Routledge
Taylor & Francis Group

Journal of Economic Methodology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rjec20

Medical epistemology meets economics: how (not) to GRADE universal basic income research

Adrian K. Yee & Kenji Hayakawa

To cite this article: Adrian K. Yee & Kenji Hayakawa (2023) Medical epistemology meets economics: how (not) to GRADE universal basic income research, Journal of Economic Methodology, 30:3, 245-264, DOI: 10.1080/1350178X.2023.2231480

To link to this article: https://doi.org/10.1080/1350178X.2023.2231480

Published online: 10 Jul 2023.

JOURNAL OF ECONOMIC METHODOLOGY
2023, VOL. 30, NO. 3, 245–264
https://doi.org/10.1080/1350178X.2023.2231480

Routledge
Taylor & Francis Group

Medical epistemology meets economics: how (not) to GRADE universal basic income research

Adrian K. Yee [a*] and Kenji Hayakawa [b]

[a]Department of Philosophy, Lingnan University, Hong Kong Catastrophic Risk Centre; [b]Independent Scholar & Professional Translator

ABSTRACT

There have recently been novel applications of medical systematic review guidelines to economic policy interventions which contain controversial methodological assumptions that require further scrutiny. A landmark 2017 Cochrane review of unconditional cash transfer (UCT) studies, based on the Grading of Recommendations Assessment, Development and Evaluation (GRADE), exemplifies both the possibilities and limitations of applying medical systematic review guidelines to UCT and universal basic income (UBI) studies. Recognizing the need to upgrade GRADE to incorporate the differences between medical and policy interventions, the GRADE Public Health Project Group (PHPG) was convened to enumerate and address these methodological challenges. However, in light of our analysis of additional methodological challenges that arise for UCT and UBI studies, we argue that the adaptation of medical systematic review guidelines to economic methodology is far from straightforward and is in fact more challenging than claimed by the PHPG.

ARTICLE HISTORY
Received 16 December 2020
Accepted 27 June 2023

KEYWORDS
Philosophy of economics; basic income studies; evidence hierarchies; development economics; experiments in economics; systematic reviews

Introduction

In the context of evidence-based policy making (EBPM), the evaluation and implementation of economic policy interventions typically rely upon disparate sources of evidence, given sampling and funding constraints, a diversity of stakeholder values, and a large variance of potential social, medical, and political outcomes. In medical epistemology, proponents of evidence-based medicine (EBM) have developed and implemented systematic review guidelines to solve the issue of evidential heterogeneity and achieve transparency and consensus regarding evidential quality for a given medical intervention. Given its success in EBM, novel attempts have been made recently in development economics to produce systematic reviews based on medical systematic review guidelines. These issues become particularly salient when applying such guidelines to the systematic review of unconditional cash transfer (UCT) and universal basic income (UBI)[1] studies. UBI is commonly defined as a universal, periodic cash transfer, often advocated as a means to eradicate poverty, enhance population well-being, and promote a form of economic egalitarianism. Since there is a clear lack of transparency and consensus regarding numerous predicted outcomes of UCT and UBI interventions in the current literature,[2] it might be tempting to attempt a straightforward application of medical systematic review guidelines to overcome this lack. However, such an application raises numerous methodological challenges at the intersection of medicine and economics.

CONTACT Adrian K. Yee adriankyle.yee@mail.utoronto.ca
*Both authors contributed equally to the contents of this paper.
This article was originally published with errors, which have now been corrected in the online version. Please see Correction (http://dx.doi.org/10.1080/1350178X.2023.2240573)

Among extant evidence hierarchies (EHs) and systematic review guidelines, the Grading of Recommendations Assessment, Development and Evaluation (GRADE) tool has been used widely in EBM and public health policy. Over 110 organizations from at least 19 countries use GRADE, including the World Health Organization (WHO), the *British Medical Journal*, and the American Academy of Family Physicians, among others. Evidence from these studies comes from a diverse array of investigative strategies[3] (e.g. randomized controlled trials (RCTs), cohort studies, case studies, field experiments, etc.) that cut across several disciplines (e.g. medicine, economics, sociology, psychology, etc.). Given its prestige and widespread use in medicine, GRADE has been seen by some as a useful tool for systematic review authors studying public policy interventions.

Such naive optimism, however, becomes called into question when considering the possibility of applying GRADE to UCT and UBI studies. As we will demonstrate throughout this essay, UBI in particular offers an edge case for GRADE that brings out many significant methodological challenges at the intersection of medical epistemology and EBPM. On the one hand, UCTs, whether non-universal or universal and periodic (i.e. UBI), have been construed by scholars as public health interventions with concrete health and well-being outcomes (Cf. Chen & Quiñonez, 2018; Dupas & Miguel, 2017). On the other hand, unlike other public health interventions, long-term UCTs and UBI possess a range of unique properties stemming from the generality of cash as a form of intervention. Those properties give rise to important differences between UCTs/UBI and other public health interventions. We catalog and elaborate on those differences in the course of our argument to demonstrate that GRADE requires extensive amendment than has previously been supposed if it were to become an adequate systematic review guideline for UBI research.

While no systematic review of UBI studies exists that follows an established guideline, Pega et al. (2017)'s Cochrane review of long-term UCT studies is a close proxy which showcases many of the possibilities and limitations of applying GRADE to economic policy interventions. Subsequently, GRADE's Public Health Group has convened a Public Health Project Group (PHPG) tasked with presenting and addressing a comprehensive list of challenges facing review authors using GRADE for public health policy interventions broadly construed. On the one hand, we argue that these methodological challenges are relevant to UBI research, yet many of them have been underplayed or neglected by UBI researchers. On the other hand, we also demonstrate in detail how the PHPG's list of challenges do not go far enough to encompass the full range of issues that systematic review authors examining UCT and UBI studies will face. Our paper explicates and then builds upon the PHPG's five-point analysis by raising important additional challenges in the case of UCT and UBI studies that, in our view, the PHPG fails to elaborate sufficiently. In doing so, we also question the optimism implied in both the PHPG's attempt to modify GRADE for economic policy interventions as well as UBI researchers' less-sophisticated attempts to draw bold conclusions from inadequate evidence and poorly conceived systematic reviews. We conclude with some suggestions for how extant UBI and UCT methodology can be improved in light of these additional challenges we pose.

1. GRADE's application in UCT studies

For at least the past 20 years, EHs have been used in EBM as a tool for supporting the evaluation of 'investigative strategies': the diverse sources of evidence regarding the expected outcomes of medical interventions. Prominent evidence hierarchies in EBM have included the Scottish Intercollegiate Guidelines Network (SIGN) (Scottish Intercollegiate Guidelines Network, 2011), the Consolidated Standards of Reporting Trials (CONSORT) (Moher et al., 2010; Montgomery et al., 2018), and GRADE (Siemieniuk & Guyatt, 2017). The use of EHs and systematic review guidelines in EBM has been heavily criticized by philosophers of science, especially philosophers of medicine,[4] leading to productive analogous discussions in EBPM (Cf. Cartwright & Hardie, 2012; Shadish et al., 2002).

What distinguishes GRADE from other systematic review guidelines is a feature which pertains to the evaluation of RCTs vis-a-vis other investigative strategies. An uncritical acceptance of RCTs as the epistemic 'gold standard' has been shown to be at odds with case studies of medical trials, where

results obtained from RCTs were later overridden by insights gained from other, allegedly 'inferior' investigative strategies (Cf. Cartwright & Deaton, 2018; Frieden, 2017). Unlike other evidence hierarchies such as SIGN, CONSORT, and the Maryland Scientific Methods Scale (Farrington et al., 2002), GRADE does not give excessive privilege to RCTs over observational studies, irrespective of the details of their design. Rather, GRADE merely ranks RCTs *provisionally* higher than observational studies, but allows for subsequent adjustment according to criteria other than risk of bias. This helps to avoid unwarranted and unconditional high regard for RCTs over other investigative strategies.

The second distinguishing feature of GRADE pertains to a less widely discussed but important topic, namely, of measuring the inconsistency among heterogeneous studies. Unlike most other extant guidelines and EHs, GRADE provides a concrete procedure by which the inconsistency among different studies should be measured using the I^2 statistic. I^2 ranges from 0% to 100% and is defined as $I^2 = \left(\frac{(Q-df)}{Q}\right) \times 100$ (interpreted as a percentage), where Q is the chi-squared (χ^2) statistic and df is the degrees of freedom (Deeks et al., 2021, §10.10.2). I^2 may also be understood less technically as the ratio of the population variance of some observed effect to the observed variance of a sample (Borenstein et al., 2017, p. 7). Among other benefits, this metric allows GRADE users to formulate recommendations that account for the heterogeneity of studies which might otherwise be overlooked.

We now proceed with a detailed exposition of GRADE in action by summarizing the only application to date of GRADE to the systematic review of long-term UCT studies: a Cochrane study conducted by Pega et al. (2017). Relevant features of both GRADE and the Cochrane review of UCT studies will be highlighted in the course of the summary, which will lay out the context for the discussions of additional challenges we pose in sections 2–6.

Pega et al. (2017) provide a synoptic, highly detailed systematic review of papers from 17 electronic databases encompassing recent research on UCTs and their impact on health outcomes in low- and middle-income countries (LMICs). This review aggregated data from 21 studies of UCTs in Africa, the Americas, and South-East Asia, and consisted of 16 cluster-RCTs, 4 controlled before-and-after studies (CBAs), and a single cohort study, involving 1,092,877 participants in total (of which 36,068 were children) (p. 2). More specifically, these studies' designs included 'parallel group and cluster-randomised controlled trials (RCTs), quasi-RCTs, cohort and controlled before-and-after studies (CBAs), and interrupted time series studies of UCT interventions in children (0–17 years) and adults (18 years or older) in LMICs' (p. 1). Their stated focus is to understand the ways in which UCTs for reducing poverty and vulnerabilities might mitigate a diversity of public health issues.

GRADE requires review authors to define the intervention. Pega et al. define a UCT as a cash transfer that satisfies the following conditions (Pega et al., 2017, p. 14):

- It must be disbursed as cash and not as credit or a voucher
- It must be unconditional
- It cannot be based on previous social insurance payments (i.e. is non-contributory)
- It must be provided by the government or a research group
- It must be provided periodically within a given time frame (i.e. is not a one-off payment)

Having defined a UCT,[5] studies were chosen which examined cash transfers that adhered to the above definition.

