On Defining and Valuing the Benefits of Health Policy Interventions: How and Why CEA in Health Morphed into CU(B)A and "Back-Door" BCA

David Salkever, Ph.D.

Final version forthcoming (2018) in *Teaching Benefit-Cost Analysis*, S. Farrow, ed. Edward Elgar, Cheltenham, UK.

On Defining and Valuing the Benefits of Health Policy Interventions: How and Why CEA in Health Morphed into CU(B)A and "Back-Door" BCA

David Salkever

Abstract

For the past five decades, the literature on economic evaluations of health programs or policies has consisted largely of cost-effectiveness analysis (CEA) rather than benefit-cost analysis (BCA). One factor contributing to this orientation was the view that we could not obtain valid estimates of consumers' monetary valuations (i.e., willingness-to-pay figures) for the benefits provided by these programs or policies. As interest in CEA methods in health expanded, and the limits of simple effectiveness measures in CEA became clearer, further refinements in effectiveness measurement have: (1) brought us closer and closer to actually conducting BCAs for heath programs, and (2) generated important new insights into defining and valuing program benefits in willingness-to-pay (WTP) terms. This chapter traces these developments in the convergence of the CEA and BCA literatures in health. A simple example is presented to highlight the major challenges to obtaining valid WTP valuations for benefits of health programs, and to compare major strategies used for generating monetary WTP benefit valuation figures.

The Core Principles of BCA: A Thumbnail Sketch

BCA, in simplest terms, involves application of accepted principles of neoclassical welfare economics that provide a systematic process for choosing between two alternative states of the world. Typically the choices are between a "status quo" state that would obtain in the absence of any changes in public policy, and an "alternative" state in which a specified new "policy" is implemented. The relevant differences between these two states constitute the "effects" of the new "policy" (relative to the status quo).

According to BCA theory, all the various effects of a new policy should be valued in monetary terms by the willingness to pay (WTP) of any and all individual citizens whose well-being is in fact altered by implementation of the new policy. Each of the separately measurable "effects" of the policy are valued in monetary WTP terms; for some individuals, WTP figures for particular effects will be negative and therefore regarded by them as "costs", while WTP figures for other effects will be positive and will be regarded by them as "benefits". The sum of all WTP figures for any individuals represents their "net benefit" for the policy (relative to the status quo). The unweighted sum of these net benefit figures across all individuals represents the overall net

benefit of the policy (relative to the status quo).

Finally, the rationale for choosing between states of the world based on the unweighted sum of net benefit figures across all individuals is that a positive (or negative) unweighted sum implies the new policy is (or is not) a potential Pareto-improvement relative to the status quo. This rationale assumes any compensation of "losers" by "winners" is feasible and costless.

Criticisms of BCA Core Principles

In the general BCA literature, a number of criticisms have been raised about strict adherence to BCA principles. Some criticisms are based on incorrect understanding of the rationale for BCA; for example, it is often asserted that BCA uses the unweighted sum of WTP figures across individuals as the measure of net benefits because it assumes that $1 of WTP for a poor person is viewed by the evaluator or policy-maker as *equal in subjective value* to $1 of WTP for a rich person. In fact, of course, BCA does not assume $1 WTP for a rich person is equivalent in subjective value to $1 WTP for a poor person. Instead, BCA separates efficiency decisions (for example, whether to adopt or not adopt a new policy) from distributional choices once the policy has been adopted. Thus, the reason for using the unweighted sum of WTP to measure net benefit in BCA derives from: 1 the assumption that costless transfers are possible and 2 the overriding efficiency objective of identifying potential Pareto improvements.[1]

Other, more valid criticisms, have focused on the obvious unreality of the assumption that compensation is costless or nearly costless. This criticism, however, can be addressed in at least some instances by developing estimates of compensation costs and using these estimates in evaluating new policy proposals. (The alternative of using “distributional weights” is often proposed but not universally accepted.)

Some other criticisms of strict adherence to core BCA principles may have validity but have been questioned as focusing on problems that could but rarely do arise. An example, is the reversibility problem of positive total WTP for both adopting a policy and for abandoning the same policy once it is adopted. Arguments that such reversals rarely occur in practice can be found in Zerbe and Bellas (2006) and Willig (1976).

Finally, there are serious criticisms about our ability to measure WTP figures that accurately reflect the preferences of well-informed individuals, who accurately perceive precisely how their well-being is in fact altered by implementation of the new policy.

Criticisms of BCA of Health Policy Interventions and the Case for Relying on CEA for Evaluating Health Policies

As Kenkel (1997) detailed in an early overview of the CEA vs. BCA controversy in health, critics of BCA in health offered criticisms ranging from "ethical" objections to the very act of "placing a monetary value on health", to the presumption that BCA "favors" interventions that help the wealthy over interventions that help the poor, to practical concerns about the validity of WTP estimates as accurate representations of individual's preferences.

