

How should Program Evaluation Standards inform the use of cost-benefit analysis in evaluation?

Julian King

Abstract

Background: Cost-benefit analysis (CBA), like any other evaluation method, should be used in ways that uphold program evaluation standards (PES) and should be subjected to meta-evaluation. In contrast to the broad remit of PES, guidelines for economic evaluation focus mainly on technical aspects aimed at ensuring precision, accuracy and reliability. Can CBA be conducted in adherence both to PES and to its own methodological principles, or are there areas where expectations conflict?

Purpose: Assess the potential for CBA to be conducted in keeping with PES.

Setting: Analysis applies to any setting in which CBA is being considered as an evaluation method.

Intervention: Not applicable

Research design: Methodological principles underpinning CBA were systematically assessed against the *Program Evaluation Standards* of the Joint Committee on Standards for Educational Evaluation, to determine the extent to which CBA is able to be conducted in a manner aligned with these PES. For each Standard, CBA was rated according to whether the Standard *can* be followed in principle, not the extent to which economists do so in practice.

Data collection and analysis: This assessment was undertaken from a theoretical perspective, through analysis of relevant literature. Ratings are evaluative: they represent the judgments of the author, made on the basis of explicit definitions.

Findings: Some ethical principles espoused in PES are also required in CBA. On the other hand, some of the PES are not explicit requirements in CBA, though they could be applied by evaluators or economists when conducting a CBA. However, there are some PES that logically cannot be met by CBA if it is used as a stand-alone method. All PES can theoretically be met when an evaluation combines CBA with other methods. In order to use CBA in adherence to PES, evaluators and economists must take an explicit interest in the effects of their analysis on people's lives. This has significant implications for the way CBA should be used, including the nature and extent of stakeholder involvement, the potential use of CBA in conjunction with other methods, and decisions about when *not* to use CBA. Deliberation is necessary over whether, when, and how to use any method in an evaluation.

Keywords: Cost-benefit analysis, program evaluation standards, meta-evaluation

Introduction

Many evaluators have for a long time accepted the principle that evaluations should be open to scrutiny, to check and ensure their quality (Scriven, 1991). There is no universal checklist for this purpose, reflecting the fact that there is no universal position on the matter of ethics in evaluation. However, many organizations have developed quality frameworks or principles defining what a good quality evaluation should look like (AEA, 2018; AES, 2013; ANZEA & Superu, 2015; OECD, 2012; Patel, 2013; UNEG, 2016; Yarbrough, Shulha, Hopson, & Caruthers, 2011). Each of these frameworks is the culmination of debate, and serves to formalise some degree of consensus about evaluation as a field of practice (King, 2019). It is widely agreed, for example, that high quality evaluations should be useful, feasible, ethical, accurate, and accountable (Yarbrough et al., 2011).

The *Program Evaluation Standards – A Guide for Evaluators and Evaluation Users* (PES) of the Joint Committee on Standards for Educational Evaluation (Yarbrough et al., 2011) are a leading example. The PES comprise 30 standards, organized under five headings of utility, feasibility, propriety, accuracy, and accountability (Table 1), together with guidance for people involved in planning, implementing, or using program evaluations. The PES argue that evaluation should take an explicit interest in its effects on people's lives. Therefore, in addition to valid evaluative reasoning and careful selection of methods, evaluation requires attention to stakeholders, concern for consequences and influence, responsive and inclusive orientation, protecting human rights and dignity, and a range of related considerations. The PES are commonly cited in program evaluation. They share many

principles in common with other evaluation standards (Coryn & Stufflebeam, 2014; Deane & Harré, 2016), and have influenced other evaluation standards internationally (Schwandt, 2015).

Cost-benefit analysis (CBA) is a method widely used by economists to evaluate the worth of policies and programs. There is a disconnect between the disciplines of evaluation and economics, and it has been argued that economic and other valuing methods should be better integrated (Davis & Frank, 1992; Julnes, 2012c; King, 2019; Yates, 2012). CBA is more widely taught and has a higher status in public policy than program evaluation, yet it is not routinely assessed against the PES (King, 2019).

Any methods employed in evaluation should be used in ways that uphold evaluation standards – and this includes economic methods of evaluation. Subjecting economic methods to meta-evaluation could facilitate their acceptance into the field of practice in program evaluation (King, 2019). There are standards to guide the design and conduct of economic evaluations, but these standards focus primarily on technical aspects concerned with ensuring precision, accuracy and reliability. For example, textbooks and guidelines devote a lot of attention to methods for valuing costs and benefits in monetary units, and determining how to adjust these valuations to take timing into account (Drummond, Sculpher, Torrance, O'Brien & Stoddard, 2005; HM Treasury, 2022; Levin & McEwan, 2001).

This paper systematically assesses the economic evaluation method of CBA against the PES. First, it introduces the inner workings of CBA. Then it describes the methods that were used to systematically compare and rate the requirements of CBA against the ethical principles of the PES. Finally, it

presents findings and discusses their implications. In particular, this paper concludes that full adherence of CBA to the PES cannot be guaranteed unless CBA is used together with other evaluation methods. In some circumstances, adherence to PES may involve a decision *not* to use some method – and CBA should not be exempt from this possibility.

Cost-benefit analysis

CBA is a method for comparing the costs and consequences of a project, program or policy with alternative courses of action (Drummond et al., 2005). It is one of a suite of economic methods of evaluation that also include cost-effectiveness and cost-utility analysis. These different methods can be distinguished by the units of measurement used to value consequences (for an illustration see Vaca and King, 2020).¹ The distinguishing feature of CBA is that it values costs and consequences in the same units.

Conducting a CBA involves systematically identifying, measuring, valuing, and comparing the costs and benefits of some intervention – either on an *ex ante* (forecasting) basis to inform decisions, or *ex post* (after the fact) to assess how well an investment performed. In simple terms, where benefits exceed costs, the intervention can be deemed worthwhile (Drummond et al., 2005). This method is often used to inform capital investment and financial decisions (Levy & Sarnat, 1994), and to appraise economic and social implications of policies in areas as diverse as transportation (Damart & Roy, 2009), healthcare (Drummond et al., 2005) and education (Levin & McEwan,

¹ <http://www.saravaca.com/project/outcome-efficiency-indicators/>

2001). Since the 1980s it has been compulsory in the US to subject all regulatory proposals to CBA (Adler & Posner, 2006).