Following GRADE, Pega et al. proceed to enumerate the primary and secondary outcomes to be examined by their review. Primary outcomes include changes in health service use, stunting, illnesses, food security, rates of depression, parenting quality, and healthcare expenditures (p. 15). Most studies, funded by either governments, international organizations, or research groups,

compared a treatment group administered with a UCT with a control group which received either a smaller UCT or no UCT. The stated cash value of the UCTs varied from study to study, ranging between 1.5% and 53.9% of the annual GDP per capita of select countries. Various effects were measured and common statistics were computed, such as risk factors for treatment versus control groups, risk ratios with 95% confidence intervals, and, most importantly, a grade for the 'quality of evidence' for each outcome in accordance with GRADE criteria.

To determine the quality of evidence, GRADE asks reviewers to make judgments about surveyed studies along eight criteria: (1) consistency of effect estimates between studies; (2) imprecision (e.g. the width of confidence intervals); (3) indirectness (i.e. whether the study is directly or indirectly relevant to the outcome of interest); (4) publication bias; (5) limitations of study design; (6) magnitude of effect; (7) residual confounders; (8) dose–response gradient. Users of GRADE are further instructed to consider reasons to downgrade a study using criteria (1) – (5) and to upgrade a study using criteria (6) – (8) (Schünemann et al., 2013, §5.1). After completing this procedure, GRADE users are expected to assign a final grade to the quality of evidence for each outcome. Those grades are either 'high', 'moderate', 'low', or 'very low'. For instance, 'high' means '[w]e are very confident that the true effect lies close to that of the estimate of the effect' and 'very low' means '[t]he true effect is likely to be substantially different from the estimate of [the] effect' (Schünemann et al., 2013, §5).

Pega et al. (2017) generally followed this procedure, listing each examined study in detail under each outcome. While the *GRADE Handbook* has a specific section dedicated to assessing the risks of bias, Pega et al. have opted to go beyond and construct their own taxonomy of bias. Specifically, Pega et al. (2017) introduced a matrix with a set of three-point criteria (positive, minus, or question mark) outlining risks of bias along several dimensions, including whether a randomization procedure was used, whether allocation was concealed from researchers, whether baseline characteristics were similar, and whether contamination had occurred (i.e. whether there was spillover from treatment to control group) (p. 27). These original criteria depart from GRADE's recommendations, which state that risk of bias should only lower the grade of a study if that study is deemed 'high quality' (in Pega et al.'s case, none of the studies surveyed were deemed to be above 'moderate' quality) and that downgrading should only occur if most studies surveyed are biased (Schünemann et al., 2013, §5.2.1).

As the above summary demonstrates, in order to fit their review process to the structure of GRADE, Pega et al. had to make several adjustments to the standard GRADE guideline, such as their introduction of custom-made criteria for evaluating the risks of bias of the UCT studies. This suggests that the application of GRADE to cash transfer studies may present additional challenges to researchers compared to its application in medical research. Indeed, a subsequently published GRADE PHPG concept paper attempts to examine and elaborate these methodological challenges for public health interventions in a more systematic way (Hilton Boon et al., 2021).[6] These challenges were grouped into five categories and presented as follows (Hilton Boon et al., 2021, p. 48):

(1) incorporating diverse perspectives
(2) selecting and prioritizing outcomes
(3) interpreting outcomes and identifying a threshold for decision-making
(4) assessing certainty of evidence from diverse sources, including non-randomized studies
(5) addressing implications for decision makers, including concerns about conditional recommendations

However, the PHPG concept paper was very brief, presenting only the bare outlines of these challenges and their solutions without sufficient elaboration, argumentation, or enough concrete examples from the UCT and UBI literature. In the following five sections, we present additional challenges which the PHPG paper fails to elaborate sufficiently and yet are crucial to keep in mind when trying to design systematic review guidelines for evaluating evidence pertaining to UCT and UBI interventions that straddle multiple disciplines.

2. Incorporating and excluding fluctuating perspectives

The first challenge the PHPG outlines is the incorporation of diverse perspectives into the review. As they note, the key here is to find a way to balance the concerns of medical stakeholders with those of non-medical stakeholders (Hilton Boon et al., 2021, p. 49). The PHPG further notes that stakeholders who read and use the systematic reviews may work within different 'cultures of evidence' and that '[t]he way that such varied audiences frame policy questions, the extent to which these audiences value health protection and improvement, and the priorities placed on various policy approaches may all differ substantially' (p. 49).

We add to this discussion in noting that Pega et al. (2017) do not explicitly explain whether, and to what extent, their review incorporates relevant perspectives, particularly of policymakers in LMICs who may be interested in UCTs as a form of public health intervention. However, the review authors did give prominence to institutional sources such as the United Nations (UN), the WHO, and the World Bank when defining a UCT as an intervention intended to reduce poverty and vulnerabilities (pp. 10–12). Partly as a consequence of this definition, 32 studies were excluded from the review 'because they did not examine an eligible UCT for reducing poverty and vulnerabilities' (p. 26). Notably, studies which focus on the efficiency of UCTs against other forms of intervention were excluded because 'no eligible outcome [was] studied' (pp. 99–100). For example, one of the excluded studies, namely Skoufias et al. (2013), compares the economic outcomes of in-kind and cash transfers in relation to Mexico's Programa de Apoyo Alimentario and finds that a cash transfer at 75% of the market value of the in-kind transfer had a poverty impact identical to that of the in-kind transfer (pp. 407–408). This type of study might be of interest to stakeholders and policymakers in LMICs who are interested in the effect of UCTs on the efficiency of resource use to fight poverty, yet the review excluded such studies and thus became less relevant to those interested in this aspect of the issue. Similar exclusions can be seen in the initial stages of systematic reviews in UBI research, as when the Irish government commissioned a review exclusively to the Low Pay Commission (Houses of the Oireachtas, 2021) and therefore neglected a wide range of other stakeholders (Social Justice Ireland, 2021). Given the broader diversity of stakeholders involved in cash transfer interventions, such exclusions may, in many cases, be inevitable in order to give a manageable scope to the review. Yet there is still a lack of sufficient *justification* for these exclusions among extant UCT and UBI reviews, which is a point that is not mentioned by the PHPG.

Another challenge of GRADE for UCT and UBI research, which the PHPG does not mention, is to account for wider fluctuations in these perspectives over time. Pega et al. (2017) trace the evolution of international perspectives on the intended outcomes of UCTs (pp. 10–11). Recent studies suggest that public opinion regarding UBI may vary according to world events such as the COVID-19 pandemic (Nettle et al., 2021), changing demographics (Vlandas, 2021), and participation in pilots and experiments (Simanainen & Kangas, 2020). Many of these fluctuations are unpredictable, yet there is arguably a need for a strategy to hedge against such events in order to preserve the long-term relevance of systematic reviews, such that these reviews may continue to inform policy even after such perspective-altering events occur. Relatedly, compared to medical interventions, the 'culture of evidence' in UCT and UBI research is highly politically charged, a difference which is not sufficiently noted in the PHPG paper. Failing to address this cultural difference may further undermine the utility of systematic reviews over time. For example, Parth and Nyby (2020) point out that welfare experiments (including UBI experiments) rarely inform policy decisions directly. Instead, these experiments tend to 'regulate discourses, shape ideological party disputes and influence agenda-setting opportunities' (p. 2). An analysis of the tweets of Finnish MPs during and after the Finnish experiment found that, while the experiment itself contributed to an increase in UBI-related tweets, Finnish MPs' opinions and perspectives on UBI remained relatively unchanged across party lines and demographics even after the preliminary empirical findings were released (Parth & Nyby, 2020). In the original 'theory-agnostic' spirit of EBPM (Favereau & Nagatsu, 2020, pp. 191–192), systematic review authors in UBI research need to find a way to respond to the

ossification of perspectival differences, so that the review may actually contribute to successful consensus-building.

3. Selecting and prioritizing causally opaque outcomes

The next challenge identified by the PHPG is that of selecting and prioritizing the outcomes to be examined in the review. In Hilton Boon et al. (2021), Stefan Lhachimi, co-author of the Cochrane review on UCTs discussed in section 1, reports that 'the Summary of Findings table went through several iterations during the review process before agreement was reached on which and how many outcomes to report' (p. 45). According to the PHPG (including Lhachimi), one major source of this difficulty was the fact that the primary outcomes studied in Pega et al. (2017) were mostly treated as secondary in the RCTs and other studies that were reviewed. Therefore, evidence was very limited on most of Pega et al.'s primary outcomes.[7] The PHPG further highlighted the challenge of balancing population health with health equity in the outcome selection process (Hilton Boon et al., 2021, pp. 49–50), although Pega et al. (2017) were able to generally overcome this challenge by following the PROGRESS-Plus guidelines.[8]

While this is a welcome start to a discussion concerning the selection and prioritization of outcomes, we argue that many additional related challenges arise for UBI research. Firstly, review authors might find it difficult to decide whether a given outcome is an outcome of UBI in the first place.[9] Unlike other, more targeted interventions, the causal pathways of UBI are inherently various and opaque.

For example, Pega et al. (2017) present Figure 1 and describe five pathways: '1. Direct consumption effects (pathway A-B-C). 2. Direct status effects (pathway A-D-E). 3. Combined consumption and status effects (pathway A-B-F-E). 4. Employment effects (pathway A-G-H). 5. Reduced financial risk (arrow I)' (p. 12). Even at this simplified level, four out of five pathways require two or more intermediate steps to be linked causally from intervention to effect, which reduces the likelihood of a successful identification and empirical observation of the full causal pathway in UBI experiments.

Moreover, we believe that such assumptions regarding causal pathways need to be stated and defended in more detail in the case of cash transfers, where socio-economic differences between communities may likely alter the way interventions work (Cf. Khosrowi, 2019). Each of the four mediating components in Figure 1 requires justification with reference to evidence, and it is the strength and character of that evidence that determines the extent to which such components may be active in different social settings (Parkhurst & Abeysinghe, 2016). For example, pathway A-B-C is supported

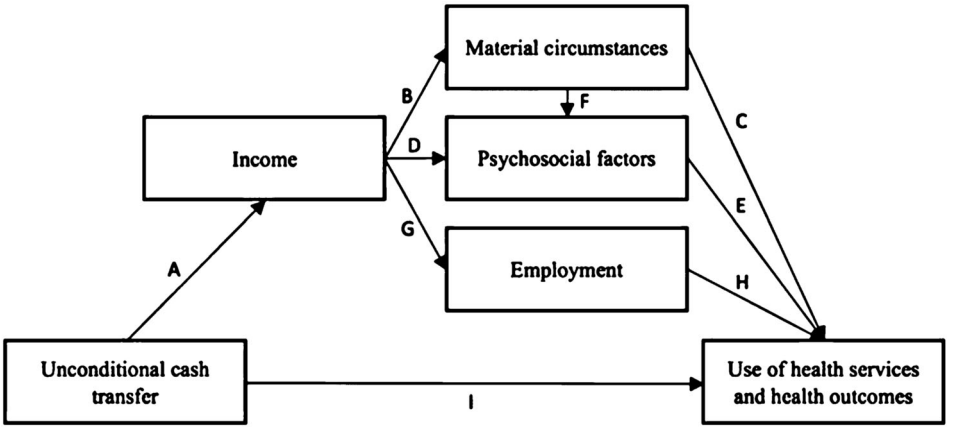

Figure 1. Conceptual framework of the causal relationship between an unconditional cash transfer for reducing poverty and vulnerabilities and the use of health services and health outcomes (Pega et al., 2017, p. 12).

by evidence specifically from Nordic countries (Lundberg et al., 2010), which suggests that this pathway may not be active in social settings that are significantly different from Nordic countries (e.g. LMICs). This approach dispenses with the idea of establishing a univocal causal relationship and instead suggests that causal pluralism ought to be considered the default mode of reasoning about the mechanisms of UCTs.

