



Commentary on “Standards of Evidence for Conducting and Reporting Economic Evaluations in Prevention Science”

Brian T. Yates¹

Published online: 12 March 2018
© Society for Prevention Research 2018

Abstract

The current paper is a commentary on the Standards of Evidence for Conducting and Reporting Economic Evaluations in Prevention Science (Crowley et al. 2018). Although the standards got a lot right, some important issues were not addressed or could be explored further. Measuring rather than modeling is encouraged whenever possible. That also is in keeping with the approach taken by many prevention researchers. Pre-program planning for collection of data on resources used by individual participants (i.e., costs) is recommended, along with devotion of evaluation resources to cost assessment throughout program implementation. A “cost study” should never be an afterthought tacked on as a later aim in a research proposal. Needing inclusion or enhancement in the standards, however, are several key concepts, starting with the often-confused distinction between *costs* and *outcomes*. The importance of collecting data on individual-level variability in resource use, i.e., costs, needs to be distinguished from simplistic disaggregation-by-division of program cost totals down to individuals. In some passages of the standards, the uniqueness of individual participants seems dismissed as error variance rather than considered a primary phenomenon for study and understanding. Standards for formatting reports of economic evaluations could themselves be more evidence-based. Missing too is an explicit call for inclusion of the standards’ recommendations in peer review of prevention research proposals, and in funding of prevention research. Finally, we can be confident that the better outcomes the standards promise will come at additional costs to prevention researchers. This commentary concludes by considering whether the *standards* themselves are cost-beneficial.

Keywords Prevention programs · Standards for economic evaluation · Cost-inclusive evaluation (CIE) · Benefit-cost analysis (BCA) · Meta-benefit-cost analysis (MBCA)

Early research in health and human services, including prevention studies, concentrated somewhat myopically on what has been called “*process*,” i.e., the implementation of a program. In the latter part of the twentieth century, research broadened its focus to include outcomes as well as processes. More recently, prevention research has begun to adopt three foci: (a) the resources consumed by a program, i.e., its *costs*, (b) the processes, i.e., *activities*, made possible by those resources, and (c) the *outcomes* produced by those activities. Moreover, research

has begun to measure the monetary value of some outcomes, i.e., *benefits*, to allow quantification of the relationship between the value of resources consumed by a program and the value of outcomes produced by that program, i.e., its *cost-benefit*.

Although some researchers and evaluators in education, health, and other human services have advocated several forms of *cost-inclusive evaluation* for decades (e.g., Herman et al. 2009; Levin 1983; Yates 1994), political and economic developments have increased the acceptability of talking frankly and openly about costs as well as benefits of prevention programs. Policy-makers across the political spectrum increasingly agree that public monies should only fund programs that work, that work well in addressing needs and entitlements, that generate more resources for society than they consume, and that work better, cost less, or work better *and* cost less, than other programs.

Standards for selecting the best investment portfolio of programs for the public good have blossomed between 2015

The content of this publication reflects only the views of the author.

✉ Brian T. Yates
brian.yates@mac.com

¹ Department of Psychology, American University, 4400 Massachusetts Avenue NW, Washington, DC 20016-8062, USA

and 2017. At least three major publications present standards for economic evaluation (Crowley et al. 2018; Neumann et al. 2017; Steuerle and Jackson 2016), while newer editions of classic texts for economic studies are more comprehensive and read much like standards themselves (e.g., Drummond et al. 2015; Levin et al. 2017). Clearly, the time is ripe for cost-inclusive standards to be introduced into prevention. So is the time for feedback on those standards.

What the Standards Got So Right

These *Standards for Economic Evaluation* (Crowley et al. 2018) present a consensus from a variety of scientific communities, not just economic evaluators, about how to conduct evaluations of prevention programs that describe and predict relationships between the value of resources invested in prevention and the value of resources saved and generated by prevention. These have been a long time coming, are sorely needed, and will be much appreciated in the long run. They provide an excellent foundation for the adaptation and evolution of economic research and program evaluation in prevention. The emphasis on multiple perspectives, including those of people who are the targets of prevention program activities, i.e., *consumers*, is wonderful, as is inclusion of hallmark examples. There is so much here that is good; I will highlight some of the best and move on to what seems to be missing and what could be enhanced.

