Public Policy Experiments Without Equipoise: When is Randomization Fair?

Douglas MacKay and Emma Cohn

Ethics & Human Research **45 no. 1 (2023): 15-28**

Abstract

Government agencies and nonprofit organizations have increasingly turned to randomized controlled trials (RCTs) to evaluate public policy interventions. Random assignment is widely understood to be fair when there is equipoise; however, some scholars and practitioners argue that random assignment is also permissible when an intervention is reasonably expected to be superior to other trial arms. For example, some argue that random assignment to such an intervention is fair when the intervention is *scarce*, for it is sometimes fair to use a lottery to allocate scarce goods. We investigate the permissibility of randomization in public policy RCTs when there is no equipoise, identifying two sets of conditions under which it is fair to allocate access to a superior intervention via random assignment. We also reject oft-made claims that alternative study designs, including stepped wedge designs and the use of uneven randomization, offer fair ways to allocate beneficial interventions.

Text

Government agencies and nonprofit organizations have increasingly turned to randomized controlled trials (RCTs) to evaluate public policy interventions.[1] Proponents of RCTs defend the use of random assignment on epistemic grounds: randomization promises to cleanly identify the causal impact of an intervention.[2] However, random assignment is also a way

of allocating access to an often-promising intervention, for example, a cash transfer or a housing voucher. For policy RCTs to be ethical therefore, it must be fair to allocate the intervention by random assignment.

Random assignment is widely understood to be fair when there is equipoise, that is, genuine uncertainty within the relevant expert community regarding which trial arm is superior for participants.[3] However, some argue that random assignment is also permissible when an intervention is reasonably expected to be superior to other trial arms. For example, some argue that random assignment to such an intervention is fair when the intervention is *scarce*, for it is sometimes fair to use a lottery to allocate scarce goods.[4] It is thus permissible, on this argument, to randomly assign people to a generous guaranteed income intervention if it is too expensive to roll out to all eligible people and a lottery is a fair way to allocate it.

In this paper, we investigate the permissibility of randomization in public policy RCTs when there is no equipoise, that is, when there is strong evidence that one intervention is superior to the alternatives. Since randomization involves the allocation of an intervention by lottery, we appeal to the philosophical literature on the use of lotteries to fairly allocate scarce goods to determine when random assignment is permissible. We identify two sets of conditions under which it is fair to allocate access to a superior intervention via random assignment. We also examine oft-made claims that alternative study designs, including stepped wedge designs and the use of uneven randomization, offer fair ways to do so. We argue that neither design offers an additional set of conditions under which random assignment is fair.

In part 1, we discuss the phenomena of randomization without equipoise and the arguments proponents offer to defend it. In part 2, we explore the argument that it is permissible

to employ randomization when a beneficial intervention is scarce. In part 3, we turn our attention to stepped wedge designs and studies that employ uneven randomization.

Our analysis is limited to the permissibility of *random assignment*. The two sets of conditions we identify are sufficient for fair randomization, not the permissibility of an RCT, since policy RCTs must satisfy several additional requirements to be permissible, including social value, informed consent, fair participant selection, and community engagement, among others. Unfortunately, a comprehensive discussion of these requirements is beyond the scope of this paper.[5]

1 Policy RCTs Without Equipoise

From December 1997 until January 2001, the New York City-based organization Pathways to Housing ran an RCT to determine the effect of the Housing First model on harm reduction for people with mental illness experiencing homelessness.[6] The trial randomized 225 participants who had a demonstrated history of chronic homelessness and a diagnosed severe mental illness into two groups. The control group entered programs that followed the standard method for addressing homelessness, the Continuum of Care model, which requires clients to complete several steps such as attending psychiatric treatment, maintaining sobriety, and spending time in transitional housing before qualifying for permanent housing. The experimental group entered a program that followed the Housing First model, which provides supportive housing without preconditions.[7] Participants completed personal interviews every six months for a two-year period and were monitored for changes in consumer choice, housing stability, substance use, treatment utilization, and psychiatric symptoms.[8]

Researchers justified the need to compare the relatively novel Housing First model against the more established Continuum of Care model by pointing to a lack of decisive evidence in support of either. As we note above, a lack of decisive evidence is often understood as a justification for the ethical permissibility of random assignment. If the expert community is collectively uncertain regarding which intervention is superior random assignment is permissible for it neither advantages nor disadvantages (ex ante) any participant.[9]

But researchers also conduct RCTs in cases where there is no meaningful uncertainty regarding the superiority of the intervention arm. For example, in the fall of 2010, the Zambian Ministry of Community Development, Mother and Child Health (MCDMCH) started implementing the Child Grant Program (CGP) in three of Zambia's poorest and most remote districts in an effort to determine the effects of cash transfers on alleviating extreme poverty.[10] This program allotted a monthly, unconditional stipend of roughly US$12 to any family with children under five years of age, approximately enough money to purchase one meal a day for an average-sized family for the duration of each payment period.[11] The experiment randomized families in each district into either an experimental group that began receiving payments in January 2011 or into a delayed control group, the latter of which was created because MCDMCH did not have sufficient resources to deliver the cash transfer to all eligible families immediately.[12]

The researchers acknowledged that cash transfers had been proven to promote the primary protective objectives of improving consumption, food security, and school enrollment, but claimed that their impact on secondary objectives, including economic and productive behaviors of households, was uncertain and understudied.[13] Given the proven effectiveness of cash transfers as a protective intervention, the researchers argued that randomization was

"ethically feasible" on the grounds that the MCDMCH did not have sufficient resources or capacity to provide the intervention to all eligible households.[14]

This appeal to *scarcity* to defend random assignment is relatively common, particularly in development economics.[15] Where a good is scarce, the argument goes, a lottery is a fair way to distribute it. Abhijit V. Banerjee and Esther Duflo put the point this way:

> Most experiments, however, are careful to avoid the potential of creating bad feeling due to the randomization. Location-level randomization is justified by budget and administrative capacity, which is precisely why the organizations often agree to randomize at that level. Limited government budgets and diverse actions by many small NGOs mean that villages or schools in most developing countries are used to the fact that some areas receive certain programs whereas others do not, and when an NGO serves only some villages they see it as a part of the organization's overall strategy. When the control areas are given the explanation that the program has enough budget for a certain number of schools only, they typically agree that a lottery is a fair way to allocate those limited resources. They are often used to such arbitrariness, and so randomization appears both transparent and legitimate.[16]

Banerjee and Duflo suggest this line of argument as a way to avoid "bad feeling" due to randomization which may lead to randomization bias. However, they are correct that lotteries are sometimes a fair way to allocate a scarce resource. Randomization may therefore be permissible not only in cases of equipoise, but also when a superior intervention is scarce.

