Cost Recommendations in the Second Edition of Cost-Effectiveness in Health and Medicine: A Review

John A. Nyman

Abstract
A 2nd edition of Cost-Effectiveness in Health and Medicine has now been published by a group of medical decision-making experts known collectively as the 2nd Panel. This is a critical review of the recommendations for how to deal with costs in cost-effectiveness analysis, recommendations that are contained in Chapter 8 of that edition, titled “Estimating Costs and Valuations of Non-Health Benefits in Cost-Effectiveness Analysis,” authored primarily but not exclusively by Anirban Basu. This review focuses on the correspondence between the costs in the numerator of the incremental cost-effectiveness ratio (ICER) and what is measured in the denominator of the ICER by the quality-adjusted life years (QALYs). Although it raises a number of issues regarding what is actually being measured in the numerator and denominator of the ICER, it primarily challenges the 2nd Panel’s recommendation that the costs of non-health consumption in any additional years of life generated by the intervention in question be accounted for in the numerator, even though no measures of the benefits are included in the QALYs in the denominator. This review is adapted from a review of the entire 2nd edition that was sent to the 2nd Panel steering committee on January 8, 2016.

Keywords
consistency of costs, cost-effectiveness recommendations, effectiveness measures, future costs

Date received: January 5, 2017; accepted: January 17, 2018

A 2nd edition of Cost-Effectiveness in Health and Medicine has now been published. The 1st edition was designed to standardize the conduct of cost-effectiveness analyses that use quality-adjusted life years as the measure of effectiveness and, to that end, established a series of recommendations for their conduct. Despite containing a number of errors, many of which were identified shortly after their publication, the recommendations from the 1st edition have remained in force for the past 20 years. With the 2nd edition, the original recommendations have now been officially superseded. The authors of the new edition, known collectively as the 2nd Panel, are to be congratulated for correcting a portion of these errors and on producing an updated and useful text. But not all the new recommendations make sense. Herein I critically review the recommendations related to costs contained in Chapter 8: “Estimating Costs and Valuations of Non-Health Benefits in Cost-Effectiveness Analysis,” authored primarily but not exclusively by Anirban Basu.

Terminology
In the “Introduction” (section 8.1) of the chapter, the 2nd Panel provides an expanded version of the diagram...
that summarizes the components of the incremental cost-effectiveness ratio (ICER) in schematic form (Figure 8.1, p. 203). The change to recommending that only the intervention’s health effects be captured in the quality-adjusted life years (QALYs) in the denominator, and none of the intervention’s costs, corrects a major (and puzzling) mistake of the 1st edition. The clarification that the numerator is to capture only the value of the resources used either in producing the intervention or that were expended as a consequence of the intervention is another welcome change. It would have helped to introduce official jargon to distinguish between these two types of costs. In the course I teach on cost-effectiveness analysis at the University of Minnesota, I use the phrases “costs of the intervention itself” and “downstream costs” to distinguish between these concepts, respectively, but an official set of terms would have been useful.

Categories and Theory

Section 8.2 on “Conceptual Cost Categories” describes the types of costs recognized by the 2nd Panel, distinguishing first between “costs within the formal health care sector” and “costs outside the health care sector.” The Panel eschews the terms “direct” and “indirect,” but again does not give analysts acceptable alternative jargon. I suspect that “inside” and “outside” will eventually fill the void, but officially sanctioned terms would have been useful.

In this section, the 2nd Panel presents another important diagram (Figure 8.2) showing how the costs during any additional years of life generated by an intervention—termed “future” costs—have different implications than the costs during the period when the time of the intervention and its comparator overlap. The 2nd Panel recommends that costs in the future years include the costs of care that is not related to the original disease addressed by the intervention, that is, the “unrelated” costs. This makes sense because, in general, the additional years of life are caused in part by the health care resources expended in treating other diseases. The 2nd Panel also recommends that the costs of non–health care consumption incurred during the future years be included because they represent resources expended that are a consequence of the intervention. Moreover, it recommends that earnings during those future years be included as negative costs because they represent resources that are generated during that period, not spent.

It is a mistake, however, to include non–health care consumption costs and earnings in future life-years in the numerator of the ICER for two reasons. First, it is simply not consistent to include the costs of non–health care consumption in the numerator of the ICER without also measuring the utility or QALY (with an additional quality-adjustment for consumption) value gained from that consumption and including it the denominator. Likewise, earnings should not enter the numerator unless the utility or value of the leisure lost also is entered in the denominator. As David Meltzer shows in his award-winning 1997 paper that is the basis for this recommendation, the ICER is an attempt to arrive at the dollar cost of a utility gain from an intervention that generates additional utility through additional health, survival, and consumption, and net of the utility lost from foregoing leisure when at work. In practice, however, ICERs that use QALYs in the denominator capture only two of these utility gains, survival and health, but not the utility gain from consumption nor the utility lost from foregoing leisure. Because QALYs do not also capture these changes in utility, the ICERs cannot also include the cost of consumption or the earnings in the same future years without biasing the ratio.

