

This is a repository copy of *Programme evaluation of population and system level policies:evidence for decision-making*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/173208/>

Version: Published Version

---

**Article:**

Walker, Simon Mark orcid.org/0000-0002-5750-3691, Fox, Aimee orcid.org/0000-0001-6944-7554, Altunkaya, James orcid.org/0000-0002-8293-3466 et al. (6 more authors) (2022) Programme evaluation of population and system level policies:evidence for decision-making. Medical Decision Making. 17–27. ISSN: 1552-681X

<https://doi.org/10.1177/0272989X211016427>

---

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution-NonCommercial (CC BY-NC) licence. This licence allows you to remix, tweak, and build upon this work non-commercially, and any new works must also acknowledge the authors and be non-commercial. You don't have to license any derivative works on the same terms. More information and the full terms of the licence here: <https://creativecommons.org/licenses/>

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



# Program Evaluation of Population- and System-Level Policies: Evidence for Decision Making

Simon Walker<sup>1</sup>, Aimee Fox<sup>1</sup>, James Altunkaya, Tim Colbourn, Mike Drummond, Susan Griffin, Nils Gutacker<sup>1</sup>, Paul Revill, and Mark Sculpher

**Background.** Policy evaluations often focus on ex post estimation of causal effects on short-term surrogate outcomes. The value of such information is limited for decision making, as the failure to reflect policy-relevant outcomes and disregard for opportunity costs prohibits the assessment of value for money. Further, these evaluations do not always consider all relevant evidence, other courses of action, or decision uncertainty. **Methods.** In this article, we explore how policy evaluation could better meet the needs of decision making. We begin by defining the evidence required to inform decision making. We then conduct a literature review of challenges in evaluating policies. Finally, we highlight potential methods available to help address these challenges. **Results.** The evidence required to inform decision making includes the impacts on the policy-relevant outcomes, the costs and associated opportunity costs, and the consequences of uncertainty. Challenges in evaluating health policies are described using 8 categories: 1) valuation space; 2) comparators; 3) time of evaluation; 4) mechanisms of action; 5) effects; 6) resources, constraints, and opportunity costs; 7) fidelity, adaptation, and level of implementation; and 8) generalizability and external validity. Methods from a broad set of disciplines are available to improve policy evaluation, relating to causal inference, decision-analytic modeling, theory of change, realist evaluation, and structured expert elicitation. **Limitations.** The targeted review may not identify all possible challenges, and the methods covered are not exhaustive. **Conclusions.** Evaluations should provide appropriate evidence to inform decision making. There are challenges in evaluating policies, but methods from multiple disciplines are available to address these challenges. **Implications.** Evaluators need to carefully consider the decision being informed, the necessary evidence to inform it, and the appropriate methods.

## Highlights

- Evaluating policies by estimating their causal effects on short-term surrogate outcomes is, in isolation, of limited value for decision making.
- Evidence for informing decision making needs to link policies to relevant outcomes, costs, and associated opportunity costs and reflect the magnitude and consequences of uncertainty in these estimations.
- Challenges in program evaluation range across defining the valuation space and comparators, understanding mechanisms of action and effects, estimating opportunity costs, and external validity.
- Methods from multiple disciplines are available to address these challenges.

## Keywords

cost-effectiveness, economic evaluation, programme evaluation

Date received: February 26, 2021; accepted: April 18, 2021

## Corresponding Author

Simon Walker, Centre for Health Economics, University of York,  
Alcuin A Block, Heslington, York, Yorkshire, YO10 5DD, UK;  
(simon.walker@york.ac.uk).

Much of the applied health economic research of population- and system-level health policies has focused on the ex post estimation of the causal effects on short-term surrogate outcomes using observational data.<sup>1,2</sup> This literature has been supported by methodological developments that facilitate causal inference.<sup>3</sup> However, the information produced by these studies is usually insufficient for decision making. The failure to reflect the impacts on policy-relevant outcomes, for which decision makers can be held accountable (e.g., population health) and the disregard for opportunity costs prohibits the assessment of value for money.<sup>1,4</sup> Further, the ex post nature limits their value for decision making, since the information they provide is necessarily available only after the decision to introduce the policy has been made, although it may be useful to inform subsequent decisions (e.g., policy redesign). Studies that focus their conclusions on what previous decisions achieved often display limited consideration of future decisions study results could inform (for example, the introduction of similar policies in other countries). Decision makers need information on a policy's expected impact at different points in its life cycle; for example, whether to introduce it, maintain it, scale it up, or withdraw it. Further, evidence-informed policy choices are not concerned only with whether a policy should be implemented but also with the timing of when it should be implemented in the face of uncertainty or whether there is value in investing in further research implementing a policy.

Economic evaluation methods, often using the framework of decision analysis, have been developed to inform a range of decisions, including whether the interventions should be introduced, continued with, or disinvested from by health care services and also to consider the value of collecting additional evidence.<sup>5,6</sup> However, these methods have not been widely used to assess health policies. Kreif et al<sup>7</sup> found that of 2419 health-related impact

evaluations identified from 2010 to 2016, only 42 (2%) included an economic evaluation, and those were generally of poor quality. In contrast to evaluations of health policies, the value for money of clinical interventions (e.g., medicines, procedures, and diagnostics) is routinely assessed in many countries,<sup>8</sup> often through health technology assessment processes, although debates continue about what are suitable methods.<sup>9</sup>

Until recently, there has been little attempt to integrate program (or impact) evaluation with economic evaluation methods. However, recent research has started to bring the two together and to integrate learning from other fields such as epidemiology,<sup>10</sup> to consider the value for money of health policies.<sup>1,4</sup> This article aims to develop this further by exploring how evaluations could better inform policy choices. We begin by defining the evidence required in terms of the impacts on the policy-relevant outcomes, the costs and associated opportunity costs, and the magnitude and consequences of uncertainty. Then we identify and categorize the key challenges encountered when evaluating policies. Finally, we highlight approaches that have been applied and a series of examples as signposts for readers who wish to explore these methods further.

