Record Type: Record

To: Renee P. Helm/OMB/EOP@EOP

cc:
Subject: Krupnick, #360

Alan J. Krupnick, PhD
Resources for the Future
1616 P Street, NW
Washington, D.C. 20036
202-328-5107; 202-939-3460 (fax)

- OMB6.DOC

Message Sent To:

John Graham/OMB/EOP@EOP
Bernie O'Brien <obrienb@mcmaster.ca>
Bill Lawrence <wlawrenc@ahrq.gov>
Carol Scrotton <cxr3@cdc.gov>
Chris Dockins <dockins.chris@epa.gov>
Clark Nardinelli <clark.nardinelli@cfsan.fda.gov>
Dan Brock <dbrock@mail.cc.nih.gov>
David Feeny <dfeeny@pharmacy.ualberta.ca>
Debbie Aiken <deborah.aiken@osha.gov>
Elise Golan <egolan@ers.usda.gov>
"F. Reed Johnson" <frjohnson@rti.org>
Jim Hammitt <jkh@hsph.harvard.edu>
Jim Schuttinga <schuttij@od1tm1.od.nih.gov>
Joe Lipscomb <lipscomj@mail.nih.gov>
Josephine Mauskopf <jmauskopf@rti.org>
Kelly Maguire <maguire.kelly@epamail.epa.gov>
Kerry Smith <vksmith@are1.cals.ncsu.edu>
"Krupnick, Alan" <Krupnick@rff.org>
Laura Blanciforti <lob9@cdc.gov>
Maureen Cropper <cropper@econ.umd.edu>
Michael Hanemann <hanemann@are.berkeley.edu>
"Morgenstern, Dick" <Morgenst@rff.org>
Peter Belenky <Peter.Belenky@ost.dot.gov>
Phaedra Corso <pas7@cdc.gov>
Scott Grosse <sgg4@cdc.gov>

"Taylor, Mike" <Taylor@rff.org>

Ted Miller <MILLER@pire.org>

Warren Prunella <wprunella@cpsc.gov>

Message Sent To:

Arthur G. Fraas/OMB/EOP@EOP

John F. Morrall III/OMB/EOP@EOP

Richard P. Theroux/OMB/EOP@EOP

Paul R. Noe/OMB/EOP@EOP

Veronica Vargas/OMB/EOP@EOP

COMMENTS ON

OMB DRAFT GUIDELINES FOR THE CONDUCT OF REGULATORY ANALYSIS AND THE FORMAT OF ACCOUNTING STATEMENTS

Alan J. Krupnick
Resources for the Future

May 30, 2003

Acknowledgments

I would like to acknowledge the insightful and careful detailed comments on drafts of this paper from Kelly Maguire and Chris Dockins, EPA; Clark Nardinelli, FDA; Peter Belenky, DOT; and Carol Scotton, Scott Grosse and Phaedra Corso, CDC, all members of the Steering Committee on Valuing Health Outcomes, as well as Reed Johnson, RTI and David Feeny, Institute of Health Economics. Thanks also to Elise Golan, USDA; Ted Miller, Pacific Institute for Research and Evaluation; my colleague at RFF, Mike Taylor; Laura Blanciforti, NIOSH (also on the Steering Committee); and Dan Brock, NIH, for their very helpful comments. In spite of their help, the content of this report is solely the responsibility of the author.

Readers should note that the author is an environmental economist producing studies that estimate the willingness-to-pay for health improvements and mortality risk reductions as well as studies comparing the benefits and costs of environmental regulations. The comments and recommendations in this paper come out of this professional experience and context, as well as the intense debates of the last few months on the relative merits of various approaches to valuing health outcomes – an unsettled and relatively unresearched debate.

I. Introduction

Regulatory and non-regulatory activities by governments at the federal, state, and even local level affect the economic and physical health of their constituents. Analytical tools and concepts are available and are often used to evaluate and compare the effectiveness, efficiency, and distributional implications of these proposed or actual activities – namely cost-benefit analysis (CBA) and cost-effectiveness analysis (CEA).

Once the decision is made to use these tools, a further decision is needed about the measures to use for health outcomes. These measures include simple counts of cases (e.g., the number of deaths) prevented and indexes to aggregate over the variety of health effects that might accompany any particular course of action contemplated or taken by the government. Broadly speaking, there are two types of indexes: monetary measures (e.g., willingness-to-pay (WTP) and cost-of-illness (COI)[1]) and indexes over health states, referred to by the general term Quality-Adjusted Life-Years (QALYs).[2] The former are used in CBA, the latter (as well as physical measures) are used primarily in CEA.[3]

The choices of tools and health benefit measures are often controversial. Some statutes (such as the Clean Air Act) forbid use of cost-benefit analysis to make certain types of decisions (for example, on the stringency of the National Ambient Air Quality Standards), while others (such as the Safe Drinking Water Act) mandate its use. Irrespective of statutory requirements, federal agencies are required under Bush's Executive Order 13258 – and Executive Orders from the Clinton to the Reagan Administrations – to evaluate the costs and benefits of major regulations, defined as those expected to have an effect on the economy of at least $100 million dollars annually. For agencies whose activities are not regulatory or whose regulations don't exceed the $100 million threshold, the choice of tools may be a matter of agency culture (although OMB has several other criteria for what actions require a CBA). While the U.S. Environmental Protection Agency (EPA) typically performs cost-benefit analyses, and may or may not supplement them with cost-effectiveness analyses, other agencies, such as the Occupational Safety and Health Administration (OSHA), generally perform cost-effectiveness analysis and do not monetize either morbidity or mortality benefits.