The first two of these objections were problematic in that they emanated from an incorrect understanding of BCA and the nature of the "benefits" offered by health policy interventions. Concern about the propriety of "placing a monetary valuation on health" is misplaced since health policies do not offer certain improvements in health; instead, they offer changes in risks of future health levels. BCA of health policies place $ values on these changes in risks rather than $ values on "health per se" (however “health” is defined).

The presumption about BCA "favoring" interventions that help the wealthy presumably derived from the mistaken impression that BCA had distributional preferences "baked in" (for example, $1 of WTP was of equal "subjective" worth for all people). In fact BCA is precisely designed to separate distributional judgments from the task of identifying potential Pareto improvements.

The most salient and challenging criticism of BCAs of health policies, however, is the argument that WTP values used in such BCA's do not accurately reflect the true preferences that would be revealed by individuals who are fully informed about the impact of these policies on their well-being. While this is, of course, a general concern in BCA across all sectors, the problem for evaluators of health policies is especially difficult because of the complexity of the beneficial effects of these policies.

As just noted, these effects are essentially changes in risks of being in various health states. The WTP figures for these effects will likely be inaccurate for two different reasons: 1 problems that individuals have in understanding and dealing with small changes in probabilities and 2 problems individuals have in fully understanding the ways in which being in a particular health state, that they have never yet experienced, impacts their well-being.

According to the critics of BCA in health, CEA offers a solution to all three of these criticisms. The CEA solution offered is simply to not place any monetary value on health policy effects that involve reducing risks of ill health (which are the main benefits of the health policy intervention). Instead these effects are only measured in "physical" (non-monetary) units such as lives saved, years of life extended, cases of a particular disease (e.g., influenza) prevented, etc. No input from consumers on the monetary valuations of these effects is required for CEA; all that is needed is to compute the relevant monetary measures of cost for implementing the policy, calibrate "effectiveness" in non-monetary terms, and compute a cost-effectiveness ratio (CER) for any new health policy relative to the status quo.

Of course, the critics of BCA in health also recognized the major limitations of CEA relative to BCA. For example, CEA generally does not provide a clear choice between a new policy and the status quo. Similarly, they recognized that even in choosing between two alternative new policies, comparisons of CERs can indicate a clear choice of one policy over the other only in rare circumstances. Notwithstanding these limitations, the critics of BCA in health have argued that the much less powerful tool of CEA is still preferable because it does not require potentially inaccurate estimates of WTP.

The Introduction and Evolution of CEA for Health Policies

The origins of applying CEA to health policies was succinctly described by Klarman (1982):

> The major impetus to applying… [CEA]… came with President Johnson's [importation] of the planning, programming, and budgeting system … from the Department of Defense… to civilian branches of the…government in 1965.…[The DoD] analysis differed from standard [BCA] in at least two respects: 1) emphasis was on ascertaining the effects [outcomes] of alternative programs; and 2) given the presence of budget constraints and similar kinds of

outcomes of competing programs, the analysis [focused on] measurement of program [monetary] cost and [estimating] program benefits in terms of physical units of outcome…(N)o attempt was made to take the next step of putting an economic value on the effects of alternative programs.

This simple version of CEA works well in many (but not all) decision problems for choosing among non-mutually exclusive alternative programs that produce the same single kind of output (for example, reduced risk of contracting influenza, reduced risk of motor vehicle traffic deaths prevented) and that only involve costs from the same constrained budget. Complications arise however when there are multi-dimensional outputs (for example., reduced risk of mortality and reduced risk of morbidity), when output units differ in kind across policy alternatives, and/or when mutually exclusive choices require choosing higher vs. lower levels of both costs and output. In these cases, the simple CEA paradigm no longer provides clear bottom-line conclusions about policy choices and additional value judgments (whether intuitive or derived from a WTP assessment) are required.

Two major concerns motivated CEA analysts to find improved ways to characterize output that would mitigate the severe limitations of the simplest CEA analyses that only used homogeneous uni-dimensional output measures. First, with multi-dimensional output measures, one cannot define a CER without somehow constructing a single effectiveness measure as the denominator, which implies the use of some weighting scheme to convert the multi-dimensional effectiveness measures. The first efforts to include weighting were somewhat arbitrary. One well-known example is the Klarman et al. (1968) CEA of chronic kidney disease which compared expected life-years for dialysis with expected life-years for transplantation but adjusted for quality of life differences by assuming the quality of a year on dialysis only equated with ¼ of a year with a transplanted kidney. A similar example by Stason and Weinstein (1977) calculated effectiveness ’in terms of increased years of life expectancy from blood-pressure control, adjusted for changes in the quality of life due to the prevention of morbid events, on the one hand, and to medication side effects, on the other.’