## What Decision Is Being Informed and What Evidence Is Required?

Economic evaluation informs policy choices by providing evidence on the benefits, costs, and opportunity costs of alternative courses of action. An intervention would be considered value for money if its benefits exceed its opportunity costs (i.e., its benefits exceed the benefits that could be generated by alternative use of the same resources).<sup>11</sup> Economic evaluation has been used within health care to consider the value of clinical interventions and by governments to consider the value of a wide range of policies.<sup>8,12</sup>

To be informative, an economic evaluation needs to report on outcome(s) of relevance to the decision maker(s) involved, based on their objectives and responsibilities.<sup>13</sup> The outcomes not only refer to an intervention's planned benefits but also include any unintended consequences that can result from policies.<sup>14</sup> To facilitate consistency in decisions and assessment of opportunity cost, it is preferable that the outcomes reflect what is considered relevant to other potential uses of the same resources (e.g., what outcomes would be considered if the resources were instead used for the provision of particular treatments). An economic evaluation should then proceed as the evidential assessment of the impact of the

---

Centre for Health Economics, University of York, York, UK (SW, MD, SG, NG, PR, MS); Adelphi Values, Bollington, Cheshire, UK (AF); Nuffield Department of Population Health, University of Oxford, Oxford, Oxfordshire, UK (JA); Institute for Global Health, University College London, London, UK (TC). The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article. The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Financial support for this study was provided by the National Institute for Health Research (NIHR; NIHR Global Health Econometrics & Economics Group) using UK aid from the UK government. The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report. The views expressed in this publication are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

policies on those outcomes, which can be used to compare the intervention's outcomes with its opportunity costs to determine if the former exceed the latter.

To evaluate value for money, it is necessary to compare an intervention against relevant alternative courses of action. This may constitute a defined set of mutually exclusive options available to pursue specific objective(s), which goes further than just comparing one policy specification with the status quo.<sup>11</sup> For example, an evaluation of a sugar-sweetened beverage tax might warrant consideration of different ways of implementing a tax on sugar. Similarly, a decision may not always be a dichotomy of whether to introduce a policy or not, and other options may be considered; for example, whether to scale up a policy or whether to invest in further research on a policy's impacts.

Evaluation is inevitably uncertain, reflecting incomplete evidence and knowledge. This uncertainty imposes costs, as choosing the suboptimal policy has negative implications for objectives. Here, we refer not to statistical significance but to decision uncertainty, the probability and consequences of incorrect decisions. Regardless of the risk of error, decisions will have to be made about what course of action to pursue at a given time. By assessing the implications of the decision uncertainty, the decision maker can assess the value of acquiring further evidence to reduce it. On this basis, evidence-informed decision making can expand on the range of decision options beyond the "yes" or "no" to a given policy to include those such as whether to delay an adoption decision until additional evidence becomes available.<sup>5,15</sup>

## Challenges in Evaluating Population and Systems Interventions: A Targeted Literature Review and Classification

We conducted a targeted literature review to identify key challenges in undertaking economic evaluation of policies, many of which apply to any form of evaluation. We identified articles by asking members of an Expert Advisory Group (see supplementary material) for key papers, both methodological and applied, on the evaluation of population- and system-level policies. We reviewed the reference list for these articles to identify further relevant publications. In total, 40 methods articles and 2 reviews of applied papers were identified (see supplementary material).

Each article was reviewed by one member of the research team responsible for reviewing (S.W., A.C., J.A., and M.D.), and the details of any methods challenges raised were extracted. Following the review of all articles, the entire research team (all coauthors) considered

the methods challenges raised and categorized them under 8 themes: 1) the valuation space; 2) comparators; 3) timing of the evaluation; 4) mechanisms of action; 5) effects; 6) resources, constraints, and opportunity costs; 7) fidelity, adaptation, and level of implementation; and 8) generalizability and external validity. Supplementary Table S1 sets out a brief description of each theme, the associated challenges, the problem(s) they entail, and key quotes. Below, we summarize the challenges grouped under each theme.

### Valuation Space

We define the valuation space as the set of policy-relevant outcomes. When evaluating health care treatments, the focus is often limited to their impact on the health of patients, typically measured using generic units of health such as the quality-adjusted life-year (QALY). This may also be true when considering some population- and system-level policies. However, broader impacts may be relevant for some policies, including those related to health (e.g., mortality) and health care processes (e.g., access to care) but also non-health related outcomes (e.g., labor market participation, financial protection).<sup>13,16–19</sup> The unit of analysis for the outcomes of interest may be at the individual level or higher (e.g., firm, hospital).

Delineating and defining the relevant outcomes for evaluation can be challenging. This is particularly the case where there are different decision makers involved with interests in varying outcomes.<sup>16–18</sup> Expanding the valuation space to multiple outcomes raises challenges related to the measurement and valuation of the outcomes (e.g., double counting) and to comparing them with the outcomes generated by other policies.<sup>13</sup> The valuation space also needs to be broad enough to capture any unintended consequences.<sup>1</sup> Economic evaluation requires evidence of not only the direct effects of the policies on all outcome(s) of interest but also of its opportunity costs, which ideally would be expressed in terms of these outcomes. Where there are improvements in some outcomes and deterioration in others, some method of aggregating across outcomes is required to decide whether the policy is beneficial (whether done informally or formally).<sup>13</sup>

Decision makers may also care about reducing unfair inequalities in the outcomes.<sup>16,20</sup> To consider these in an evaluation, evidence on both the baseline distribution of the outcomes and distribution of the policy effects (both direct and opportunity costs) on the outcomes across equity-relevant groups of interest is required.<sup>21</sup> Further information will be needed on the degree of inequality aversion for the outcomes.<sup>22,23</sup>

### *Comparators*

We define the comparators as the set of alternative courses of actions subject to evaluation. To establish whether a policy is of value requires it to be compared with other potential courses of action that could be used to pursue the same objective(s). Evaluations of health care interventions often focus on alternative options for a single component within a treatment pathway, for example, whether to provide drug A or B. However, health care policies are typically more complex, potentially consisting of multiple components that have an impact across different points in multiple pathways.<sup>24,25</sup> As such, there are, in principle, a large number of potential comparators, and establishing the expected impacts of each on the relevant outcomes is analytically and computationally challenging, if not impossible. One potential comparator is the status quo, although this may differ between settings and over time.<sup>25</sup>