The Office of Management and Budget (OMB) is now attempting to change how government agencies do their analyses with its "OMB Draft Guidelines for the Conduct of Regulatory Analysis and the Format of Accounting Statements" (termed Draft

[1] COI measures typically include medical costs, loss in productivity from illness, and premature death losses valued as lost productivity (termed the human capital approach).

[2] In this report, QALYs are a term used for a wide variety of indexes that permit aggregation of morbidity effects in terms of health states and the duration they are experienced, as well as life-years lost.

[3] Use of QALYs in a cost-effectiveness analysis is sometimes termed cost-utility analysis (CUA), referencing the fact that certain types of health indexes are based on Von Neumann-Morgenstern cardinal utility theory.

Guidance here), which is Appendix C in its *Draft 2003 Report to Congress on the Costs and Benefits of Federal Regulations*. This Guidance calls for expanded use of cost-effectiveness analysis for regulatory analyses. The effectiveness measures are not spelled out, but it is thought by some observers that this is a call for more analyses based on cost per QALY or cost per life-year gained.

The primary purpose of this paper is to provide commentary to OMB on its Draft Guidance. Practitioners in government and stakeholders interested in agencies' Regulatory Impact Analysis (RIA) procedures may also find this document useful.

The responsibility for the ideas and commentary in this paper lies wholly with the author as a member of RFF. In large measure the intellectual underpinnings for this paper can be attributed to a conference and workshop held at the behest of an Interagency Steering Committee headed by the author and Michael Taylor, also of RFF. The Steering Committee members include: Kelly Maguire and Chris Dockins, EPA; William Lawrence, AHRQ; Elise Golan, USDA; James Schuttinga, NIH; Clark Nardinelli, FDA; Jim Schaub, ORACBA, USDA; Joseph Lipscomb, NIH; Debbie Aiken, OSHA; Peter Belenky, DOT; Greg Rodgers, CPSC; Laura Blanciforti, NIOSH and Carol Scotton and Phaedra Corso, CDC.

The conference on *Valuing Health Outcomes* was held February 13-14, 2003 at the RFF Conference Center in Washington, D.C., and involved some 25 speakers and 175 participants who contributed a great deal of information describing the conceptual and empirical foundations of the health valuation measures and their use in CBA and CEA. The follow-up workshop was held April 29, 2003, and involved a subset of this group in detailed discussions on a major paper written by the author, which analyzed CEA, CBA, WTP, QALY and COI measures in detail. At this workshop it was decided that a separate, more focused paper – this paper – would be written by the author to provide comments to OMB on its Draft Guidance.

The recommendations made in this paper are not necessarily consistent with views of any of the steering committee members or participants in the conference and workshop. Indeed, some members and participants disagree with some of these recommendations. Also, some topics in this paper lie at the fringes of topics discussed at the conference or workshop.

The remainder of the paper describes the provisions of the OMB Draft Guidance that are relevant to the debate over CEA and CBA analyses and appropriate health outcome measures and provides commentary and recommendations associated with these provisions.

Note that none of these particular OMB provisions speak to the decision context in which CBA or CEA are placed or to the use of these and other analyses to address the distribution of effects across populations. Both of these factors are vitally important in understanding how regulatory decisions should and do get made. Further, the measures of health value discussed below embed implications for equity effects that must be

understood to choose among them. A paper discussing these and many related technical issues will be released soon, drawing on the conference and workshop presentations and discussions. However, these issues are largely absent from the paper below.

II. OMB Draft Guidance, Commentary and Recommendations

OMB's Draft Guidance proposes a number of changes to current procedures (or a reemphasis of current procedures) for the conduct of RIAs. There are seven provisions in the Guidance that overlap the interest of our conference and workshop. With each is included a brief summary of the provision, a commentary, and recommendations.

The OMB provisions are:

(1) Use CEA and CBA in the same regulatory impact analysis.
(2) Give agencies discretion on the choice of effectiveness measures in CEA, but require justification.
(3) Use a value of statistical life (VSL) that matches the case under study when monetizing mortality risk reductions.
(4) WTP measures based on revealed preference approaches or passing scope tests (if based on stated preference approaches) are favored for their enhanced credibility.
(5) Agencies must provide OMB with data to permit cross-rule calculations and comparisons.
(6) Rules with costs over $1 billion must use Monte Carlo simulation approaches to represent uncertainties.
(7) Draw on a literature that combines WTP and QALY estimates where gaps in WTP values exist.

1. Use CEA and CBA

The revised guidelines, with some exceptions, call for agencies to present both CEA and CBA in support of major proposed rules. The guidelines emphasize a call for CEA for all major rulemakings where the majority of benefits are improvements in health and safety. In addition, CBA should be conducted for major health and safety rulemakings "to the extent that valid monetary values can be assigned to the expected health and safety outcomes" (p.5516). CBA is justified because it "a) provides some indication of what the public is willing to pay for improvements in health and safety and b) offers additional information on preferences for health using a different research design than is used in CEA." (p. 5520). For regulations that are not focused on health and safety, CBA must be done, and a CEA should be performed when some of the benefits cannot be expressed in monetary units.[4]

[4] Performing a CEA if one is already doing a CBA is easy if you have a single physical health endpoint that generates the benefits part of the CBA. The complications for CEA begin when multiple physical health outcomes must be incorporated

Implementation of the proposed guidelines would have far-reaching effects. EPA rarely performs CEA, for instance, and the agency would need to start doing so. At the same time, some of the health agencies do not routinely perform CBA. Our reading of the OMB guidance is that it seeks to place both of these tools in a more equal position at all agencies whose regulatory decisions affect public health. Nevertheless, OMB is not expecting agencies to perform CBA where "valid monetary values" do not exist.