These both illustrated the general idea that when interventions alter the probability of survival but also alter the probability among survivors of having varying degrees of morbidity, downward

or upward adjustments, for more vs. less severe levels of morbidity, provides a weighting scheme to yield a single effectiveness measure even if outcomes are heterogeneous (or multi-dimensional). Following the terminology of Klarman et al. (1968), Stason and Weinstein (1977), and Weinstein and Stason (1977), the label of "Quality Adjusted Life Years" (QALYs) has been applied to these effectiveness measures.

Second, the economists doing health care CEA's realized that the only source for these QALY weights that was ultimately non-paternalistic was the preferences of the individuals whose risks for various morbidity levels were being impacted by the policies being evaluated. The idea here was to make the process for deriving weights at least roughly consistent with the BCA stricture that WTP's for health benefits should be based on the expressed preferences of the individuals experiencing these benefits (assuming they are also fully informed of the nature of these benefits). This second concern led to numerous efforts to apply preference assessment methods for deriving the weight for each relevant health state (i.e., morbidity level). These methods (for example, using standard gamble (sg) and time tradeoff (tto) questions, or visual analog scale (vas) designations) seek to measure the "health utility" of a particular health state relative to extreme health states (for example, perfect health, death).

It is also important to note that this process involves two separate analytic steps: 1 defining each specific health state and 2 applying an assessment method (e.g., standard gamble, time tradeoff, visual analog scale) to derive health-state specific utility value (HSSU) for one unit of time (for example, one year) in that specific health state. Various strategies have been used for constructing a "scenario" that defines each specific health state. One common approach is to use the separate dimensions used in a health utility scale for this purpose. For example, using the EuroQOL EQ-5D dimensions (mobility, self-care, usual activities, pain/discomfort, anxiety/depression), a specific health state could be described by specific levels of function vs. dis-function (1=best to 3=worst) for each of the dimensions.

By applying these assessment methods in CEA for each health state, we can obtain the resulting unidimensional measure of effectiveness that is simply the increase in expected "health utility" produced by a new health policy relative to the status quo, or that same increase compared across

multiple health policy alternatives. (When multiple time periods are considered and discount rates are used to weight present vs. future "health utility" gains, the corresponding effectiveness measures is simply the discounted expected health utility gain.)

This mode of analysis, which is now termed cost-utility analysis (CUA), offers several advantages over the original simple CEA approach that ignored heterogeneity in outcomes. First, it is in some sense consumer-preference-based and does not involve paternalistically imposed weights from "experts" or policy-makers.[2] Second, it provides a measure of effectiveness which abstracts from any specific outcomes produced by specific types of interventions (e.g., cases of influenza prevented); this means it can be used to compare CU ratios across many differing types of health policy interventions.

A Simple Example of Using Health-State-Specific Utility (HSSU) Weights to Calculate Gains in Expected Health Utility

We can illustrate the application of HSSU weights to compute a single effectiveness measure with a simplified example based in part on Drummond et al. (2015, Chap. 9). In this example, we consider two alternative interventions: 1 a policy of treating HIV-positive persons with a single drug therapy ("monotherapy") and 2 a policy of treating HIV-positive persons with a multiple drug "cocktail" ("combotherapy"). We assume there are just five alternative health states defined by HIV-positive status and CD4 cell count in the blood: (1) healthy and not HIV-positive, (2) HIV-positive but fairly healthy (CD4 count >200 and < 500 cells per mm^3), (3) HIV-positive and less healthy (CD4 count <200 cells per mm^3), (4) full-blown AIDS, and (5) dead. (Note that we also assume persons not HIV-positive at the start of the policy may become HIV-positive in later years of the policy; thus health benefits are reckoned for these persons as well rather than for person who happen to be HIV-positive at the start of the policy. The general point is that the expected health benefits represented by the change in health-state probabilities extend beyond current patients to a presumably much larger group of potential patient beneficiaries of the policy.)

Assume the health-state transition probabilities for one year to the next for monotherapy and combotherapy policies are given by the two panels in Tables 12.1 below:

TABLE 12.1: YEAR-TO-YEAR TRANSITION PROBABILITIES					
MONOTHERAPY					
Health State:	Healthy	200<CD4<500	CD4<200	Aids	Dead
Healthy	0.8	0.194	0	0	0.006
200<CD4<500		0.721	0.202	0.067	0.01
CD4<200			0.581	0.407	0.012
Aids				0.75	0.25
Dead					1
COMBOTHERAPY					
Health State:	Healthy	200<CD4<500	CD4<200	Aids	Dead
Healthy	0.8	0.194	0	0	0.006
200<CD4<500		0.86	0.103	0.03	0.007
CD4<200			0.581	0.407	0.012
Aids				0.75	0.25
Dead					1

Note that the transition probabilities for moving from "healthy" to any other state are the same for both therapies (since they have no primary prevention effect), but the combination therapy reduces the rate of transition from 200<CD4<500 to CD4<200, from 200<CD4<500 to full-blown AIDS, and from 200<CD4<500 to dead.