In order to evaluate CBA against the PES, it is important to first understand what CBA sets out to do, how it is structured, and how it functions as an evaluation method. While the theoretical roots of CBA stem from the field of welfare economics, it can also be understood as a form of evaluative reasoning (King, 2017; 2019) – that is, a way to reach evaluative judgments from evidence as described by the general logic of evaluation (Scriven, 1980; 1991; 1994; 1995; 2012). This general logic is shared by all evaluation approaches and is often characterized in terms of four steps: i) establishing criteria of merit or worth; ii) constructing standards; iii) gathering and analyzing evidence; and iv) synthesizing and integrating the criteria, standards, and evidence to reach an evaluative judgment (Fournier, 1995). However, the logic can be applied in a diversity of ways (Schwandt, 2015) and “what counts as criteria or evidence or how the evidence is weighed varies from one approach to another” (Fournier, 1995, p. 17).

In general, **criteria** of merit and worth are determined according to context (Schwandt, 2015). In CBA, however, there is one criterion, which comes packaged with the method and is not always made explicit (King, 2019). CBA is primarily concerned with whether an intervention increases or decreases overall social welfare (Sunstein, 2018). Any change that increases net welfare (makes society better off) is deemed desirable, regardless of how the changes in welfare are distributed between people (Adler & Posner, 2006). This is often an important factor in decision-making and CBA is unique in evaluating policies and programs from this perspective (Sunstein,

2018). This criterion is formally defined as *Kaldor-Hicks efficiency*.² Kaldor-Hicks efficiency aligns with positivist paradigms in economics and represents an attempt at establishing an ‘objective’ criterion for evaluating the efficiency of resource allocations (Backhouse, 2016). However, it cannot escape being normative – for example, its ostensibly neutral treatment of equity is still a normative position (King, 2019).

The **standard** in CBA is that the intervention being evaluated should improve Kaldor-Hicks efficiency to a greater degree than alternative interventions. These alternatives do not have to be directly evaluated (though they can be); the universe of hypothetical alternatives is represented within the structure of CBA by an input variable called the *discount rate* (King, 2019).³ When using CBA to evaluate a financial decision, the discount rate represents the *opportunity cost of capital* – that is, the rate of return that could be expected from alternative investments with similar characteristics. When

² Kaldor-Hicks efficiency is an elaboration on Pareto efficiency. An allocation of resources is said to be Pareto-efficient if there is no alternative allocation in which one person can be made better off without making somebody else worse off (Drummond et al., 2005). The Pareto criterion is too restrictive, however, to be practical for evaluating real-world policy proposals, which usually have distributive implications. Kaldor (1939) modified the Pareto criterion by arguing that for an action to be in the public interest, those who gain from it must be able to compensate those who lose from it, and still find the action worthwhile. Crucially, the compensation does not have to actually occur (Drummond et al., 2005). Hicks (1939) added that the losers must not be able to bribe the gainers to forgo the action.

³ The discount rate reflects the *time value of money* – the observation that people would prefer to receive a dollar now than later (Destremau & Wilson, 2017). A dollar we have to wait for is less desirable, and therefore less valuable, than a dollar we receive straight away. The lower value represents the opportunity cost associated with the delay – that is, the ways we could have used the dollar in the intervening period. For example, if we could invest one dollar at 8% cent per annum, turning it into \$1.08 after one year, then the present value of a dollar we have to wait one year to receive is $1/1.08 = 93$ cents. The higher the discount rate, or the further in the future a cost or benefit occurs, the lower its present value.

conducting a social CBA, the discount rate represents either the *social opportunity cost of capital* (the expected rate of return from similar social investments) or the *social rate of time preference* (the premium we collectively place on having an asset available to us now rather than later). This premium would exist even if there were no interest or inflation, because having something sooner gives us more options (Drummond et al., 2005). Neither the opportunity cost of capital nor the social rate of time preference can be directly measured, but they can be contextually determined through modelling, using a combination of empirical data and analyst judgment (Creedy & Passi, 2017). Whichever approach is used, the greater the discount rate, the higher the hurdle that must be met for a project to meet the Kaldor-Hicks criterion (Drummond et al., 2005) – and as such, the discount rate represents the standard in a CBA (King, 2019).

The **evidence**, in CBA, takes the form of monetary valuations of costs and benefits. Though concerned ultimately with social welfare, CBA uses a construct called *utility* to approximate changes in welfare. Under this construct, individuals have preferences, and they enjoy an increase in utility when their preferences are satisfied. Everybody affected by a program experiences increases or decreases in utility. CBA provides analytical structure for estimating and aggregating those gains and losses in individual utility (Adler & Posner, 2006). A construct called *money* is used to value the gains and losses in utility. In principle, the value of anything can be expressed in monetary units (Nicholls et al., 2012; Svistak & Pritchard, 2014), and an extensive suite of methods has evolved for valuing things monetarily, ranging from market prices to empirical measurement. These methods of monetization

imperfectly value changes in utility, and utility imperfectly represents welfare. Alternative approaches to quantifying and monetising wellbeing, based on subjective measures such as life satisfaction, have also emerged (Fujiwara & Campbell, 2011; MacLennan, Stead, & Little, 2021). Selecting contextually appropriate valuation methods is a matter for deliberation (Drummond et al., 2005), and issues such as construct validity and measurement error must be carefully considered (King, 2019). Nonetheless, the evidence in CBA should be recognized as evaluative data: If wellbeing measures are used, the evidence directly represents evaluations by individuals about their own lives (Fujiwara & Campbell, 2011). Even if market prices or other proxies are used to monetize impacts, these valuations indirectly represent the preferences of individuals (Adler & Posner, 2006).

The **synthesis** step in CBA brings together the criterion (Kaldor-Hicks efficiency), the standard (the discount rate) and the evidence (monetary valuations of costs and benefits) to arrive at an estimation of the program's worth. In CBA, the synthesis step is technocratic and tightly prescribed. The methodological prescription specifies the factors to be taken into account in the synthesis, namely monetary valuations of costs and benefits, the points in time at which they occur, and the chosen discount rate. The synthesis step has been summarised algebraically as follows:

If there are i possible social investments (where $i = 1 \dots I$), then the *net present value* (NPV) of project i is:

$$NPV_i = \sum_{t=1}^n \frac{b_i(t) - c_i(t)}{(1+r)^{t-1}}$$

Where:

$b_i(t)$ = benefits, in monetary terms, in year t

$c_i(t)$ = costs, in monetary terms, in year t

$1/(1+r)$ = a discount factor at annual interest rate r and

n = the lifetime of the project (Drummond et al., 2005).