Political debates have been fierce about the use of CBA in government agencies and have resulted in suspicion that requirements for more analysis mask an intent for government paralysis. In this paper, these concerns are not addressed.

Ignoring the politics and focusing on informing decisions, it is hard to argue with the basic idea of providing additional information in a regulatory analysis, i.e., a CEA to supplement a CBA and a CBA to supplement a CEA. More information is generally desirable because it covers more perspectives. However, if the information provided is misleading or results in information overload to the point that decision makers fall back on heuristics, then more information may be undesirable. For instance, if these different analyses yield different policy prescriptions, decision makers will need to know in detail the meaning of the two analyses. If such meanings, which are complex and are based on many, often implicit, assumptions, cannot be clearly and succinctly conveyed, decision makers may simply ignore the least familiar analysis or "split the difference," or follow some other rule of thumb that may simplify or ignore vital pieces of information.

Several issues need clarification for this provision to be useful and workable, including: (i) the information provided by the two tools; (ii) the meaning of "valid monetary values," (iii) the practicality of developing an inventory of such measures; (iv) how to convey this added complexity to decision makers; and (v) the degree of harmonization implied by the draft guidance.

(i) Information Provided by the Tools

CBA, with WTP-based values used to value health improvements,[5] is clearly based on social utility and, thus can in theory answer the question of whether a particular course of action has net benefits to society. CEA itself cannot do this.[6] Physical effectiveness measures, such as life years or the number of lives saved, provide no information regarding society's preferences. However, both measures can rank-order options, albeit with different indices. CBA rank-orders cardinally according to a money metric; CEA rank-orders with a cost-per-unit-of-effectiveness metric.

[5] CBA using WTP metrics is to be distinguished from CBA using cost-of-illness metrics, which cannot make the claim of unambiguously measuring utility. More will be said about this difference below. Unless otherwise specified, CBA refers to analyses done with WTP metrics.

[6] Using the "price per QALY" approach to monetize QALYs can provide a net benefits measure, but this approach is problematic. See below.

CBA has the ability to combine mortality and morbidity improvements into one (monetary) measure. That is, willingness-to-pay measures can capture all of the endpoints associated with a particular policy option – health and non-health. This is appealing because many (if not all) policies address multiple endpoints. In reality, willingness-to-pay measures for morbidity endpoints are often unavailable (as well as non-health endpoints).

CEA can also combine mortality and morbidity effects if a QALY metric is chosen, but this metric has not been used to aggregate non-health endpoints (although non-health benefits could be subtracted from costs and used in a net cost-per-QALY analysis). Morbidity values are more available in QALY terms than as WTP measures but the former may be defined too narrowly to match epidemiological endpoints. If CEA is used with a physical measure of effectiveness instead, the dominant measure of program output is chosen, for example, lives saved or life-years saved. In this case, other health endpoints are either ignored[7] or monetized and netted out of cost. It may be impossible, however, to find values for the monetization, and ignoring other effects may seriously bias the analysis.

Recommendations

Agencies should develop a common understanding of the advantages and disadvantages of these analytical tools and develop facility in using these tools.

(ii) "Valid Monetary Values"

This term applies only to the use of CBA and is meant to apply to the benefits side of such an analysis, although it could apply to the cost side as well.

The term "valid" can apply in theory and empirically. Turning first to theory, WTP-based values would qualify as "valid monetary values," being based on the generally accepted paradigm of neoclassical welfare economics.

COI measures might not. Harrington and Portney (1987) and others show how COI measures fit into the neoclassical economics paradigm and conclude that they are generally considered a lower bound to WTP measures, because they leave out pain and suffering, for instance. Yet, in more complex and realistic models, this conclusion will not hold. Where individuals are insured, and do not see the marginal cost of their treatment decisions, medical costs may exceed willingness to pay, even without the COI measure capturing pain and suffering. COI could also overstate WTP because of technical inefficiencies in health care provision. The market for health care services is

[7] Consider the case of ranking alternative approaches to reducing fatality risks from automobile accidents using cost per life saved. The per life saved measure would ignore morbidity effects of automobile accidents, effects that might well dominate deaths in terms of public preferences.

not competitive, meaning prices for health care services may not reflect social opportunity costs.

Another approach for possibly obtaining valid monetary values is to compute QALYs and convert them to a monetary metric using a "price per QALY" approach. This situation is problematic. First, the theory for such a conversion rests on the validity of using the revealed preferences of government or health care decision makers as a proxy for individual (public) preferences. As such decisions rest (rightly) on many factors beyond public preferences, there is no reason to think that this proxy is at all accurate. Second, CBA is meant to be used as an indicator of social welfare. If QALYs are not such an indicator (see the discussion under provision 2 below) then converting them to monetary terms does not make them any more of an indicator.

