI are susceptible to the assumption that it provides a direct measure of disease severity. This is not true. Both the human capital component and the direct cost component are influenced by a number of factors besides disease severity. A number of these factors are examined below.

Most glaringly, the human capital component of the COI reflects the current distribution of earnings, which in turn reflects the current distribution of education and job skills. In other words, indirect costs are greatly influenced by socio-economic characteristics including race and sex. As a result, value-of-life estimates calculated with the human capital approach indicate values for women, minorities, and the unskilled trailing behind those of white males. Robinson (1986) quotes a study by Cooper and Brody (1976) in which they estimate the value of a college-educated white man between the ages of 25 and 29 at $475,000, a similarly aged white male high-school dropout at $248,000, and a similarly aged African-American male high-school dropout at $165,000. They value a white female high-school dropout at $140,000 and an African-American high-school dropout at $108,000. COI estimates would therefore indicate that a disease that strikes only white males is more severe than a disease that strikes only African-American males or only females, even when the incidence and symptoms of the diseases are identical.[7]

The COI approach might also indicate that illnesses in economically developed countries are more severe

---

[7]Quoting the Old Testament, Berndt (1991) observes that differential valuations of human capital have been around for a long time.

> The Lord said to Moses, 'Say to the people of Israel,...your valuation of male from twenty years old up to sixty years old shall be fifty shekels of silver, according to the shekel of the sanctuary. If the person is a female, your valuation shall be thirty shekels....And if the person is sixty years old and upward, then your valuation for a male shall be fifteen shekels, and for a female ten shekels.' *The Bible*, Revised Standard Version, Leviticus 27:3-7

than similar illnesses in economically developing countries. For example, a study on the impact of global warming assigned lower dollar values to the lives of residents of lesser-developed nations than to residents of industrialized nations: the human capital approach led to values differing by an order of magnitude (cited in Pearce, 1995). Global cost-benefit analysis incorporating this valuation bias would lead to equally biased policy recommendations. That is, value choice could influence whether storm barriers are cost-effective in Bangladesh or whether nations that use relatively larger quantities of fossil fuels should plant trees to reduce CO2 levels in the atmosphere. Differing value-of-life estimates result in recommendations favoring the most highly valued population, a fact that was not missed in recent international climate negotiations. News reports commented that these negotiations were threatened due to the unequal valuation of lives used in the background analysis (Pearce, 1995).

Direct expenditures are also influenced by the distribution of income. Health care is a normal good, and increases in income will be accompanied by increased consumption of health care. Viscusi (1994a) summarized studies estimating, at the margin, individual willingness to consume health-related services out of income. He found that the different studies and methodologies all yielded low, but decidedly positive, marginal propensities to consume health care, although, on average, results from the international studies were three times those from cross-section studies.

Because health care is a normal good, an illness that strikes low-income individuals (perhaps caused by an opportunistic microorganism attacking individuals whose health is already compromised) would cause smaller direct health expenditures than a disease that strikes randomly throughout the population, even if disease incidence and symptoms were similar. Again, COI calculations would show larger costs for the randomly striking disease than for the low-income disease. If COI were used to judge severity, an analyst would conclude that the randomly striking disease was more severe than the low-income disease.

Direct expenditures also reflect the ability of current medical techniques to treat the disease under consideration. For example, treatment of the common cold generates enormous expenditures on cold medicines each year, while a disease like malaria may generate relatively few expenditures because there are few remedies. COI estimates for each disease might indicate that the cost of a cold is greater than that of a debilitating disease like malaria. Both direct and indirect costs would contribute to this conclusion because malaria incidence is highest in low-income countries. If a treatment for malaria is discovered, the recalculated COI would soar with purchases of the newly discovered treatment. Advances in medical science can simultaneously improve individual welfare and increase calculated COI.

A variety of factors influence earnings, health, and health-care consumption. As a result, the severity of an illness is not identical to the severity of the economic consequences of an illness. The COI of a particular disease that targets a particular population not only measures the severity of the disease, but also the population's education, skill level, income, sick-leave benefits, and insurance coverage, as well as the types of medical interventions currently available.

Is the COI Approach Ever a Useful Tool?

Though the cost-of-illness approach is not a useful tool for measuring social or individual welfare changes or for measuring disease severity, it can provide economists and policymakers with useful information. The COI approach traces the economic flows associated with an adverse health outcome. It accounts for the drop in productivity resulting from illness, accident, or premature death, and it accounts for the shift in consumer expenditure from more general consumption goods, and savings and investment, to medical goods and services. Cost-of-illness (COI) estimates provide an accounting of the dollars spent on medical expenses and the dollars of employment compensation that are forgone as a result of illnesses, accidents, or premature deaths. Such an accounting provides useful information to economists and policymakers interested in gauging the pure economic impact of government policy to reduce adverse health outcomes.

In addition, when combined with a general equilibrium analysis, such as a Social Accounting Matrix, the COI approach provides the first step in deciphering the full economic impact of illness and premature

death. For example, Golan et al. (1998) use a Social Accounting Matrix model to gauge the extent and distribution of the costs of foodborne illness due to meat and poultry. With the SAM model they trace the economic ramifications of the dollar costs of foodborne illness. They find that though the human capital costs of foodborne illness result in a general decline in economic activity, the direct costs trigger growth in the medical support industries and decline in general consumption goods and services. This redistribution of economic activity results in a redistribution of income extending past those individuals who actually contract a foodborne illness.