Secondly, due to the sheer variety of outcomes as well as the aforementioned challenge of establishing reliable accounts of causal pathways, the selection of outcomes in UBI research is especially prone to the 'streetlight effect'. The streetlight effect obtains when research focuses on outcomes that are easier to study but less important, rather than on outcomes that are more important but harder to study (Widerquist, 2018, p. 32). Outcomes whose causal pathways are relatively well-established, and outcomes which are relatively easier to study or are more frequently examined in extant studies, could risk being blindly prioritized over other, equally important outcomes. For instance, UBI advocates have argued that UBI is an optimal way to harness the value created from technological innovation (Cf. Bregman, 2017; Varoufakis, 2020; Yang, 2018). Yet no empirical study exists to date that studies this outcome, mainly because it is challenging to design an experiment for it compared to other outcomes such as employment. In addition to the streetlight effect, political bias can also distort the outcome selection process in ways that are not common in the case of medical systematic reviews. For example, a typical selection pattern among pro-UBI researchers is to include employment and health effects in the same study. This serves to generate a positive image of UBI insofar as 'minor' declines in employment can be presented as being 'outweighed' by 'major' improvements in physical and psychological health (Cf. Allegri & Foschi, 2020, pp. 12–13; German Institute of Economic Research, 2020, p. 24; p. 47; McDowell & Ferdosi, 2021, p. 3). These biasing factors at the outcome selection stage are not sufficiently noted by the PHPG.

Thirdly, we argue that due to its generality as an intervention and the high potential magnitude of its effects, UBI touches on a wider range of national and international policy priorities compared to other, more targeted economic interventions. A sample list of UBI-related priorities expressed by governments include mitigating the negative impacts of the COVID-19 pandemic (Gentilini et al., 2021), maximizing the utility of social protection budgets (UNDP China, 2020), enhancing human agency (Government of India, 2017), encouraging all residents to participate in public life (Federal Council of Switzerland, 2015), fostering social inclusion and equality (Department of the Taoiseach, 2020), and, more simply, guaranteeing an income floor for all (Arriba González de Durana & Rodríguez-Cabrero, 2021; Segal et al., 2020). Moreover, although extant studies and reviews focus overwhelmingly on the 'People' goals (Goals 1–6) of the UN Sustainable Development Goals, UBI has been claimed to have an impact on the 'Prosperity' (7–12) and 'Planet' (13–15) goals too, yet discussions of these aspects remain at a theoretical level due to the lack of empirical research (MacNeill & Vibert, 2019). As with the previous challenge in section 2, here again UCT and UBI review authors are faced with the need to provide a sufficient justification for excluding certain outcomes that touch on these diverse policy priorities, a need which is not mentioned by the PHPG.

4. Interpreting outcomes and acknowledging social equipoise

The PHPG raises several challenges regarding the *interpretation* of the selected outcomes in GRADE-based systematic reviews (Hilton Boon et al., 2021, p. 50). The first challenge noted is that ostensibly small changes of outcome at the individual level can lead to significant changes at a population level, such as in a finding from a salt-reduction program in which 'a[n individual] reduction in salt explains between one fifth and one third of [population-level] cases of hypertension' (Webster et al., 2017, p. 570). Relatedly, Pega et al. (2017) note how a comparatively small UCT can lead to large increases in dietary diversity and protein intake (p. 36). Moreover, the PHPG notes that different stakeholders may desire different degrees of granularity in the reporting of outcomes (Hilton Boon et al., 2021, p. 50). This is a valid point in the context of UCT and UBI studies which Pega et al. (2017) do not sufficiently incorporate. Different stakeholders may not require or agree

to a uniform level of granularity and may only be interested in, for example, whether UCTs will lead to a general increase or decrease in health service use. Counterfactually, the same studies reviewed in Pega et al. (2017) might possibly have provided moderate or even high-quality evidence if the outcome was characterized in a more coarse-grained fashion.[10]

Furthermore, citing Petticrew et al. (2013), the PHPG called for greater attention to the problem of 'social equipoise', that is, equipoise for non-health outcomes (Hilton Boon et al., 2021, p. 50). Petticrew et al. (2013) define equipoise as the 'open acknowledgment of uncertainty about effectiveness' which obtains when there exists 'no consensus about the comparative merits of the alternatives to be tested' (p. 1). Particular challenges that arise with social equipoise include the lack of clarity on the full list of alternative interventions, the multiplicity of outcomes, and the unwillingness of stakeholders (particularly politicians) to acknowledge equipoise (ibid, pp. 1–2).

Two additional challenges, not noted by the PHPG, arise in the context of this discussion. Firstly, review authors in UCT and UBI lack a shared standard for amalgamating evidence and translating the findings into decision thresholds. On the one hand, this is not necessarily a vice. Allowing each team of review authors to define their own standards might help make the review relevant to the particular stakeholders they are working with. On the other hand, a lack of standards might encourage reviewers to 'fill in the gap' and translate their findings into decision thresholds in an exaggerated manner. As a notable example, a recent systematic review examines the effect of UBI on labor supply, finds that 'no evidence has been found of most of the negative [labor supply] effects attributed to a UBI' (De Paz-Báñez et al., 2020, p. 18), and concludes that the review 'settles the discussion on the subject' (p. 4). This is a bold claim which fails to account for causal pluralism (as discussed in section 3) and other factors that make evidence amalgamation difficult in the case of UBI studies. The challenge, then, is to set standards that are sufficiently flexible as to encourage review authors to present the idiosyncrasies of individual studies in detail, while also discouraging review authors from making exaggerated claims to comprehensiveness and impartiality.

Secondly, acknowledging social equipoise for UBI requires review authors to account for feedback effects (such as inflation, the Hawthorne effect,[11] and reflexivity) of the intervention, a challenge which is not mentioned by the PHPG. Extant UBI studies tend to fail to account for such effects and thus wrongly assume that certain outcomes follow from the intervention by definition. For example, Widerquist (2018) lists as many as five claims that 'do not need a test' (p. 105). These include the 'poverty claim', which states that 'UBI set at or above the poverty line necessarily eliminates poverty' (p. 106). Moreover, the 'freedom claim' (UBI will reduce recipients' dependency on employers), the 'compensation claim' (UBI transfers resources from the unjustly wealthy to the unjustly poor), and the 'reciprocity claim' (UBI enables recipients to consume products without providing a corresponding amount of labor) are asserted to be 'true by definition' (ibid). In a similar vein, Wells (2019), who won the 2017 *Basic Income Studies* Essay Prize, boldly claims that '[a] global basic income programme that transferred $1 per day from the rich world to each poor person would eliminate extreme poverty directly and at negligible cost' (p. 1) without fully acknowledging the possible feedbacks (such as global-scale fluctuations in the value of currencies) that might occur as a result of such an intervention. Giving a sufficient account of such feedback effects, however, is challenging for UCT and UBI interventions, making it that much more difficult to convince review authors that certain outcomes might not follow as readily as they assume, and that therefore social equipoise ought to be acknowledged.

5. Assessing certainty of evidence concordantly with the intervention

Under this rubric, the PHPG highlights the challenge of finding sources of even moderate-quality evidence when randomized studies are unavailable (Hilton Boon et al., 2021, p. 50). For instance, Pega et al. (2017) report that while randomized trials were widely available (16 out of 21 studies reviewed), due to the nature of the intervention (i.e. a UCT), these RCTs lacked appropriate blinding. Therefore, all RCTs were downgraded according to the Cochrane Risk of Bias Tool's criterion of performance

bias (Pega et al., 2017, pp. 27–28). For this reason, even when randomized studies were available, Pega et al. still faced the challenge of finding sources of moderate-quality evidence. Relatedly, the PHPG further argues that there needs to be a more fine-grained way of evaluating the strengths and weaknesses of different study designs within the category of non-randomized studies (NRSs), noting the need to further develop the tool for assessing the 'risk of bias in non-randomized studies of interventions' (ROBINS-I)[12] (Hilton Boon et al., 2021, p. 50).

Designing a ROBINS-I-like tool for UBI presents additional challenges which are underexplored by the PHPG. For a given outcome, ROBINS-I requires review authors to construct a 'target trial', which is defined as an ideal RCT against which the actual NRSs are evaluated using a checklist of 'bias domains' (Sterne et al., 2016). However, this procedure is arguably problematic for UBI research, since randomization may not be the most important criterion for evaluating UBI studies in the first place. The fact that most of the outcomes of UBI are *community effects* (as opposed to purely individual effects) suggests that the criteria of (1) geographical saturation[13] and (2) scale of the intervention and treatment groups are more important than (3) randomization, in terms of making the empirical findings closer to what might be expected to happen in an actual implementation of UBI.

One proposed solution to this challenge, which the PHPG does not mention, is an alternative methodological schema developed by Widerquist (2018) which prioritizes criterion (1) over (2) and (2) over (3) when evaluating field experiment methods. For Widerquist, UBI's 'target trial' (in the ROBINS-I sense) is what he calls a 'dream test', which is an ideal international-scale cluster RCT with unlimited resources (e.g. time, money, political will) and unanimous participation by all nations in the world. A variety of UBI systems, existing social institutions, and other alternative policies would then be distributed among countries in such a way that researchers can 'test the effect of UBI on a *nation* the way medical researchers test the effect of medicines on individuals' (Widerquist, 2018, p. 131). Using the dream test as the 'target trial', Widerquist follows the above three criteria and rank-orders national tests, cluster RCTs, saturation studies, and individual RCTs. For example, a nation, a cluster, and a saturation site may have the same degree of geographical saturation, thus satisfying (1). By virtue of its geographical scale, the national test is ranked above the other two due to (2), which is then followed by a cluster RCT due to (3), while a saturation study, which satisfies neither (2) nor (3), is ranked third. Individual RCTs are very weak on (1) and are thus ranked below saturation studies (Table 1).