From an empirical perspective, it is helpful to begin from current state-of-the-art practice of CBA, say in the most recent EPA RIAs under the Clean Air Act (USEPA, 2003). There, WTP measures are used to value arguably the most important as well as the most numerous effects, i.e., mortality risk reductions, chronic respiratory disease risk reductions, and reductions in acute health effects. COI estimates are used to value effects not covered elsewhere, such as hospitalizations. Thus, the state-of-the-art incorporates a mix of WTP and COI approaches. Some analysts also use factors to convert COI to a WTP measure by computing the ratio of observed WTP values to COI values for health endpoints that have these values available. WTP to COI ratios range from 2-3 for the endpoints generally used in air quality CBAs. CBAs are also performed, primarily for use in medical settings, by converting QALYs to monetary measures, generally using a single conversion factor. Such factors vary from $25,000 to $100,000 per QALY and higher. Sometimes, these factors are not actually termed "prices" or conversion factors, but are used as benchmarks to determine whether specific medical interventions are covered by insurance.

The use of factors to convert COI to WTP and to convert QALYs to monetary terms, while easy to do empirically, do not, in the author's judgment, pass the validity test. The COI conversion factor is an empirical observation over a narrow range of endpoints and might not hold up over a broader range. The QALY conversion factors are judgmental and arbitrary (as a measure of public preferences). Further, conferees agreed that there is no reason to think that a single conversion factor could be used to make such a conversion sensibly. How multiple conversion factors could be used or what they would be remains to be seen.

As for COI as a valid empirical measure, procedures for estimating these values have been refined over a long period and are routinely used in government analyses. There is reasonable consensus that COI will be lower than WTP for the types of endpoints examined for food safety and air quality improvements, particularly for the more serious illnesses. In this case the direction of the bias when using COI is known and regulations where benefits exceed costs can then be reliably described as delivering welfare improvements. But whether these relationships would hold more generally is an open question.

Empirically, the validity of both WTP and QALY measures remains highly contentious. Clearly, WTP measures are less often available for health endpoints of interest than QALY measures. A future paper will take up this issue in detail.

Recommendations

Members of the Steering Committee and other expert participants are particularly divided on the use of price per QALY factors in CBA. The author suggests that OMB discourage the use in CBA of both "price per QALY" factors and COI to WTP factors unless and until credible estimates are available.

Research is needed to develop WTP-based values for endpoints currently covered by COI estimates.

(iii) Practical Considerations

It will generally be easier for agencies doing CBA to do supplemental CEA analysis than for agencies doing CEA to do CBA, if the effectiveness measure is a physical one.

For agencies doing CBA already, CEA with a physical effectiveness measure is simply a subset of a regulatory analysis that estimates health effects and then monetizes them. If an agency is supplementing CBA with CEA expressed with a QALY effectiveness measure, the transformation of physical effects measures to QALYs may not be as straightforward, because there are many different health status indices to choose from and, as noted above, the degree to which one can map QALYs into specific epidemiological outcomes is unclear.

For agencies doing CEA, to do a CBA requires a means of converting the physical measures to monetary measures. In addressing morbidity effects, no WTP measures may be available. COI measures are generally available or can be estimated at fairly low cost.

With some agencies familiar with QALYs and in need of monetary measures, and other agencies familiar with monetary measures and in need of QALYs, there will be a big increase in demand for procedures and values (WTP, COI, and QALY weights) for use in such analyses. At the same time there is skepticism that the supply of appropriate and credible values is close to being adequate. For any of these measures it can be difficult to judge their generalizability. For instance, QALYs are not always based on a representative sample of the population with a particular problem.

Recommendations

It may be useful for OMB (or perhaps a special Review Board) to examine the supply of values by establishing data repositories or by endorsing integrated assessment/benefit models that serve to provide repositories and procedures for using the data. To a certain

extent, existing OMB (and other agency) guidance and standard operating procedures do just this.

Models are reasonably well established for estimating WTP for the health benefits of air quality improvements and for estimating the cost-of-illness associated with foodborne disease. Possibly the best example of the former is the recently released BENMAP model developed by OAQPS, EPA. A latter example is the recently released Foodborne Illness Cost Calculator from the Economic Research Service, USDA (http://www.ers.usda.gov/data/foodborneillness/). So-called league tables of QALY weights and scores associated with various health states and medical interventions are widely available (see http://www.hcra.harvard.edu/pdf/ preferencescores.pdf), although QALYs (and cost/QALYs) associated with medical interventions are far more numerous than those for particular health states, the latter probably being more useful for national policy initiatives.

Use of such tables and models should not be in any way mandated by OMB, but rather offered by OMB as a service to agencies doing RIAs and as an attempt to encourage harmonization of methods across agencies and to give RIAs more weight with decision makers. The Environmental Valuation Reference Inventory (EVRI, http://www.evri.ca/) developed by Environment Canada (with USEPA support) is a good example of such a database. Further, agencies could add to this database in some controlled manner, perhaps through peer review, so that their work would be available to and benefit all. This approach should avoid freezing the state-of-the-art and would recognize the significant heterogeneity of contexts in which agencies write rules.