### *Time of the Evaluation and Decisions*

We refer to the time of the evaluation in respect to the policy's life cycle, which runs from its conception prior to implementation, through initial roll out, to full implementation, and finally refinement and maturation. Different decisions are required at different time points in its life cycle. Dependent on the timing of the evaluation, there may be different evidential challenges to address. For example, early on, the challenges may stem from limited understanding of the possible outcomes.<sup>1</sup> Later, challenges may stem from lack of data on what the outcomes would be without the policy (counterfactual). The value of a policy may also change over time. For example, if evaluating a policy with large upfront (potentially sunk) costs, these would need to be considered when evaluating whether to introduce it, but following its introduction, these costs may be irrecoverable and do not result in further opportunity costs when considering whether to continue with the policy. The level of decision uncertainty will also vary over time, and decision makers may need to consider whether they require further information to help inform the current decision and whether a decision to implement the policy would prevent further evidence generation or alternatively lead to further learning about its effects.<sup>6</sup>

### *Mechanisms of Action*

The mechanisms of action are the causal chains and processes on which a policy acts to produce its effects. Health systems are complex with multiple interacting constituent parts,<sup>26,27</sup> many of which may be affected by

the introduction of a policy. The various potential mechanisms of action are unlikely to be independent of each other, and the overall impact of a policy will reflect their joint effect, including any interactions or spillovers. To evaluate the overall effect of a policy, it may be necessary to understand how the changes resulting from its introduction affect each constituent part and their interactions, with the failure to do so potentially leading to incorrect estimation of effects. For example, the introduction of co-payments intended to reduce unnecessary care was shown to reduce demand for health care for those patients incurring the co-payments that, if considered in isolation, suggested it had the desired effect.<sup>28</sup> However, further research demonstrated that physicians responded by increasing care and costs to other patients at the same practice, thus increasing unnecessary care.<sup>29</sup>

### *Estimating Effects*

Establishing a causal relationship between a policy and policy-relevant outcomes is necessary to determine its value; however, it is challenging.<sup>1</sup> For example, as noted above, the effects of a policy are potentially dependent on its impact and interaction with numerous constituent parts,<sup>1</sup> or it may not be easy to separate the impacts of the policy from other changes occurring contemporaneously.<sup>4,30</sup> This is further complicated by the effects of the policy evolving over time, for example, due to learning or bedding in processes.<sup>30,31</sup> Health care policies may also have important unintended consequences that will need to be captured.<sup>1,4,25,29,30,32</sup> The above all raise challenges around the appropriate methods to estimate a policy's impacts, including estimation problems associated with causal inference, understanding the processes by which a policy affects outcomes, capturing any changes over time, and having an appropriately broad perspective to capture both intended and unintended impacts.

### *Resources, Constraints, and Opportunity Costs*

For clinical interventions, economic evaluation has used cost-effectiveness thresholds to represent the maximum a health care system might pay for additional health outcomes. The threshold can represent an estimate at the margin of how much health is lost elsewhere for a given amount of expenditure being reallocated (i.e., it represents the opportunity cost).<sup>33</sup> Using this approach, the monetary cost of an intervention can be converted into opportunity costs measured in the same metric as the benefits (i.e., policy-relevant outcomes) to see whether the benefits exceed the opportunity costs. Resource constraints are often depicted as a singular monetary

constraint, for example, the health care budget. However, population- and system-level policies may affect multiple budgets (e.g., primary and secondary care) and other nonfinancial constraints (e.g., staff, capital equipment, capacity), which can result in differential opportunity costs per unit of expenditure depending on the particular limits of each constraint.<sup>13,34–36</sup>

Because of their scale, some policies may have non-marginal impacts, requiring very large amount of resources (e.g., a new policy requiring a significant percentage of the health care budget) and/or inducing a change in the overall production function (e.g., a new technology that significantly affects the overall health produced by the health care system) such that estimates of the marginal productivity of health care expenditure no longer reflect opportunity costs.<sup>37</sup> Estimating the monetary costs of a policy may itself be challenging, particularly if they are subject to economics of scale (where the marginal resource requirements of a policy decrease as its scale increases) or economies of scope (where marginal resource requirements are dependent on the other services being provided). Finally, the extent of the valuation space may require understanding of opportunity costs on a wide set of outcomes and that an estimate of opportunity cost related to health and health care resources may not be sufficient.<sup>38</sup>

#### *Fidelity, Adaptation, and Level of Implementation*

A health care policy's impacts reflect the way in which it is applied and may be altered by factors such as the fidelity of its application, the intensity of effort to apply it, the comprehensiveness with which it is applied, the level of uptake, and the scale of implementation.<sup>30,39,40</sup> These may all differ over the life cycle of the policy. For example, experience from the early stages of implementation of a policy is frequently used to inform implementation in the later stages, policies are not always clearly defined and may evolve over time, and staff may adapt their behavior to them.<sup>4</sup> Evaluations should strive to take these factors into consideration when assessing a policy.

#### *Generalizability and External Validity*

Generalizability refers to the utility of evidence outside the time or setting in which it was generated. Program evaluation methods that are highly focused on establishing the internal validity of the evaluation are tied to the setting in which the observations were made. However, for decision making, external validity is vitally important, particularly when decision makers are interested in applying policies to alternative settings.<sup>41</sup> The effects of policies are likely to vary across settings and over time.

For example, one might expect financial incentives for improving care to have different impacts in the United Kingdom as compared with the United States, given differences in the remuneration structure for clinicians. Careful consideration needs to be given to whether the evidence produced by an evaluation is applicable to the decision being addressed.

### **What Are the Methods Available to Address These Issues?**

Many, if not all, of the challenges outlined are not unique to population- or system-level interventions and arise also in the evaluation of other interventions. Different evaluations will be affected by these challenges to differing degrees, and there are no universal approaches or methods to address all of them. In this section, we highlight some potential approaches and methods available to tackle these challenges and provide some examples in which these approaches have been applied as signposts for readers who wish to see further details.