To date, Widerquist's discussion is the only attempt within the UBI research community to provide a systematic evaluation of the different field experiment methods for studying UBI according to a set of common criteria. This schema is an improvement on ROBINS-I insofar as randomization is not blindly privileged, and that Widerquist's criteria (1) and (2) capture features that have higher relevance when translating findings from UBI studies into recommendations and policy decisions.

However, we argue that Widerquist's schema remains problematic insofar as non-field experiments are not included and that it rank-orders experiments solely by the general study design without paying sufficient attention to other features. In this context, two additional sets of challenges arise. The first set of challenges pertains to finding a way to systematically evaluate other investigative strategies, such as laboratory experiments,[14] qualitative methods, and, most especially, computer simulations. Computer simulations using high-quality empirical big data, obtained from a saturation site, would satisfy criteria (1) and (2). However, the reliability of findings from such simulations will also depend on the adequacy of the model. Models that have been used in the UBI literature include EUROMOD (Martinelli, 2020; Sutherland & Figari, 2013), Canada's SPSD/M (Ammar et al., 2020; MacDonald, 2016; Statistics Canada, 2016), PolicyMod (Henderson et al., 2020), customized models tested for overall empirical fit (Clavet et al., 2013), customized models based on experimental findings in the UBI literature (Nikiforos et al., 2017), and cruder equations that are not tested for fit (Hoynes & Rothstein, 2019). Within the UBI literature, critics have pointed out that such models are subject to the Lucas critique (Delsen, 2019, p. 16; Spermann, 2017, p. 4).[15] By comparison, computer simulations using more sophisticated models in other fields have actively informed policymakers, most notably the Integrated Assessment Models (IAMs) for evaluating climate policies

Table 1. Sample categorization of UBI experiments according to Widerquist (2018)'s schema.[a]

Rank	Method	UBI experiments	Quasi-UBI experiments[b]
–	Dream Test	–	–
1	National Test	Human Development Fund (Mongolia)	Oil Dividend (Iran) Amiri Grant (Kuwait)
2	Cluster RCT	GiveDirectly (Kenya) SEWA (Madhya Pradesh) Programa de Apoyo Alimentario Pilot (Mexico)	–
3	Saturation Study	Wealth Partaking Scheme (Macau SAR) Namibia BIG (Otjivero-Omitara) Permanent Fund Dividend (Alaska)	Mincome (Manitoba) Reddito di Cittadinanza (Italy) Telangana (India)
4	Individual RCT	B-Mincome (Barcelona) SEED (Stockton, CA)	OBIP (Ontario) Kela (Finland) Government Pilot (New Delhi) USA NIT Experiments (IN, IA, NJ, NC, Seattle/Denver)
–	Miscellaneous	Cherokee Dividend (USA) ReCivitas (Brazil)	Yusaku Maezawa (Japan) Satoru Mochizuki (Japan) Gyeonggi Pilot (Korea) Freedom Dividend Pilot (USA)

[a]We constructed this table on the basis of the following sources: Human Development Fund (Gentilini et al., 2019), GiveDirectly (GiveDirectly, 2019), SEWA (Davala et al., 2015), Programa de Apoyo Alimentario (Cunha et al., 2019), Wealth Partaking Scheme (Kwong, 2013), Namibia BIG (Haarmann et al., 2009), Alaska Permanent Fund Dividend (Widerquist & Howard, 2012), B-Mincome (Lain, 2019), SEED (Baker et al., 2019, 2021), Cherokee Dividend (Akee et al., 2010), ReCivitas (Augusto & Brancaglione, 2018), Iran Oil Dividend (Gentilini et al., 2019), Amiri Grant (Gentilini et al., 2019), Mincome (Forget, 2011; Hum & Simpson, 1993), Reddito di Cittadinanza (Gentilini et al., 2019), Telangana (Gentilini et al., 2019), OBIP (McDowell & Ferdosi, 2021), Kela (Kangas et al., 2019), Indian Government Pilot (Gentilini et al., 2019), USA NIT Experiments (Gentilini et al., 2019), Gyeonggi Pilot (GRI Basic Income Research Group, 2019), and Yusaku Maezawa (Unayama, 2020). Information for the Satoru Mochizuki Pilot and the Freedom Dividend Pilot were obtained from social media.

[b]Under this heading, we include field experiments which are improperly called 'UBI experiments' in popular discourse. The cash transfers examined in these experiments fail to satisfy the criteria of universality, unconditionality, and periodicity. Mincome, the OBIP, and the USA NIT Experiments were structured as NIT experiments, where for every dollar earned through employment, a certain fixed amount was deducted from the benefit transfer. As already mentioned in the introduction, NITs contain hidden conditionalities and are thus distinct from UBIs (Green et al., 2021; Vanderborght & Van Parijs, 2017, pp. 32–40). The Kela experiment is commonly treated as a UBI experiment, yet as Kangas et al. (2019) report, a closer examination of the experiment setup shows that this is more accurately characterized as a conditional cash transfer experiment.

and calculating the social cost of carbon (Forster et al., 2018, pp. 8–9). We believe that the IAM evaluation literature, for example, gives a detailed list of challenges (Wilson et al., 2021)[16] which may be useful for developing additional guidelines for evaluating the quality of computer simulation models in UBI.

The second set of challenges pertains to accounting for features of individual studies other than their general study design. Relatedly, ROBINS-I-like tools for UBI require additional flexibility to account for how the importance of different types of evidence changes depending on the outcome being studied. For example, when assessing whether a UBI reduces social stigma compared to conditional welfare (Calnitsky, 2016; Roelen, 2020), qualitative evidence arguably plays a central role in providing review authors and decision makers with a realistic social-psychological account. Reviews that focus on other, more quantitative outcomes, such as general equilibrium effects, may see purely quantitative data related to consumer spending, consumer price, investment levels, per capita income, and other such factors to be more important (Cf. Egger et al., 2019). Furthermore, the procedure of evaluating actual studies against a 'target trial' should be abandoned in the case of UBI research. This is because in the case of UBI, studies focusing on different outcomes will frequently end up using the same 'target trial', namely Widerquist's 'dream test'. The differences between the dream test and actual studies will necessarily be too great, leading to most if not all studies being downgraded to 'very low quality', resulting in a failure to make informative assessments. These pitfalls need to be avoided when developing guidelines for evaluating NRSs for UBI research.

6. Formulating recommendations for review users

The last set of challenges presented in the PHPG paper pertains to finding a way 'to reconcile the tension between the methodologically correct presentation of evidence and recommendations as per GRADE and the implications of strong vs. conditional recommendations from the perspective of decision makers in political environments' (Hilton Boon et al., 2021, pp. 50–51).[17] The PHPG states that there are many cases in which there is pressure to give strong recommendations to public health interventions even if the quality of evidence is low, such as in the presence of life-threatening situations and/or unclear benefits but clear threat of harm if the intervention is not introduced (ibid). To illustrate this point with reference to Pega et al. (2017), the review authors claim that they 'did not identify any harms from UCTs' (p. 2), yet it was unclear as to whether situations would exist in which not implementing a UCT might lead to serious harms. As an example emblematic of this ambiguity, the effect of UCTs on 'parenting quality' ranged from 'a meaningful harm to no meaningful change' (pp. 7–8). Likewise, the effect of UCTs on levels of depression also contained 'very serious imprecision' in the form of 'the 95% confidence estimate or estimates ranging from a meaningful benefit to a meaningful harm' (p. 6). Given this situation, Pega et al. were unable to formulate any recommendations regarding these and other primary outcomes.[18]

Contrary to what the PHPG paper suggests, we claim that GRADE-like guidelines for UBI will not be able to reconcile epistemic correctness with the high political demand for exaggerated recommendations. Given the politically charged nature of discourse around UBI, claims resembling a strong recommendation have been made and will continue to be made by influential figures regardless of whether they are sufficiently warranted by evidence. For example, US President Joe Biden has argued that UBI will not enable the unemployed to live a dignified life because it ends up 'selling American workers short', a claim which he translates into a 'strong recommendation' against UBI as a way to combat technological unemployment (Biden, 2017). Greek Member of Parliament Yanis Varoufakis has argued that UBI will increase the bargaining power of workers and thus 'single-handedly' make the exploitation of gig-workers 'impossible', an argument which he converts to a 'strong recommendation' for UBI as a way to fight worker precarity (Varoufakis, 2020, p. 54). While this is partly a problem with the communication of UBI research, unlike medical systematic reviews, UBI studies garner a much broader public interest, thus exposing itself to risks of miscommunication in ways that go far beyond what is anticipated by GRADE and even the PHPG. Therefore, it is *prima facie* unclear as to how systematic review guidelines can provide meaningful guidance to review authors on how to pre-empt and mitigate such issues when formulating recommendations in the case of UBI.

Moreover, due to the variety of active causal pathways in different social settings, it is practically challenging to specify the conditions under which UBI can be recommended as a way to produce a given outcome. Contrary to the PHPG's advocacy of conditional recommendations, this implies that reviews of UBI need to find a way to be policy-relevant without necessarily presenting a generalized recommendation. As one model for integrating evidence from systematic reviews with local settings, Cowen and Cartwright (2019) present a case study of a policing policy intervention, namely alley-gating, in which local and context-specific pieces of evidence play a role that is arguably more important than RCTs in helping policymakers to decide whether alley-gating should be implemented to prevent burglary. Even for this integrated approach to work, however, conclusions produced by UBI experiments still need to have a degree of generalizability to other settings. The challenge is made even more serious by the fact that most extant UBI studies do not contain a section discussing the generalizability, or lack thereof, of their findings to other social contexts (either similar or dissimilar).

Another challenge related to addressing implications for decision makers arises in the debate as to whether Frequentist statistical inference methods are to be preferred over Bayesian decision procedures for enhancing decision-making. This is especially pertinent when studies lack a sufficient sample size to have high confidence in the accuracy of effect estimates. Frequentist inferences in

the context of public health interventions typically require larger sample sizes compared to Bayesian inferences (Fischer et al., 2013, p. 489). Moreover, Johannesson (2020) has recently shown that the Bayesian and the Frequentist cannot, in certain non-trivial cases, even agree upon the assignment of probabilities to the same set of data, forcing anyone employing these statistical paradigms to take a side on the matter. In the context of UBI research, Pega et al. (2017, p. 18) employ the formalism of confidence intervals, a hallmark of Frequentist statistical inference, following the *GRADE Handbook* (Schünemann et al., 2013, §7.5.2).[19] Bayesian statisticians would reject this method, opting instead for using 'credible intervals', which have a different epistemic structure tied to subjective credence rather than the frequency of events. Indeed, Bayesian methods are seldom used in the UBI literature[20] and it is unclear why when sample sizes may not be large enough in some cases to justify extant Frequentist methods. Systematic review guidelines for cash transfer studies ought to address the Frequentist vs. Bayesianism debate and help review authors navigate these methodological issues.