Quasiexperimental Evaluation Can Be OK

Acknowledgement that randomized clinical trials may not be possible, acceptable, or ethical for some prevention programs is a move forward that may be contentious but certainly is needed. It is laudable that the standards explicitly mention the possibility of quasiexperimental and naturalistic experimental designs in the section “Describe the Evaluation of the Prevention Program’s Efficacy or Effectiveness in Terms of Its Impact on Behavioral and Other Outcomes.”

Measuring Is Superior to Modeling

The standards prefers an empirical, measurement-oriented approach to research over a modeling approach for which a few point estimates become hooks on which to hang whole distributions for costs or outcomes. Explicitly advocating testing for statistical significance as well as effect size is long overdue in economic evaluation (cf. Barber and Thompson 1998), a field populated by reviewers who routinely dismiss statistical tests as poor substitutes for model parameters, and who often deny the need for nonparametric statistical analyses even when assumptions of parametric analyses cannot be maintained despite valiant transformations. This preference for

measurement instead of only modeling should engage many prevention researchers whose primary training is in research design and analysis of data collected from participants.

Defining Costs with a Comprehensive List of Program Ingredients

Going beyond the idealized figures in program budgets to collect real data on the actual types, amounts, and monetary values of resources used in program activities is a crucial step forward in economic evaluation. Some researchers have long advocated for this (e.g., Carter and Newman 1976; Yates 1980). In a way, measuring actual use of program resources by individual participants is giving cost evaluation “its due” and it is about time, given the rather exclusive prior concentration on activities and outcomes in fields from education to mental health to substance abuse. Including in cost assessment real measurement of resources used by prevention programs for adoption, implementation, sustainability enhancement, and evaluation capacity-building is a major step toward enhancing the replicability and dissemination of evidence-based prevention. These are relatively new, important additions to the basic lists of those essential “ingredients” that needing costing.

Beyond Total Ingredients: Resources Used in Specific Program Activities

Amounts and types of resources actually used, *measured specifically for each activity of the program*, can be more useful for program administration and adaptation than summary prices drawn from multiyear budgets for an entire program. Some of us have advocated for this more fine-grained approach to resource activity assessment for decades, despite resistance from colleagues. This goes beyond the “ingredients approach” to cost assessment, providing the foundation not only for common-sense management decisions but also for applying the quantitative optimization methods of *operations research* (e.g., Yates 1980, 1996) used in business and some branches of government.

Avoiding Over-Counting Benefits

As recommended in the standards, moving beyond a simple summing approach to assessing benefits as well as costs also is important for prevention research. The advocacy that one often feels when working with a program can blind one to the need to use an evidence-based logic model of causal sequences to avoid double-counting monetary outcomes and misrepresenting the cost-benefit of prevention efforts.

What Is Missing, Needed, and Next in Standards for Economic Evaluation of Prevention Programs

Missing

Easy Reading

The standards is not written in a consistently approachable style that balances clarity for a broad professional readership with understanding of economic terminology and concepts. This is difficult, of course, but important. For example, although *shadow pricing* is defined well and at length in the standards, the definition given for *net present value* (NPV) is too general for many readers to realize that, as the time horizon for total benefits and total costs moves deeper into the future, later benefits are substantially diminished in their value—as are future, delayed expenditures. Readers could have been informed of this by a simple example, but instead are directed to consult references. The reader also encounters in the standards the varied writing styles typical of a manuscript with multiple authors.

Also missing is an upbeat message that the additional efforts called for by these standards will help us achieve something worthwhile, making our world the best place possible for the most people via informed decision-making. Perhaps a standards is not the place for this message. However, the evocation toward the end of the document that, “Economic evaluation is never a substitute for democratic and administrative decision making...” is at least the beginning of an inspirational appeal. Why not go further to suggest the formative role that economic evaluation can play in guiding and funding prevention research? The methods evoked in these standards need not lead to a “dismal science,” as economics is sometimes called. Economic evaluation promises to generate more societal health and productivity, and more social wealth and personal wellness, for more people throughout our world. Let us hope to *inspire* researchers to follow the standards.