#### *What Outcomes to Capture and How Should They Be Valued? Defining the Valuation Space*

The choice of outcomes to capture represents a value judgment about the key issues of consequence, and this judgment should reflect the views of the relevant decision makers, not the analyst.<sup>13</sup> When considering which outcomes to include, one potential approach is to determine a set of outcomes and to estimate the impact on each of them.<sup>13</sup> When multiple outcomes are considered, an aggregation function (a further value judgment) will need to be imposed to determine if a policy is beneficial overall if there are winners and losers.<sup>13</sup> A cost-benefit analysis or social return on investment approach would aggregate outcomes by attaching monetary values, often derived from individuals' preferences, to each.<sup>11</sup> Multicriteria decision analysis can also be used to help reach consensus on the impacts on multiple outcomes; however, consideration needs to be given to consistency across decisions and whether opportunity costs are appropriately considered.<sup>42–44</sup> An alternative approach is to define a single (universal) outcome measure; examples include approaches such as the QALY, capability approaches, or extended QALYs.<sup>45–48</sup> However, such an approach imposes a value judgment that all issues of consequence and tradeoffs between them are appropriately captured within the measure.

If a broad set of outcomes are relevant for decision making, the approach of capturing the impacts on multiple outcomes may be preferable. Further, by capturing the outcomes individually, alternative value judgments

around the appropriate methods of aggregation can be considered. Recent examples, looking at policies for alcohol use disorder and air pollution policies, have shown how the impacts on multiple outcomes can be estimated, and different value judgments in their aggregation considered.<sup>13,49–51</sup>

### *What Decision Are We Trying to Inform?*

Determining the decision options is likely to involve significant scoping work and interactions with the decision makers and other relevant stakeholders to determine the range of options available.<sup>30</sup>

Deferring a decision on the implementation of a policy to conduct further research to inform a later decision may be possible.<sup>5,52</sup> The ability of a decision maker to consider further research will depend on his or her responsibilities and the nature of the policy being considered.<sup>5,6,53</sup> Any evaluation will be undertaken in the face of uncertainty, and reducing this uncertainty generates benefits in terms of better decisions.<sup>54</sup> Value-of-information methods consider the value of further research to reduce decision uncertainty. Research into the adoption of new interventions expanded decision options away from only “approve” or “reject” to include options such as “only in research,” in which interventions are funded only for those included in research studies or coverage with evidence development, whereby an intervention is funded for all while research is conducted.<sup>5,6,52</sup> These approaches could be used to inform policy decisions, for example, before a policy’s introduction, a program of further research including only a pilot or partial roll out may be feasible to generate additional evidence.<sup>5,6</sup>

### *Vehicles for Evaluation*

A key challenge in evaluating policies is establishing their effects on relevant outcomes. This may require bringing together evidence from a range of sources. There is a continuum of approaches for evaluation, ranging from empirical analysis of 1 or more data sets to decision analytic modeling approaches that synthesize evidence from multiple sources.

Empirical approaches rely on having data available, which includes measurements that can be used to establish the impacts of a policy over an appropriate time horizon. The analysis by nature will be a retrospective assessment of the value of the policy over that period in that setting. As such, careful consideration needs to be given to the relevance of the evidence for any subsequent decisions within that or a different setting. In addition, there may be several data sets, each of which has

potential relevance to a decision problem and should be reflected in the evaluation.

Decision analytic modeling brings together evidence from the range of available sources to estimate the impacts of policies,<sup>11,55</sup> often in unison with statistical synthesis methods.<sup>56,57</sup> Modeling approaches to valuing policies can be broadly categorized in 2 ways: 1) linking short-term impacts of the policies from empirical analyses to longer-term outcomes modeled by the impact on individual(s) and 2) modeling the impacts on the systems and individuals jointly. The first of these involves no explicit modeling of the impact of the policy on different constituent parts of the system, instead taking an estimate of the overall impact of these system effects on a surrogate outcome (e.g., on mortality, length of stay) from empirical or other studies and extrapolating the change in the surrogate to the overall impacts on individuals (e.g., QALYs). Recent evaluations using such an approach include assessment of payment for performance mechanisms with mortality impacts linked to lifetime health outcomes and costs.<sup>58–60</sup> The second approach models the system and its constituent parts. Approaches under this range from mathematical programming looking at aggregating the independent impacts on the costs and outcomes resulting from different constituent parts (e.g., treatments received), to complex system modeling reflecting the interactions between multiple constituent parts.<sup>34,61–63</sup> The mathematical programming approach to the allocation of health care resources has been used to consider budgetary policies, health system strengthening, and investment in resource constraints.<sup>34,35,63–66</sup> Dynamic simulation modeling, such as system dynamic modeling and agent-based modeling, have been used to simulate the impacts of health care policies on health care systems and their multiple constituent parts.<sup>67,68</sup> A recent review identified 39 studies that have used such approaches to examine the impacts of health care policies on targets such as overstretched resources, length of stay, and undesirable patient outcomes.<sup>68</sup>

### *What Are the Expected Mechanisms of Action of the Policy?*

A key challenge in evaluation is understanding the way a policy is expected to affect the relevant outcomes. Theory-based approaches can be used to build this understanding by defining the hypothesized causal chains.<sup>69,70</sup> Approaches such as mechanism mapping, in which potential mechanism-context interactions are identified, or group model building, in which a diverse set of practitioners collaborate to develop causal models, can help to establish the expected effects and to provide

suitable context-specific adaptations.<sup>71,72</sup> Realist evaluation can also play a role, focusing on the context-specific nature of mechanisms of action to understand “what works, for whom, under what circumstances.”<sup>73,74</sup> Artificial intelligence methods may also have a role in generating information on the mechanisms of action.<sup>75</sup>

### *Estimating Effects*

Establishing a causal relationship between a policy and the outcomes is necessary to determine its value. The estimate of the impact could be on a singular overall measure of effect, which could be a key driver of value or relate to the impacts across different constituent parts of the system. Before a policy is introduced, evidence on its impacts may have to be taken from the literature or other settings, with careful consideration given to its transferability to the context being considered. Methods are available to help adapt estimates in a transparent way (e.g., midrange theory<sup>76</sup>). If no relevant empirical evidence exists, methods such as structured expert elicitation are available.<sup>77</sup>