The principal output of a CBA is an indicator of the investment's worth. In the formula above, the indicator is NPV. If the present value of benefits is greater than the present value of costs, then NPV will be greater than zero. Where this is the case, the project is judged to meet the Kaldor-Hicks criterion, implying it is better than alternatives and therefore worth doing.⁴

To illustrate, imagine we are tasked with conducting an *ex-ante* CBA on a proposed initiative to introduce healthy eating guidelines, based on the principle of consuming real or minimally processed foods and avoiding ultra-processed foods (Lustig, 2021; Ministry of Health of Brazil, 2014). For **evidence**, we estimate the costs of producing and promoting the guidelines, the change in resource costs of producing more real foods and less processed foods, and benefits such as improvements in population health, reduced demand for health care, and associated improvements in national productivity. Our **criterion** is Kaldor-Hicks efficiency; that is, we want to know if the net effect of the various costs and benefits across our society is positive or

⁴ Another indicator often used is the benefit:cost ratio (BCR) – the present value of benefits divided by the present value of costs. This is also seen in Social Return on Investment, where it may be expressed as a statement such as: ‘For every \$1 invested in the program, value of \$2.75 is created’. If the present value of benefits is greater than the present value of costs, BCR will be greater than one.

negative overall. To set a **standard**, we select a discount rate of 3.5% in line with UK governmental guidance (HM Treasury, 2022), with sensitivity analysis at discount rates of 0-7%. **Synthesis** is performed using a spreadsheet to calculate NPV, taking into account estimated costs and benefits over a selected time horizon of 20 years. NPV is greater than zero, indicating that the proposed initiative is expected to create benefits that exceed costs. We conclude the initiative is worthwhile.

As some of the input variables in a CBA may be estimates or assumptions subject to risk or uncertainty, it is often a good practice to conduct sensitivity analysis, varying input values to test whether they have a material bearing on the results of the analysis and the conclusions that might be reached (King, 2015). Appraisal of distributional effects can also be carried out as an adjunct to CBA (HM Treasury, 2022).

Alongside NPV, wider considerations, including qualitative considerations, may also be examined. For example, some benefits and costs may be too hard to monetize (Adler & Posner, 2006; MacLennan et al., 2021). An investment may have social, cultural, environmental, or ethical considerations that cannot be included in the analysis but can be described qualitatively (Executive Order No. 13563, 2011, p. 3821; HM Treasury, 2022; Sinden et al., 2009). Historically, no guidance has existed for integrating these additional considerations with the core findings of a CBA. An interdisciplinary methodology involving evaluative reasoning and mixed methods was recently developed to address this gap (King, 2019).

CBA, then, is a form of evaluation. Any evaluation should be conducted in ways that uphold evaluation standards. Accordingly,

deliberation about whether, when, and how economic methods are used in an evaluation is necessary. For example, careful selection of methods would include checking whether the assumptions, criteria, metrics, and processes of reaching conclusions in a CBA are explicitly justified in the cultures and contexts where the evaluation has consequences (Yarbrough et al., 2011) – and if they are not, then alternative methods should be considered.

Some ethical principles espoused in the PES are not explicitly prescribed in economic evaluation. Can CBA be conducted in adherence both to PES and to its own methodological prescriptions, or are there areas where the two sets of standards conflict? Analysis was undertaken to investigate.

Methods

Gap analysis was undertaken to systematically assess the capacity of CBA to adhere to PES. Notwithstanding the multitude of program evaluation standards in use globally, the *Program Evaluation Standards* of the Joint Committee on Standards for Educational Evaluation (Yarbrough et al., 2011) were used as a checklist to systematically assess CBA, acknowledging their longstanding and wide usage in program evaluation, their influence on other evaluation standards internationally (Schwandt, 2015), and the degree of commonality they share with other evaluation standards (Coryn & Stufflebeam, 2014; Deane & Harré, 2016).

The methodological prescription for CBA has been widely published in numerous texts, three of which were the core publications used in this analysis: a textbook commonly used in teaching economic evaluation to program evaluators (Levin & McEwan, 2001); a seminal text for health economists (Drummond et al., 2005); and CBA guidance set out in the latest

edition of the *Green Book* (HM Treasury, 2022). The *Green Book* is the principal document setting out government guidance on the appraisal of public investments in the United Kingdom. It is cited widely in international guidance on CBA (Argyrous, 2013). The prescriptions within these three texts cohere with each other in regard to the core requirements and principles underpinning CBA (King, 2019).

Additional sources were consulted to supplement the three core texts (Adler & Posner, 2006; Allan, Kerr & Grimes, 2013; Chapple, 2017; Destremau & Wilson, 2017; Frank, 2000; Fujiwara & Campbell, 2011; Husereau et al., 2013; Julnes, 2012b; King, 2017; MacLennan et al., 2021; Pinkerton et al., 2002). A literature search did not identify a directly comparable set of standards in economics that would correspond with program evaluation standards. The American Economic Association first adopted a code of professional conduct in April, 2018, and this code was included in the analysis (American Economic Association, 2018).

A rating system was developed, taking the perspective of *efficacy* (addressing the question, ‘*can* CBA meet the PES?’) as distinct from real-world *effectiveness* (‘*does* CBA meet the PES?’) which depends on a wider set of contextual factors. The rating system assesses the inherent capacity of CBA to be conducted in keeping with the PES in principle, rather than the extent to which CBAs have actually met the PES in practice (King, 2019). Three types of rating are assigned: prescribed, permitted, and precluded. *Prescribed* means there is an explicit expectation that CBA should fulfil the relevant PES – it is explicitly stated in any one of the texts consulted. *Precluded* means that adherence to CBA would logically or practically make

adherence to the program evaluation standard impossible. *Permitted* means that meeting the program evaluation standard is neither prescribed nor precluded by the CBA texts consulted (King, 2019). The ratings are evaluative: they represent the judgments of the author, made on the basis of the rating definitions and the literature cited above.

Findings⁵

The methodological prescription for CBA focuses primarily on technical aspects of economic methods to ensure precision, accuracy and reliability (King, 2019). In particular, textbooks and guidelines focus on fidelity to methods for the valuation and discounting of costs and benefits (Drummond et al., 2005; HM Treasury, 2022; Levin & McEwan, 2001). Standards for reporting findings of economic studies emphasise transparency and replicability. For example, reports from health economic evaluations should include a description of relevant context such as setting and location, the study scope and perspective, justification for methodological decisions such as choice of outcome measures, time horizon and discount rate, and discussion of study findings, limitations, generalisability, and how the findings fit with current knowledge (Husereau et al., 2013). Absent from economic texts is guidance on ethical aspects of evaluation such as concern for consequences and influence, contextual viability, human rights and respect as articulated in the PES (Yarbrough, et al., 2011).