Finally, given a comparative lack of direct empirical studies testing UBI policies, any recommendation for or against a full-scale UBI will need to take into account its fundamental uncertainty. The ethical and political stakes of failing to prepare for unforeseen problems is quite high. For example, based on a survey conducted among participants of the OBIP, 95% of respondents have reported that 'they put their life plans on hold or abandoned them altogether as a result of the cancelation of the pilot' (McDowell & Ferdosi, 2021, p. 32). If the cancelation of a pilot project has already impacted the lives of former recipients, it is reasonable to assume that the cancelation of a full-scale UBI due to unforeseen circumstances would have an even greater negative impact on the population. Whether and to what extent a systematic review employing GRADE would be able to help policymakers prepare for such 'unknown unknowns' remains to be seen.

7. Conclusion

The long-form Cochrane review by Pega et al. as well as the subsequent discussion by the GRADE PHPG may create the impression that GRADE is an adequate foundation for designing a systematic review guideline for reviewing UBI and other cash transfer studies. However, this paper suggests that the adaptation of medical systematic review guidelines to economic methodology is far from straightforward and is in fact more challenging than claimed by the PHPG. While an updated and more nuanced GRADE-like guideline for UBI is forthcoming, methodologists will do well to note several under-discussed factors which we argued for in this paper: the intrinsic diversity of stakeholders, the exclusion of relevant perspectives from their reviews, the potential biases that informed the selection and prioritization of UBI-related outcomes as well as their interpretations, the specific ways in which NRSs, including computer simulations and qualitative studies, were or were not incorporated into the review, and the potential risks of miscommunication that might occur when the findings of the review are presented to the general public.

Solutions to the challenges raised in this paper are not obvious and will vary depending on one's role, expertise, and the specificities of the task at hand. As demonstrated in sections 2 through 6, the GRADE PHPG underestimates or overlooks many of the additional challenges that we raised in the context of designing and implementing an adequate systematic review guideline for UCT and UBI studies. Therefore, those involved in the development of systematic review guidelines, including GRADE, will need to pay closer attention to the details of these challenges. If sufficiently adequate solutions are forthcoming, GRADE and GRADE-like guidelines will need to at least flag the existence of these unresolved challenges so as to inform review authors of the potential limitations of their findings.

Finally, we note that the foregoing discussion does not imply the rejection of adapting medical systematic review guidelines *tout court* for UCT and UBI research. Once their methodological challenges and limitations are taken into account, successful guidelines in medical research such as

GRADE may continue to provide valuable guidance and models for producing high-quality systematic reviews for economic policy interventions.

Notes

1. We define 'UBI' as an unconditional, universal, periodic cash transfer, while noting that there is a lack of consensus on an exact definition in the literature. Note that according to our definition, negative income tax (NIT) policies (Cf. Friedman, 1962 [2002]) are not UBIs due to implicit conditionalities (Green et al., 2021; Vanderborght & Van Parijs, 2017, pp. 32–40). We also slightly depart from the Basic Income Earth Network's (BIEN) definition – the most commonly referenced definition – by dropping the criterion of individuality, since this criterion, in our view, unduly excludes salient household-based studies (e.g. B-Mincome) from the evidence base. Our definition is furthermore consonant with an important recent systematic review of UBI studies conducted by the Stanford Basic Income Lab (Hasdell, 2020).
2. In particular, there is a clear lack of transparency and consensus regarding the *evidence base* for UBI-related claims in a variety of different arenas, including the peer-reviewed literature on UBI (Cf. Delsen, 2019; Widerquist, 2018), popular literature on UBI (Cf. Bregman, 2017; Forget, 2020; Lowrey, 2018; Standing, 2020; Yang, 2018), commentary in scientific journals (Cf. Arnold, 2020; Kariel & Patel, 2021), policy documents (Cf. Ammar et al., 2020; Segal et al., 2020; UNDP China, 2020), and public petitions (European Basic Income Network, 2020). Moreover, evidence hierarchies (EHs) proposed in the context of EBPM tend to be too simplistic (Cf. Farrington et al., 2002; Leigh, 2010, p. 224; Madaleno & Waights, 2015; Pawson, 2006, p. 49) for the purpose of evaluating UCT and UBI policy interventions. We explore the reasons in detail throughout the essay.
3. We follow Jerkert (2021) who argues that the term 'investigative strategy' is the best phrase to employ when describing the highly diverse sources of evidence that medical researchers draw upon. As Jerkert points out, EHs typically list not just study designs and methods but other forms of investigative strategies such as mechanistic reasoning and clinical experience (p. 3). We agree with this holistic perspective and reflect it in our terminology.
4. For example, Jerkert (2021) argues that attempts to rank order evidence have been consistently unclear on what exactly is being ranked, especially when such investigative strategies employ probabilistic evidence. Mercuri et al. (2018) argue that many metrics of epistemic rigor in GRADE are arbitrary and ill-defined. Stegenga (2014) has gone so far as to argue that there is ultimately little justification for EHs at all in medicine and that they should be avoided as a means of assessing the evidence for causal claims. Goldenberg (2006) argues EHs in medicine have typically obscured unavoidably subjective elements in all medical epistemology that are detrimental to sound medical policymaking.
5. The above definition is not only consonant with our definition of UBI but also captures all of the core features of a UCT that distinguishes it from other welfare interventions, while also remaining broad enough to apply to a variety of studies.
6. It is worth noting that Pega et al. (2017) was featured in the PHPG concept paper as one of four case studies in the application of GRADE to public health. Moreover, one of Pega et al. (2017)'s coauthors, Stefan Lhachimi, was also a coauthor of the PHPG paper. Many of the challenges highlighted by the PHPG paper, therefore, represent those that were also faced by GRADE users reviewing UCT studies.
7. For example, as already noted in section 1, the review authors found 'no evidence' of the effect of UCTs on the number of deaths (Pega et al., 2017, p. 34), which is emblematic of the challenge of aligning the review authors' priorities with what is available for review
8. PROGRESS-Plus stands for Place of residence, Race/ethnicity/culture/language, Occupation, Gender/sex, Religion, Education, Socioeconomic status, and Social capital. 'Plus' refers to characteristics associated with discrimination (e.g. age), social relationships or status known to compromise good health (e.g. grew up in a household with parents who smoke), and time-dependent relations that make people temporarily disadvantaged (Cochrane Equity Methods, 2021). Along those lines, Pega et al. found that studies on the following five outcomes reported equity-related measurements that were relevant to PROGRESS-Plus: 'height for age' (negligible effects on inequality reported), 'disease or illness' (not statistically significant), 'food security' (not statistically significant; uncertain evidence), 'dietary diversity' (modest reduction in inequality; uncertain evidence), and 'depression' (negligible effects on inequality reported) (2017, pp. 39–40).
9. In the methodology literature, this problem is often discussed in the context of mechanisms and mechanistic evidence (Cf. Clarke et al., 2014; Illari, 2011). Russo and Williamson (2007) argue that establishing a causal relationship between an intervention and its effect requires both probabilistic and mechanistic evidence. Favereau and Nagatsu (2020) further argue that the lack of generalizability of most RCTs in social policy can be overcome by supplementing the evidence base with reliable sources of mechanistic evidence, such as 'lab-like field experiments', as well as insights from behavioral economics. However, Marchionni and Reijula (2019) point out that the distinction between these two types of evidence (i.e. probabilistic and mechanistic) does not hold for studies of social policy interventions.

10. The counterfactual scenario mentioned here is inspired by Parker (2020, p. 473). Although Parker (2020)'s chief focus is on finding a way to evaluate the adequacy-for-purpose of climate models, similar discussions can be fruitful in the context of UBI studies, since the latter also needs to inform policy while also often failing to achieve a fully accurate representation of the causal pathways from intervention to outcome.
11. Relatedly, Teira (2013) argues that the practical and ethical challenges around designing and implementing a successful blinding and masking procedure may be insurmountable in the case of randomized field experiments. This implies that RCTs in EBPM research have a high chance of violating the non-interference assumption, which states that the outcome of the trial should not be influenced by factors other than the intervention itself (Teira, 2013). Moreover, Sedgwick and Greenwood (2015) point out that the Hawthorne effect may still obtain in double-blinded (or even triple-blinded) trials, in the sense that the results obtained in the study may still not generalize to the target population due to the fact that study participants' behavior might collectively be influenced by the mere knowledge of being part of a trial. These issues become particularly challenging for UBI research, where the potential psychological effects of taking part in the trial of a 'big idea' like UBI may be considerable (Cf. Widerquist, 2018). See also Jiménez-Buedo and Guala (2015) for a detailed discussion of the 'artificiality' of experiments and its potential effects on the generalizability of results.
12. When selecting eligible studies, Pega et al. (2017) referenced the Cochrane Effective Practice and Organisation of Care (EPOC) criteria for CBAs and used self-defined criteria for cohort studies (pp. 13–14). Moreover, when assessing the risk of bias, Pega et al. did not use the ROBINS-I tool, claiming that 'no credible, standardised tool for assessing the risk of bias in cohort studies currently exists' (p. 17) at the time when their review was drafted. Instead, the review authors followed the 2009 Centre for Reviews and Dissemination (CRD) guideline, which was mostly identical to the Cochrane guideline for assessing the risk of bias for RCTs.
13. 'Saturation' refers to the degree to which a given community is 'saturated' by the intervention, with the highest saturation reached when everyone within the relevant community receives the intervention. In the UBI literature, a 'saturation study' refers to a study comparing two communities, where all members of one of the communities receive the intervention (Widerquist, 2018, p. 22).
14. See Jokipalo (2019) and Haigner et al. (2012) for concrete examples. Also see Noguera and Wispelaere (2006) for a discussion on the need for laboratory experiments in UBI research.
15. The relevant version of the Lucas critique states that econometric models ignore the fact that 'when economic policy changes, the structure of the relationships between economic variables shifts due to changes in agents' behavior … This would invalidate [long-term] econometric inferences between past data and the forecasted effects of a new policy' (Goutsmedt et al., 2015, p. 3). As Goutsmedt et al. (2015) point out, there is an ongoing discussion on whether and to what extent dynamic stochastic general equilibrium models adequately overcome the Lucas critique.
16. For example, Weyant (2017) identifies seven challenges related to the improvement of the design and use of IAMs:

    > (1) what to count and how to count it; (2) the inclusion of extreme and discontinuous outcomes; (3) the treatment of regional, national, and international equity; (4) the treatment of intertemporal discounting and intergenerational equity; (5) projections of baseline drivers; (6) capturing interactions between impact sectors and feedbacks to the climate system; and (7) dealing with uncertainty and risk. (p. 124)

    Much of this applies directly to UBI simulations too, especially when the term 'climate system' is replaced by the term 'social system'. Other challenges include decisions on assumptions regarding market rebound effects (Colmenares et al., 2020), monetary theory (Pollitt & Mercure, 2017), and GDP growth (Keyßer & Lenzen, 2021).
17. According to the GRADE Evidence to Decision framework,

    > [a] strong recommendation means the panel is confident that the desirable consequences outweigh the undesirable, or vice versa; a conditional recommendation means the panel is less confident. When panels make a conditional recommendation, they should provide clear guidance regarding the specific conditions that favour implementing or rejecting the option. (Moberg et al., 2018, p. 11)

18. It is worth noting that Pega et al. (2017) nonetheless highlight certain pieces of information that might help decision makers specify the conditions which change the effectiveness of UCTs. For example, the review authors highlight how several of the reviewed studies fail to sufficiently account for the diminishing marginal utility of cash transfers on those who are wealthy compared to those who are poor: 'USD 10 provided to a participant with an annual income of USD 15,000 is not equivalent to USD 10 provided to a participant with an annual income of USD 50,000' (p. 18). Nevertheless, the broader difficulty of formulating conditional recommendations remains.
19. See Higgins et al. (2021), specifically section 10.13, for a discussion of Bayesian meta-analysis techniques in the context of Cochrane. However, the *GRADE Handbook* does not have a section dedicated to the discussion of Bayesian methods.
20. A notable exception is Dorsett (2020), who opts for a Bayesian method to study the effect of the Alaska Permanent Fund Dividend on crime rates in Alaska. More specifically, Dorsett employs counterfactual reasoning in their

linear regression model as a means of simulating what would have happened were the Alaskan dividend not given to recipients. See Widerquist and Howard (2012) for details on the Alaskan dividend program.

Acknowledgements

We thank Anna Alexandrova, Lukas Beck, Joseph Berkovitz, Olesya Bondnarenko, Andrew Jones, Isaac Kean, Christian Philippi-Larroulet, Rebecca Livernois, Soroush Marouzi, Matt Mercuri, Karl Widerquist, and two anonymous referees for constructive feedback on ideas in this paper.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

Adrian acknowledges the Ontario Graduate Scholarship and the University of Toronto FAST Scholarship as sources of funding.

Notes on contributors

Kenji Hayakawa (MA; BA) is a translator, writer, and independent researcher based in Dublin, Ireland. He has translated the works of Noam Chomsky, Dipesh Chakrabarty, Naomi Oreskes, Yoko Tawada, Yanis Varoufakis, and Andrew Yang, among others. He is an expert on James Joyce's Finnegans Wake and the epistemology of basic income, and has also published on social, political, and economic topics related to climate change and the Anthropocene.

Adrian K. Yee (PhD; MA; BA) completed his doctoral studies at the University of Toronto in history & philosophy of science. He is Research Assistant Professor at Lingnan University and faculty member of the Hong Kong Catastrophic Risk Centre. He specializes in the intersection of philosophy of science, philosophy of economics, and political philosophy, with active research interests in artificial intelligence, attention economics, and military intelligence analysis. He has also previously published on misinformation, universal basic income, and the curious historical connections between models in physics and finance.

ORCID

Adrian K. Yee http://orcid.org/0000-0002-4170-9257

References

Akee, R., Copeland, W., Keller, G., Angold, A., & Costello, E. J. (2010). Parents' incomes and children's outcomes: A quasi-experiment using transfer payments from casino profits. *American Economic Journal: Applied Economics*, *2*(1), 86–115. https://doi.org/10.1257/app.2.1.86

Allegri, G., & Foschi, R. (2020). Universal basic income as a promoter of real freedom in a digital future. *World Futures*, *77* (1), 1–22. https://doi.org/10.1080/02604027.2020.1792600

Ammar, N., Busby, C., & Ahmed, S. (2020). *Costing a guaranteed basic income during the COVID pandemic*. Office of the Parliamentary Budget Officer.

Arnold, C. (2020). Pandemic speeds major test of universal basic income. *Nature*, *583*(7817), 502–503. https://doi.org/10.1038/d41586-020-01993-3

Arriba González de Durana, A., & Rodríguez-Cabrero, G. (2021). *New developments in the national guaranteed minimum income scheme in Spain: ESPN flash report 2021/15*. European Social Policy Network (ESPN). European Commission.

Augusto, B., & Brancaglione, M. (2018). Brazil: A basic income experiment as a citizen-to-citizen project. In A. Downes & S. Lansley (Eds.), *It's basic income: The global debate* (pp. 193–198). Policy Press.

Baker, A., Balakrishnan, S., Martin-West, S., Rao, K., & You, G. (2019). *Pre-analysis plan: Stockton economic empowerment demonstration*. Retrieved January 24, 2020, from https://static1.squarespace.com/static/6039d612b17d055cac14070f/t/605029f652a6b53e3dd39044/1615866358804/SEED+Pre-analysis+Plan.pdf.

Baker, A., Coltrera, E., Samra, S., & West, S. (2021). *Preliminary analysis: SEED's first year*. Retrieved May 5, 2021, from https://static1.squarespace.com/static/6039d612b17d055cac14070f/t/6050294a1212aa40fdaf773a/1615866187890/SEED_Preliminary(Analysis-SEEDs(First(Year_Final(Report_Individual(Pages(.pdf

Biden, J. (2017). *Let's choose a future that puts work first*. Retrieved May 25, 2021, from https://www.bidenschool.udel.edu/bideninstitute/research-policy/biden-institute-blog/Let%E2%80%99s-Choose-a-Future-That-Puts-Work-First

Borenstein, M., Higgins, J., Hedges, L., & Rothenstein, L. (2017). Basics of meta-analysis: I^2 is not an absolute measure of heterogeneity. *Research Synthesis Methods*, *8*(1), 5–18. https://doi.org/10.1002/jrsm.1230

Bregman, R. (2017). *Utopia for realists: And how we can get there*. Elizabeth Manton (trans.). Bloomsbury.

Calnitsky, D. (2016). "More normal than welfare": The mincome experiment, stigma, and community experience. *Canadian Review of Sociology*, *53*(1), 26–27. https://doi.org/10.1111/cars.12091

Cartwright, N., & Deaton, A. (2018). Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine*, *210*, 2–21. https://doi.org/10.1016/j.socscimed.2017.12.005

Cartwright, N., & Hardie, J. (2012). *Evidence-based policy: A practical guide to doing it better*. Oxford University Press.

Chen, Y.-A., & Quiñonez, C. (2018). Basic income guarantee: A review of implication for oral health. *Journal of Public Health Dentistry*, *78*(1), 56–62. https://doi.org/10.1111/jphd.12239

Clarke, B., Gillies, D., Illari, P., Russo, F., & Williamson, J. (2014). Mechanisms and the evidence hierarchy. *Topoi*, *33*(2), 339–360. https://doi.org/10.1007/s11245-013-9220-9

Clavet, N., Duclo, J., & Lacroix, G. (2013). Fighting poverty: Assessing the effect of guaranteed minimum income proposals in Québec. *Canadian Public Policy*, *39*(4), 491–516. https://doi.org/10.3138/CPP.39.4.491

Cochrane Equity Methods. (2021). PROGRESS-Plus. *Cochrane Collaboration*. Retrieved May 27, 2021, from https://methods.cochrane.org/equity/projects/evidence-equity/progress-plus

Colmenares, G., Löschel, A., & Madlener, R. (2020). The rebound effect representation in climate and energy models. *Environmental Research Letters*, *15*(123010), 1–35.

Cowen, N., & Cartwright, N. (2019). Street-level theories of change: Adapting the medical model of evidence-based practice for policing. In N. Fielding, K. Bullock, & S. Holdaway (Eds.), *Critical reflections on evidence-based policing* (pp. 52–71). Routledge.

Cunha, J., De Giorgi, G., & Jayachandran, S. (2019). The price effects of cash versus in-kind transfers. *The Review of Economic Studies*, *86*(1), 240–281. https://doi.org/10.1093/restud/rdy018

Davala, S., Jhabvala, R., Standing, G., & Mehta, S. (2015). *Basic income: A transformative policy for India*. Bloomsbury. Retrieved December 10, 2019, from https://www.bloomsburycollections.com/book/basic-income-a-transformative-policy-for-india/

Deeks, J., Higgins, J., & Altman, D. (2021). Chapter 10: Analysing data and undertaking meta-analyses. *Cochrane Handbook for Systematic Reviews of Interventions.* Retrieved May 23, 2021, from https://training.cochrane.org/handbook/current/chapter-10

Delsen, L. (ed.). (2019). *Empirical research on an unconditional basic income in Europe*. Springer.

Department of the Taoiseach. (2020). *Programme for Government: Our shared future*. Retrieved May 28, 2021, from https://www.gov.ie/en/publication/7e05d-programme-for-government-our-shared-future/

De Paz-Báñez, M., Asensio-Coto, M., Sánchez-López, C., & Aceytuno, M.-T. (2020). Is there empirical evidence on how the implementation of a universal basic income (UBI) affects labour supply? A systematic review. *Sustainability*, *12*(22), 9459, 1–36. https://doi.org/10.3390/su12229459

Dorsett, R. (2020). A Bayesian structural time series analysis of the effect of basic income on crime: Evidence from the Alaska permanent fund. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *184*(1), 179–200. https://doi.org/10.1111/rssa.12619

Dupas, P., & Miguel, E. (2017). Impacts and determinants of health levels in low-income countries. *Handbook of Economic Field Experiments*, *2*, 3–93. https://doi.org/10.1016/bs.hefe.2016.09.003

Egger, D., Haushofer, J., Miguel, E., Niehaus, P., & Walker, M. (2019). General equilibrium effects of cash transfers: experimental evidence from Kenya. Working Paper. Retrieved 15, 2020, from https://www.givedirectly.org/wp-content/uploads/2019/11/General-Equilibrium-Effects-of-Cash-Transfers.pdf

European Basic Income Network. (2020). *Petitions for any type of basic income or helicopter money*. Retrieved March 30, 2020, from https://ubi-europe.net/ubi/petitions-for-any-type-of-basicincome-or-helicoptermoney/

Farrington, D., Gottfredson, D., Welsh, B., & Sherman, L. (2002). Maryland scientific methods scale. In L. W. Sherman (Ed.), *Evidence-based crime prevention* (pp. 13–21). Routledge.