If a program of research is planned to be conducted as part of the roll out of a policy to inform a future decision, randomized controlled trials (RCTs) or similar approaches may be feasible.<sup>53</sup> Pandya et al.<sup>60</sup> provided a recent example of the use of an RCT to evaluate financial incentives to manage cholesterol levels. However, RCTs have been subject to criticism for the evaluation of policies, notably regarding limits on their external validity and whether the causal effects estimated in the experimental setting generalize to “routine” contexts.<sup>78–80</sup> If a randomized comparison is not feasible, an organized program of data collection using a study design that provides an opportunity for rigorous causal inference methods could be considered, although such methods are also subject to issues of external validity.<sup>81</sup> It may be desirable to improve the external validity by considering the results of different causal inference evaluations relevant to different contexts using an analytical modeling approach. Choice of study design can be informed by an explicit consideration of the value of the evidence generated using value-of-information methods.<sup>82,83</sup>

For analyses following the introduction of a policy, the use of quasi-experimental approaches to establishing causality has been a key focus of the program evaluation literature.<sup>3</sup> These rely on the ability to identify a suitable control against which to judge the policy. A recent example used causal inference methods (synthetic controls) to estimate a mortality impact and a decision analytic model to extrapolate the results to estimate the cost-effectiveness of the UK Quality and Outcomes framework.<sup>59</sup>

### *Generating Evidence on Costs and Opportunity Costs*

To establish value for money of a health policy, it is essential to estimate the opportunity costs. Challenges arising from multiple budgets, nonfinancial constraints, and nonmarginal impacts do not require significant deviations from standard cost-effectiveness methods. For multiple budgets or nonfinancial constraints, estimates of the opportunity costs can be captured through marginal productivities for the outcome(s) of interest for each budget or constraint.<sup>13,33,84,85</sup> An evaluation of air pollution strategies considered budgets in the National Health Service (NHS), public health, and social care, whereas nonfinancial constraints have been considered in evaluations of eye care services in Zambia and viral load testing in sub-Saharan Africa.<sup>34,35,54</sup> When there are nonmarginal impacts, evidence is required on the scale of the impact and the change in the opportunity costs per unit of expenditure.<sup>37</sup> Lomas et al. considered the implications of nonmarginal budget impacts on value for money of new treatments for hepatitis C, which had been estimated to cost the NHS more than £700m per annum (0.7% of total NHS budget).<sup>37,86</sup>

The challenges raised here around financial costs being used to estimate opportunity costs also raise issues with the use of cost-benefit analysis methods (whereby the benefits measured in monetary terms are directly compared with the monetary costs),<sup>11</sup> which typically assume that opportunity cost is equal to monetary cost. Recent research has considered the extension of cost-benefit analysis approaches for evaluating public expenditure by considering the marginal value of public funds, which is equal to the ratio of beneficiaries’ willingness to pay and the net cost to the public sector.<sup>87</sup>

### *Reflecting System Complexity*

Where there are significant interactions within the system, estimating financial cost implications and the effectiveness of policies in isolation may not be sufficient for establishing value for money. Previous work has shown how extensions to the mathematical programming approach on which cost-effectiveness analysis is based can be used to evaluate different types of system-strengthening interventions.<sup>63,66</sup> For example, Morton et al.<sup>66</sup> considered the optimal spend on system strengthening, which increased the effectiveness of treatments in an HIV prevention program. However, these approaches are informationally expensive, requiring estimates of the costs and effects of all the independent interventions in the system and the impact that the policies have on each of these. Alternatively, others have highlighted the need to develop



whole-system models to evaluate policy outcomes and opportunity costs. The Thanzi La Onse project is developing an individual-based comprehensive whole system and an all-disease model of the Malawian health care system, which will be capable of examining the impacts with and without population- and system-level health policies.<sup>62</sup>

### Considerations for the Analyst

In this section, we have signposted a range of approaches of which we are aware to tackle challenges in evaluating population- and system-level health policies. Many of the approaches are not mutually exclusive and can be used in combination, and the choice of methods requires considering their pros and cons in light of the specific evaluation being undertaken. Further, the choice of methods may be limited by the agency of the decision maker and analyst and by the structural circumstances of the evaluation. For example, where the decision maker lacks agency in research, producing information on the value of further research may be unwarranted. Evaluations are not costless, and analysts will have to use their scientific judgment to select approaches that are both appropriate and feasible within the resources available. Regardless of the approach taken, the analyst should engage with the decision makers involved to ensure the evaluation will provide appropriate evidence.

### Conclusions


When evaluating health care policies, it is essential to consider both their effects and their opportunity costs to establish whether they represent value for money. However, program evaluation of population- and system-level health policies has often focused on estimating causal effects on short-term surrogate outcomes. These evaluations are of limited value for decision making as they fail to reflect the policy-relevant outcomes and disregards opportunity costs. This article has aimed to show how the evaluations of such policies could better inform decision making by defining the evidence required to inform policy choices in terms of the impacts on the policy-relevant outcomes of interest, the costs and associated opportunity costs, and the magnitude and consequences of uncertainty. The article also identified key challenges described in evaluating population- and system-level policies and examined methods that can address those challenges. We would advocate that further method development is necessary but that a multidisciplinary approach bringing together health economics and adjacent fields such as epidemiology and mathematical


modeling will improve the evaluation of population- and system-level policies.


### Acknowledgments

We would like to acknowledge the members of our Expert Advisory Group for providing suggestions for the articles on which our targeted review was based and for providing comments on the article. We would also like to acknowledge Anna Heard (Poverty Action) for providing comments on the article and Angela Bates (Northumbria University) and members of the audience for discussing an earlier draft of the article at the Health Economics Study Group in Newcastle in 2020.

### ORCID iDs

Simon Walker  <https://orcid.org/0000-0002-5750-3691>

Aimee Fox  <https://orcid.org/0000-0001-6944-7554>

Nils Gutacker  <https://orcid.org/0000-0002-2833-0621>

### Supplemental Material

Supplementary material for this article is available on the *Medical Decision Making* website at <http://journals.sagepub.com/home/mdm>.