Since Table 12.1 only gives the initial one-year transition probabilities, we can extend the description of the benefits of combination therapy to longer periods by repeating the transition process for a series of additional yars. Assuming a 10-year life span, the corresponding 10-year sequences of health state probabilities are given by Appendix Tables A12.1 and A12.2 respectively. The 10-year sequence of changes in these probabilities due to adopting combotherapy is shown in Appendix Table A12.3. These changes represent the health benefits of combotherapy (relative to monotherapy), and a monetary WTP figure for these changes is needed for a BCA of adopting combotherapy.

Suppose we assume the following HSSU QALY weights: 1 for state (1), 0.8 for state (2), 0.6 for state (3), 0.4 for state (4), and 0 for state (5). The results of applying these weights to the change in health-state probabilities implied by switching from monotherapy to combotherapy are shown in Table A12.2:

TABLE 12.2: QALY VALUES OF CHANGES IN YEARLY HEALTH-STATE PROBABILITIES: MONO VS. COMBO THERAPIES

Health State	Healthy	200<CD4<500	CD4<200	Aids	Dead	Yrly. Total	PV of Yrly. Tot. QALYs w. Discount Factor =1/1.03	Discount
HSSU Weight (QALY)	1	0.8	0.6	0.4	0			Factor
QALY Change								0.9709
Year 1	0	0	0	0	0	0	0	1
Year 2	0	0.0216	-0.0115	-0.0029	0.0000	0.0072	0.0070	0.9709
Year 3	0	0.0514	-0.0226	-0.0072	0.0000	0.0216	0.0216	0.9426
Year 4	0	0.0816	-0.0291	-0.0157	0.0000	0.0368	0.0357	0.9151
Year 5	0	0.1083	-0.0310	-0.0235	0.0000	0.0538	0.0507	0.8885
Year 6	0	0.1294	-0.0290	-0.0293	0.0000	0.0711	0.0651	0.8626
Year 7	0	0.1445	-0.0246	-0.0323	0.0000	0.0876	0.0778	0.8375
Year 8	0	0.1539	-0.0189	-0.0327	0.0000	0.1023	0.0882	0.8131
Year 9	0	0.1582	-0.0129	-0.0308	0.0000	0.1145	0.0959	0.7894
Year 10	0	0.1583	-0.0072	-0.0272	0.0000	0.1239	0.1008	0.7664
10-yr. Total	0	1.0071	-0.1869	-0.2015	0.0000	0.6188	0.5428	

The expected QALY gain per person for combotherapy relative to monotherapy is 0.6188 QALYs undiscounted; assuming a 3 per cent annual discount rate, the corresponding present value of the expected QALY gain is 0.5428. Assuming there are N potential beneficiaries with a non-zero probability of ever receiving combotherapy (over the time horizon of the policy), and that the QALY weights, discount rate, and health state probabilities are the same across these N persons, the expected QALY gain from the program is simply N times the per person expected QALY-gain figure.

From CEAs to CUAs to "League Tables" and "Acceptability" Thresholds

Using preference-based effectiveness measures (expected QALY gains), CUA moved beyond the restrictions of the original CEA applications that relied on unidimensional "physical" outcome measures (e.g., numbers of influenza cases prevented) which could only compare interventions producing the same types of outcome units and were often funded out of the same constrained public agency budgets. Instead, CUAs could compute cost-utility ratios (CURs) with monetary cost measures in the numerator and effectiveness measures of gains in expected "health utilities" in the denominator and could treat these CURs as commensurable across virtually any types of health programs. This implied that choices between qualitatively very different health programs could be compared to one another via "league tables" which simply ranked many different and diverse programs on the basis of their CURs (with lower CURs indicating more "efficient" programs).

Table 12.3 shows an example of a league table. Each row in the table shows a comparison of two different policies – a "new" policy and a "comparator" (which could be a status quo of doing nothing). The second and third columns of the table show the difference in costs between the two policies as well as the difference in "effectiveness" measured in QALYs per person.