One might object here that there is no need to consider an additional justification for random assignment. If an intervention is truly superior to the status quo, there is no reason to study it and it would be wrong to do so, violating the social value requirement. If there is some

uncertainty regarding its effects, then it is not superior and so equipoise is satisfied. In the Zambian cash grant study, one might argue, the researchers are wrong to claim that randomization is ethical because of scarcity; instead, it is ethical because there is some uncertainty regarding the program's impact. As they note, while there is strong evidence to suggest the transfers will improve recipients' food security, consumption, and school enrollment, there is uncertainty regarding whether it will positively impact recipients' healthcare usage, learning outcomes, or economic activity.[17]

There is a larger question here regarding when equipoise is and is not disturbed, one which is deserving of more consideration than we can give here. But however this question is ultimately answered, it is unreasonable to hold that equipoise is satisfied whenever there is meaningful uncertainty regarding one or more outcomes of an intervention. As such, there will be cases where equipoise is not satisfied, but there is good scientific reason to study an intervention using random assignment.

Public policy interventions have what we might call 'primary' outcomes. These are the outcomes that a policy is designed to realize. Education policies are thus designed to improve student learning and public health policies are designed to secure and promote population health. But policies may also have 'secondary' outcomes. For example, education policies may also improve students' health or promote voting. When it comes to public policy research, it would be wrong, all else being equal, to withhold a feasible intervention proven to realize primary outcomes – i.e. to deny people access to an effective, feasible intervention that is not scarce– to study its effects on secondary outcomes. So, in the Zambian child cash grant study, it would be wrong to withhold the grant from eligible families, given its proven effectiveness in realizing the primary outcomes of consumption, food security, and school enrollment, simply for the purposes

of studying its impact on the study's secondary outcomes, the productive and economic behaviors of families. This is true even though these impacts are worthy of study.

Edward Asiedu et al. thus argue that judgments of whether one intervention is superior to another should only consider the "problem at hand" - i.e. the primary outcomes of the interventions in question.[18] Whether this is right or not, there will be cases, such as the Zambian cash grant RCT where (1) there is no meaningful uncertainty that intervention A is superior to intervention B on primary outcomes, but (2) there is meaningful uncertainty regarding the precise magnitude of this superiority, and/or the impacts of A and B on secondary outcomes. Since an RCT comparing A and B should be understood, in our view, to violate equipoise since there is no meaningful uncertainty regarding A and B on primary outcomes, it is worth considering whether it is fair to conduct the RCT when there is no equipoise. We explore this next, turning to the question of when it is permissible to allocate a scarce good by lottery.

2 Lotteries and Fair Distribution of Scarce Goods

Lotteries are often held up as a fair way to allocate a scarce good, for example, a donated organ, enrollment in a desirable school, or a seat in a lifeboat. After all, they seem to embody the values of equality and impartiality. However, the mere fact that a good is scarce does not mean that a lottery is fair. Some argue, for example, that scarce resources should go to those who are worse off.[19]

We argue here that there are conditions under which it is permissible to allocate scarce goods using a lottery. This conclusion has important implications for the permissibility of RCTs without equipoise since randomization involves the assignment of participants to interventions via lottery. We argue, first, that it is permissible to use random assignment when people have

equal claims to a scarce good. We argue, second, that it is also permissible to use random assignment when people have *no* claims to a scarce good.

2.1 Random Assignment and Equal Claims to a Superior but Scarce Intervention

Lotteries are processes that yield unpredictable outcomes, including coin flips and die tosses. More technically, following Peter Stone, we understand a lottery to be a "process capable of generating a set of outcomes, in which the particular outcome to be expected whenever the process occurs is unpredictable given available information."[20] A "fair lottery" is one where the outcomes are equally probable; a "weighted lottery" is one where the outcomes occur with unequal probabilities.[21] In what follows, we shall call fair lotteries 'lotteries,' and weighted lotteries 'weighted' or 'uneven' lotteries.

Many philosophers agree that lotteries are morally required when a benefit is scarce, indivisible, and subject to equally strong claims.[22] Stone offers a precise formulation of this claim with the "Just Lottery Rule:"

> Under conditions of indeterminacy, if an agent must allocate a scarce homogeneous lumpy good amongst a group of parties with homogeneous claims, then that agent must do so using a fair lottery.[23]

Conditions of indeterminacy hold, Stone suggests, when the parties have equally strong claims to the good.[24] A good is "lumpy" if it cannot be further subdivided without compromising its value. The good is "homogeneous" if units of the good are interchangeable such that one unit is as beneficial for the recipient as another, and claims to the good are "homogeneous" if each party has a claim to only one unit of the good.[25] For example, the just lottery rule is satisfied when one

must allocate a scarce organ among prospective recipients who do not differ in any morally relevant way and so have equally strong claims to it.

Philosophers offer different accounts of why a lottery is required in cases that satisfy these conditions. Stone appeals to the principle of impartiality, according to which agents ought to allocate goods "according to claim strength, and only claim strength."[26] In cases where people have equally strong claims to a scarce good, agents have no reason to favor one prospective recipient over another. A lottery is required, Stone argues, for lotteries sanitize the decision-making process from the influence of any reasons, ensuring that no illegitimate reasons influence the allocation of the good, and so violate the principle of impartiality.[27]

By contrast, John Broome argues that lotteries are required in cases that satisfy the above conditions because they are ways of treating people equally.[28] For Broome, fairness requires that "*claims should be satisfied in proportion to their strength*."[29] In cases where people have equal claims to a scarce good, some unfairness is inevitable since not all claims can be satisfied. However, a lottery offers a fair way to allocate the good since it gives each person an equal chance of receiving the good, thus offering them "a sort of surrogate satisfaction" of their claims.[30]

As we discuss below, this difference in justification for the just lottery rule matters for cases where people have unequal claims to a scarce good. But, for now we may conclude that it is permissible - indeed obligatory - to use a fair lottery to allocate a scarce good when the conditions of the just lottery rule are met.