Any COI estimate can be disaggregated (as in the Golan et al. study) to examine the direction of the economic flows resulting from illness and premature death. If this step is taken, the COI approach can reveal not just the magnitude, but the distributional consequences of illness. COI is therefore a useful tool for gauging the extent and distribution of the costs of adverse health outcomes. It is a first step in deciphering the economic distortions triggered by illness and premature death.

Empirical Considerations

It is widely accepted in the health economics literature that the direct and indirect expenses incorporated in COI measures are relatively easy to estimate, and that therefore, despite its flaws, the COI approach is preferable to the other approaches, particularly the willingness-to-pay approach. The assumed empirical superiority of the COI approach prompted Mishan (1975) to make his much-quoted observation:

> In view of the existing quantomania, one may be forgiven for asserting that there is more to be said for rough estimates of the precise concept [willingness-to-pay] than precise estimates of economically irrelevant concepts [COI]. (p. 320)

How precise are calculated COI estimates? The alleged straightforwardness of the empirical estimation is only apparent to those who have never tried it. In reality, it is quite difficult to decipher what the direct and indirect costs associated with an illness are. There is no COI template to follow, and data are guaranteed to be insufficient and inexact.

The primary problem with empirically estimating direct costs is that the prices charged to health care consumers are usually distorted and rarely reflect true economic value. For example, the consumer price of medicine or medical services is typically much lower than the true cost. Interactions between insurance plans and the medical regulatory system yield a gap between accounting costs and economic costs throughout the medical system (Sox et al., 1988; Finkler, 1982; Hildred and Watkins, 1996). As a result, the empirical researcher could be faced with three or four prices for the same good or service. This abundance of prices makes comparisons across COI studies almost impossible.

Even if a consistent approach to determining cost is developed, it remains difficult to decipher exactly what treatments are being purchased and for whom. The standard procedure in comprehensive empirical studies of COI, is to use estimates from the Health Care Financing Administration (HCFA) on total health care expenditures as the basis for estimating specific expenditures by disease (e.g., physician services, hospital services, pharmaceuticals, medical equipment). This procedure is subject to numerous sources of error, many of which are summarized by Scitovsky (1982) in her review of the empirical literature.[8]

One of the primary difficulties that arises in estimating COI involves determining the type of medical expenditure. It is particularly difficult to disaggregate hospital payments. These expenditures typically include drugs administered on the premises plus salaries paid to health professionals and staff meaning that "professional medical services" and "drugs and medical sundries" are underestimated while "hospital services" and "nursing home services" are overestimated.

Another problem arises due to inaccuracies in hospital diagnostic data and the fact that expenses might

---

[8] Most of the bias mentioned by Scitovsky stems from use of the HCFA and other specific data sets, however, since these data sets are the primary sources of information on medical expense, her observations are pertinent to any COI study (see Kenkel, 1994 for an overview of Scitovsky's critiques and other critiques of the HCFA data set).

not be attributed to the correct illness. Similarly, a number of illnesses might be grouped under one diagnostic code making it impossible to decipher individual expenses. For example, in the National Health Interview Survey, most symptoms potentially due to foodborne pathogens are coded in four general disease categories, "intestinal infections due to other organisms, not elsewhere classified," "food poisoning–unspecified," "infectious colitis, enteritis, and gastroenteritis," or "infectious diarrhea." This level of generality makes it difficult to differentiate specific illnesses such as salmonellosis from campylobacteriosis.

Another difficulty with many of the large data sets is that they typically assume the same charge for all types of physician services, when in fact a visit to a physician for a routine physical does not cost the same as a visit for cancer. Another problem concerns the treatment of multiple conditions. The convention is to allocate all expenses to the patient's primary diagnosis, a practice that leads to substantial overestimation of some expenses and underestimation of others. Scitovsky (1982) estimates that 52 percent of all hospital patients have multiple conditions.

A number of difficulties also arise in calculating the indirect costs of illness. Not only is it difficult to accurately establish the number of work-loss days through the use of survey data, but it is also a challenge to determine the cost of these days. It is difficult to account for the cost of non-paid labor (for example, human capital costs of stay-at-home parents), and it is often equally difficult to accurately estimate forgone earnings. An employee's compensation typically includes more than wages. Pension plans, health insurance, flexible hours, etc., can all contribute to compensation and should be included in an accurate measure of human capital costs. Failure to account for these benefits will result in underestimation of indirect costs, especially for wealthier income groups.

In light of the myriad of difficulties listed above, it is clear that empirical estimation of COI is not as straightforward as advertised. The empirical researcher faces a number of difficult decisions in determining direct and indirect costs and there is little chance for conformity across studies. Little about COI estimates is mechanical, and judgment and interpolation are often the analyst's principal function.

Conclusion

Since its inception in the middle of the 20th century, the cost-of-illness approach to measuring the cost of adverse health outcomes has been cast in many roles: as a direct measure of societal welfare; as a measure of individual welfare change resulting from changes in health status; and as an indicator of outcome severity. In the discussion above we have shown that the COI approach is not a valid tool for welfare analysis because it does not provide adequate estimates of individual or social welfare. We have also demonstrated that COI estimates are not reliable measure of disease severity. In addition we have illustrated some of the difficulties that arise in calculating COI.

However, despite its shortcomings for welfare analysis and as a measure of disease severity, the COI approach is still a useful economic tool. The COI approach provides an *accounting* of the dollars spent on medical expenses and the wage dollars that are forgone as a result of illness, accident, or premature death. Such an accounting provides useful information to economists and policymakers because it indicates the direction and magnitude of the economic flows resulting from health shocks to the economy.