(iv) Reporting

Because of the need to explain and report the results of the application of several different benefit measures, the burden on the agencies to report clearly on their methods and assumptions is great. Yet, overhaul of agency reporting protocols is needed for many reasons. First, the agencies are still not reporting uncertainties clearly[8] and will have an even larger burden once they start performing Monte Carlo simulations (see below). Second, they are not reporting enough detail for even professionals to understand how certain key estimates were derived. For instance, the discussion of estimating benefits of the EPA's very recent Off-Road Diesel rule (USEPA, 2003) omitted mention of the assumption that people benefiting from the rule would have otherwise had 6 months to live if they had chronic respiratory disease and five years to live if they died of other causes. Third, agencies are already using base case and alternate analyses with sensitivity analyses layered on top of these in appendices. Now, with multiple benefit measures and entirely separate analyses being used, these protocols will need to be revised.

---

[8] See, for instance, National Research Council. 2002. *Estimating the Public Health Benefits of Proposed Air Pollution Regulations* (NRC, Washington, D.C.)

Recommendations

OMB should encourage studies of decision makers at agencies concerning how they use complex information and how such information could be better reported to make its use easier.

Uniformity in reporting formats in RIAs should be encouraged. As in provision 7, below, this uniformity should include how uncertainties should be addressed (NRC, 2002).

(v) Harmonization of methods

Historically, OMB's guidance documents enforce a relatively loose regime for the conduct of CBA, as evidenced by the wide variety of choices made by the agencies in their choice of specific measures for such important elements as the value of statistical life. The same may be true of agencies using CEA.

Recommendations

For several reasons a tighter regime is recommended, where the guidance is written with more specificity about the options the agencies have and what the default is to do both types of analyses. When the guidance is not followed, agencies would have a high burden of proof for showing why they were not followed.

Why this tighter regime? First, agency cultures are entrenched when it comes to use of CBA or CEA. At EPA, for instance, standard procedures are for using CBA and there is likely to be a costly learning curve for using QALYs. At some of the public health agencies the learning curve for monetization techniques is likewise high; more importantly, at these agencies there is a traditional reluctance to monetize reduced health effects. These agencies may not easily buck this culture and need a strong push from OMB to make such a change happen. Second, OMB wants the ability to conduct its own analyses of agency analyses and to compare findings across rules both within and across agencies. The only way to make this happen – and the author endorses the idea of cross agency comparisons (see provision 5) – is for the agencies to follow procedures in greater faithfulness and in greater detail than previously.

A corollary to this recommendation is that OMB consider whether this threshold is low enough to result in pressure to change agency culture (granting that performing such analyses on very minor regulations would in itself waste resources and would play into the suspicion of paralysis by analysis raised above. One option would be to drop all thresholds, but state that the degree of sophistication of the analysis match what is at stake with the rule. Defining these terms would be a challenge, of course.

These recommendations are offered with two caveats. First, because legislative mandates and requirements as well as rulemaking contexts may differ markedly across regulatory

efforts both within and across agencies, such agencies should have the option of doing analyses their preferred way as a supplement to meeting the OMB guidelines and to do so without need of justification.

Second, because learning to do analyses in new ways takes time and money, and because building the inventory of values would take time as well, these requirements should be phased in over several years and special budgetary allotments should be made for such training.

2. Appropriate Effectiveness Metric and Agency Justification

The OMB Guidance draft does not endorse any particular effectiveness metric for CEA. These metrics can be in physical terms, e.g., lives saved or life-years-saved, or in terms of QALYs. If the latter, they may be drawn from particular health status indices, such as the Quality of Well Being Scale (QWB) or the Health Utility Index (HUI). The indices can be based on four preference survey approaches: standard gamble (SG), rating scales (RS), time tradeoff (TTO), or person tradeoff (PTO). Direct elicitation of preferences using these survey approaches can also occur.

(i) Physical Effectiveness Measures

The technical requirements associated with presenting physical effectiveness measures in a CEA are far less than those with QALYs. Generally, epidemiological studies present outcomes in terms of lives saved, cases reduced, days of illness reduced, among others. These measures, when divided into costs, provide a CEA ratio directly and simply.

There are several issues associated with physical effectiveness measures. First, as noted above, only one effectiveness measure can be used at a time. For example, lives saved cannot also be combined with reduced hospital visits. Other physical effects must then be either monetized and subtracted from costs or ignored. If key effects are ignored in a CEA, there is obviously the potential for the analysis to be misleading. Without direct use of preference-based information to determine which effects are "key" and which can be safely ignored, the problems for doing a useful CEA analysis are magnified.

Nevertheless, where key output measures have been identified *a priori,* whether or not based on preference-based information, use of CEA with physical effectiveness measures could yield important and useful information. EPA's RIA for reducing fine particulate concentrations, for instance, could usefully be enhanced with cost-effectiveness analysis based on cost per life saved (or life-year-saved, see below), because it is widely accepted that the mortality effects dominate these analyses (based on WTP studies over the last decade).

Second, the controversy between using lives saved versus life-years saved as the physical effects measure is an issue. A life-years saved measure presumes that saving younger

people's lives is "worth" more than saving older people's lives, or that saving a year of life is equivalent irrespective of one's age. Using the lives saved measure, on the other hand, presumes that age doesn't matter.

Third, really a generalization of the second point above, the quality that makes physical measures of effectiveness in CEA attractive – their simplicity – is also their weakness. People may wrongly, but understandably attach normative significance to cost-effectiveness rankings based on them, not realizing that preferences for health outcomes may depend on far more than their number.