Favereau, J., & Nagatsu, M. (2020). Holding back from theory: Limits and methodological alternatives of randomized field experiments in development economics. *Journal of Economic Methodology*, *27*(3), 191–211. https://doi.org/10.1080/1350178X.2020.1717585

Federal Council of Switzerland. (2015). *BBI 2015 9553: Bundesbeschluss über die Volksinitiative «Für ein bedingungsloses Grundeinkommen»*. Retrieved May 28, 2021, from https://www.fedlex.admin.ch/eli/fga/2015/2197/de

Fischer, A., Threlfall, A., Meah, S., Cookson, R., Rutter, H., & Kelly, M. P. (2013). The appraisal of public health interventions: An overview. *Journal of Public Health*, *35*(4), 488–494. https://doi.org/10.1093/pubmed/fdt076

Forget, E. (2011). The town with no poverty: The health effects of a Canadian guaranteed annual income field experiment. *Canadian Public Policy*, *37*(3), 283–305. https://doi.org/10.3138/cpp.37.3.283

Forget, E. (2020). *Basic income for Canadians: From the COVID-19 emergency to financial security for All*. Lorimer.

Forster, P., Huppmann, D., Kriegler, E., Mundaca, L., Smith, C., Rogelj, J., & Séférian, R. (2018). Mitigation pathways compatible with 1.5? in the context of sustainable development, supplementary material. In V. Masson-Delmotte, P.

Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P. R. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J. B. R. Matthews, Y. Chen, X. Zhou, M. I. Gomis, E. Lonnoy, T. Maycock, M. Tignor, & T. Waterfield (Eds.), *Global warming of 1.5?: An IPCC special report on the impacts of global warming of 1.5? above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty*. Retrieved May 26, 2021, from https://www.ipcc.ch/site/assets/uploads/sites/2/2019/05/SR15_Chapter2_SM_Low_Res.pdf

Frieden, T. R. (2017). Evidence for health decision making: Beyond randomized, controlled trials. *The New England Journal of Medicine*, *377*(5), 465–475. https://doi.org/10.1056/NEJMra1614394

Friedman, M. (1962 [2002]). *Capitalism and freedom*. University of Chicago Press.

Gentilini, U., Almenfi, M., Orton, I., & Dale, P. (2021). *Social protection and jobs responses to COVID-19: A real-time review of country measures, "living paper" version 15 (May 14, 2021)*. World Bank. Retrieved May 28, 2021, from https://openknowledge.worldbank.org/handle/10986/33635

Gentilini, U., Grosh, M., Rigolini, I., & Yemtsov, R. (2019). *Exploring universal basic income: A guide to navigating concepts, evidence, and practices*. World Bank Group. Retrieved December 10, 2019, from http://documents.worldbank.org/curated/en/993911574784667955/Exploring-Universal-Basic-Income-A-Guide-to-Navigating-Concepts-Evidence-and-Practices

German Institute of Economic Research. (2020). *Basic income pilot project*. Retrieved May, 2021, from https://www.diw.de/documents/dokumentenarchiv/17/diw_01.c.797177.de/basic_income_pilot_project_magazine.pdf

GiveDirectly. (2019). *Basic income Kenya study*. Retrieved November 21, 2020, from https://www.givedirectly.org/ubi-study/

Goldenberg, M. (2006). On evidence and evidence-based medicine: Lessons from the philosophy of science. *Social Science & Medicine*, *62*(11), 2621–2632. https://doi.org/10.1016/j.socscimed.2005.11.031

Goutsmedt, A., Pinzón-Fuchs, E., Renault, M., & Sergi, F. (2015). Criticizing the Lucas critique: Macroeconometricians' response to Robert Lucas. *Centre d'Economie de la Sorbonne Working Papers*. https://doi.org/10.2139/ssrn.2837766

Government of India. (2017). Universal basic income: A conversation With and within the mahatma. *Economic Survey 2016–17*, 172–212. Retrieved December 10, 2019, from https://www.indiabudget.gov.in/budget2017-2018/es2016-17/echapter.pdf

Green, D., Gutierrez, P., Milligan, K., & Snowberg, E. (2021). Basic income: Characteristics related to presence in and absence from the tax system. Research paper commissioned by the Expert Panel on Basic Income, British Columbia. Retrieved June 1, 2021, from https://econ2017.sites.olt.ubc.ca/files/2021/03/Basic-Income-Characteristics-Related-to-Presence-in-and-Absence-From-the-Tax-System_DGreen_PGutierrez_KMilligan_ESnowberg-0021.pdf

GRI Basic Income Research Group. (2019). *1Q 2019 satisfaction survey report on the youth basic income in gyeonggi province*. Retrieved December 10, 2019, from https://basicincomekorea.org/wp-content/uploads/2019/10/1Q-2019-Satisfaction-Survey-Report-on-Gyeonggi-Youth-Basic-Income.pdf

Haarmann, C., Haarmann, D., Jauch, H., Shindondola-Mote, H., Nattrass, N., van Niekerk, I., & Samson, M. (2009). *Making the difference: The BIG in Namibia: Basic income grant pilot project assessment report*. Basic Income Grant Coalition. Retrieved December 10, 2019, from http://www.bignam.org/Publications/BIG_Assessment_report_08b.pdf

Haigner, S., Höchtl, W., Schneider, F., Wakolbinger, F., & Jenewein, S. (2012). Keep on working: Unconditional basic income in the Lab. *Basic Income Studies*, *7*(1), 1–19. https://doi.org/10.1515/1932-0183.1230

Hasdell, R. (2020). What we know about universal basic income: A cross-synthesis of reviews. *Stanford Basic Income Lab*. Retrieved July 8, 2022, from https://basicincome.stanford.edu/research/papers/what-we-know-about-universal-basic-income/

Henderson, T., Spies-Butcher, B., & Philips, B. (2020). Between universalism and targeting: Exploring policy pathways for an Australian basic income. *The Economic and Labour Relations Review*, *31*(4), 502–523. https://doi.org/10.1177/1035304620964272

Higgins, J., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (2021). *Cochrane Handbook for systematic reviews of interventions, Version 6.2*. Retrieved May 19, 2021, from http://www.training.cochrane.org/handbook

Hilton Boon, M., Thomson, H., Shaw, B., Akl, E. A., Lhachimi, S. K., López-Alcalde, J., Klugar, M., Choi, L., Saz-Parkinson, Z., Mustafa, R. A., Langendam, M. W., Crane, O., Morgan, R. L., Rehfuess, E., Johnston, B. C., Chong, L. Y., Guyatt, G. H., Schünemann, H. J., & Katikireddi, S. V. (2021). Challenges in applying the GRADE approach in public health guidelines and systematic reviews: A concept article from the GRADE public health group. *Journal of Clinical Epidemiology*, *135*, 42–53. https://doi.org/10.1016/j.jclinepi.2021.01.001

Houses of the Oireachtas. (2021). *Question 201, Dáil Éireann Debate, Wednesday – 21 April 2021*. Retrieved May 25, 2021, from https://www.oireachtas.ie/en/debates/question/2021-04-21/201/

Hoynes, H., & Rothstein, J. (2019). *Universal basic income in the US and advanced countries*. National Bureau of Economic Research. https://doi.org/10.3386/w25538.

Hum, D., & Simpson, W. (1993). Whatever happened to Canada's guaranteed income project? *Canadian Public Administration/Administration Publique du Canada*, *36*(3), 442–450. https://doi.org/10.1111/j.1754-7121.1993.tb01963.x

Illari, P. (2011). Mechanistic evidence: Disambiguating the Russo–Williamson thesis. *International Studies in the Philosophy of Science*, *25*(1), 139–157. https://doi.org/10.1080/02698595.2011.574856

Jerkert, J. (2021). On the meaning of medical evidence hierarchies. *Philosophy of Medicine*, *2*(1), 1–21. https://doi.org/10.5195/philmed.2021.31

Jiménez-Buedo, M., & Guala, F. (2015). Artificiality, reactivity, and demand effects in experimental economics. *Philosophy of the Social Sciences*, *46*(1), 3–23. https://doi.org/10.1177/0048393115618015

Johannesson, E. (2020). Classical versus Bayesian statistics. *Philosophy of Science*, *87*(2), 302–318. https://doi.org/10.1086/707588

Jokipalo, V. (2019). Basic income, wages, and productivity: A laboratory experiment. *Basic Income Studies*, *14*(2), 1–13. https://doi.org/10.1515/bis-2019-0016

Kangas, O., Jauhiainen, S., Simanainen, M., & Ylikännö, M. (2019). *The basic income experiment 2017–2018 in Finland: Preliminary results*. Ministry of Social Affairs and Health. Retrieved December 10, 2019, from http://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/161361/Report_The%20Basic%20Income%20Experiment%2020172018%20in%20Finland.pdf

Kariel, J., & Patel, S. (2021). Universal basic income and COVID-19 pandemic. *British Medical Journal*, *372*(193), 1–2.

Keyßer, L., & Lenzen, M. (2021). 1.5°C degrowth scenarios suggest the need for new mitigation pathways. *Nature Communications*, *12*(2676), 1–16. https://doi.org/10.1038/s41467-021-22884-9

Khosrowi, D. (2019). Extrapolation of causal effects: Hopes, assumptions, and the extrapolator's circle. *Journal of Economic Methodology*, *26*(1), 45–58. https://doi.org/10.1080/1350178X.2018.1561078

Kwong, B. (2013). A comparative analysis of the cash handout policy of Hong Kong and Macau. *Journal of Current Chinese Affairs*, *42*(3), 87–100. https://doi.org/10.1177/186810261304200305

Lain, B. (2019). *Report on the preliminary results of the B-MINCOME project (2017–2018)*. Barcelona City Council. Retrieved December 10, 2019, from https://ajuntament.barcelona.cat/dretssocials/sites/default/files/arxius-documents/results_bmincome_eng.pdf

Leigh, A. (2010). Evidence-based policy: Summon the randomistas? In *Strengthening evidence-based policy in the Australian federation, roundtable proceedings* (pp. 215–226). Productivity Commission.

Lowrey, A. (2018). *Give people money: How a universal basic income would End poverty, revolutionize work, and remake the world*. Crown.

Lundberg, O., Fritzell, J., Yngwe, M., & Kölegård, M. (2010). The potential power of social policy programmes: Income redistribution, economic resources and health. *International Journal of Social Welfare*, *19*(s1), S2–S13. https://doi.org/10.1111/j.1468-2397.2010.00727.x

MacDonald, D. (2016). *A policymaker's guide to basic income*. Canadian Center for Policy Alternatives.