This conclusion has important implications for the permissibility of randomization, implying that it is in principle permissible to employ random assignment amid scarcity in a

policy RCT. But investigators must ensure that the conditions of the just lottery rule are satisfied throughout the course of the experiment.

First, the intervention must be lumpy, such that it cannot be further subdivided without compromising its value. Some policy interventions are clearly lumpy, for example, places in a charter school or access to a health insurance program. Others may not be. For example, cash transfers may not be lumpy in cases where the proposed transfer would still positively impact people's lives even if subdivided and dividing it would not pose any severe bureaucratic problems for the roll-out of the program. It may not therefore be permissible to randomly assign people to a cash transfer if all otherwise eligible people could instead receive some portion of the initial transfers.[31]

Second, study participants must have equally strong claims to the superior intervention. Participants have *claims* to an intervention if the distributing agent has a duty to provide it to them. They have *equally strong* claims to an intervention if they do not differ in any ways that are morally relevant for the distribution of the good. In the case of RCTs authorized or funded by governments, participants have a claim to an intervention if the government agency in question has a duty of justice to implement it. All too briefly, following Douglas MacKay we suggest that governments have a duty to implement policies that (1) realize legitimate government purposes; (2) are evidence-based; (3) are consistent with people's rights; and (4) are attainable and sustainable given the government's level of fiscal resources and bureaucratic expertise.[32] Participants therefore have claims to be subject to policies which satisfy (1) - (4), for example, the Zambian Child Grant Program which offers an affordable and evidence-based intervention by which the Zambian government can address deep poverty.

Third, participants must have stronger claims to the intervention than nonparticipants. If some nonparticipants have claims to the intervention that are as strong as or stronger than those of participants, it is wrong to exclude them from random assignment. For example, if all people below a certain income threshold have an equally strong claim to access a cash transfer intervention, the randomization procedure must grant each an equal chance of receiving the intervention. In the case of the Zambian Child Grant RCT, researchers acted rightly by randomly assigning communities in the *poorest* districts to the intervention, recognizing that they had the strongest claim to the intervention.

Finally, no person's claim may be left unmet, for any period of time, longer than is necessary due to legitimate scarcity. Resource constraints must be such that the good in question - e.g. access to a beneficial social program - will remain sufficiently scarce over the course of the study so that the intervention and control arms have enough participants to ensure adequate statistical power and the experiment runs for long enough to generate meaningful data. In other words, scarcity should not be manufactured to create the conditions for a well-designed RCT; people in the control arm and nonparticipants should not be denied access to an intervention longer than is required by legitimate resource constraints.

Summing up therefore, it is fair to randomly assign people to a superior but scarce intervention when four conditions are satisfied:

1. The intervention may not be divided without compromising its value or significantly interfering with the roll-out of the program;
2. The people subject to randomization have equally strong claims to the intervention;
3. The people subject to randomization have stronger claims to the intervention than those not subject to randomization; and,

4. No person's claim is left unmet longer than is necessary due to legitimate scarcity.

Perhaps the chief difficulty in determining whether random assignment is permissible under these conditions is determining when the people subject to randomization have an equally strong claim to the scarce intervention and a stronger claim than those not so subject. People clearly have equally strong claims to a good if they are the same along all dimensions that are morally relevant to the distribution of the good. But which dimensions are morally relevant? Utilitarians might argue that A has a stronger claim to a good than B if the world in which A receives the good is likely to contain more happiness than the world in which B receives it. Prioritarians, by contrast, would argue that the strength of A's claim is a function of both the size of the benefit to A and A's level of wellbeing.[33] Sufficientarians might argue that the strength of A's claim is a function of whether A is below some specified level of wellbeing, and perhaps also how far below this threshold A is. Desert theorists might argue that the strength of A's claim also depends on questions of desert, for example, whether A's need for the scarce good is due to A's irresponsible behavior or circumstances beyond A's control.

Researchers may not need to take a position in these long-standing normative debates for they may be limited in what they can come to know about prospective participants. In some circumstances, investigators may not have the data necessary to make fine-grained distinctions regarding the strength of people's claims, for example, to determine which allocation scheme is likely to yield the most happiness, or, in a large population of people who fall below some income threshold, to determine which households are worse off. Where it is not possible to make such fine-grained distinctions, investigators may justifiably treat people who are similarly situated as having equally strong claims to a good.

But investigators may not always face such epistemic limits and may even have a duty to gather the data necessary to ensure a fair distribution when doing so is not unreasonably burdensome.[34] In cases where researchers have good reason to draw distinctions among people's claims, condition 2 is not satisfied and so it is unfair to randomly assign participants to a beneficial intervention. This point has also been emphasized by development economists. As Christopher B. Barrett and Michael R. Carter put it, the use of random assignment too often requires researchers to deviate from the "targeting principle upon which most development interventions are appropriately founded."[35] They continue,

> Given the scarce resources and fiduciary obligations of donors, governments and charitable organizations entrusted with resources provided (voluntarily or involuntarily) by others, there is a strong case to be made for exploiting local information to improve the targeting of interventions to reach intended beneficiaries…By explicitly refusing to exploit private information held by study participants, randomized interventions routinely treat individuals known not to require the intervention instead of those known to be in need, thereby predictably wasting scarce resources. Indeed, in our experience the unfairness and wastefulness implied by strict randomization in social experiments often sows the seeds of some implementers' breach of research design.[36]