Social Return on Investment (SROI)

After stating that they will not provide a primer on specific types of economic evaluation, the authors do—well and succinctly—in the section *A Brief Overview of Economic Evaluation Methods*. Missing from this, however, are several newer and increasingly popular forms of economic evaluation, including Social Return On Investment (SROI; Yates and Marra 2017a). This is odd, given the standard’s long section on Pay for Success, an approach closely related to SROI.

Another significant omission in the standards’ list of economic evaluation types is a form of cost-utility analysis that would seem particularly relevant to prevention research in health: *cost per Quality-Adjusted Life Year Gained*

(\$/QALYG) (e.g., Neumann et al. 2017, among others). Cost-utility analysis is mentioned later in the standards, apparently without definition: “... the standards described here are complementary to efforts that have largely focused on cost-effectiveness or cost-utility analyses ...,” and only refers to disability-adjusted life years (DALYs). This is itself buried in a discussion of shadow prices. QALY-based cost-utility analyses are forms of evaluation that certainly are economic and already have a history of use at national levels in funding decisions for health interventions (cf. Neumann et al. 2017). QALYs as well as DALYs have long promised to be universal measures of outcome that are independent of natural units of measurement such as myocardial infarctions averted or HIV transmissions prevented. Both can be more appealing to some stakeholders than outcomes monetary and imperfectly monetized. Cost per QALY or DALY gained still is cost-inclusive in that resources consumed are valued monetarily. Both still address needs of decision-makers and funders.

Needed

Conceptual Foundations Clarified and Refined

A fundamental confusion about economic evaluation that one encounters when working with prevention scientists, health researchers, and program evaluators is between costs and outcomes, and especially between costs and monetary outcomes. Too often those well-trained to measure outcomes for alternative prevention programs in rigorously controlled clinical trials find it difficult to think of “costs” as anything other than another type of outcome, albeit one that seems measured in monetary units with methods that seem to some both mildly exciting and vaguely repulsive. The standards offers a brief, potentially effective way of describing the essential distinction between costs and outcomes: costs are the value of investments made in prevention efforts; outcomes are the results of those efforts. Some of those results are not monetary and some are. But usually it takes more than this to get across the basic idea that, “*costs go in, outcomes come out.*”

The authors also recommend adding to positive outcomes such as distal improved earnings those more proximal outcomes that are negative, such as short-term loss of income during receipt of higher education. The authors note that including the negative monetary outcomes among costs rather than outcomes is conceptually incorrect. Indeed. That some evaluators have been confused about this demonstrates the need for a clear conceptual distinction upfront and often between costs and outcomes.

Adjustment of Future Uncertainties Beyond Present-Valuing

Present-valuing may not be a sufficient adjustment for delayed benefits and delayed costs, especially for prevention programs

with outcomes and costs which can extend decades into the future. These adjustments might consider not only the value lost at least temporarily when an investment is made now but returns are delayed by years, and not only the cumulative opportunity cost of making an investment now rather than waiting to make that investment later. Also to be anticipated in cost-inclusive evaluations of prevention programs is the uncertainty of outcomes in a future fraught with upheavals that may diminish, eliminate, or possibly reverse the outcomes predicted. Just choosing a higher discount rate may not be enough. Economic analyses of prevention programs also might consider the likelihood that even better prevention programs will be developed in the future, including ones that have lower costs for similar or better outcomes. Programs that have an ever larger or quicker returns on investment would, in fact, be expected with increased Pay for Success funding of prevention programs.

Specification of the preferred level(s) of specificity

A fundamental decision to be made in any prospective cost-inclusive evaluation is setting the level or levels of specificity at which costs, activities, and outcomes will be assessed. Macro? Micro? Some combination of these? Common practice in economics and accounting assumes that costs are measured initially at the macro level of the entire program, often using data from financial reports, and are then disaggregated by simple division of total program costs by the number of participants. This procedure assigns the same cost to all participants. It does not assign different costs to participants who spend more or less time and effort in the same program activities, or who otherwise consume varying amounts of program or personal resources. Some researchers in education, behavioral health, and medicine would be satisfied assuming costs are identical for each participant. These same researchers would, however, be reluctant to assume that outcomes for each participant were similarly identical, as when total program benefits are divided by number of participants. So should we be reluctant to assume that costs are the same for each participant.