### References

1. Meacock R. Methods for the economic evaluation of changes to the organisation and delivery of health services: principal challenges and recommendations. *Health Econ Policy Law*. 2019;1–16. doi:10.1017/S1744133118000063
2. Gertler PJ, Martinez S, Premand P, et al. *Impact Evaluation in Practice*. 2nd ed. Washington (DC): Inter-American Development Bank and World Bank; 2016. doi:10.1596/978-1-4648-0779-4
3. Hernán MA, Robins JM. *Causal Inference: What If*. Boca Raton (FL): Chapman & Hall/CRC; 2019.
4. Sutton M, Garfield-Birkbeck S, Martin G, et al. Economic analysis of service and delivery interventions in health care. *Health Serv Deliv Res*. 2018;6:1–16. doi:10.3310/hsdr06050
5. Walker S, Sculpher M, Claxton K, et al. Coverage with evidence development, only in research, risk sharing, or patient access scheme? A framework for coverage decisions. *Value Health*. 2012;15:570–9. doi:10.1016/j.jval.2011.12.013
6. Claxton K, Palmer S, Longworth L, et al. A comprehensive algorithm for approval of health technologies with, without, or only in research: the key principles for informing coverage decisions. *Value Health*. 2016;19:885–91. doi:10.1016/j.jval.2016.03.2003
7. Kreif N, Mirelman AJ, Kim S, et al. From impact evaluation to decision-analysis: assessing the extent and quality of evidence on ‘value for money’ in health impact evaluations in low- and middle-income countries. *Gates Open Res*. 2021;5:1.

8. Torbica A. HTA around the world: broadening our understanding of cross-country differences. *Value Health*. 2020;23:1–2. doi:10.1016/j.jval.2019.12.001
9. Sculpher MJ, Claxton K, Drummond M, et al. Whither trial-based economic evaluation for health care decision making? *Health Econ*. 2006;15:677–87. doi:10.1002/hec.1093
10. Phillips A, Shroufi A, Vojnov L, et al. Sustainable HIV treatment in Africa through viral-load-informed differentiated care. *Nature*. 2015;528:S68–76. doi:10.1038/nature16046
11. Drummond M, Sculpher MJ, Claxton K, et al. *Methods for the Economic Evaluation of Health Care Programmes*. Oxford (UK): OUP; 2015.
12. HM Treasury. *The Green Book Appraisal and Evaluation in Central Government Treasury Guidance*. 2011. Available from: [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/220541/green\\_book\\_complete.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/220541/green_book_complete.pdf). Accessed March 28, 2017.
13. Walker S, Griffin S, Asaria M, et al. Striving for a societal perspective: a framework for economic evaluations when costs and effects fall on multiple sectors and decision makers. *Appl Health Econ Health Policy*. 2019;17(5):577–90. doi:10.1007/s40258-019-00481-8
14. Bonell C, Jamal F, Melendez-Torres GJ, et al. ‘Dark logic’: theorising the harmful consequences of public health interventions. *J Epidemiol Community Health*. 2015;69:95–8. doi:10.1136/jech-2014-204671
15. Claxton K, Griffin S, Koffijberg H, et al. How to estimate the health benefits of additional research and changing clinical practice. *BMJ*. 2015;351:h5987. doi:10.1136/bmj.h5987
16. Frew E, Breheny K. Methods for public health economic evaluation: a Delphi survey of decision makers in English and Welsh local government. *Health Econ*. 2019;28:1052–63. doi:10.1002/hec.3916
17. Frew E, Breheny K. Health economics methods for public health resource allocation: a qualitative interview study of decision makers from an English local authority. *Health Econ Policy Law*. 2020;15:128–40. doi:10.1017/S174413311800052X
18. Wildman J, Wildman JM. Combining health and outcomes beyond health in complex evaluations of complex interventions: suggestions for economic evaluation. *Value Health*. 2019;22:511–7. doi:10.1016/j.jval.2019.01.002
19. Kutzin J. Health financing for universal coverage and health system performance: concepts and implications for policy. *Bull World Health Organ*. 2013;91:602–11. doi:10.2471/BLT.12.113985
20. Asaria M, Griffin S, Cookson R. Distributional cost-effectiveness analysis: a tutorial. *Med Decis Making*. 2016;36(1):8–19. doi:10.1177/0272989X15583266
21. Asaria M, Griffin S, Cookson R, et al. Distributional cost-effectiveness analysis of health care programmes—a methodological case study of the UK Bowel Cancer Screening Programme. *Health Econ*. 2015;24:742–54. doi:10.1002/hec.3058
22. Robson M, Asaria M, Cookson R, et al. Eliciting the level of health inequality aversion in England. *Health Econ*. 2017;26:1328–34. doi:10.1002/hec.3430
23. McNamara S, Holmes J, Stevely AK, et al. How averse are the UK general public to inequalities in health between socioeconomic groups? A systematic review. *Eur J Health Econ*. 2020;21(2):275–85. doi:10.1007/s10198-019-01126-2
24. Deidda M, Geue C, Kreif N, et al. A framework for conducting economic evaluations alongside natural experiments. *Soc Sci Med*. 2019;220:353–61. doi:10.1016/j.socsci.med.2018.11.032
25. Tsiachristas A, Stein KV, Evers S, et al. Performing economic evaluation of integrated care: highway to hell or stairway to heaven? *Int J Integr Care*. 2016;16:3. doi:10.5334/ijic.2472
26. Smith PC, Yip W. The economics of health system design. *Oxford Rev Econ Policy*. 2016;32:21–40. doi:10.1093/oxrep/grv018
27. World Health Organization. Everybody’s business—strengthening health systems to improve health outcomes: WHO’s framework for action. Available from: <https://apps.who.int/iris/handle/10665/43918>
28. Keeler EB. *Effects of Cost Sharing on Use of Medical Services and Health*. Santa Monica (CA): RAND Corp; 1992. Available from: <https://www.rand.org/pubs/reprints/RP1114.html>. Accessed October 18, 2020.
29. Fahs MC. Physician response to the United Mine workers’ cost-sharing program: the other side of the coin. *Health Serv Res*. 1992;27:25–45.
30. Raine R, Fitzpatrick R, Barratt H, et al. Challenges, solutions and future directions in the evaluation of service innovations in health care and public health. *Health Serv Deliv Res*. 2016;4:1–136. doi:10.3310/hsdr04160
31. Hanson K, Ranson MK, Oliveira-Cruz V, et al. Expanding access to priority health interventions: a framework for understanding the constraints to scaling-up. *J Int Dev*. 2003;15:1–14. doi:10.1002/jid.963
32. Craig P, Dieppe P, Macintyre S, et al. Developing and evaluating complex interventions: The new Medical Research Council guidance. *BMJ*. 2008;337:979–83. doi:10.1136/bmj.a1655
33. Claxton K, Martin S, Soares M, et al. Methods for the estimation of the National Institute for Health and Care Excellence cost-effectiveness threshold. *Health Technol Assess*. 2015;19(14):1–503, v–vi. doi:10.3310/hta19140
34. van Baal P, Morton A, Severens JL. Health care input constraints and cost effectiveness analysis decision rules. *Soc Sci Med*. 2018;200:59–64. doi:10.1016/J.SOCSCIMED.2018.01.026
35. Revill P, Walker S, Cambiano V, et al. Reflecting the real value of health care resources in modelling and cost-effectiveness studies—the example of viral load informed differentiated care. *PLoS One*. 2018;13:e0190283. doi:10.1371/journal.pone.0190283
36. Vassall A, Mangham-Jeffries L, Gomez GB, et al. Incorporating demand and supply constraints into economic