Table 1 sets out the systematic assessment of CBA against the PES, showing which standards are prescribed, permitted, or precluded when

⁵ The analysis presented here is summarized from doctoral research (King, 2019).

following the economic guidance for a CBA. The first column of ratings assesses CBA when used as a stand-alone method. The second column explores the potential use of CBA when combined with any complementary methods, quantitative and/or qualitative (King, 2019).

Generally, there is nothing in the methodological guidance for CBA that would preclude CBA from being conducted in adherence to program evaluation standards, provided there is freedom to choose not to conduct a CBA, or to conduct a CBA in combination with other methods. If, however, a decision is made from the outset to conduct a CBA as a complete, stand-alone evaluation, then there are risks that the evaluation will fall short of standards concerned with *explicit values, negotiated purposes, meaningful purposes and products, concern for consequences and influence, contextual viability, and responsive and inclusive orientation* (King, 2019). These risks are summarised below. Although they overlap in some respects, the following analysis approaches each standard discretely.

The **explicit values** standard states: “Evaluations should clarify and specify the individual and cultural values underpinning purposes, processes, and judgments” (Yarbrough et al., 2011, p. 37). This standard recognises that valuing is central to the decisions and judgments made in evaluation, and that stakeholders are more likely to find an evaluation credible and useful if they see their own values reflected in it, and are able to see and understand the value perspectives of others (King, 2019). This standard encourages evaluators to work in open and inclusive ways, ensuring the values guiding the evaluation are not only the values of those in power. Evaluators should avoid imposing their own values, including the values they might place on

different methods and evidence, and should not purport to be objective or values-free (Yarbrough et al., 2011).

CBA seeks to measure the values of stakeholders by way of *compensating variations*, defined as the monetary value that should change hands to make a person as well-off under the proposed change as they would be in the status quo, based on their preferences (Adler & Posner, 2006). Compensating variations can be revealed, for example, by observing prices set by well-functioning markets or through surveys and other methods aimed at eliciting or revealing people's 'willingness to pay' (Drummond et al., 2005). These approaches are capable of being used to estimate intangible values (Allan, Kerr & Grimes, 2013). CBA, in this sense, is a rigorous and disciplined approach to making values explicit.

For example, in our CBA of the proposed introduction of healthy eating guidelines, we could conduct a survey to estimate people's willingness to pay for living longer and in better health. Alternatively, we could infer the value people place on longevity through the use of proxy values such as the *value of a statistical life year* (HM Treasury, 2022), or we could estimate and monetise changes in life satisfaction (MacLennan et al., 2021), informed by a literature review to derive suitable estimates from past studies.

Valuing all costs and benefits in the same units is both a strength and a limitation of CBA. This feature enables CBA to directly compare and reconcile value consumed and value created, reducing them to a single number. However, it also means that once monetized, all values are *fungible* – they look the same and appear interchangeable, though they may be qualitatively different.

A significant hazard to the explicit values standard is the overarching criterion of Kaldor-Hicks efficiency, which comes bundled with the method and includes the position that any resource allocation that increases overall welfare is desirable, regardless of its distributive effects – a remarkable value judgment that is seldom made explicit nor questioned in CBA (King, 2019). Additional criteria such as relevance, sustainability, distributive justice, cultural and historical significance, and deontological ethics, may be important in social policies and programs (Adler & Posner, 2006; Boston, 2017; Pinkerton et al., 2002), are peripheral to CBA, and should not be subordinated to efficiency (King, 2019).

Moreover, by focusing on costs and consequences, CBA runs the risk of missing significant values underpinning program processes (Julnes, 2012b). In this sense, CBA falls short of the full remit of the ‘explicit values’ standard.

For example, perhaps the healthy eating guidelines include recipe suggestions that reflect values and preferences of Western cultures and are incompatible with the cultural traditions of some minority groups. These conflicting values, invisible in the CBA, are likely to contribute to disparities in health outcomes between different groups. The CBA would ascertain that the guidelines confer a net social benefit, but may fail to recognize that the benefit is unevenly distributed and below its full potential.

Bearing all of these considerations in mind, CBA alone would be too restrictive a framework in which to meet the ‘explicit values’ standard, unless the evaluation design was negotiated and agreed in advance with all

stakeholders, and unless the possibility existed to include a wider set of values (King, 2019).

The **negotiated purposes** standard states: “Evaluation purposes should be identified and continually negotiated based on the needs of stakeholders” (Yarbrough et al., 2011, p. 29). Evaluation purposes guide decisions about the design and implementation of an evaluation. Program evaluation stakeholders will have different needs which may not always be aligned with the needs of those responsible for decision-making or resourcing the program and its evaluation (King, 2019).

For example, in contrast to the goal of maximising aggregate welfare, “resources may also be allocated for reasons of equity, which itself has multiple possible conceptions” (Chapple, 2017). Questions of equity require evaluators to adopt a normative position on how resources and opportunities should be distributed. Although distributional analysis can be conducted within a CBA framework (HM Treasury, 2022), this is still fundamentally applying the lens of an efficiency-oriented method to questions of equity, and a more comprehensive treatment of equity and social justice may require an entirely different set of evaluation methods (King, 2019).

For example, it is known that people in poor neighborhoods often live in “food deserts” where ultra-processed food is more accessible and affordable than real food (Lustig, 2021). Introducing the healthy eating guidelines could make society better off overall while at the same time increasing health disparities between rich and poor people. Consumer advocates raise concerns that CBA could justify approval of the proposal despite its inability to improve nutrition in low income neighborhoods. To

mitigate these concerns, complementary strategies could be implemented to address structural factors that give rise to food deserts. The anticipated costs and benefits of these strategies can be incorporated in the CBA. However, determining their acceptability to consumers and producers will require consultative methods that fall outside the remit of CBA. Conducting CBA alone would fail to meet the needs and expectations of affected stakeholders. Negotiating evaluation purposes might lead to an agreement to use multiple methods.

Agreeing to implement any particular method prior to clarifying evaluation purposes is a hazard to meeting the ‘negotiated purposes’ standard. This does not point to any inherent shortcoming in CBA but indicates that, as with all methods, its use should be negotiated and not preordained (King, 2019).