For example, consider the Moving to Opportunity experiment. Beginning in 1994, the U.S. Department of Housing and Urban Development (HUD) ran the Moving to Opportunity for Fair Housing (MTO) Demonstration Program, an experiment designed to evaluate whether offering housing vouchers to low-income families to help them move from high-poverty neighborhoods to lower-poverty neighborhoods would improve their lives and the lives of their children.[37] The trial tracked 4,604 low-income households who had children and lived in public

housing projects in Baltimore, Boston, Chicago, Los Angeles, and New York. It randomized them into three groups: 1) an experimental group that received Section 8 vouchers that were only valid in neighborhoods with poverty rates under 10 percent, as well as counseling and help in renting a new unit; 2) a Section 8 only group that did not receive counseling and were given non-specific vouchers that could be used anywhere; and 3) a control group that did not receive any vouchers or counseling beyond services for which they were already eligible.[38] The initial evaluation was conducted four to seven years after families entered the program, whereas the final evaluation measured longer-term effects 10 to 15 years after families enrolled in MTO. Both evaluations measured myriad aspects of change, including social networks, physical and mental health, and economic self-sufficiency.[39]

HUD does not have sufficient resources to offer Section 8 housing vouchers to all eligible households in the U.S., and so the standard voucher and experimental voucher can be understood to be scarce interventions. For this experiment to satisfy condition 2, all households randomly assigned to one of the three interventions would need to have equally strong claims to one of the voucher arms. However, it's not clear that this condition was satisfied. First, investigators and HUD officials no doubt had data regarding eligible households that would have enabled them to identify some households as worse off than others with respect to securing safe and adequate housing. For example, households differ in terms of number of children, how much they spend on housing, and whether they are homeless or living in substandard housing, and data regarding these differences was likely readily available. Second, it is highly plausible that such differences are relevant to the strength of people's claims, for some households are not only worse off than others, but also more likely to benefit than better off households from a voucher. Such differences matter morally whether one is a utilitarian, prioritarian, or sufficientarian.

Third, it is feasible for local housing agencies to allocate Section 8 housing vouchers based on such factors rather than simply relying on first-come, first-served or a lottery. HUD currently recommends that regional planning bodies allocate vouchers on the basis of need, where need includes consideration of: chronic homelessness; significant health or behavioral health challenges or functional impairments which require a significant level of support in order to maintain permanent housing; high utilization of crisis or emergency services, including emergency rooms, jails, and psychiatric facilities, to meet basic needs; the extent to which people, especially youth and children, are unsheltered; vulnerability to illness or death; risk of continued homelessness; and vulnerability to victimization, including physical assault or engaging in trafficking or sex work.[40] Less emphasized is time spent on a regional body's waiting list. This means that households who score moderately high on these factors may never receive assistance if there are scarce resources and they are consistently displaced by higher scoring households.

Our point here is not to criticize the designers of the Moving to Opportunity study, but only to illustrate how condition 2 may be quite easily violated. Where researchers and policymakers can draw morally relevant distinctions among prospective recipients of a beneficial but scarce social program, and where it is feasible to allocate access to the program in a way that takes these distinctions into account, condition 2 will not be satisfied and it will not be permissible to use random assignment.

2.2 Random Assignment and No Claims to a Superior but Scarce Intervention

The set of conditions we identify above is *sufficient* for permissible random assignment to a superior but scarce intervention but not *necessary*. This means there may be other conditions

under which random assignment is permissible. We argue here that random assignment is also sometimes permissible when participants have *no claims* to a superior but scarce intervention[41].

Consider the Canadian province of Ontario's recent Basic Income Pilot Project. Participants in three locations were randomly assigned to a generous basic income intervention or the comparatively meager status quo policy.[42] Single recipients of the basic income intervention were eligible to receive up to $16,989 per year for a single person (less 50% of any earned income), while couples were eligible to receive $24,027 per year (less 50% of any earned income). By contrast, the status quo policy offered recipients a maximum transfer of $8,472 for individuals and $13,140 for couples, and recipients were required to actively seek out employment. The basic income intervention is clearly superior for recipients along several measures, including food security, income security, and housing stability, but researchers and policymakers wanted to know the effects of the intervention on people's participation in the labor market, health outcomes, and healthcare usage, in part to determine the affordability of rolling out the intervention to all eligible Ontarians.

The Ontario government has no obligation to roll out the basic income intervention to low-income Ontarians. While no doubt superior to the status quo policy in promoting important outcomes, the intervention is potentially unaffordable and its cost-effectiveness is unclear. As such, the basic income intervention does not satisfy the four conditions we outline above and so is not the subject of a governmental duty to implement. Low-income Ontarians therefore have no claim of justice to access the program. This does not mean that the government of Ontario may distribute access to the basic income intervention however it likes. Ontario has a duty to treat residents equally and so it may not grant access to the intervention based on arbitrary factors

such as race, sexual orientation, or religion. However, distributing access to the intervention by means of a lottery will often be consistent with this duty.

As we understand it, the duty to treat people equally is not a duty to treat people the same, but rather a duty not to treat people differently based on some morally arbitrary ground. A 'morally arbitrary ground' is some basis of differential treatment that is irrelevant to the treating agent's realization of one or more of its morally valuable purposes. A government agency thus acts wrongly by favoring potential employees based on sexual orientation or religion for such features of a person's identity are irrelevant to the furthering of the agency's mission. Potential employees' education level or work experience, by contrast, are not morally arbitrary features of their identity, at least when it comes to hiring decisions. In the case of RCTs studying a superior but scarce intervention, while the use of random assignment treats people differently, giving some access to the intervention while withholding it from others, this differential treatment is defensible when necessary to further the research sponsor's mission of producing policy relevant knowledge.[43] Random assignment of participants to a superior but scarce intervention is thus also permissible when the following two conditions are satisfied:

1. Participants have no claims to the intervention; and
2. The use of random assignment furthers the research sponsor's goal of producing policy-relevant knowledge.