- evaluations in low-income and middle-income countries. *Health Econ.* 2016;25:95–115. doi:10.1002/hec.3306
37. Lomas J, Claxton K, Martin S, et al. Resolving the “cost-effective but unaffordable” paradox: estimating the health opportunity costs of nonmarginal budget impacts. *Value Health.* 2018;21(3):266–75. doi:10.1016/J.JVAL.2017.10.006
  38. Walker S, Gutacker N, Sculpher M. A scoping review on the production of different aspects of quality of health care. Policy Research Unit in Economic Evaluation of Health and Care Interventions. Universities of Sheffield and York. EEP RU Research Report 053, February 2017.
  39. Hargreaves JRM, Goodman C, Davey C, et al. Measuring implementation strength: lessons from the evaluation of public health strategies in low- and middle-income settings. *Health Policy Plan.* 2016;31:860–7. doi:10.1093/heapol/czw001
  40. Damschroder LJ, Aron DC, Keith RE, et al. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. *Implement Sci.* 2009;4:50. doi:10.1186/1748-5908-4-50
  41. Lant P, Justin S. Context matters for size: why external validity claims and development practice do not mix. *J Glob Dev.* 2014;4:161–97.
  42. Munda G, Albrecht D, Becker W, et al. The use of quantitative methods in the policy cycle. In: In: Šucha V, Sienkiewicz M, eds. *Science for Policy Handbook*. Philadelphia: Elsevier; 2020:206–22. doi:10.1016/B978-0-12-822596-7.00018-8
  43. Thokala P, Duenas A. Multiple criteria decision analysis for health technology assessment. *Value Health.* 2012;15:1172–81. doi:10.1016/J.JVAL.2012.06.015
  44. Marsh KD, Sculpher M, Caro JJ, et al. The use of MCDA in HTA: great potential, but more effort needed. *Value Health.* 2018;21:394–7. doi:10.1016/j.jval.2017.10.001
  45. Al-Janabi H, Peters TJ, Brazier J, et al. An investigation of the construct validity of the ICECAP-A capability measure. *Qual Life Res.* 2013;22:1831–40. doi:10.1007/s11136-012-0293-5
  46. Flynn TN, Huynh E, Peters TJ, et al. Scoring the Icecap-a capability instrument: estimation of a UK general population tariff. *Health Econ.* 2015;24:258–69. doi:10.1002/hec.3014
  47. University of Sheffield. About the Project. Extending the QALY. Available from: <https://scharr.dept.shef.ac.uk/e-qaly/about-the-project/>. Accessed August 3, 2020.
  48. Broadway R, Bruce N. *Welfare Economics*. Oxford (UK): Basil Blackwell; 1984.
  49. Ramponi F, Walker S, Griffin S, et al. Cost-effectiveness analysis of public health interventions with impacts on health and criminal justice: an applied cross-sectoral analysis of an alcohol misuse intervention *Health Econ.* 2021;30(5):972–88.
  50. Neumann PJ, Sanders GD, Russell LB, et al. *Cost Effectiveness in Health and Medicine*. Oxford (UK): Oxford University Press; 2016.
  51. Griffin S, Walker S, Sculpher M. Distributional cost effectiveness analysis of West Yorkshire low emission zone policies. *Health Econ.* 2020;29:567–79. doi:10.1002/hec.4003
  52. McKenna C, Soares M, Claxton K, et al. Unifying research and reimbursement decisions: case studies demonstrating the sequence of assessment and judgments required. *Value Health.* 2015;18:865–75. doi:10.1016/j.jval.2015.05.003
  53. Venkataramani AS, Underhill K, Volpp KG. Moving toward evidence-based policy: the value of randomization for program and policy implementation. *JAMA.* 2020;323(1):21–2. doi:10.1001/jama.2019.18061
  54. Claxton K. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. *J Health Econ.* 1999;18:341–64. doi:10.1016/S0167-6296(98)00039-3
  55. Briggs A, Claxton K, Sculpher M. *Decision Modelling for Health Economic Evaluation*. Oxford (UK): OUP; 2006.
  56. Welton NJ, Sutton AJ, Cooper NJ, et al. *Evidence Synthesis for Decision Making in Healthcare*. Chichester (UK): John Wiley & Sons; 2012. doi:10.1002/9781119942986
  57. Dias S, Ades AE, Welton NJ, et al. *Network Meta-Analysis for Decision Making*. Chichester (UK): John Wiley & Sons; 2018. doi:10.1002/9781118951651
  58. Meacock R, Sutton M, Kristensen SR, et al. Using survival analysis to improve estimates of life year gains in policy evaluations. *Med Decis Making.* 2017;37:415–26. doi:10.1177/0272989X16654444
  59. Pandya A, Doran T, Zhu J, et al. Modelling the cost-effectiveness of pay-for-performance in primary care in the UK. *BMC Med.* 2018;16:135. doi:10.1186/s12916-018-1126-3
  60. Pandya A, Asch DA, Volpp KG, et al. Cost-effectiveness of financial incentives for patients and physicians to manage low-density lipoprotein cholesterol levels. *JAMA Netw Open.* 2018;1:e182008. doi:10.1001/jamanetworkopen.2018.2008
  61. Mielczarek B, Uziako-Mydlkowska J. Application of computer simulation modeling in the health care sector: a survey. *Simulation.* 2012;88:197–216. doi:10.1177/0037549710387802
  62. Thanzi la Onse. Epidemiology and modelling. Available from: <https://thanzi.org/research/epidemiology-and-modeling/>. Accessed November 22, 2019.
  63. Hauck K, Morton A, Chalkidou K, et al. How can we evaluate the cost-effectiveness of health system strengthening? A typology and illustrations. *Soc Sci Med.* 2019;220:141–9. doi:10.1016/j.socscimed.2018.10.030
  64. Stinnett AA, Paltiel AD. Mathematical programming for the efficient allocation of health care resources. *J Health Econ.* 1996;15:641–53. doi:10.1016/S0167-6296(96)00493-6
  65. Epstein DM, Chalabi Z, Claxton K, et al. Efficiency, equity, and budgetary policies: informing decisions using mathematical programming. *Med Decis Making.* 2007;27:128–37. doi:10.1177/0272989X06297396
  66. Morton A, Thomas R, Smith PC. Decision rules for allocation of finances to health systems strengthening. *J Health Econ.* 2016;49:97–108. doi:10.1016/J.JHEALECO.2016.06.001