The **meaningful processes and products** standard states:

“Evaluations should construct activities, descriptions, and judgments in ways that encourage participants to rediscover, reinterpret, or revise their understandings and behaviors” (Yarbrough et al., 2011, p. 51). It is argued that this standard is a necessary part of ensuring stakeholders’ needs are met in an evaluation. Risks to meeting the standard include using the evaluation terms of reference to stipulate and impose upon stakeholders the approach that will be taken, and proceeding without regard to stakeholders’ reactions to the evaluation.

CBA could be adapted to meet the ‘meaningful processes and products’ standard. In particular, this would require the flexibility to engage stakeholders in determining whether or not to use CBA, whether to use CBA

alone or in combination with other methods, and in making sense of findings. This practice could enhance the use of CBA by facilitating understanding, ownership and use of the evaluation (King, 2019). However, this standard would not be met if CBA were mandated or imposed.

For example, before conducting a CBA on the introduction of healthy eating guidelines, we could establish a citizen panel. We could explain to the panel the objectives of the study, how CBA works, what CBA can tell us that we can't get from other methods, what CBA can't tell us, and discuss options for conducting a wider analysis not limited to economic methods and metrics alone. One of the possible outcomes of this consultation is that citizens may reject the use of CBA. We could also engage with the panel when reviewing the findings of our study and considering implications for policy making. None of these processes are recommended in CBA guidelines. The *Green Book* suggests engagement with stakeholders as a source of inside knowledge to inform scoping of policy options, costs and benefits (HM Treasury, 2022) but not directly in evaluation design, methods or reaching conclusions.

Concern for consequences and influence encourages those conducting evaluations to “promote responsible and adaptive use while guarding against unintended negative consequences and misuse” (Yarbrough et al., 2011, p. 65). This standard seeks to ensure evaluation contributes to social betterment by catalysing improvements in policies, programs and contexts, and recognises that evaluations have potential to do harm – for example, by jeopardising democratic participation, equity, social justice or truth. The standard states that it is important not to assume “that a technically

excellent evaluation is sufficient for positive use and effective influence” (p. 67).

For example, consider a scenario in which decision-makers commission a CBA as the exclusive method to evaluate the proposal to implement healthy eating guidelines. Although we point out the limitations and risks of conducting CBA alone, the commissioners consider CBA to be all that is needed. Being PES-carrying evaluators, we decline to undertake the study. The work is picked up by an economist who conducts the analysis on a desktop basis and in keeping with CBA standards. Stakeholders are not consulted and are unaware of the study until its findings are released.

The ‘concern for consequences and influence’ standard again underscores the risk of affording CBA ‘gold standard’ status and argues for a more flexible and responsive approach to evaluation design and methods (King, 2019).

The **contextual viability** standard states that “Evaluations should recognize, monitor, and balance the cultural and political interests and needs of individuals and groups” (Yarbrough et al., 2011, p. 93). This standard recognises the potential power imbalances between stakeholders and the need to understand the different cultural, political and economic interests held by different stakeholder groups. It is important to take care to respond to all stakeholder needs in a balanced way and not, for example, to be perceived as placing the needs of one group (such as decision-makers) ahead of others. Adherence to this standard requires appropriate mechanisms for stakeholders to have input, which again reinforces the need for flexibility to determine an appropriate method or mix of methods (King, 2019).

The method of synthesis in CBA is a hazard to contextual viability, because valuing all costs and benefits in commensurable units and aggregating them using the CBA formula can serve to obscure clarity where a divergence of values are at play, such as power imbalances or incompatible worldviews (Julnes, 2012b). Indeed, one of the criticisms of CBA is that it can act in effect like a voting system in which the interests of a powerful majority override those of disadvantaged minorities (King, 2019). In some contexts, deliberation on values that diverge or are in tension may be more appropriate than attempting to aggregate those values (House & Howe, 1999).

For example, experts anticipate that the food and pharmaceutical industries, both of which profit from widespread consumption of ultra-processed foods, might oppose attempts to improve population nutrition (Lustig, 2021). We adapt the CBA to examine costs and benefits of healthy eating guidelines from the perspectives of consumers and producers respectively, and find that consumers stand to benefit while industry faces a reduction in profits. To the extent that the values of industry lobby groups conflict with the health and wellbeing interests of the population, we cannot calculate our way to an evaluative judgment using CBA. However, by separating out the perspectives of different groups, the CBA contributes important information to a wider process of deliberation by clarifying tensions between different sets of interests.

There are also important ethical implications associated with some seemingly technical decisions within the design of a CBA, such as the perspective taken (i.e. costs and benefits to whom?) and the choice of discount rate. For example, taking a societal perspective in the identification

of costs and benefits, as is commonplace in CBA, implicitly treats all members of society as equally interested parties, sidestepping the fundamental democratic question of “who should have how much say about what?” (Mulgan, 1984). The choice of discount rate affects the valuation of long-term costs and benefits that have impacts for future generations who are not able to have a say at the time of the decision (Destremau & Wilson, 2017).

CBA also privileges quantitative evidence. In the view of many evaluators, multiple forms of evidence and ways of creating knowledge should contribute to evaluative judgments about complex social issues (Deane & Harré, 2016; Greene, 2005; Mertens & Hesse-Biber, 2013; Wehipeihana & McKegg, 2018).⁶

The conduct of CBA in accordance with the ‘contextual viability’ standard demands stakeholder input into evaluation design decisions (such as perspective and discount rate) that have traditionally been the domain of the analyst, and requires remaining open to the possibility that CBA may not be a viable evaluation approach in some contexts.

Responsive and inclusive orientation demands that “Evaluations should be responsive to stakeholders and their communities” (Yarbrough et al., 2011, p. 113). Evaluators have a “moral professional duty” to ensure stakeholders are included and attended to in a proportionate, systematic and transparent way. Meeting this standard requires evaluators to build

⁶ It is also noteworthy that considerable qualitative judgment goes into the process of quantification in CBA (for example, selecting methods of monetization). If these judgments are not made transparent, this would also fall short of the ‘explicit values’ standard.

meaningful relationships and seek stakeholder contributions to the evaluation. It requires an openness to contradictory views and interests. It may, on occasion, involve deliberative and democratic processes. Hazards to this standard include “always favouring a specific evaluation method or approach without proper regard for the needs of the actual stakeholders in the current setting and the purposes of the evaluation”, “not attending adequately to context or culture in evaluation designs and practices” and “ignoring the political vibrancy and inherent value of stakeholder positions and value judgments” (p. 116). This standard would be difficult to meet using CBA alone, for many of the reasons already canvassed (King, 2019).