This set of conditions is relevant not only in cases where governments wish to learn more about a novel beneficial intervention whose cost-effectiveness is in doubt, but also in cases where non-governmental organizations (NGOs) wish to evaluate an intervention. NGOs are organizations constituted by private individuals and groups to further a socially valuable goal. Most importantly for our purposes, NGOs have a constitutive duty to their donors to use the

donated money to realize their stated mission and a duty of beneficence to all people.[44] The latter, in turn, consists of two obligations: the duty of easy rescue and the imperfect duty of beneficence. The former is the duty to help others in extreme need in cases where doing so requires little personal sacrifice, does not conflict with a weighty moral aim, and is likely to significantly benefit the recipient.[45] The latter is the duty to benefit others, with the agent having discretion over whom and how to help (Pierson and Millum 2018, 8). The imperfect duty of beneficence is thus different from principle of utility, which requires agents to choose the action with the greatest benefit.

Many NGOs, including GiveWell, GiveDirectly, Innovations for Poverty Action, 3ie, IDInsight, and the Bill and Melinda Gates Foundation, design, fund, and/or conduct RCTs. This activity is unlikely to fall under the duty to rescue for there are few cases of research that are highly likely to significantly benefit people (Pierson and Millum 2018, 8). When designing and planning an RCT therefore, NGOs are therefore largely operating under the imperfect duty of beneficence. As such, they do not have strict obligations to use their resources in specific ways for the benefit of particular people; and potential recipients do not have claims of justice on them. In these cases, therefore, it may be permissible for NGOs to employ random assignment. NGOs have discretion over whom and how to help and so violate no perfect duty when they withhold an effective intervention from some but not others. Nor do they violate a duty of justice to treat people equally for, as we discuss above, random assignment is consistent with such a duty when a component of a socially valuable study.

Importantly, there may be circumstances where NGOs have a perfect duty to provide the beneficial intervention to as many people as possible. As we note above, NGOs have a duty of easy rescue and so may not withhold a beneficial intervention when providing the intervention

will significantly benefit people, requires little sacrifice on behalf of the NGO, and does not conflict with a weighty moral aim. NGOs may be bound by such a duty when governments fail egregiously to fulfill their obligations.[46]

Thus far we have identified two sets of conditions under which it is permissible for researchers to randomly assign participants to a beneficial intervention. Since each set of conditions offers a *sufficient* condition for permissible randomization, these sets may not be exhaustive. We next consider two possible additional sets of sufficient conditions which do not appeal to the scarcity of an intervention.

3 Fairness and Alternative Study Designs

Researchers have proposed alternative study designs to ameliorate the fairness issues associated with randomization and the withholding of a beneficial intervention. In this section of the paper, we focus on two: stepped wedge designs and uneven randomization. Our aim is to determine whether these designs offer further sets of sufficient conditions for fair randomization.

3.1 Fairness and Stepped Wedge Designs

In stepped wedge RCTs, all participants will eventually receive the intervention. Rather than randomly allocate *participants* to the intervention, stepped wedge RCTs randomly assign the *time* at which participants will receive it.[47] Stepped wedge designs are often implemented on the grounds that they are fairer than traditional RCTs since for participants in the control arm, the intervention is merely delayed, not withheld.[48] In a systematic review of stepped wedge RCTs conducted between 2010 and 2014, Emma Beard et al. write that out of 37 studies included in the review, in "16 studies, authors described a lack of equipoise for the intervention based on

positive pilot study results or prior literature, and felt it would be unethical to deny the intervention to some groups."[49]

Proponents of stepped wedge RCTs are correct that delaying access to a beneficial intervention is ethically superior to withholding it. But this does not mean that delaying access is always fair. If people have a claim to an intervention, and it is feasible to provide it to them, it is unfair to withhold it, even for a short period of time, and even if the withholding is necessary to conduct an RCT and produce socially valuable knowledge. As Ariella Binik puts it:

> There is no ethical principle that explains why it might be permissible to deprive any participants of an effective intervention even temporarily...if it is impermissible to withhold the intervention in the first place, then it is far from clear why depriving it for a shorter period of time would be ethical.[50]

For a stepped wedge design to be permissible then, it must not involve withholding the intervention from people who have a claim to it for any period longer than is dictated by scarcity. That is, the beneficial intervention must be scarce, and the crossover of participants from the control arm to the intervention arm must be dictated by the availability of the intervention in question, not the epistemic demands of the study. People's access to the intervention must not be delayed for even a short time for the purposes of yielding valid results. For example, to return to the Zambian child grant stepped wedge RCT, it would have been wrong to delay households' access to the cash transfer for any reason other than legitimate scarcity. If it became possible to roll out the cash grant to eligible households faster than expected, policymakers should have done so, regardless of whether it would compromise the study.

Stepped wedge RCTs should not therefore be seen as a design that addresses the fairness concerns with withholding a beneficial intervention. Where people have a claim to the

intervention, delaying access to it is permissible only to the extent dictated by legitimate resource constraints. At best, stepped wedge designs offer a way to conduct RCTs while progressively rolling out the intervention to more and more people as resources become available.

3.2 Fairness and Uneven Randomization

A second alternative design involves uneven randomization. Thus far we have been understanding randomization on the model of a 'fair lottery,' wherein each participant has an equal probability of ending up in the treatment arm of an RCT, or an equal probability of receiving the treatment at a particular point in time.[51] But it is also possible to conduct an RCT using uneven randomization or a 'weighted lottery' where the probability of ending up in the treatment or control arm varies by individual. For example, type A individuals would be given a 70% chance of receiving the intervention while type B individuals would be given a 30% chance. This type of weighted lottery may seem to introduce a form of selection bias since people subject to the intervention would be systematically different from those subject to the control, but this problem can be avoided if the lottery is conducted around an eligibility cutoff.[52]

For example, suppose a government has sufficient resources to offer housing vouchers to all households below the poverty line. Suppose further that it wants to evaluate some aspect of the program, for example, its effects on self-reported health and labor market participation. To carry out an RCT, it could conduct a lottery among marginally eligible households - households just below the poverty line - and marginally ineligible households - households just above the poverty line. Assuming these households are similar enough, researchers could be confident that any difference in outcomes among the treatment and control group is due to the voucher. Since eligible households have a stronger claim to the voucher, researchers could recognize this by

giving them a higher chance of receiving it compared to marginally ineligible households. For example, the former could be given a 60% chance of receiving the intervention compared to a 40% chance for the latter.