67. Marshall DA, Burgos-Liz L, Ijzerman MJ, et al. Applying dynamic simulation modeling methods in health care delivery research: the SIMULATE checklist: report of the ISPOR simulation modeling emerging good practices task force. *Value Health*. 2015;18:5–16. doi:10.1016/j.jval.2014.12.001
68. Cassidy R, Singh NS, Schiratti PR, et al. Mathematical modelling for health systems research: a systematic review of system dynamics and agent-based models. *BMC Health Serv Res*. 2019;19:845. doi:10.1186/s12913-019-4627-7
69. De Silva MJ, Breuer E, Lee L, et al. Theory of change: a theory-driven approach to enhance the Medical Research Council's framework for complex interventions. *Trials*. 2014;15:267. doi:10.1186/1745-6215-15-267
70. Weiss CH. How can theory-based evaluation make greater headway? *Eval Rev*. 1997;21:501–24. doi:10.1177/0193841X9702100405
71. Williams MJ. External validity and policy adaptation: from impact evaluation to policy design. *World Bank Res Obs*. 2020;35(2):158–91. doi:10.1093/wbro/lky010
72. Siokou C, Morgan R, Shiell A. Group model building: a participatory approach to understanding and acting on systems. *Public Heal Res Pract*. 2014;25. doi:10.17061/phrp2511404
73. Pawson R, Tilley N. *Realistic Evaluation*. Thousand Oaks (CA): SAGE; 1997. Available from: <https://uk.sagepub.com/en-gb/eur/realistic-evaluation/book205276>. Accessed June 24, 2020.
74. Moore GF, Audrey S, Barker M, et al. Process evaluation of complex interventions: Medical Research Council guidance. *BMJ*. 2015;350:h1258. doi:10.1136/bmj.h1258
75. Michie S, Thomas J, Mac Aonghusa P, et al. The Human Behaviour-Change Project: an artificial intelligence system to answer questions about changing behaviour. *Wellcome Open Res*. 2020;5:122. doi:10.12688/wellcomeopenres.15900.1
76. Vaessen J, Pawson R. Middle range theory and program theory evaluation: from provenance to practice 1. In: Vaessen J, Leeuw FL, eds. *Mind the Gap*. London: Routledge; 2018. p 171–202. doi:10.4324/9781315124537-11
77. Soares MO, Sharples L, Morton A, et al. Experiences of structured elicitation for model-based cost-effectiveness analyses. *Value Health*. 2018;21:715–23. doi:10.1016/j.jval.2018.01.019
78. Deaton A, Cartwright N. Understanding and misunderstanding randomized controlled trials. *Soc Sci Med*. 2018;210:2–21. doi:10.1016/j.socscimed.2017.12.005
79. Chelwa G, Muller S. The poverty of poor economics. Available from: <https://africasacountry.com/2019/10/the-poverty-of-poor-economics>. Accessed June 24, 2020.
80. Muller SM. Causal interaction and external validity: obstacles to the policy relevance of randomized evaluations. *World Bank Econ Rev*. 2015;29:217–25. doi:10.1093/wber/lhv027
81. Deaton A. Instruments, randomization, and learning about development. *J Econ Lit*. 2010;48:424–55. doi:10.1257/jel.48.2.424
82. Fenwick E, Steuten L, Knies S, et al. Value of information analysis for research decisions—an introduction: report 1 of the ISPOR Value of Information Analysis Emerging Good Practices Task Force. *Value Health*. 2020;23:139–50. doi:10.1016/j.jval.2020.01.001
83. Rothery C, Strong M, Koffijberg HE, et al. Value of information analytical methods: report 2 of the ISPOR Value of Information Analysis Emerging Good Practices Task Force. *Value Health*. 2020;23:277–286. doi:10.1016/j.jval.2020.01.004
84. Martin S, Lomas J, Claxton K. Is an ounce of prevention worth a pound of cure? A cross-sectional study of the impact of English public health grant on mortality and morbidity. *BMJ Open*. 2020;10:e036411. doi:10.1136/bmjopen-2019-036411
85. Longo F, Claxton K, Lomas J, et al. Does public long-term care expenditure improve care-related quality of life in England? 2020. Available from: [www.york.ac.uk/che](http://www.york.ac.uk/che). Accessed January 20, 2021.
86. Hawkes N. NICE approval of new hepatitis drug could result in £700m bill for NHS. *BMJ*. 2015;351:h5554. doi:10.1136/bmj.h5554
87. Finkelstein A, Hendren N. Welfare analysis meets causal inference. *J Econ Perspect*. 2020;34:146–67. doi:10.1257/JEP.34.4.146