How to merge macro- and micro-level data on costs and outcomes is a major decision on which the standards seems mute. There is a real danger that an exclusively macro-level view could encourage dismissal as “unknown variance” or “uncertainty” the very real individual variations that differences in race, gender, class, and culture can produce in prevention program outcomes, in program activity participation, and in the devotion of often scarce personal resources such as consumers’ time, energy, personal transportation to program activities. This issue is addressed only obliquely and briefly, in one paragraph with the heading “Identify Moderating Factors of Economic Impact.”

The exclusive focus of at least some of these standards on the macro level is revealed in the following passage:

The NPV can be estimated on a per-participant basis or for the entire sample for which the BCA was conducted. The per-participant value is easier to grasp and usually the level at which model testing and statistical significance are computed, but the sample-based value provides a more complete picture of the overall loss or gain resulting from the program.

Unfortunately, this passage can be interpreted as casting individual participants’ costs and benefits as just the mathematical averages for all participants in the program. Variability of different individuals’ responses to prevention efforts, whether those responses are participants’ use of program resources or participants’ outcomes, is more than random error. Different participants experience different costs and different benefits as a result of interactions between even highly standardized program activities and participants’ unique abilities, expectancies, self-regulatory skills, and other personality characteristics (cf. Mischel 2007). The individual *micro* level of the participant is where many prevention programs have their effects. The behaviors, thoughts, and feelings of these individuals are of interest to many prevention scientists. For decision-makers, it is at the micro level that many taxes are paid and all votes are cast. Their variability is their individuality, their culture, their religion, and who they are. It seems prudent to avoid the possible perception that economic prevention studies average away individuality to make findings more readily grasped. The individual is the beginning level at which the phenomena of primary interest in prevention—birth, life, education, work, wellness, illness, death—occur. Everything else is a group statistic.

What About Variability in Program Participation?

A common problem in prevention programs is that some participants do not participate fully in all the activities to which they were “assigned” or “invited.” This variability in participation has been called many things in many fields. The standards includes both the common “intent to treat” (ITT, which these standards might have expressed as “intent to prevent” or ITP) and “treatment on the treated” or TOT (a term which is both unusual in human services research and a poor fit for prevention programs at least). “As treated” (AT) is more common in mental health and substance abuse trials, although “as prevented” (AP) would be even better. Dryly noting that, “The validity of the TOT estimate depends on participation being exogenous to any characteristic of the assigned participants ...” not only neglects potential program-by-individual interactions, but also does not address the likelihood that those who participate less or drop out sooner may not benefit as much as those who participate more or stay longer. The primary concern of some decision-makers is how many “successes” actually emerge from a program in which they invested, i.e., *cost-effectiveness*, not *cost-efficacy*.

Next Steps

Separating Prevention Payload from Delivery Focus, Site, and System

Most program implementers realize that the manipulations they conduct can be focused on one person at a time, one family at a time, one community or company at a time, and so on, in the home, clinic, street, office, or factory, and via more or less expensive means, e.g., one-on-one therapy versus group therapy versus influencing entire communities via internet-based media (cf. Yates 2011). The delivery system is not always one easily isolated activity inside a program. The delivery focus, delivery site, and delivery system used by a program to provide its content can affect costs and outcomes more than the well-researched content being delivered (e.g., Yates et al. 2011). This is only implicit in the standards: only “delivery setting” is mentioned.

What Really is Missing: Better Data on Program Costs

The standards describes a variety of ways to deal with the uncertainty introduced by the numerous assumptions typically needed when assessing costs of prevention programs. As recommended in the standards, data on costs should be collected throughout the prevention study, after careful planning, and with methods similar to those used to collect data on outcomes. Data on costs as well as outcomes should have high reliability and validity.

Unfortunately, there is a long history of making as many assumptions as necessary in economic studies to get the cost data in the analyses. There is little of this in other social sciences, especially if the data are for outcomes. To assume the same outcome for everyone participating in a program would be viewed by most educators and psychologists as absurd. So, too, it should be for costs. Most researchers in education and psychology, for example, wince if they have to assume the same distance and time for transportation to and from a clinic or school for participants who simply could have been asked for those cost data during the study. Too often, the myriad assumptions in most cost evaluations seem to follow from a post hoc post-funding “let us add a cost study” decision. As we engage in the more prospective, planful economic evaluations called for by these standards, we should need fewer sensitivity analyses and fewer Monte Carlo simulations of the likely effects of assumptions about costs and outcomes, gaining better knowledge of the actual values and distributions of costs as well as outcomes.