The 2nd Panel was aware of this argument but countered that the maximization of QALYs can be viewed as equivalent to maximization of overall welfare under Bleichrodt and Quiggin’s assumptions that overall utility is measured by the product of QALYs and a constant utility of consumption. If these assumptions hold, then the 2nd Panel argues that it does not matter whether QALYs also measure the utility from consumption or not: increases in QALYs or increases in QALYs times a constant utility from consumption would increase monotonically and so measure the same concept. Therefore, consumption costs during future years could be included in the numerator without their utility also being included in the denominator.

However, if utility increases with consumption but is assumed to be constant, Bleichrodt and Quiggin add that “it follows that consumption itself is constant over the individual’s life-cycle.” This, of course, is not supported by the data, as the 2nd Panel’s own Table 8.2 (showing how average annual consumption costs varies with age) makes clear. Indeed, one of the main points of Meltzer’s 1997 paper (that originally argued for inclusion of future consumption costs in the ICER) was that consumption varies over the life cycle and that taking account of consumption costs that vary with age would affect the ICER calculations of interventions that generate additional life years at different ages. So the 2nd Panel’s assumption that allows for the inclusion of consumption costs in the numerator is not only empirically unsubstantiable, but it would also negate the original explanation for
why it is important to include future consumption costs in the numerator of ICERs in the first place. An analogous argument is true for earnings.

Moreover, the assumption that welfare can be characterized by the product of QALYs and a (constant) utility from consumption implies that the rich deserve more health care spending (or more precisely, should have a higher threshold value of a QALY) because they consume more non–health care goods and services. The 2nd Panel should ask themselves if they really want to base their recommendations on what Bleichrodt and Quiggin characterize as an “ethically troubling” assumption.

The second reason why including consumption costs during future years is a mistake is that we already know the welfare implications of future consumption and they are positive. That is, we conduct CEAs and they are positive. That is, we conduct CEAs to include the costs of non-health consumption in the numerator would increase the non-health consumption of $50,000, the inclusion of these costs in the numerator would increase the numerator of the ICER not only represents a bias against it. The 2nd Panel’s statement is misleading.

Moreover, the assumption that welfare can be characterized by the product of QALYs and a (constant) utility from consumption implies that the rich deserve more health care spending (or more precisely, should have a higher threshold value of a QALY) because they consume more non–health care goods and services. The 2nd Panel should ask themselves if they really want to base their recommendations on what Bleichrodt and Quiggin characterize as an “ethically troubling” assumption.

The second reason why including consumption costs during future years is a mistake is that we already know the welfare implications of future consumption and they are positive. That is, we conduct CEAs because we are faced with a new treatment and the effects of the treatment on the subject’s utility are complex, and so are not obvious to the patient, the payer, or society. Accordingly, it is not clear whether the treatment is worth the added cost. In contrast, non-health consumption decisions are much less complicated for the consumer to evaluate, and because they are voluntary, we already know that the utility gain (as represented by the willingness to pay for them) is at least as great as the additional cost. Similarly, when we make a voluntary decision to work, the earnings exceed the value of the utility lost by giving up leisure, otherwise we would not work. Because both decisions are voluntary, we can assume that there is net welfare gain from any decision to consume or to work. Therefore, the recommendation to include the costs of non-health consumption in the numerator of the ICER not only represents a bias because a corresponding utility effect is missing from the denominator, but also because we know that the corresponding utility effect would generate a welfare gain in terms of willingness to pay that exceeds the cost of the consumption. Thus, this recommendation creates an ICER-increasing bias when actually the underlying economics should generate a reduction of the ICER. An analogous but opposite argument can be made for the inclusion of voluntary earnings in future years in the numerator of the ICER.

As an example of the consumption cost bias, assume a retired 70-year-old receives a new treatment that causes him to live another year with a quality of life of 1.0, and that the ICER for that treatment is $80,000/QALY. If instead that member of society has no earnings but has non-health consumption spending of $50,000, the inclusion of these costs in the numerator would increase the ICER to $130,000, and bias the ICER so that the intervention would no longer be cost-effective if the value of a QALY were $100,000. If the welfare gain were captured by a willingness-to-pay for this voluntary non–health care consumption of $70,000, the true ICER should actually decline by $20,000 to $60,000/QALY (or alternatively would generate a commensurate increase of a broader measure of the QALY benefits of the intervention in the denominator). So by taking account of the welfare gain from the voluntary non-health consumption, the treatment should be seen as even more cost-effective than it was originally. Thus, the recommendation to account for consumption in future years by simply including non-health consumption costs in the numerator would doubly bias the ICER.