Table 12.3: Selected Examples from a League Table			
Intervention	Change in Cost (2010 $'s)	Change in QALY	Cost-Effectiveness Ratio (2010 $'s)
Coumadin (Warfirin) compared to aspirin for 70 year-olds with atrial fibrillation	3,000	0.81	3,704
Diabetes education and self management compared to standard care for patients newly diagnosed with type 2 diabetes	200	0.04	5,000
Daily dialysis compared to dialysis every other day for 60 year-old critically ill men with kidney injury	13,000	2.14	6,075
ICD (implantable cardioverter defibrillator) compared to current standard of care to prevent sudden cardiac death for patient who are at risk for sudden death due to left ventricular systolic dysfunction	113,000	3	37,667
HIV counseling, testing, and referral compared to current standard of care in high risk populations (HIV annual incidence 1.2% and prevalence of undiagnosed HIV 0.3%)	1,000	0.03	33,333
Spine surgery compared to nonoperative treatment for adult patients with confirmed spinal stenosis and spinal nerve-based (radicular) leg pain	15,000	0.17	88,235
Annual CT screening compare to no screening for 60 year-old heavy smokers who are eligible for lung reduction surgery	6,000	0.04	150,000
Screening for osteoporosis with a bone densitometry and osteoporosis treatment compared to no densitometry or treatment for men age 65 and older with no prior fracture	4,000	0.03	133,333
Source - http://healtheconomics.tuftsmedicalcenter.org/cear4/Resources/LeagueTable.aspx			

The original idea of the league table was that all the comparisons in the tables demonstrated alternative ways to spend funds from a "health sector budget" on the "new" policy. Looking at a single row in the table, you will typically see a "new" policy that involves additional costs but produces additional effectiveness; incremental costs are positive, incremental effectiveness is

positive, and the incremental cost effectiveness ratio (ICER) is also positive. The problem for the policy-maker is to decide whether the incremental effectiveness is large enough to warrant spending the incremental costs. Looking at this problem for only a single row (i.e., policy choice), the policy-maker can only make an intuitive judgment since she(he) does not have a valuation of incremental effectiveness in monetary terms that would be provided by a BCA. The idea of presenting a league table with multiple rows was devised as a way to facilitate the policy-maker's choices on individual programs. All policy choices in the tables are ranked in ascending order of the ICER; the various policy choices involve a wide range of dissimilar interventions, but this is not problematic since effectiveness is measured in commensurable units (QALYs per person) across all choices. The policy choices typically range widely in their ICER values, and they also typically include at least some choices on which policy decisions have already been made. Thus, policy-makers facing the need to make one specific decision can see from the table whether other previous policy choices that involved higher or lower ICERs were or were not in fact adopted.

For example, if the choice in question has a lower ICER than other choices that were previously adopted, the decision makers could view this as supporting the adoption in the current choice since this would not implicitly place a higher value (in terms of dollars per QALY gained) than other interventions that had previously been chosen. By similar reasoning, the retrospective use of the league table could be employed to identify particular previous policy choices that implied a much higher dollars per QALY value than other interventions that were not adopted. In the broadest case, one could view all recent health policy choices as subject to re-evaluation, and thus use a comprehensive league table to identify the set of non-mutually exclusive policy choices that would maximize overall effectiveness (in terms of QALYs) given the health sector "budget" constraint.[3]

Finally, many commentators have suggested that there is a rough consensus (at least within a particular country) as to the highest ICER value that would be "acceptable" for any policy choice. This ICER "threshold" then becomes the recommended marginal monetary valuation for an incremental gain in effectiveness, and the implication is that the "net monetary benefit" is positive for any intervention with an ICER below this threshold value. In short, the use of CUA

coupled with an "acceptable" threshold has allowed us to obtain conclusions about net benefits in dollar terms of policy choices and thereby circumvented the purported limitations and impracticalities of the "text-book" BCA that required measurement of WTP in monetary terms for all relevant effects. As CEA of health policies morphed into CUA of health policies and was paired with "consensus" dollars-per-QALY "threshold" values, the ultimate result was a revival of BCA in health, albeit in the form of cost-utility-benefit-analysis (CU(B)A).

Progress in Measuring the WTP for the Benefits of Health Policies?

As noted above, the original impetus for use of CEA rather than BCA in health resulted not only from invalid presumptions about BCA "placing a dollar value on health" and biasing policy choices toward the wealthy, but also from valid concerns about the difficulty of obtaining accurate and valid WTP figures for benefits of health policies. The subsequent development, in CUA, of a preference-based approach to effectiveness measurement required a clearer and more useful strategy for defining the nature of the benefits of health policies. Specifically, it led to explicit representation of these benefits as a set of changes in the probabilities of being in any potentially relevant health states over the potential beneficiary's life span.

Thus, in our simplified example (in Section V above), we described the health benefits of combination therapy for HIV-AIDS relative to monotherapy as the difference between two "lifetime" sequences of probabilities of being in each of 5 distinct health states over 10 different years (the differences between Appendix Tables A12.3 and A12.4). Note that this approach of describing health benefits as changes in the lifetime sequences of health state probabilities can be used for <u>any</u> health policy intervention; "real-world" examples are more complex.

With the development of CU(B)A, we now have at least two different paradigms for deriving a monetary valuation of these health benefits:

> A) The "pure" BCA paradigm bases valuation on: 1 informing potential beneficiary consumers of the specific nature of each relevant health state, 2 informing them of the year-by-year changes in the probabilities of being in each of these relevant health states, and 3 eliciting a monetary WTP figure from each consumer for the entire set of these changes in probabilities, and 4 adding up these WTP figures (appropriately discounted) across the population of potential beneficiaries.