To end on a positive note, consider a scenario in which we successfully advocate for a multi-method approach which combines CBA with democratic deliberative citizen engagement (House & Howe, 1999). Explicit use of the general logic of evaluation facilitates transparent balancing of NPV with other criteria (King, 2019) such as equity, cultural acceptability, and commercial interests. The evaluation meets the decision-maker’s requirement for an overall assessment of the net social benefit of implementing healthy eating guidelines, while also identifying unanticipated issues that, if addressed, are likely to improve the acceptability and adoption of the guidelines, ultimately leading to greater and more equitable impacts.

Collectively, the analysis summarised in Table 1 indicates that CBA can be conducted in keeping with the PES, provided its use as a stand-alone method is not predetermined or imposed on stakeholders. This finding supports the proposition that CBA should be regarded as one tool in an

evaluator's toolbox, to be used in contextually responsive ways, and in combination with other methods (King, 2019).

Conclusion

For CBA to be accepted into the field of practice of program evaluation, it should adhere to PES and be subjected to meta-evaluation against such standards. Although such standards are open to ongoing debate and interpretation, they nevertheless provide a widely accepted point of reference for assessing the legitimacy and quality of evaluation designs, practices and products (King, 2019).

Systematic analysis of CBA against the PES found that some ethical principles espoused in the PES are not explicitly prescribed in economic evaluation. Nevertheless, CBA is generally capable of being conducted in accordance with the standards, provided the options of not using CBA, or combining CBA with other methods, are available. Applying the PES to economic evaluations may therefore enhance the use of CBA. While many of these ethical principles may already be well accepted and followed by people conducting economic evaluations, there is value in making them explicit (Yarbrough et al., 2011).

There are significant risks that the PES cannot be adhered to in situations where CBA is chosen in advance as the sole evaluation method. Where an evaluation commissioner prescribes the use of CBA, they run the risk that the evaluation will fall short of standards for explicit values, negotiated purposes, meaningful purposes and products, concern for consequences and influence, contextual viability, and responsive and

inclusive orientation. Use of PES should help to clarify when and how to use economic methods in evaluation.

CBA is a valid form of evaluation. It estimates something important – aggregate welfare (Sunstein, 2018). It may do so imperfectly (Adler & Posner, 2006; Julnes, 2012b) but it is the most fit-for-purpose method available to address this criterion (Adler & Posner, 2006). “Whether or not an analysis of costs and benefits tells us everything we need to know, at least it tells us a great deal that we need to know” (Sunstein, 2018, p. xi). Evaluators should use CBA more (Gargani, 2017; Julnes, 2012c; King, 2019; Yates, 1996).

However, CBA is also a limited form of evaluation because it: privileges a single criterion and a single standard (without consultation and often without making them explicit); focuses on costs and consequences while remaining agnostic about how consequences are achieved (processes) or distributed (equity); values all costs and consequences in a single metric; and aggregates them using a tightly prescribed formula.

For these reasons, evaluation theorists have argued that if CBA is conducted in isolation from other forms of evaluation, there is a risk that it may provide an incomplete picture of a program’s worth, “a distorted understanding of the public interest and a diminished capacity for evaluation in general to serve that interest” (Julnes, 2012a, p. 1).

Given these considerations, CBA will often be insufficient as a stand-alone method to provide a complete evaluation. CBA has great potential to inform evaluations and should be used more widely for this purpose.

However, in general CBA should be used in a supporting role to a wider

process of evaluative reasoning and in conjunction with other methods (King, 2019).

From this analysis, it is concluded that CBA can be conducted in adherence to program evaluation standards, provided its use is negotiated with stakeholders. If CBA is used in combination with other methods, it is possible to work in an inclusive and responsive way, with the full range of stakeholder values, and conduct evaluations that are contextually viable and meaningful for stakeholders.

Acknowledgement

I am grateful to Aaron Schiff and Charles Sullivan for their feedback on this paper. Any errors or omissions are mine alone.

Table 1: Assessment of CBA against program evaluation standards

Program Evaluation Standards (Yarbrough et al., 2011)	Efficacy of CBA as a stand-alone evaluation method	Efficacy of CBA as part of a mixed methods evaluation
Utility standards		
U1: Evaluator credibility Evaluations should be conducted by qualified people who establish and maintain credibility in the evaluation context	Prescribed ⁷	Prescribed
U2: Attention to stakeholders Evaluations should devote attention to the full range of individuals and groups invested in the program and affected by its evaluation	Prescribed	Prescribed
U3: Negotiated purposes Evaluation purposes should be identified and continually negotiated based on the needs of stakeholders	<u>Precluded</u>	Permitted
U4: Explicit values Evaluations should clarify and specify the individual and cultural values underpinning purposes, processes, and judgments	<u>Precluded</u>	Permitted
U5: Relevant information Evaluation information should serve the identified and emergent needs of stakeholders	Permitted	Permitted
U6: Meaningful processes and products Evaluations should construct activities, descriptions, and judgments in ways that encourage participants to rediscover, reinterpret, or revise their understandings and behaviors	<u>Precluded</u>	Permitted
U7: Timely and appropriate communicating and reporting Evaluations should attend to the continuing information needs of their multiple audiences	Prescribed	Prescribed
U8: Concern for consequences and influence Evaluations should promote responsible and adaptive use while guarding against unintended negative consequences and misuse	<u>Precluded</u>	Permitted
Feasibility standards		
F1: Project management Evaluations should use effective project management strategies	Permitted	Permitted
F2: Practical procedures	Permitted	Permitted

⁷ *Prescribed* means there is an explicit expectation that CBA should fulfil the relevant PES – it is explicitly stated in any of the reference texts. *Precluded* means that adherence to CBA would logically or practically make adherence to the program evaluation standard impossible. *Permitted* means that meeting the program evaluation standard is neither prescribed nor precluded (King, 2019). The ratings are evaluative: they represent the judgments of the author, on the basis of these definitions and the cited literature.