Even though the utility from consumption and from the loss of leisure are included in Meltzer’s original equation, Meltzer also ignores the issue of whether the utilities from consumption and from the loss of leisure are actually captured in a QALY when making his recommendation to include consumption costs and earnings in the numerators of ICERS. As was recognized at the time, Meltzer’s is an extreme position. Just after discussing how not including non-health consumption costs would bias the calculation, the 2nd Panel attempts to justify its recommendation by noting that “the inclusion of future costs (health care and non-healthcare) is now standard in many governmental agencies” (p. 214). Actually, only Sweden has recommended the inclusion of non–health care future consumption costs. All other countries that have considered this issue have either declined to make a recommendation or decided to recommend against it. The 2nd Panel’s statement is misleading.

Societal Cost Categories

Because QALYs measure health-related quality of life and survival, there is a logical justification for including health-related consumption costs (such as travel costs) in the numerator of the ICER. While a general utility function would also include non-health consumption and leisure, other arguments might also contribute to utility. For example, police protection and the criminal justice system contribute to welfare and might have a better claim of having a constant utility over the life cycle. If non-health consumption costs and earnings were included, there would be no reason not to include an individual’s share of the costs of these other utility-increasing commodities, too. That is, with the Bleichrodt and Quiggin assumption of a constant effect on utility, there is no logical way to delineate what costs should be included in the numerator of the ICER and what costs
to exclude. The costs of any resources that increase utility in future years of life could be included under this assumption, regardless of whether their benefits were picked up in the denominator or not.

Indeed, in the very next section of the chapter (“Deciding on Which Categories to Account for in Reference Case Analysis,” section 8.3), the 2nd Panel includes a table (Table 8.1, p. 217) listing the other non-health care sector costs to be included under a societal perspective: “Social services,” “Legal or criminal justice,” “Education,” “Housing,” “Environment,” and “Other (e.g., friction costs).” These costs are “to be included in the calculation of the ICER” under a societal perspective.3(p217) (Why and how “housing” costs are distinguished from non-health “consumption” costs is not explained, and the same is true for many educational and legal costs.) Because these additional costs are listed along with “consumption” and “earnings,” but were not discussed earlier when the arguments for inclusion of non-health consumption cost and earnings were originally introduced, this comes as a surprise to the reader. That they were not discussed earlier further raises the question of whether the 2nd Panel really intends for practitioners to include all these costs in the societal perspective ICER if they occur in future years. Piling all these additional costs onto the numerator of the ICER when not balanced by a measure of the utility value of the gains in the denominator could bias the cost-effectiveness conclusions dramatically.

Finding Unit Cost Values

In the next section (“Methods for Valuing Resources in Monetary Terms for Specific Categories,” section 8.4), the 2nd Panel discusses the identification, measurement, and valuation of unit costs. Like the original panel, the 2nd Panel distinguishes micro-costing from gross-costing. With respect to gross-costing, the 2nd Panel seems to expect that a full-fledged cost-function study would be done to model the incremental individual-level health care costs. They cite the well-known econometric issues surrounding such analyses (nonnormality of the distribution of costs, difficulty of retransforming log costs into costs, the importance of explaining as much of the variance in costs as possible, and issues of censoring).

Knowledge of these issues, however, would not help the analyst who has mean utilization data for each intervention but who needs a unit cost value and does not have a cost-function analysis from the literature for that value or the data to conduct such a study anew. In the United Kingdom and other advanced countries, this gross-costing issue is solved by using the unit payments under the country’s national health system. In the United States, the closest we have to a national pricing system is Medicare. Moreover, most Medicare unit costs are based on a previous micro-costing study (e.g., William Hsiao and his colleagues9 conducted a micro-costing study to determine the Resource-Based Relative Value Scale for paying physicians under Medicare). And, Medicare unit costs are readily available (e.g., a Medicare “PC Pricer,” generating fee-for-service unit values for a variety of services can be accessed at https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/PCPricer/10). Although the 2nd Panel recommends the use of Medicare Cost Reports to obtain cost-to-charge ratios in micro-costing, it does not also recommend the use of Medicare unit payments in gross-costing, especially when no other nationally representative unit cost figures are available. Why this gross-costing solution is not recommended, let alone discussed, is a puzzle and represents an important omission from the chapter.