We Need More Resources for Better Cost and Benefit Assessment

Even if mandated in the call for proposals, too often “the cost study” is budgeted for the last year of the contract or grant.

Data on resource expenditures and outcomes monetary as well as nonmonetary need to be collected throughout program implementation. Cost studies also may blossom only in the last year of a program budget when the funder realizes that program sustainability beyond the brief period of federal support is unlikely unless costs and benefits are assessed carefully, reported fully, and understood readily by potential local and private funders. To realize the recommendations made in the standards, funding for collection of cost as well as activity and benefit data needs to be apportioned over the duration of the prevention study.

But what will *that cost*? And who will pay for inclusion of comprehensive cost and benefit measurements throughout a prevention study? The standards can be viewed as an intervention, prescribing activities for economic evaluation, with hoped-for outcomes of better utilization and funding of prevention research but with no specification of the types or amounts of resources needed to conduct the prescribed activities. “Hire some economists” may not solve the problem and certainly would require additional funds. To conduct the activities required by the standards, the field of prevention needs sufficient additional resources. Otherwise the standards cannot be enacted, not without sacrificing quality of research on program activities or outcomes or both. This is, perhaps, the responsibility of peer grant review committees: to identify as having heavily-weighted weaknesses those proposals that inadequately budget for cost-inclusive prevention research.

Monetizing Resources and Outcomes with More Attention to Demographic Differences

The ethics of emphasizing monetary outcomes is addressed only tangentially—so much so that it is easily skipped by readers eager to complete their task. Common practices of monetizing resources of participant time according to their actual income, and of monetizing productivity outcomes as increments in lifetime earnings, can favor those who make more not because of their superior value as more productive employees or entrepreneurs, but because of their gender, race, class, age, or culture. As noted by others, including Yates (2012), use of personal income to value individuals’ times and lives when evaluating outcomes of health, mental health, and substance abuse interventions can maintain or exacerbate existing inequities in cumulative earnings associated with purely demographic differences rather than differences in potential to contribute to society. One alternative is to assume that all of our hours spent working, and all the years of all of our lives, are equally valuable—and to monetize their worth accordingly, even though the marketplace may not always concur. Another is to value the worth of one’s time and effort after removing effects of discrimination against genders, races, ages, classes, and cultures. Challenging, difficult? Yes. Necessary? Consider what economic prevention

evaluations could foster if we do not make these adjustments: funding might flow more toward prevention of health and mental health problems for those individuals who have greater earning potential solely because of their gender, race, class, age, or culture.

Evidence-Based Readability for Better Report Use

Although the authors acknowledge the need for transparency in reporting, they seem to hope that economic evaluation will be utilized because it includes variables of particular import to some decision-makers. That probably is not sufficient. Understandability of reports can be costly to assess, but seems necessary to both measure and optimize. Reports, articles, and presentations are the delivery systems for research findings. Formats themselves can be evidence-based, using a balance of text and white space, sentence length, fonts, heading sizes, clear graphics, and organization to communicate with a clarity that is measurable, hopefully at a cost that is reasonable. Publication and presentation formats can be interventions themselves guided by data.

Convince Others that Economic Evaluation Is Itself Cost-Beneficial

As one discovers quite soon when trying to persuade program designers, researchers, and providers to use economic evaluation methods (cf. Yates 2010), a primary objection to cost-inclusive evaluation is that it costs too much in time and effort relative to its anticipated outcomes, i.e., that economic research has not *itself* been shown to be cost-beneficial or cost-effective relative to alternatives. Although economic researchers' initial reaction may be frustration followed by dismissive encouragement to "have faith," this meta-benefit-cost analysis (MBCA) is a form of meta-evaluation (Scriven 1969) which is important for those of us averse to hypocrisy. Can we show that our research, our studies, our evaluations, guided by these standards, are themselves "worth it" (Herman et al. 2009; Yates and Marra 2017b)? I look forward to joining you in finding an answer to this question.

Funding No funding supported the writing of this manuscript.

Compliance with Ethical Standards

Disclosure of Potential Conflicts of Interest The author has no conflicts of interest.

Ethical Approval No data were collected or analyzed for this manuscript, so no ethics approval was needed.