Evaluation procedures should be practical and responsive to the way the program operates		
F3: Contextual viability Evaluations should recognize, monitor, and balance the cultural and political interests and needs of individuals and groups	<u>Precluded</u>	Permitted
F4: Resource use Evaluations should use resources effectively and efficiently	Prescribed	Prescribed
Propriety standards		
P1: Responsive and inclusive orientation Evaluations should be responsive to stakeholders and their communities	<u>Precluded</u>	Permitted
P2: Formal agreements Evaluation agreements should be negotiated to make obligations explicit and take into account the needs, expectations, and cultural contexts of clients and other stakeholders	Permitted	Permitted
P3: Human rights and respect Evaluations should be designed and conducted to protect human and legal rights and maintain the dignity of participants and other stakeholders	Permitted	Permitted
P4: Clarity and fairness Evaluations should be understandable and fair in addressing stakeholder needs and purposes	Permitted	Permitted
P5: Transparency and disclosure Evaluations should provide complete descriptions of findings, limitations, and conclusions to all stakeholders, unless doing so would violate legal and propriety obligations	Prescribed	Prescribed
P6: Conflicts of interest Evaluations should openly and honestly identify and address real or perceived conflicts of interests that may compromise the evaluation.	Prescribed	Prescribed
P7: Fiscal responsibility Evaluations should account for all expended resources and comply with sound fiscal procedures and processes	Permitted	Permitted
Accuracy standards		
A1: Justified conclusions and decisions Evaluation conclusions and decisions should be explicitly justified in the cultures and contexts where they have consequences	Permitted	Permitted
A2: Valid information Evaluation information should serve the intended purposes and support valid interpretations	Prescribed	Prescribed
A3: Reliable information Evaluation procedures should yield sufficiently dependable and consistent information for the intended uses	Prescribed	Prescribed
A4: Explicit program and context descriptions Evaluations should document programs and their contexts with appropriate detail and scope for the evaluation purposes	Prescribed	Prescribed

A5: Information management Evaluations should employ systematic information collection, review, verification, and storage methods	Permitted	Permitted
A6: Sound designs and analyses Evaluations should employ technically adequate designs and analyses that are appropriate for the evaluation purposes	Prescribed	Prescribed
A7: Explicit evaluation reasoning Evaluation reasoning leading from information and analyses to findings, interpretations, conclusions, and judgments should be clearly and completely documented	Permitted	Permitted
A8: Communication and reporting Evaluation communications should have adequate scope and guard against misconceptions, biases, distortions, and errors	Prescribed	Prescribed
Evaluation accountability standards		
E1: Evaluation documentation Evaluations should fully document their negotiated purposes and implemented designs, procedures, data, and outcomes	Prescribed	Prescribed
E2: Internal meta-evaluation Internal: Evaluators should use these and other applicable standards to examine the accountability of the evaluation design, procedures employed, information collected, and outcomes.	Permitted	Permitted
E3: External meta-evaluation Program evaluation sponsors, clients, evaluators, and other stakeholders should encourage the conduct of external meta-evaluations using these and other applicable standards	Permitted	Permitted

References

- Adler, M.D., & Posner, E.A. (2006). *New Foundations of Cost-Benefit Analysis*. Cambridge, Mass: Harvard University Press.
- Allan C., Kerr, S., Grimes, A. (2013). *Value and Culture. Motu Working Paper 13-09*. Wellington: Motu Economic and Public Policy Research.
- American Economic Association. (2018). AEA Code of Professional Conduct, adopted April 20, 2018. [web page]. Retrieved from: <https://www.aeaweb.org/about-aea/code-of-conduct>
- American Evaluation Association. (2018). *Guiding Principles for Evaluators*. [web page]. Retrieved from: <https://www.eval.org/About/Guiding-Principles>
- ANZEA & Superu. (2015). *Evaluation standards for Aotearoa New Zealand*. Wellington, NZ: Aotearoa New Zealand Evaluation Association and Social Policy Evaluation and Research Unit.
- Argyrous, G. (2013). *A review of government cost-benefit analysis guidelines*. SSC/ANZSOG Occasional Paper. Canberra, Australia: Australia and New Zealand School of Government.
- Australian Evaluation Society. (2013). *Guidelines for the Ethical Conduct of Evaluations*. Retrieved from: https://www.aes.asn.au/images/AES_Guidelines_web_v2.pdf?type=file
- Backhouse, R.E. (2016, August). *The origins of New Welfare Economics. Preliminary draft written to provoke discussion at a workshop in welfare economics*, Hitotsubashi University, Japan.
- Boston, J., & Gill, D. (Eds.). (2017). *Social investment: A New Zealand policy experiment*. Wellington, New Zealand: Bridget Williams Books.

Chapple, S. (2017). Corked wine in a cracked bottle. In J. Boston & D. Gill (Eds.), *Social investment: A New Zealand policy experiment* (pp. 355- 379). Wellington, New Zealand: Bridget Williams Books.

Coryn, C.L.S. & Stufflebeam, D.L. (2014). *Evaluation Theory, Models, & Applications*. San Francisco, CA: Jossey-Bass.

Creedy, J., & Passi, H. (2017). *Public sector discount rates: a comparison of alternative approaches*. Working Paper 17/02. Wellington, New Zealand: NZ Treasury.

Damart, S., Roy, B. (2009). The uses of cost-benefit analysis in public transportation decision-making in France. *Transport Policy*, 16, 200- 212.

Davis, K.E., Frank, R.G. (1992). Integrating costs and outcomes. *New Directions for Evaluation* (54), 69-84.

Deane, K.L., & Harré, N. (2016). Developing a thoughtful approach to evaluation: Values-driven guidelines for novice evaluators. *Evaluation Matters—He Take Tō Te Aromatawai*, 2, 53-78.

Destremau, K., & Wilson, P. (2017). Defining social investment, Kiwi-style. In J. Boston & D. Gill (Eds.), *Social investment: A New Zealand policy experiment* (pp. 32-79). Wellington, New Zealand: Bridget Williams Books.

Drummond, M. F., Sculpher, M. J., Torrance, G. W., O'Brien, B. J., & Stoddard, G. L. (2005). *Methods for the economic evaluation of health care programs*. Oxford, England: Oxford University Press.

Exec. Order No. 13563, 3 C.F.R. Page 3821 (2011).

Fournier, D. M. (1995). Establishing evaluative conclusions: A distinction between general and working logic. In D. M. Fournier (Ed.), *Reasoning in Evaluation: Inferential Links and Leaps*. *New Directions for Evaluation*, (58), 15-32.

Frank, R. (2000). Why is cost-benefit analysis so controversial? In M.D. Adler & E.A. Posner (Eds.), *Cost-Benefit Analysis: Legal, Economic, and Philosophical Perspectives* (pp. 77-94). Chicago, IL: University of Chicago Press.