Broome provides a case for why a weighted lottery is fair in this type of situation. Recall from above that for Broome, fairness requires that "*claims should be satisfied in proportion to their strength*."[53] He continues:

> I do not mean 'proportion' to be taken too precisely. But I do mean that equal claims require equal satisfaction, that strong claims require more satisfaction than weaker ones, and also - very importantly - that weaker claims require some satisfaction. Weaker claims must not simply be overridden by stronger ones.[54]

Where there is not enough of the good to go around and everyone has equally strong claims to the good, it is thus fair to conduct a fair lottery for doing so gives each an equal chance of receiving the good.[55] Not everyone's claim will be satisfied and so some unfairness is inevitable, but a lottery offers each person "a sort of surrogate satisfaction."[56] In cases where people have claims of differing strength, however, Broome argues that fairness implies that each be given a chance of receiving the good that is proportional to the strength of their claim.[57] If the good is simply given to those with the strongest claims, those with weaker claims would be treated unfairly since their claims would be overridden, receiving no satisfaction. In the hypothetical case above therefore, a weighted lottery should be conducted.

Broome's view offers a defense of the use of weighted lotteries in cases of eligibility cutoffs, provided that the marginally ineligible have a claim to the intervention. If they do not, it is unfair to conduct even a weighted lottery. There may be cases, moreover, where the marginally ineligible do not have such a claim. For example, in the above case of housing

vouchers, the specification of the eligibility cutoff may reflect an informed decision by policymakers that people below the cutoff have a claim to the program while those above it do not.

Unfortunately, Broome's view faces serious objections. To start, the concept of surrogate satisfaction is confused and so cannot support the claim that offering claimants a chance of receiving the good - e.g. through a weighted lottery - is fairer to some extent than not doing so. As Iwao Hirose notes (and Broome would likely agree), people's claims are claims to a *good*, not to a *chance* of receiving a good.[58] But, since satisfying a person's claim involves giving them the good, there is no sense in which giving them a chance of receiving the good offers them any degree or kind of claim satisfaction.[59] People of course prefer to have a chance of receiving a benefit than not, but receiving a chance of a benefit does not constitute a partial or surrogate satisfaction of their claim to the benefit.

Gerard Vong provides a more accurate way to understand people's claims and how lotteries do or do not satisfy them. He distinguishes between *benefit claims* and *procedural claims*.[60] The former are claims to the good in question and are satisfied when the claimant receives it. The latter are claims to be treated fairly in the allocation of the good. In cases of scarcity, not everyone's benefit claims can be satisfied, however their procedural claims can be, for example, by conducting a fair lottery when claimants have equally strong claims or providing it to those with the strongest claims. In cases where people have equal claims to a scarce good, it is thus fairer to conduct a lottery not because doing so offers a partial satisfaction of everyone's claim, but rather because doing so is procedurally fair. Broome's concept of surrogate satisfaction arguably conflates the concepts of benefit claims and procedural claims, holding that lotteries offer a way to satisfy people's benefit claims to a fair extent.

This problem with Broome's concept of surrogate satisfaction matters for he cannot argue that weighted lotteries are fairer than allocation to those with stronger claims simply because the former offer claimants some satisfaction of their claims while the latter does not. Since benefit claims cannot be partially satisfied, there is no sense in which weighted lotteries offer surrogate satisfaction of people's claims to a good.

In response to this objection, Broome may simply decide to reformulate his view using Vong's distinction, arguing that in cases where people have claims of differing strength to a scarce good, it is procedurally unfair to give the scarce good to people with the stronger claims.[61] Instead, procedural fairness requires conducting a weighted lottery for giving people a chance of having their benefit claims satisfied is a way of recognizing their claims rather than allowing them to be overridden.

But consider first that there is nothing obviously unfair with the procedural rule 'satisfy stronger claims prior to weaker ones.' If A's claim to a good is indeed stronger than B's, why isn't it fair to give the good to A?[62] A procedure that follows this rule also recognizes B's benefit claim for it requires the distributive agent to consider and weigh all benefit claims to the good. Giving B's benefit claim some chance of satisfaction is thus not the only way to recognize it.

Consider second that Broome's view of procedural fairness has counterintuitive implications.[63] Conducting a weighted lottery among the marginally eligible and the marginally ineligible may seem fair since the strength of their claims are comparable. But Broome's view implies that any person's claim must receive some chance of satisfaction, regardless of how weak it is. For example, suppose policymakers wish to roll out the above-mentioned housing voucher program to all people living under the poverty line, but only have enough resources over the first two years of the program to offer it to 50% of this population. Assuming all people

under the poverty line have a claim to the program, Broome's view implies that it is fairer in at least one respect to conduct a weighted lottery, giving those closest to the poverty line, say, a very low chance of receiving it, and those experiencing deep poverty a much higher chance. In our view, it would be wrong to give the marginally eligible any chance at all to receive the intervention; instead, it should be provided to the poorest people, since they have the strongest claims to it. More generally, distributing scarce goods by means of a weighted lottery runs the risk of people with weaker claims receiving the good in question while those with stronger claims go without.

Finally, although we may be torn in cases of marginally stronger claims, we need to be certain that the case under consideration is not in fact one where people have equally strong claims. In addition, where A's claim is marginally stronger than B's, our claim is only that it is *marginally* more fair to give the good to A; we can still accept that it is seriously unfair that B must go without.[64] In sum, if we know who has the stronger claim to the good, there's no reason to conduct a lottery; conducting a lottery only introduces the possibility that a weaker claim will be satisfied instead of a stronger one.