2) The typical CU(B)A bases valuation on: 1 obtaining the HSSU QALY- weight for the "average" consumer for each relevant health states, 2 computing the change in discounted expected QALYs based on these HSSU weights, 3 applying the "average" consumer's monetary value per QALY to this change in discounted expected QALYs, and 4 multiplying the result (appropriately discounted) by the number of potential beneficiaries.

We expect that the results obtained from these two methods will not necessarily be approximately equal. Several major differences between them should be noted.
CU(B)A in its simplest form – focusing on a single "average" potential beneficiary - does not account for heterogeneity among potential beneficiaries in the following:

a) their HSSU preference weights for any specific health states,
b) their monetary valuations per QALY[4],
c) their risk preferences,[5] or
d) the influence of the sequence in which specific health states occur on their HSSU weights.[6]

To some extent, heterogeneity among potential beneficiaries could be accounted for in CU(B)A by doing separate analyses for sub-groups of beneficiaries. This would, however, not eliminate the assumptions about risk-neutrality or sequence independence of preferences (items (c) and (d) above). Such assumptions are required in order for the CB(U)A approach to allow the application of a single set of HSSU weights.[7]

The "pure" BCA paradigm does not require these simplifications and assumptions, but the practicalities are daunting. One could attempt to elicit a single WTP figure directly from a representative sample of potential beneficiaries for valuing all of the relevant changes in health state probabilities. This could be done using any one of a variety of WTP question formats. It would involve presenting, to a representative sample of individual respondents: 1 detailed descriptions (sometimes called "scenarios") of all relevant health states and 2 the full set of changes in probabilities of being in each relevant state over the respondents' lifetimes.

Is it reasonable to expect that the respondents could understand the scenarios and the changes in

probabilities well enough to provide responses to WTP questions that accurately reflected their true preferences? Several considerations suggest that the answer to this question is no. These include the fact that many respondents will never have actually experienced the scenarios corresponding to the relevant health states. It may also be presumed that respondents will have difficulty in processing the meanings of small changes in probabilities and assigning monetary values to these changes that reflect their true preferences.

These practical problems with the "pure" BCA paradigm provide the principal argument for CU(B)A, namely that it provides a simplified strategy for eliciting individuals' preferences for various health states and for coming up with a "shadow price" of QALYs (i.e., a "threshold" dollar value for a QALY). On the other hand, the CU(B)A approach involves the series of strong assumptions and simplifications noted above.

Finally, note that coupling of the CU(B)A approach with the use of league tables and "consensus" threshold values does not entirely avoid the CU(B)A simplifications and assumptions. In addition it raises several other major questions. Should we be looking at past decisions on health policy interventions as at all determinative for current health policy decisions? Should we base policy choice decisions on a single "consensus" threshold value in all cases? Is there any consensus at this point about the correct threshold value? Should we instead recognize the possibility that the threshold value depends on the specifics of the policy choice and therefore could vary across differing kinds of decisions? In response to the numerous unanswered questions about the use and level of CEA thresholds, the report of the Second Panel on Cost-Effectiveness in Health and Medicine has labeled further inquiry into these questions as a "key area for future research." (Neumann et al., 2017, Chap. 2).

Unanswered Questions and Thoughts on Next Steps

As a result of the development of CEA into CUA into CU(B)A described above, we are now left with two similar yet different paradigms for devising monetary valuations of health program benefits. In both paradigms, the program benefits are described in terms of year-by-year changes in the probabilities of being in specific health states. One paradigm, CU(B)A, involves simplifying assumptions that may not be valid. The other paradigm, "pure" BCA, presents

serious practical challenges in the WTP elicitation process that also raise validity questions. Thus, comparison of results from applying each paradigm to the same policy may be of interest, but is not a test of accuracy per se.

At this point, research toward improving accuracy of results has therefore focused on testing the validity of assumptions in the CU(B)A. Several recent studies by Bobinac and colleagues illustrate some interesting new directions in estimating WTP figures for QALYs. In particular, they published two recent studies (Bobinac et al., 2010; Bobinac et al., 2014) that allow comparisons of WTP-per-QALY values obtained from elicitation methods that specifically include risk vs. results obtained with conventional elicitation methods that involve valuation of an outcome that is certain. In particular, they compared WTP per QALY estimates from elicitation questions that explicitly included uncertainty with their own previous results with similar questions about prospects that were certain, a substantial increase in implied WTP per QALY was observed. This result: 1 reinforces our conjecture that risk preferences are in fact important in assessing WTP of health policies, while 2 raising concerns about deficiencies in the CU(B)A reliance on HSSU weights based on elicitation questions that assume risk neutrality.