(ii) QALYs

Where a key effect hasn't been established or monetization of key effects is either impossible or has been otherwise ruled out for a net cost-effectiveness analysis, agencies are likely to use QALYs. At the conference, and, in particular, at the following workshop, significant time was spent examining the validity, reliability and practicality of using alternative QALY indices.

At these meetings, QALYs were often termed a measure that can be "utility-based" (if based on standard gamble or time-tradeoff approaches to deriving weights) while not being a representation of utility. If this were the accepted interpretation of QALYs, then many of the conceptual objections raised about this measure (e.g., its empirical violation of necessary assumptions to be a measure of welfare) would disappear. In this case, QALYs may be regarded as a convenient way of aggregating a variety of physical health effects for the purposes of ordering regulatory options in a CEA. This CEA measure would be little different in meaning than a cost per life saved or other physical effectiveness measure, except it would account for morbidity effects as well. As such, this measure would not directly capture preferences related to risk "qualities," such as stigma, fear, controllability and the like.

But empirical issues remain. For instance, QALYs are often derived from small, specialized samples. They also are often based on survey methods that have remained largely unchanged for many years and have not been subjected to rigorous validity tests, raising the question of their robustness and unbiasedness. It was generally agreed at the workshop that indices based on SG and TTO surveys more closely mirrored the policy-relevant choice situations faced by individuals and better measured preferences than other weighting approaches. At the same time, it was acknowledged that even these approaches had significant issues. Both of these approaches were criticized as placing individuals in unrealistic choice situations, particularly with respect to SG questions for measuring preferences to reduce acute morbidity. However, participants in the workshop cited at least one study that has captured such effects well (Feeny et al., 2002).

It is worth noting here that WTP measures face many empirical issues as well. See provision 4 below.

Recommendations

OMB should be silent on the agency choice of effectiveness measure, but laying out the tradeoffs between using physical measures vs. QALYs. OMB should note that the physical measures are very simple but should have no normative significance attached to them.

OMB should require agencies to provide an explanation for their choice of effectiveness measure. This requirement would help OMB and other agencies and stakeholders understand the agency's rationale for this important set of choices. It could also help make consistency across agencies easier to achieve.

At the workshop, there was discussion about the idea of agencies electing to use QALYs being asked to "unpack" their estimates to increase the transparency of this measure. For instance, rather than presenting one "score"[9] for the change in QALYs, the analysis would present a score for the life-years-saved and a score for the change in the morbidity portion of the QALYs. The sum total of these two categories would be the total QALYs gained from the option under consideration.[10]

A way to unpack the morbidity portion of QALYs is to separate duration from the morbidity preference weight. The weights come from the original SG, PTO, TTO or RS surveys, while duration information may have many origins and be quite uncertain. Unpacking estimates in this manner can help focus attention on these two components.[11] At the workshop this unpacking idea lacked consensus.

3. VSL Heterogeneity

OMB does not suggest a particular VSL for CBA. Rather, OMB suggests that "You should not use a VSL estimate without considering whether it is appropriate for the size and type of risks addressed by your rule" (p. 5521).

OMB is asking that VSLs be applied where they are appropriate to the context being considered. For instance, OMB does not want VSLs derived for fatal heart attacks to be applied to reductions in cancer mortality risk. The guidance document notes that if a

[9] It is convenient here to refer to a score or QALY score as the number of QALYs estimated to be related to some state or intervention. The term QALYs is used for the technique for assigning and adding up life-years and quality-adjusted life-years and also the total score.

[10] One potential problem with this approach is that adding up QALYs for life-years gained only is transparent where each life-year takes a value of one or some other constant value. If life-years take on different values according to the health status of the life saved, then transparency will be harder to achieve.

[11] The idea of unpacking could also apply to WTP measures. If income matters, computations could be based on different income groups with different VSLs, but also, computations could be based on average income. The same could also be done for age and health status, for instance.

VSL is being applied in a different context from which it was derived, appropriate adjustments should be made, and the differences in contexts should be outlined. This guidance, although not altogether new, is in marked contrast to procedures generally followed in RIAs, where one VSL is chosen, albeit after much analysis of the existing literature.

While OMB is aware of the newer VSL literature suggesting that VSLs vary by risk size and type as well as population characteristics, use of varying VSLs is highly controversial. Administrator Whitman has vowed that a "senior discount" will never be used at EPA to make decisions (National Public Radio, May 21, 2003).

Recommendations

The practical issue about encouraging VSL heterogeneity is that such context-specific values are often unavailable. In this case, analysts typically use benefit transfer and meta-analysis techniques. Thus, the guidance needs to reflect this practical reality but could require agencies to include a justification.

OMB should support additional research in valuing mortality risk reductions in contexts where estimates are currently unavailable. It is no secret that getting survey work done within the federal government or with federal funding is difficult at best because of the holdups often experienced in the Information Collection Request approval process. Addressing this significant hurdle will result in great strides in our abilities to develop WTP estimates using survey techniques.

It is worth noting that this recommendation applies just as well to any survey work, including surveillance surveys that can be used to obtain data on prevalence, incidence, duration, and severity, which are needed to estimate QALYs as well as WTP measures of morbidity. It would also apply to surveys used to develop QALYs themselves.

How alternative life-saving programs (and health-improving programs, in general) are to be ranked is an important public policy question that needs open debate. OMB should develop or support the development of a forum for such a debate.