Broome's view of lotteries offers a possible justification for the fairness of uneven randomization in policy RCTs evaluating a beneficial but scarce intervention. Unfortunately, this view faces serious objections. As we note above, Stone offers a different view of allocative justice and the role of lotteries. For him, allocative justice demands that scarce goods be distributed "according to claim strength, and only claim strength."[65] Lotteries are necessary in cases where people have equally strong claims to ensure impartiality - i.e. that goods are not allocated on the basis of something other than claim strength such as racial bias. But this also implies that weighted lotteries are not permissible in treatment cutoff cases where people have

unequal claims to the intervention. Instead, people with the stronger claims to the intervention should receive it first, for their claims are stronger.

Conclusion

Researchers and policymakers sometimes argue that it is permissible to randomly assign people to beneficial intervention even though there is no equipoise. Our aim in this paper has been to explore the lines of argument that have been offered to support this claim. Appealing to the philosophical literature on the use of lotteries in allocation, we identified two sets of conditions under which it is fair to randomly assign people to a beneficial but scarce intervention. Since these sets offer sufficient conditions for fair randomization, there may be other circumstances when random assignment to such an intervention is permissible.

Researchers have also argued that stepped wedge designs and uneven randomization offer additional ways to address the fairness problem with random assignment. We concluded, first, that while stepped wedge designs are permissible when they satisfy one of the two sets of conditions we identify above, they do not offer an additional set of conditions under which random assignment is fair. We concluded, second, that while uneven randomization is permissible under Broome's view of the ethics of lotteries, this view faces compelling objections.

We hope to have contributed to the growing literature on the ethics of public policy RCTs. There is much work to be done in this area, and our paper alone leaves two important tasks for future work. First, we have presupposed throughout that it is possible to identify which policies it is to which people have claims. Given reasonable disagreement on questions of justice,

this task is a challenging one, and it would be helpful to provide more guidance to researchers and government agencies on how to make such judgments in pluralistic societies. Second, it would be helpful to articulate a more refined conception of equipoise so that researchers can more easily determine when equipoise has been disturbed and so when they must appeal to a different set of criteria to determine if random assignment is permissible. Unfortunately, we must leave these tasks for future work.

---

[1] For good overviews of the increasing use of RCTs, see Levitt, S.D., and J. A. List, "Field Experiments in Economics: The Past, the Present, and the Future," *European Economics Review* 53, no. 1 (2009): 1-18; Gueron, J.M., and H. Rolston, *Fighting for Reliable Evidence* (New York: Russell Sage Foundation, 2013); Baldassarri, D., and M. Abascal, "Field Experiments Across the Social Sciences," *Annual Review of Sociology* 43 (2017): 41-73; Leigh, A., *Randomistas: How Radical Researchers are Changing Our World* (New Haven: Yale University Press, 2018); and Leão, L.D.S., and Eyal, G., "The Rise of Randomized Controlled Trials (RCTs) in International Development in Historical Perspective," *Theory and Society* 48, no. 3 (2019): 383-418.

[2] Shadish, W.R., T. D. Cook, and D. T. Campbell, *Experimental and Quasi-Experimental Designs for Generalized Causal Inference* (Boston: Houghton Mifflin, 2002), 226-278; Banerjee, A.V. and E. Duflo, *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty* (New York: Public Affairs, 2011); and Karlan, D., and J. Appel, *More Than Good Intentions: How a New Economics is Helping to Solve Global Poverty* (New York: Dutton, 2011).

[3] MacKay, D., "The Ethics of Public Policy RCTs: The Principle of Policy Equipoise," *Bioethics* 32, no. 1 (2018): 59-67; MacKay, D., "Government Policy Experiments and the Ethics of Randomization," *Philosophy and Public Affairs* 48, no. 1 (2020): 319-352; Asiedu, E. et al., "A Call for Structured Ethics Appendices in Social Science Papers," *PNAS* 118, no. 29 (2021): e2024570118; and London, A.J., "Equipoise: Integrating Social Value and Equal Respect in Research with Humans," in *The Oxford Handbook of Research Ethics*, edited by Ana Iltis and Douglas MacKay (New York: Oxford University Press, Forthcoming).

[4] MacKay, "Government Policy Experiments and the Ethics of Randomization," 333-348; Asiedu et al., "A Call for Structured Ethics Appendices in Social Science Papers," 2-4; and Banerjee, A., and E. Duflo, "The Experimental Approach to Development Economics," in *Field Experiments and Their Critics: Essays on the Uses and Abuses of Experimentation in the Social Sciences*, Edited by Dawn Langan Teele (New Haven: Yale University Press, 2014), 101.

[5] For more comprehensive discussions of the ethics of policy RCTs see MacKay, D., "The Ethics of Public Policy Research: Lessons from Clinical Research Ethics." In *The Oxford Handbook of Research Ethics*, edited by Ana S. Iltis and Douglas MacKay. New York: Oxford University Press, forthcoming; Asiedu et al., "A Call for Structured Ethics Appendices in Social Science Papers;" Phillips, T., "Ethics of Field Experiments." *Annual Review of Political Science* 24

(2021): 277-300; Glennerster, R., and S. Powers, "Balancing Risk and Benefit: Ethical Tradeoffs in Running Randomized Evaluations," in *The Oxford Handbook of Professional Economic Ethics*, edited by George F. DeMartino and Deirdre McCloskey, 367-401. New York: Oxford University Press, 2019; Teele, D. L., "Reflections on the Ethics of Field Experiments," in *Field Experiments and Their Critics*, edited by Dawn L. Teele, 115-140. New Haven: Yale University Press, 2014; and Baele, S. J., "The Ethics of New Development Economics: Is the Experimental Approach to Development Economics Morally Wrong?" *The Journal of Philosophical Economics* 7, no. 1 (2013): 2-42.

[6] Tsemberis, S., L. Gulcur, and M. Nakae, "Housing First, Consumer Choice, and Harm Reduction for Homeless Individuals With a Dual Diagnosis," *American Journal of Public Health* 94, no. 4 (April 2004), 651.

[7] Ibid, 651-652.

[8] Ibid, 652.

[9] MacKay, "Government Policy Experiments and the Ethics of Randomization," 328-329.

[10] Handa, S., et al., "The Social and Productive Impacts of Zambia's Child Grant," *Journal of Policy Analysis and Management* 35, no. 2 (Spring 2016): 360-361.