Bobinac et al. (2010) also find substantial variation among survey respondents in their expressed WTP per QALY values based on income differences; this implies that the use of a single "correct" monetary QALY value in the standard CU(B)A approach is at variance with consumer preferences and reinforces the need to look explicitly at distributional equity issues in CU(B)A analyses. They also report empirical results (Bobinac et al., 2012) that do not support an assumption that individuals' WTP per QALY figures are invariant to the size of the QALY gains provided by an intervention; this is at variance with the standard CU(B)A procedure of using a single WTP-per-QALY figure to compute "monetary net benefit" of an intervention.

In conclusion, further methodological development is needed both: 1 to devise modifications to the current CU(B)A procedures that are more consistent with this recent empirical evidence, and 2 to better cope with the practical difficulties of implementing BCA valuations that at least approximate the results from applying the BCA paradigm.

References and Recommendations for Further Reading

Bellinger W.K. (2016), *The Economics Analysis of Public Policy* (2nd edition). New York: Routledge.

Bobinac A, N. J. A. van Exel F. F. H.Rutten, W. B. F. Brouwer(2010), 'Willingness to pay for a QALY: the individual perspective', *Value Health*.;**13**,1046–55.

_____.(2012), 'Get more pay more? An elaborate test of construct validity of willingness to pay per QALY estimates obtained through contingent valuation', *J Health Econ*. **31**,158–68.

_____ (2014), 'The Value of a QALY: Individual Willingness to Pay for Health Gains Under Risk', *PharmacoEconomics* **32**,75–86

Drummond M. F., M. J. Sculpher , K. Claxton G. L. Stoddart, and G. W. Torrance (2015), *Methods for the Economic Evaluation of Health Programs* (4th ed.). Oxford: Oxford University Press.

Huang, E. S., Q. Zhang, Q., S. E. S. Brown, , M. L. Drum, ., D. O. Meltzer, and M. H. Chin (2007), 'The Cost-Effectiveness of Improving Diabetes Care in U.S. Federally Qualified Community Health Centers', *Health Services Research*, **42**, 2174–2193

Johannesson M., D. O. Meltzer and R. M. O'Conor (1997), 'Incorporating Future Costs in Medical Cost-Effectiveness Analysis: Implications for the Cost-Effectiveness of the Treatment of Hypertension', *Med Decis Making* **17**,382-389.

Kenkel D.(1997), 'On valuing morbidity, cost-effectiveness analysis, and being rude', *Journal of Health Economics* **16**,749-57.

Klarman HE, J. O. S. Francis, G. D. and Rosenthal(1968), 'Cost effectiveness analysis applied to the treatment of chronic renal disease', *Medical Care* **6**(1)(Jan. - Feb., , 48-54.

Klarman H.E. (1982), 'The road to cost-effectiveness analysis', *The Milbank Memorial Fund Quarterly. Health and Society*, **60**(4)Autumn, . 585-603.

Meltzer, D.O. (1997), 'Accounting for future costs in medical cost-effectiveness analysis,' *Journal of Health Economics* **16**(1), 33-64.

Meltzer, DO, Egleston B, Stoffel, D, and Dasbach, E. Effect of Future Costs on Cost-Effectiveness of Medical Interventions Among Young Adults: The Example of Intensive Therapy for Type 1 Diabetes Mellitus. *Medical Care*: June 2000 - Volume 38 - Issue 6 - pp 679-685.

Neumann P.J., G. D. Sanders L. B. Russell J. E. Siegeland T. G. Ganiats, (eds.). (2017),*Cost-Effectiveness in Health and Medicine*, 2nd ed. Oxford: Oxford University Press,

Salkever, David. (2013), 'Social costs of expanding access to evidence-based supported employment: concepts and interpretive review of evidence', *Psychiatric Services* **64**(2), , 111-119.

Stason W.B., and M. C. Weinstein (1977), 'Allocation of resources to manage hypertension', *New England Journal of Medicine*, **296**(13),732-9, March 31.

Weinstein M.C, and W. B. Stason (1977), 'Foundations of Cost-Effectiveness Analysis for Health and Medical Practices' *New England Journal of Medicine*, **296**(13),716-721 (March 31).

Willig R. (1976), 'Consumer surplus without apology', *American Economic Review* **66**(4): 589-97.

Zerbe R.O. and A. S. Bellas (2006) *A Primer for Benefit-Cost Analysis*. Northampton MA: Elgar.

NOTES

[1] This criticism of using an unweighted sum of WTP figures to measure net benefit is also often based on the argument that the marginal utility of income declines with the level of income so the utility value of one dollar of WTP for the poor is greater than it is for the rich. (See, for example, Bellinger, Chaps. 3 and 6.) From the neoclassical welfare economics perspective, this argument is flawed since it requires interpersonal comparisons of utility (e.g., that 1 "util" is the same for a poor person as for a rich person) that are not empirically verifiable.