4. Enhancing the Credibility of Measures

OMB emphasizes, as it has done in previous guidelines, that the credibility of WTP measures is enhanced when they are derived from revealed preference (RP) studies. The guidelines note, however, that RP data are often not available for health and safety risks or are derived from studies far from the appropriate regulatory context. In this case, values from stated preference (SP) studies can be used, with those passing external scope tests judged to be more reliable than values from other studies.

These statements regarding WTP measures should be revisited to place stated preference methods on an equal, if not superior, footing with RP measures as SP practices improve

in quality or if research shows that the inability of RP measures to match the appropriate regulatory context is a significant impediment to their applicability.

Further, these comments regarding WTP measures should apply equally to QALY measures, with OMB setting out validity tests on such measures that would enhance their credibility. Preference weights underlying QALY calculations need to be subjected to more validity tests and guidelines, just like WTP studies.

The OMB Guidance is silent on the question of whether QALYs or WTP measures are preferable for use in RIAs. Implicitly, WTP is endorsed because CBA is endorsed and QALYs are not ruled out as an effectiveness measure in CEA.

The participants in the conference and workshop agreed that both measures have advantages and disadvantages, implying that one measure cannot be unambiguously preferred to the other. Participants also felt that an important contribution of further efforts would be to determine when a WTP measure would be useful and when a QALY measure would be useful to a policy analysis.

To give some flavor for the issues discussed in the workshop and conference, a paper summarizing what was learned is in the preparation stage. In this paper, the following points are made. Both measures provide an index useful for aggregating diverse health effects, with one index based on a monetary unit and the other based on a life-year. Both measures reflect something about preferences across health effects, with some approaches to eliciting those preferences being very simple or even simplistic and others being very complex, sophisticated and realistic. Both measures carry significant "baggage," such as the controversy over monetizing "life," and the recent controversy over the "senior discount" that would also follow from use of a QALY measure. Both sets of measures have problems demonstrating credibility and both have problems of data availability.

Recommendations

OMB should rewrite the Guidance to demand equal burden of credibility/validity for WTP and effectiveness measures such as QALYs and should revisit its bias towards revealed preference approaches to measuring WTP.

5. Data provision to OMB

The guidelines require agencies to provide OMB with all relevant data, including morbidity and mortality, population data, and the severity and duration of the conditions evaluated, so that OMB can make comparisons across rulemakings that employ different measures.

CEA comparisons may be problematic unless the effectiveness measures are standardized. Requiring that each agency use lives saved may not be enough for physical measure comparability if it is deemed that life-years saved is important. Neither is

sufficient where morbidity is important (unless the morbidity effects can be monetized and subtracted from costs). Requiring the agency to use QALYs as the effectiveness measure is not enough to assure comparisons are possible because there are so many competing indices.

Recommendations

Requiring agencies to provide underlying data and models in sufficient detail that OMB can do the calculations according to its own dictates is likely to be challenging because of the many specialized data and specific decisions needed to estimate QALYs. On the other hand, to the extent life-years dominate the QALY calculations, QALY scores across alternative indices may not be much different. Therefore, more study is needed about whether OMB's recommendation is practical.

6. Monte Carlo Analysis

The guidelines note that relevant outcomes should be reported as probability distributions where possible, and that for rules with an economic impact greater than $1 billion, a formal quantitative analysis of uncertainty is required. For rules below the $1 billion threshold and without sufficient information to report probability distributions, sensitivity analysis can be used to assess uncertainty. For rules above the threshold, however, probabilistic analysis should be applied throughout the entire range of calculations, such that uncertainty is carried through to the endpoints.

Improving and standardizing the characterization of uncertainty in RIAs is of major importance, given the significant uncertainties associated with all parts of CEAs and CBAs in support of major rules. Use of Monte Carlo analysis is becoming more routine in the literature and among practitioners of these analyses, partly because the computer tools are becoming more widely available and easier to use.

At the same time, the new guidance is silent on techniques for reporting uncertainties. One suggestion is for OMB to mandate incorporation of tables listing unquantifiable uncertainties, e.g., model uncertainties as opposed to statistical uncertainties, and the direction of possible biases.

A more fundamental issue involving uncertainty and reporting concerns the procedures for defining and presenting main and alternative results and sensitivity analyses in the RIA. The recent *Off-Road Diesel Engine RIA* issued by EPA (2003) is a case in point. In this report, EPA presented a Base case and an Alternate case, plus sensitivity analyses in an appendix. The Base case incorporated what EPA considered their best guess at appropriate model selection (e.g., epidemiological studies) and parameter choices. The Alternate case incorporated models and parameters that were towards the lower range of plausible values. The resulting benefit estimates from the Base and Alternate cases were not offered with uncertainty distributions. Furthermore, although a reasonable case could be made for the validity of models and parameter values towards the upper range, benefits based on these choices were relegated to sensitivity analyses in the appendix.

Thus, all but the most dogged reader would believe that the mass of uncertainty lay below the base case benefit estimate (which it very well may, but if so, this is not clear). The use of Monte Carlo simulation techniques would improve upon these procedures but not eliminate them.

Beyond these "macro-uncertainty" issues are some "micro" issues. Because of the significant uncertainties in WTP, QALY and physical effects (as well as cost) estimates, it is essential to quantify such uncertainties. This is standard procedure with WTP estimates in cost-benefit analysis and is also generally followed in the better cost-effectiveness analyses. For instance, duration and the preference weights in QALYs are encoded as distributions in the better analyses.