Fujiwara, D., & Campbell, R. (2011). *Valuation Techniques for Social Cost-Benefit Analysis: Stated Preference, Revealed Preference and Subjective Well-Being Approaches – A Discussion of the Current Issues*. Government of the United Kingdom. Retrieved from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/209107/greenbook_valuationtechniques.pdf

Greene, J.C. (2005). The generative potential of mixed methods inquiry. *Westminster Studies in Education*, 28(2), 207-211.

HM Treasury. (2022). *The Green Book: Central government guidance on appraisal and evaluation*. Government of the United Kingdom. Retrieved from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1063330/Green_Book_2022.pdf

House, E.R., Howe, K.R. (1999). *Values in Evaluation and Social Research*. Sage.

Husereau, D., Drummond, M., Petrou, S., Carswell, C., Moher, D., Greenberg, D., Augustovski, F., Briggs, A.H., Mauskopf, J., and Loder, E. (2013). *Consolidated Health Economic Evaluation Reporting Standards (CHEERS) Statement*. Cambridge University Press.

Julnes, G. (2012a). Editor's Notes. In G. Julnes (Ed). *Promoting Valuation in the Public Interest: Informing Policies for Judging Value in Evaluation. New Directions for Evaluation*, 133, pp 1-2.

Julnes, G. (2012b). Managing valuation. In G. Julnes (Ed). *Promoting Valuation in the Public Interest: Informing Policies for Judging Value in Evaluation. New Directions for Evaluation*, 133, pp 3-15.

Julnes, G. (2012c). Promoting valuation in the public interest. In G. Julnes (Ed). *Promoting Valuation in the Public Interest: Informing Policies for Judging Value in Evaluation. New Directions for Evaluation, 133*, pp 109-129.

King, J. (2015). Use of cost-benefit analysis in evaluation. Letter to the editor. *Evaluation Journal of Australasia, 15*(3), 37-41.

King, J. (2017). Using Economic Methods Evaluatively. *American Journal of Evaluation, 38*(1), 101-113.

King, J. (2019). *Evaluation and Value for Money: Development of an approach using explicit evaluative reasoning*. (Doctoral dissertation). Melbourne, Australia: University of Melbourne.

Levin, H.M., McEwan, P.J. (2001). *Cost-Effectiveness Analysis*. 2nd Ed. Thousand Oaks: Sage.

Levy, H., & Sarnat, M. (1994). *Capital Investment & Financial Decisions (5th Ed)*. Hertfordshire, United Kingdom: Prentice Hall.

MacLennan, S., Stead, I., Little, A. (2021). *Wellbeing Guidance for Appraisal: Supplementary Green Book Guidance*. HM Treasury, Government of the United Kingdom.

Retrieved from:

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1005388/Wellbeing_guidance_for_appraisal_-_supplementary_Green_Book_guidance.pdf

Mertens, D.M., Hessie-Biber, S. (2013). Mixed methods and credibility of evidence in evaluation. In D.M. Mertens & S. Hessie-Biber (Eds.), *Mixed methods and credibility of evidence in evaluation. New Directions For Evaluation, 138*, 5-13.

Mulgan, R.G. (1984). Who should have how much say about what? Some Problems in Pluralist Democracy. *Political Science, 36*(2), 112-124.

Nicholls, J., Lawlor, E., Neitzert, E., Goodspeed, T. (2012). *A Guide to Social Return on Investment*. January 2012. Haddington, England: The SROI Network.

OECD DAC. (2012). *DAC Guidelines and Reference Series. Quality Standards for Development Evaluation*. Evaluation Network of the Development Assistance Committee (DAC) of the Organisation for Economic Co-operation and Development (OECD). Retrieved from <http://www.afdb.org/fileadmin/uploads/afdb/Documents/Evaluation-Reports/Quality%20Standards%20for%20Development%20Evaluation.pdf>

Patel, M. (2013). African Evaluation Guidelines. *African Evaluation Journal*, 1(1), 1-5.

Patton, M.Q. (2017). *Evaluation Flash Cards: Embedding evaluative thinking in organizational culture*. St Paul, MN: Otto Bremer Trust.

Pinkerton, S.D., Johnson-Masotti, A.P., Derse, A., Layde, P.M., (2002). Ethical issues in cost-effectiveness analysis. *Evaluation and Program Planning* 25, 71-83.

Schwandt, T. (2015). *Evaluation Foundations Revisited: Cultivating a Life of the Mind for Practice*. Redwood City: Stanford University Press.

Scriven, M. (1980). *The logic of evaluation*. Thousand Oaks, CA: Sage.

Scriven, M. (1991). *Evaluation thesaurus*. Newbury Park, CA: Sage.

Scriven, M. (1994). The Final Synthesis. *Evaluation Practice*, 15(3), 367-382.

Scriven, M. (1995). The Logic of Evaluation and Evaluation Practice. *New Directions For Evaluation*, 68, 49-70.

Scriven, M. (2012). The logic of valuing. In G. Julnes (Ed.), *Promoting valuation in the public interest: Informing policies for judging value in evaluation. New directions for evaluation*, 133, 17-28.

Sinden, A., Kysar, D. A., & Driesen, D. M. (2009). Cost-benefit analysis: New foundations on shifting sand. *Regulation & Governance*, 3, 48-71.

Sunstein, C. R. (2018). *The Cost-Benefit Revolution*. Cambridge, MA: MIT Press.

Svistak, M., & Pritchard, D. (2014). *Economic evaluation: What is it good for? A guide for deciding whether to conduct an economic evaluation*. London, England: New Philanthropy Capital.

United Nations Evaluation Group. (2016). *Norms and Standards for Evaluation*. New York, NY: UNEG.

Vaca, S., King, J. (2020). *Outcome Efficiency Indicators*. [web page]. Retrieved from: <http://www.saravaca.com/project/outcome-efficiency-indicators/>

Wehipeihana, N., & McKegg, K. (2018). Values and culture in evaluative thinking: Insights from Aotearoa New Zealand. In A.T. Vo & T. Archibald (Eds.). *Evaluative Thinking. New Directions for Evaluation*, 158, 93-107.

Yarbrough, D. B., Shulha, L. M., Hopson, R. K., and Caruthers, F. A. (2011). *The program evaluation standards: A guide for evaluators and evaluation users (3rd ed.)*. Thousand Oaks, CA: Sage

Yates, B.T., (2012). Step arounds for common pitfalls when valuing resources used versus resources produced. *Promoting Valuation in the Public Interest: Informing Policies for Judging Value in Evaluation. New Directions for Evaluation*, 133, 3-52.