Informed Consent As no participants provided data for this manuscript, no informed consent was possible.

References

- Barber, J. A., & Thompson, S. G. (1998). Analysis and interpretation of cost data in randomised controlled trials: Review of published studies. *British Medical Journal*, *317*, 1195–1200.
- Carter, D. E., & Newman, F. L. (1976). *A client-oriented system of mental health service delivery and program management: A workbook and guide*. Washington, DC: DHEW Publication No. (ADM) 76-307, Superintendent of Documents, US Government Printing Office.
- Crowley, D. M., Dodge, K. A., Barnett, W. S., Corso, P., Duffy, S., Graham, P., et al. (2018). Standards of evidence for conducting and reporting economic evaluations in prevention science. *Prevention Science*. <https://doi.org/10.1007/s11121-017-0858-1>.
- Drummond, M. F., Sculpher, M. J., Claxton, K., Stoddart, G. L., & Torrance, G. W. (2015). *Methods for the economic evaluation of health care programmes* (4th ed.). Oxford: Oxford University Press.
- Herman, P. M., Avery, D. J., Schemp, C. S., & Walsh, M. E. (2009). Are cost-inclusive evaluations worth the effort? *Evaluation and Program Planning*, *32*, 55–61.
- Levin, H. M. (1983). *Cost-effectiveness analysis*. Beverly Hills, CA: Sage.
- Levin, H. M., McEwan, P. J., Belfield, C. R., Bowden, A. B., & Shand, R. D. (2017). *Economic evaluation in education: Cost-effectiveness and benefit-cost analysis* (3rd ed.). Los Angeles: Sage.
- Mischel, W. (2007). Toward a cognitive social learning reconceptualization of personality. In Y. Shoda, D. Cervone, & G. Downey (Eds.), *Persons in context: Building a science of the individual* (pp. 278–326). New York: Guilford Press.
- Neumann, P. J., Sanders, G. D., Russell, L. B., Siegel, J. E., & Ganiats, T. G. (2017). *Cost-effectiveness in health and medicine* (2nd ed.). New York: Oxford University Press.
- Scriven, M. (1969). An introduction to meta-evaluation. *Educational Products Report*, *2*, 36–38.
- Steuerle, E., & Jackson, L. M. (Eds.). (2016). *Advancing the power of economic evidence to inform investments in children, youth, and families*. Washington, DC: National Academies of Sciences, Engineering, Medicine.
- Yates, B. T. (1980). *Improving effectiveness and reducing costs in mental health*. Springfield: Thomas.
- Yates, B. T. (1994). Toward the incorporation of costs, cost-effectiveness analysis, and cost-benefit analysis into clinical research. *Journal of Consulting and Clinical Psychology*, *62*, 729–736.
- Yates, B. T. (1996). *Analyzing costs, procedures, processes, and outcomes in human services*. Thousand Oaks: Sage.
- Yates, B. T. (2010). Evaluating costs and benefits of consumer-operated services: Unexpected resistance, unanticipated insights, and déjà vu all over again. Case 7 in J. A. Morell (Ed.), *Evaluation in the face of uncertainty: Anticipating surprise and responding to the inevitable*. New York: Guilford Press.
- Yates, B. T. (2011). Delivery systems can determine therapy costs, and effectiveness, more than type of therapy. *Perspectives on Psychological Science*, *6*, 498–502.
- Yates, B. T. (2012). Step arounds for common pitfalls when valuing resources used versus resources produced. In G. Julnes (Ed.), *Promoting valuation in the public interest: Informing policies for judging value in evaluation*. *New Directions in Program Evaluation*, *133*, 43–52.
- Yates, B. T., & Marra, M. (2017a). Introduction to social return on investment (SROI). *Evaluation and Program Planning*, *64*, 95–97.
- Yates, B. T., & Marra, M. (2017b). Social return on investment (SROI): Problems, solutions ... and is SROI a good investment? *Evaluation and Program Planning*, *64*, 136–144.
- Yates, B. T., Mannix, D., Freed, M. C., Campbell, J., Johnsen, M., Jones, K., & Blyler, C. (2011). Consumer-operated service programs: Monetary and donated costs and cost-effectiveness. *Psychiatric Rehabilitation Journal*, *35*(2), 91–99.