MacNeill, T., & Vibert, A. (2019). Universal basic income and the natural environment: Theory and policy. *Basic Income Studies*, *14*(1), 1–15. https://doi.org/10.1515/bis-2018-0026

Madaleno, M., & Waights, S. (2015). *Guide to scoring methods using the Maryland scientific methods scale*. What Works Centre for Local Economic Growth. Retrieved May 13, 2021, from https://whatworksgrowth.org/public/files/Scoring-Guide.pdf

Marchionni, C., & Reijula, S. (2019). What is mechanistic evidence, and why do we need it for evidence-based policy? *Studies in History and Philosophy of Science*, *73*, 54–63. https://doi.org/10.1016/j.shpsa.2018.08.003

Martinelli, L. (2020). A basic income trilemma: Affordability, adequacy, and the advantages of radically simplified welfare. *Journal of Social Policy*, *49*(3), 461–482. https://doi.org/10.1017/S0047279419000424

McDowell, T., & Ferdosi, M. (2021). The impacts of the Ontario basic income pilot: A comparative analysis of the findings from the Hamilton region. *Basic Income Studies*, *16*, 1–48. https://doi.org/10.1515/bis-2020-0034

Mercuri, M., Baigrie, B., & Upshur, R. E. G. (2018). Going from evidence to recommendations: Can GRADE get US there? *Journal of Evaluation and Clinical Practice*, *24*(5), 1232–1239. https://doi.org/10.1111/jep.12857

Moberg, J., Oxman, A., Rosenbaum, S., Schünemann, H. J., Guyatt, G., Flottorp, S., Glenton, C., Lewin, S., Morelli, A., Rada, G., & Alonso-Coello, P. (2018). The GRADE evidence to decision (EtD) framework for health system and public health decisions. *Health Research Policy and Systems*, *16*(45), 1–15. https://doi.org/10.1186/s12961-018-0320-2

Moher, D., Hopewell, S., Schulz, K., Montori, V., Gøtzsche, P. C., Devereaux, P. J., Elbourne, D., Egger, M., & Altman, D. G. (2010). CONSORT 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. *British Medical Journal*, *340*(869), 1–28.

Montgomery, P., Grant, S., Mayo-Wilson, E., Macdonald, G., Michie, S., Hopewell, S., & Moher, D. (2018). Reporting randomised trials of social and psychological interventions: The CONSORT-SPI 2018 extension. *Trials*, *19*(407), 1–14. https://doi.org/10.1186/s13063-018-2733-1

Nettle, D., Johnson, E., Johnson, M., & Saxe, R. (2021). Why has the COVID-19 pandemic increased support for universal basic income? *Humanities and Social Science Communications*, *8*(79), 1–12. https://doi.org/10.1057/s41599-021-00760-7

Nikiforos, M., Steinbaum, M., & Zezza, G. (2017). *Modeling the macroeconomic effects of a universal basic income*. Roosevelt Institute. Retrieved June 3, 2021, from https://rooseveltinstitute.org/publications/macroeconomic-effects-universal-basic-income-ubi/

Noguera, J., & Wispelaere, J. (2006). A plea for the use of laboratory experiments in basic income research. *Basic Income Studies*, *1*(2), 1–8.

Parker, W. (2020). Model evaluation: An adequacy-for-purpose view. *Philosophy of Science*, *87*(3), 457–477. https://doi.org/10.1086/708691

Parkhurst, J., & Abeysinghe, S. (2016). What constitutes "good" evidence for public health and social policy-making? From hierarchies to appropriateness. *Social Epistemology*, *30*(5–6), 665–679. https://doi.org/10.1080/02691728.2016.1172365

Parth, A.-M., & Nyby, J. (2020). Welfare experiments as tools for evidence-based policy making? The political debate on Twitter about the basic income trial in Finland. *Policy Studies*, *43*, 1–19. https://doi.org/10.1080/01442872.2020.1772217

Pawson, R. (2006). *Evidence-based policy: A realist perspective*. Sage.

Pega, F., Liu, S., Walter, S., Pabayo, R., Saith, R., & Lhachimi, S. K. (2017). Unconditional cash transfers for reducing poverty and vulnerabilities: Effect on Use of health services and health outcomes in Low- and middle-income countries. *Cochrane Database of Systematic Reviews*, *11*, 1–137. https://doi.org/10.1002/14651858.CD011135.pub2

Petticrew, M., McKee, M., Lock, K., Green, J., & Philips, G. (2013). In search of social equipoise. *British Medical Journal*, *346* (4016), 1–3.

Pollitt, H., & Mercure, J.-F. (2017). The role of money and the financial sector in energy-economy models used for assessing climate and energy policy. *Climate Policy*, *18*(2), 184–197. https://doi.org/10.1080/14693062.2016.1277685

Roelen, K. (2020). Receiving social assistance in low- and middle-income countries: Negating shame or producing stigma? *Journal of Social Policy*, *49*(4), 705–723. https://doi.org/10.1017/S0047279419000709

Russo, F., & Williamson, J. (2007). Interpreting causality in the health sciences. *International Studies in Philosophy of Science*, *21*(2), 157–170. https://doi.org/10.1080/02698590701498084

Schünemann, H., Brożek, J., Guyatt, G., & Oxman, A. (2013). *GRADE handbook: Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach*. Retrieved May 4, 2021, from https://gdt.gradepro.org/app/handbook/handbook.html

Scottish Intercollegiate Guidelines Network (SIGN). (2011). *Sign 50: A guideline developer's handbook*. Elliott House.

Sedgwick, P., & Greenwood, N. (2015). Understanding the Hawthorne effect. *British Medical Journal*, *351*(4672), 1–2. https://doi.org/10.1136/bmj.h4672

Segal, H., Forget, E., & Banting, K. (2020). *A federal basic income within the post COVID-19 economic recovery plan*. Royal Society of Canada.

Shadish, W., Campbell, D., & Cook, T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.

Siemieniuk, R., & Guyatt, G. (2017). What is GRADE? *British Medical Journal*, Retrieved April 29, 2021, from https://bestpractice.bmj.com/info/us/toolkit/learn-ebm/what-is-grade/

Simanainen, M., & Kangas, O. (2020). Speaking to those who know it best: Does participation in an experiment explain citizens' attitudes to basic income? *Journal of International and Comparative Social Policy*, *36*(3), 269–283. https://doi.org/10.1017/ics.2020.14

Skoufias, E., Unar, M., & Gonzalez de Cossio, T. (2013). The poverty impacts of cash and in-kind transfers: Experimental evidence from rural Mexico. *Journal of Development Effectiveness*, *5*(4), 401–429. https://doi.org/10.1080/19439342.2013.843578

Social Justice Ireland. (2021). *Delivering a Basic Income Pilot: A pathway to a Basic Income Pilot for Ireland*. Retrieved May 25, 2021, from https://www.socialjustice.ie/sites/default/files/attach/publication/6493/2021-05-12deliveringabasicincomepilotforirelanddocx.pdf

Spermann, A. (2017). Basic income in Germany: Proposals for randomised controlled trials using nudges. *Basic Income Studies*, *12*(2), 1–9. https://doi.org/10.1515/bis-2016-0017

Standing, G. (2020). *Battling eight giants: Basic income Now*. Bloomsbury.

Statistics Canada. (2016). *SPSD/M: Product overview*. Retrieved May 10, 2021, from https://www.statcan.gc.ca/eng/microsimulation/spsdm/spsdm/spsd_m-overview.pdf

Stegenga, J. (2014). Down with hierarchies. *Topoi*, *33*(2), 313–322. https://doi.org/10.1007/s11245-013-9189-4

Sterne, J., Hernán, M., Reeves, B., Savović, J., Berkman, N. D., Viswanathan, M., Henry, D., Altman, D. G., Ansari, M. T., Boutron, I., Carpenter, J. R., Chan, A.-W., Churchill, R., Deeks, J. J., Hróbjartsson, A., Kirkham, J., Jüni, P., Loke, Y. K., Pigott, T. D., … Higgins, J. P. (2016). ROBINS-I: A tool for assessing risk of bias in non-randomised studies of interventions. *British Medical Journal*, *355*, i4919. https://doi.org/10.1136/bmj.i4919

Sutherland, H., & Figari, F. (2013). Euromod: The European Union tax-benefit microsimulation model. EUROMOD Working Paper No. EM 8/13.

Teira, D. (2013). Blinding and the non-interference assumption in medical and social trials. *Philosophy of the Social Sciences*, *43*(3), 358–372. https://doi.org/10.1177/0048393113488871

Unayama, T. (2020). *Maezawa basic income social experiment 2020 and basic income social experiment survey*. Retrieved December 15, 2020, from https://www.yusakumaezawa.com/assets/pdf/BasicIncome_Social_Experiment_Survey_en.pdf

UNDP China Office. (2020). *Universal basic income in China*. Retrieved May 4, 2021, from https://www.cn.undp.org/content/china/en/home/library/innovation-/universal-basic-income-in-china.html

Vanderborght, Y., & Van Parijs, P. (2017). *Basic income: A radical proposal for a free society and a sane economy*. Harvard University Press.
Varoufakis, Y. (2020). *Another now: Dispatches from an alternative present*. Bodley Head, Penguin.
Vlandas, T. (2021). The political economy of individual-level support for the basic income in Europe. *Journal of European Social Policy*, *31*(1), 62–77. https://doi.org/10.1177/0958928720923596
Webster, J., Waqanivalu, T., Arcand, J., Trieu, K., Cappuccio, F. P., Appel, L. J., Woodward, M., Campbell, N. R. C., & McLean, R. (2017). Understanding the science that supports population-wide salt reduction programs. *Journal of Clinical Hypertension*, *19*(6), 569–576. https://doi.org/10.1111/jch.12994
Wells, T. (2019). Just end poverty now: The case for a global minimum income. *Basic Income Studies*, *14*(2), 1–13. https://doi.org/10.1515/bis-2019-0009
Weyant, J. (2017). Some contributions of integrated assessment models of global climate change. *Review of Environmental Economics and Policy*, *11*(1), 115–137. https://doi.org/10.1093/reep/rew018
Widerquist, K. (2018). *A critical analysis of basic income experiments for researchers, policymakers, and citizens*. Palgrave MacMillan.
Widerquist, K. & Howard, M. (eds.). (2012). *Alaska's permanent fund dividend: Examining its suitability as a model*. Palgrave Macmillan.
Wilson, C., Guivarch, C., Kriegler, E., van Ruijven, B., van Vuuren, D. P., Krey, V., Schwanitz, V. J., & Thompson, E. L. (2021). Evaluating process-based integrated assessment models of climate change mitigation. *Climatic Change*, *166*(1–2), 3. https://doi.org/10.1007/s10584-021-03099-9
Yang, A. (2018). *The war on normal people*. Hachette Books.