[11] Ibid, 361.

[12] Ibid.

[13] Ibid, 360.

[14] Ibid, 361.

[15] See Diener, E., and R. Crandall, *Ethics in Social and Behavioral Research* (Chicago: University of Chicago Press, 1978), 136; Banerjee and Duflo, "The Experimental Approach to Development Economics;" Gueron, J.M., "The Politics and Practice of Social Experiments: Seeds of a Revolution," in *Handbook of Economic Field Experiments,* Volume 1, edited by Abhijit Banerjee and Esther Duflo, (Amsterdam: North-Holland, 2017), 35, 49; and Glennerster and Powers, "Balancing Risk and Benefit," 380-381.

[16] Banerjee and Duflo, "The Experimental Approach to Development Economics," 101.

[17] Handa et al., "The Social and Productive Impacts of Zambia's Child Grant," 359-360.

[18] Asiedu et al., "A Call for Structured Ethics Appendices in Social Science Papers," 3.

[19] Parfit, D., "Equality and Priority," *Ratio* 10, no. 3 (1997): 202-221.

[20] Stone, P., *The Luck of the Draw: The Role of Lotteries in Decision Making* (New York: Oxford University Press, 2011), 20.

[21] Ibid, 24.

[22] For example, see Rawls, J., "Outline of a Decision Procedure for Ethics," in *John Rawls: Collected Papers*, edited by Samuel Freeman (Cambridge: Harvard University Press, 1999), 15-16; Elster, J., *Solomonic Judgments: Studies in the Limitations of Rationality* (Cambridge: Cambridge University Press, 1989); Broome, J., "Fairness," *Proceedings of the Aristotelian Society* 91 (1990-1991), 97-98; Stone, *The Luck of the Draw*, 53.

[23] Ibid.

[24] Ibid, 50.

[25] Ibid.

[26] Ibid, 78.

[27] Ibid, 82-83.

[28] Broome, "Fairness," 97-98.

[29] Ibid, 95.

[30] Ibid, 98.

[31] For discussion on this point, we are grateful to Chad Horne.
[32] MacKay, "Government Policy Experiments and the Ethics of Randomization," 323-324.
[33] Parfit, "Equality and Priority," 213.
[34] Thanks to an anonymous reviewer for raising this possibility.
[35] Barrett, C.B., and M. R. Carter, "The Power and Pitfalls of Experiments in Development Economics: Some Non-random Reflections," *Applied Economic Perspectives and Policy* 32, no. 4 (2010), 521.
[36] Ibid, 521.
[37] Sanbonmatsu, L., et al., *Moving to Opportunity for Fair Housing Demonstration Program: Final Impacts Evaluation* (New York: U.S. Department of Housing and Urban Development, October 2011), xiv.
[38] Ibid.
[39] Ibid, xv.
[40] Department of Housing and Urban Development, *Notice Establishing Additional Requirements for a Continuum of Care Centralized or Coordinated Assessment System* (Washington, DC: U.S. Department of Housing and Urban Development, January 2017), 10.
[41] Stone, *The Luck of the Draw*, 67-68.
[42] Government of Ontario, *Ontario Basic Income Pilot*, https://www.ontario.ca/page/ontario-basic-income-pilot
[43] For further development of this argument, see MacKay, "Government Policy Experiments and the Ethics of Randomization," 335-341.
[44] Pierson, L., and J. Millum, "Health Research Priority Setting: The Duties of Individual Funders," *The American Journal of Bioethics* 18, no. 11 (2018), 12.
[45] MacKay, D., and T. Rulli, "The Duty to Rescue and Investigators' Obligations," *Kennedy Institute of Ethics Journal* 27, no. 1 (2017), 71.
[46] Thanks to an anonymous reviewer for raising this point.
[47] Beard, E., et al., "Stepped Wedge Randomised Controlled Trials: Systematic Review of Studies Published Between 2010 and 2014," *Trials* 16 (2015): 1-14.
[48] See Brown, C.A., and R. J. Lilford, "The Stepped Wedge Trial Design: A Systematic Review," *BMC Medical Research Methodology* 6, no. 54 (2006); and Beard et al., "Stepped Wedge Randomised Controlled Trials," 4.
[49] Ibid.
[50] Binik, A., "Delaying and Withholding Interventions: Ethics and the Stepped Wedge Trial," *Journal of Medical Ethics* 45 (2019), 664.
[51] Stone, *The Luck of the Draw*, 24.
[52] Glennerster, R., and K. Takavarasha, *Running Randomized Evaluations: A Practical Guide* (Princeton: Princeton University Press, 2013), 124.
[53] Broome, "Fairness," 95.
[54] Ibid, 95.
[55] Ibid, 98.
[56] Ibid.
[57] Ibid.
[58] Hirose, I., "Weighted Lotteries in Life and Death Cases," *Ratio* 20, no. 1 (2007), 52-54. See also Henning, T., "From Choice to Chance? Saving People, Fairness, and Lotteries," *The Philosophical Review* 124, no. 2 (2015), 171-177.

[59] For an additional problem with Broome's notion of surrogate satisfaction, see Vong, G., "Fairness, Benefiting by Lottery and the Chancy Satisfaction of Moral Claims," *Utilitas* 27, no. 4 (2015): 470-486.
[60] Ibid, 479.
[61] Christian Piller argues that this is actually what Broome has in mind. Lotteries offer a form of surrogate satisfaction of people's claims in the sense that they treat people fairly, not in the sense that a chance of receiving the good is a surrogate or replacement for the good itself. Piller, C., "Treating Broome Fairly," *Utilitas* 29, no. 2 (2017), 225-226.
[62] Lazenby, H., "Broome on Fairness and Lotteries," *Utilitas* 26, no. 4 (2014): 331-345.
[63] Hooker, B., "Fairness," *Ethical Theory and Moral Practice* 8, no. 4 (2005), 349; Hirose, "Weighted Lotteries in Life and Death Cases," 54-56.
[64] Lazenby, "Broome on Fairness and Lotteries," 342-343.
[65] Stone, *The Luck of the Draw*, 78.