In this same section, the 2nd Panel goes on to consider productivity cost for those who are prohibited from working in the formal labor market due to illness or recuperation from a treatment. The options considered by the 2nd Panel are to use a “human capital” approach where the productivity losses for the absent worker accumulate for as long as the ill worker is absent, or a “frictional” approach where the productivity losses accrue only until a replacement worker has been hired. The 2nd Panel argues that, because a replacement worker was likely to be working full-time either informally or in household production, there would really be no production saved if that worker started to work formally by replacing a sick worker. Therefore, the 2nd Panel recommends the human capital approach.

However, whether a newly hired replacement worker would have spent all his or her replacement time engaged in informal work or household production, and none on leisure activities, is an empirical question and one that has not been satisfactorily researched. Interestingly, when a similar question arises later in this section regarding whether a retiree’s productivity costs should reflect informal labor production or leisure, the 2nd Panel states that “[m]ore research is needed to appropriately characterize the time spent by retirees in doing informal productive work” (p. 224). The 2nd Panel should be consistent and make a similar recommendation with regard to a formal worker who becomes ill, or at least explain why a different approach is needed.

The 2nd Panel goes on to point out that a “more nuanced micro-costing approach to productivity”
(p. 222) would consider presenteeism as well as absenteeism and how those factors affect the production and production forgone. A more nuanced approach would also include a discussion of the other side of the labor market transaction, but this issue is ignored even though it appears in the literature. That is, if a worker works, she or he generates production (with a value equal to wages and fringe benefits in a competitive market), but she or he also relinquishes leisure to do so, which represents the other side of the transaction. The utility from leisure forgone is not captured in the QALYs, but this would be acceptable if that leisure is also not available for someone who is ill. That is, for a healthy worker, utility from leisure is sacrificed for wages and fringe benefits, which represent together the value of production. For an ill person, utility from leisure cannot be sacrificed because, if one cannot work because of illness, it has been assumed that one cannot also play golf or travel or do any of the things that generate utility during leisure time.

Nevertheless, some illnesses allow persons to do the same things—reading and watching TV—they would do in their leisure time, and as a result there may be a utility gain from leisure when ill and cannot work. Therefore, focusing on the production lost alone without also taking into account the value of any leisure time gained would bias the calculations. While this is a complex issue that would require additional information on the overlap between illness and leisure activities (or additional assumptions explicitly noted), it is as important as the nuance of determining to what extent a worker (who does not work because of illness) would actually have been engaged in productive work. This issue should have been acknowledged in the chapter and discussed.

As already noted, the 2nd Panel includes Table 8.2 that presents average consumption costs in 2014 by age, broken down by health care and non-health care expenditures. These numbers are intended as a useful source for values of non-health care consumption costs in future years. The Panel also includes Table 8.3 that presents average earnings by age, intended to capture productivity costs from missing work because of illness, and the productivity gained in future years, presumably because of being able to work. (In the text on pp. 223-224, the table number labels are erroneously switched.) For example, Table 8.3 shows that a person age 75 years earned $37,300 in 2014, thus the implication is that by living a 75th year (that the person would not have otherwise lived, were it not for the treatment in question), $37,300 worth of production is accrued by society. Real-world earnings at this age, however, reflect a large portion of investment earnings. Investment earnings would presumably exist, regardless of whether the original owner were alive, or had died and bequeathed the investment instrument to an heir. So, in addition to the issue of the utility loss of earnings not being acknowledged in the QALYs, there is also the issue that these future earnings may be largely due to investments. How unearned income like this should be treated in the ICER is not addressed.

Under the subsection titled “Other Sector Costs,” the 2nd Panel writes,

In addition, for a Reference Case analysis from a societal perspective, intrinsic non-health benefits, such as the psychological benefits of lower crime, should be valued in monetary terms and included in the numerator of the ICER. However, the methods of valuation would depend on the jurisdiction of the analysis. In the United States, willingness to-pay would be an acceptable method. (p. 227)

This is wrong in a number of ways. First, one of the most attractive aspects of the 2nd Panel’s approach to costs, and a major improvement over the original panel’s approach, was the apparent desire to reserve the numerator of the ICER for only the value of the resources expended or gained as a consequence of the intervention. Why is that approach violated in the case of “other sector costs” and nowhere else? Second, if (as we are told) the Bleichrodt and Quiggin assumptions (that QALYs represent welfare and that the utility of non-health consumption is constant) hold for non-health consumption costs, making inclusion of benefits from non–health care consumption in the denominator unnecessary in survivor years, why do the same assumptions not now hold for items that generate “other sector costs” as well? Third, if (as we are told) the willingness to pay for other sector (intrinsic non-health consumption) benefits is to be included in the numerator (along with the cost of these other sector items), why is the willingness to pay for non-health consumption goods during future life years to be ignored and not also included (along with the costs of non-health consumption) in the numerator of the ICER? Fourth, why do the same issues not arise for earnings and the willingness to accept for foregoing leisure?