Finally, an issue was raised in the workshop about the appropriate treatment of uncertainties from multiple links in a chain of causation. It may be that the uncertainties in one variable in the calculation are correlated with uncertainties in another variable. Not accounting for this dependence, but rather assuming that the uncertainties are independent, would result in wider confidence intervals (if the correlations are positive).

Recommendations

OMB guidance on the appropriate development of scenarios to fairly represent statistical and model uncertainties, given that Monte Carlo simulation will be used, would be a useful addition to the draft guidance document.

An explication of why a cost threshold of $1 billion would be used for probabilistic CBA/CEA would be useful, as this figure is arbitrary. Why not a lower figure, say $100 million? Because of uncertainties in the WTP and QALY measures and historic problems within the agencies in fairly and completely representing uncertainties, a lower threshold would assure that more analyses include this important added information.

As Monte Carlo analysis tends to be performed with the assumption of independence in uncertainty across variables in a causal chain, it would be useful for OMB to flag this issue and suggest that attention should be paid to modeling correlated uncertainties where there is good reason to think they exist.

7. Use literature combining WTP and QALYs

OMB offers a suggestion to agencies faced with gaps in their WTP information. The Guidance says: "If data are not available to support monetization, you might consider an alternative approach that makes use of health-utility studies... This health utility [QALY] information may be combined with known monetary values for well-defined health states to estimate monetary values for a wide range of health states of different severity and duration. If you use this approach, you should acknowledge your assumptions and the limitations of your estimates" (p. 5521).

With this comment OMB advances the idea of using both measures together, of combining WTP and QALY measures ("QALYfying the WTP"). Because there is as yet a small literature in this area (e.g., Johnson et al, 1997; Van Houtven et al, 2003), this suggestion serves to underline the need for more research, rather than propose something the agencies could immediately operationalize in constructing their regulatory impact analyses. Indeed, researchers at the conference and workshop talked much more about mounting different research efforts.[12]

Recommendations

OMB guidance should acknowledge that combining WTP and QALYs is not something that can be operationalized without additional research and should not limit or favor ways in which WTP and QALY concepts and measures can be combined.

[12] The conference and workshop participants generated many ideas for research, which will be presented in a future report. A few ideas are presented here: One would involve implementing a conjoint analysis stated preference survey, where the attributes of the survey would include dimensions that would apply to a wide variety of morbidities (see Johnson, et al, 2000 for the model and results for a small Canadian sample). One of the issues with this type of study is that the many different dimensions of morbidity may exceed respondent ability to tradeoff among the necessary attributes. Another idea (Johnson 2002) is to construct "super-QALYs," by estimating willingness-to-wait estimates for morbidity and convert them to time-tradeoff QALY weights. Another (Smith,Van Houtven and Pattanayak, 2003) propose to use results of QALY elicitation along with economic information about behavior to calibrate the health-related components of individual preference functions.

References

Feeny, David, Marie Townsend, William Furlong, Darrell J. Tomkins, Gail Erlick Robinson, George W. Torrance, Patrick T. Mohide, and Qinan Wang, 2002. "Health-Related Quality of Life Assessment of Prenatal Diagnosis: Chorionic Villus Sampling and Amniocentesis," *Genetic Testing* **6** (1), 39-46.

Harrington, Winston and Paul Portney. 1987. "Valuing the Benefits of Health and Safety Regulations," *Journal of Urban Economics* **22** (1), 101-112.

Johnson, F. Reed, M.R. Banzhaf, and W.H. Devouges et al. 2000. "Willingness to Pay for Improved Respiratory and Cardiovascular Health: A Multiple Format, Stated Preference Approach." *Health Economics* 9: 295-317.

Johnson, F.R., E.E. Fries, and H.S. Banzhaf. 1997. "Valuing Morbidity: An Integration of the Willingness-To-Pay and Health-Status Index Literatures." *Journal of Health Economics* 16(97):641–665.

Johnson, F.R. J. Mauskopf, and A.B. Hauber. 2002. "Measuring Health-State Utility: Rulers, Thermometers, or Shoe Sizes? Working Paper. RTI Health Solutions.

National Public Radio. 2003. "EPA Administrator Whitman to Step Down." *All Things Considered*, May 21, 2003.

National Research Council (NRC). 2002. *Estimating the Public Health Benefits of Proposed Air Pollution Regulations*. Washington, DC.

Office of Management and Budget (OMB). 2003. Draft Report to Congress on the Costs and Benefits of Federal Regulations. Available online: http://www.whitehouse.gov/omb/fedreg/2003draft_cost-benefit_rpt.pdf

Smith, Kerry, George Van Houtven, and Subhredu Pattanayak. 2003. Preference Calibration with QALYs, working paper.

U.S. Environmental Protection Agency (USEPA) 2003. "Chapter 9: Cost-Benefit Analysis". *Off-Road Diesel Engine RIA* (draft). Available online: http://www.epa.gov/nonroad/r03008j.pdf

Van Houtven, George, et al. 2003. *Valuation of Morbidity Losses: A Meta-Analysis of Willingness-to-Pay and Health Status Measures*. Final Report. Report prepared for Food and Drug Administration Center for Food Safety and Applied Nutrition. March.