[2] Also note, however, that CUA does typically treat units of effectiveness in "utility" terms as equivalent across all persons; such interpersonal utility comparisons are not required in BCA.

[3] The notion of the relevant budget constraint, however, becomes problematic in CUA. Funding for treatments in the league table come from sources besides a single public agency's budget (e.g., patient-payments, private insurance, etc.) and span multiple years (over which agency budgets change). Similar complications arise when the "health-sector" cost of treatment is defined to include impacts on all other health services received by the patients undergoing treatment (including costs for "related" and "unrelated" health problems). Inclusion of other "societal" costs in the C-U ratio denominator of "societal" CUA's (e.g., patients' time and travel costs, impacts on patients' earnings and non-health resource consumption) led to Meltzer's (1997) fundamental contribution on defining the relevant "societal" resource constraint in constructing the incremental C-U ratio denominator for health interventions. More recent studies have applied Meltzer's societal-cost conceptualization to specific treatment/policy interventions (Huang et al. , 2007; Johannesson et al., 1997; Meltzer at al., 2000, and Salkever, 2013). This approach is also recommended by the Second Panel on Cost Effectiveness in Health and Medicine (Neumann et al., 2017, Chap. 3).

[4] Alternatively, this valuation is often a "threshold" value from external sources, that may be quite different from the monetary valuations per QALY of the "average" consumer.

[5] In the usual s-g approach, respondents are assumed to be risk-neutral. The t-t-o and v-a-s elicitation methods assume that uncertainty is irrelevant, since they entail eliciting responses to choices between certain health states. Risk preferences are, however, of potential importance for assessing interventions with different risk profiles (for example., comparing a high-risk surgical interventions with high expected levels of effectiveness vs. lower-risk non-surgical interventions with lower expected levels of effectiveness).

[6] CU(B)A assumes that the relevant utility weight for any specific health state is the same regardless of the health states which preceded it in time and/or followed it in time.

[7] Also note that the s-g, t-t-o, and v-a-s elicitation methods assume the resulting utility weights are invariant to the time durations used in the elicitation questions. A disconcerting feature of this assumption is that it implies a discount rate of zero.

Appendix Tables

TABLE A12.1:
10-YR. SEQUENCE OF YEARLY HEALTH-STATE PROBABILITIES--
MONOTHERAPY

Health State	Healthy	200<CD4<500	CD4<200	Aids	Dead
Year 1	0.800	0.194	0.000	0.000	0.006
Year 2	0.640	0.295	0.039	0.013	0.013
Year 3	0.512	0.337	0.082	0.036	0.023
Year 4	0.410	0.342	0.116	0.083	0.040
Year 5	0.328	0.326	0.136	0.132	0.068
Year 6	0.262	0.299	0.145	0.177	0.108
Year 7	0.210	0.266	0.145	0.212	0.158
Year 8	0.168	0.233	0.138	0.235	0.217
Year 9	0.134	0.200	0.127	0.248	0.280
Year 10	0.107	0.170	0.114	0.251	0.347

TABLE A12.2 :
SEQUENCE OF YEARLY HEALTH-STATE PROBABILITIES: COMBOTHERAPY

Health State	Healthy	200<CD4<500	CD4<200	Aids	Dead
Year 1	0.800	0.194	0.000	0.000	0.006
Year 2	0.640	0.322	0.020	0.006	0.012
Year 3	0.512	0.401	0.045	0.018	0.020
Year 4	0.410	0.444	0.067	0.044	0.031
Year 5	0.328	0.462	0.085	0.073	0.048
Year 6	0.262	0.461	0.097	0.103	0.073
Year 7	0.210	0.447	0.104	0.131	0.104
Year 8	0.168	0.425	0.106	0.154	0.143
Year 9	0.134	0.398	0.106	0.171	0.187
Year 10	0.107	0.368	0.102	0.183	0.234

TABLE A12.3:

CHANGES IN YEARLY HEALTH-STATE PROBABILITIES: COMBO THERAPY vs. STATUS QUO (MONOTHERAPY)

Health State	Healthy	200<CD4<500	CD4<200	Aids	Dead
Year 1	0.000	0.000	0.000	0.000	0.000
Year 2	0.000	0.027	-0.019	-0.007	-0.001
Year 3	0.000	0.064	-0.038	-0.018	-0.003

Year 4	0.000	0.102	-0.049	-0.039	-0.009
Year 5	0.000	0.135	-0.052	-0.059	-0.020
Year 6	0.000	0.162	-0.048	-0.073	-0.035
Year 7	0.000	0.181	-0.041	-0.081	-0.054
Year 8	0.000	0.192	-0.032	-0.082	-0.074
Year 9	0.000	0.198	-0.022	-0.077	-0.094
Year 10	0.000	0.198	-0.012	-0.068	-0.113