Miscellaneous Issues

The final section of the chapter addresses “Other Issues” (section 8.5) and covers price inflation in the health care sector, returns to scale and (briefly) discounting. The
recommendations for how to deal with inflation in the health care sector are particularly detailed and useful.

One issue missing from the discussion is how to address a cost incurred that has no compensating benefit. For example, a treatment for a sleep disorder may cause drowsiness and could result in a car accident. If the patient is injured, this would be considered an adverse event and its effect on QALYs measured accordingly. However, the car accident may not cause injury and therefore have no effect other than through the loss of resources (the wrecked car) itself. If so, the reduction in the value of the car would represent part of the resources expended as a result of the treatment and included in the numerator of the ICER. If the consistency between the numerator and denominator of an ICER were appropriately valued by the 2nd Panel, it would discuss and make recommendations for such a scenario.

Revision

Shortly after the original panel’s recommendations were published, it was clear that errors had been made. Yet these mistaken recommendations continued to dictate the standard by which cost-effectiveness analyses were conducted and evaluated for the next 20 years. In this review, a number of important mistakes have also been identified. It is likely that other mistakes will be identified as the 2nd Panel’s recommendations receive further scrutiny from the academic community. A vehicle should exist for a new consensus to be reached and for the official recommendations to be changed, and it should be done in a much more timely fashion.

Note

i. An anonymous reviewer of this article made a number of comments that I have interpreted as suggesting that non-health consumption should be divided into a subsistence portion that is fixed per year and a discretionary portion that is variable. The subsistence consumption costs would be required for an additional year of life to exist; therefore, their utility is already captured by the QALYs. Thus, these subsistence consumption costs would appear to be another type of health-related consumption costs, just like travel and time costs (see next section), and be included when future years of life are generated by an intervention. Meltzer does not recognize this type of consumption as he makes survival a function of health care alone, not health care and consumption. The 2nd Panel also does not recognize subsistence consumption. One issue is empirical: what minimum level of consumption costs would be required to survive another year? While dividing this consumption this way may eliminate some consumption costs from requiring a corresponding utility gain measure, it does not eliminate the discretionary consumption costs from needing one. Again, it was this variation in consumption across the life cycle that represented Meltzer’s original motivation for including such costs in the ICER. If this amount were added to consumption without additional assumptions, there is no similar argument for excluding any of the earnings in the denominator.

References

1. Neumann PJ, Sanders GD, Russell LB, Siegel JE, Ganiats TG, eds. Cost-Effectiveness in Health and Medicine. 2nd ed. New York: Oxford University Press; 2017.
2. Gold M, Siegel JE, Russell LB, Weinstein MC, eds. Cost-Effectiveness in Health and Medicine. New York: Oxford University Press; 1996.
3. Basu A. Estimating costs and valuations of non-health benefits in cost-effectiveness analysis. In: Neumann PJ, Sanders GD, Russell LB, Siegel JE, Ganiats TG, eds. Cost-Effectiveness in Health and Medicine. 2nd ed. New York: Oxford University Press; 2017. Chapter 8.
4. Nyman JA. More on survival consumption costs in cost-utility analysis. Health Econ. 2006;15(3):319–22.
5. Nyman JA. Should the consumption of survivors be included as a cost in cost-utility analysis? Health Econ. 2004;13(5):417–27.
6. Meltzer D. Accounting for future costs in medical cost-effectiveness analysis. J Health Econ. 1997;16(1):33–64.
7. Bleichrodt H, Quiggin J. Life-cycle preferences over consumption and health: when is cost-effectiveness analysis equivalent to cost-benefit analysis. J Health Econ. 1999;18:681–708.
8. Nyman JA. Measurement of QALYs and the welfare implications of survivor consumption and leisure foregone. Health Econ. 2011;20(1):56–67.
9. Hsiao WC, Braun P, Yntema D, Becker ER. Estimating physicians’ work for a resource-based relative-value scale. N Engl J Med. 1988;319:835–41.
10. Centers for Medicare & Medicaid Services. PC Pricer [cited 5 January 2017]. Available from: https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/PCPricer/.
11. Nyman JA. Productivity costs revisited: towards a new U.S. policy. Health Econ. 2012;21(12):1387–401.