Douglas MacKay

The Ethics of Public Policy Experiments: Lessons from Clinical Research Ethics

Abstract

Social scientists and research ethicists have begun, somewhat belatedly, to confront and address the ethical challenges raised by public policy experiments. In doing so however, they have not fully availed themselves of the large and sophisticated literature on the ethics of clinical research which has developed over the past 40 years. While clinical and public policy research are different, I argue that the clinical research ethics literature yields valuable insights for discussions of the ethics of policy experiments. Focusing on seven ethical issues which have received a good deal of attention in public and scholarly discussions of the ethics of policy experiments, I make use of the history of reflection on the ethics of clinical research to provide guidance to researchers planning and conducting policy experiments.

Keywords: Public policy experiments; research ethics; field experiments; randomized controlled trials; informed consent; community engagement.

Text

Beginning in 1970, the U.S. Department of Health, Education, and Welfare, in partnership with the states of Washington and Colorado, authorized and funded the Seattle-Denver Income Maintenance Experiment to study negative income tax plans (United States 1983). Such plans provide households with a guaranteed income in the form of a cash transfer, which is taxed at a specific rate as income from other sources increases. Almost 5,000 low- and middle-income households in Seattle and Denver were randomly assigned to one of four treatments, lasting either 3 or 5 years: (1) a negative income tax plan; (2) employment counseling/training; (3) a negative income tax plan and employment counseling/training; and (4) the status quo (United States 1983). The Department of Health, Education, and Welfare was interested in reforming existing transfer programs for disadvantaged households and wanted to determine whether a negative income tax policy would disincentivize participation in the labor market (United States 1983). The Seattle-Denver Income Maintenance Experiment was the largest of four income maintenance experiments undertaken in the U.S. at the time, with others taking place in New Jersey, rural Iowa and North Carolina, and Gary, Indiana (United States 1983). These experiments were remarkable not only for their evaluation of a promising and provocative new policy – the negative income tax – but also because they involved the first use of random assignment on a large scale to determine the causal efficacy of a policy intervention (Gueron and Rolston 2013, 6).

Since 1970, governments have increasingly used randomized controlled trials (RCTs) to evaluate policy interventions, and the use of field experiments has become more prevalent in economics (particularly development economics) and political science (Desposato Forthcoming; de Souza Leão and Eyal 2019; Leigh 2018; Baldassarri and Abascal 2017; Gueron and Rolston 2013; Karlan and Appel 2011; Levitt and List 2009). Social scientists and research ethicists have also

begun, somewhat belatedly, to confront and address the ethical challenges raised by these experiments (see Desposato Forthcoming; Asiedu et al. 2021; Evans 2021; Phillips 2021; Abramowicz and Szafarz 2020; McDermott and Hatemi 2020; MacKay 2020; MacKay and Chakrabarti 2019; Whitfield 2019; MacKay 2018; Glennerster and Powers 2016; Teele 2014; Baele 2013). However, while this work is valuable, scholars have not fully availed themselves of the large and sophisticated literature on the ethics of *clinical* research which has developed over the past 40 years. For example, some studies of the ethics of field experiments appeal only to *The Belmont Report*, applying the principles of respect for persons, beneficence, and justice to the conduct of field experiments (McDermott and Hatemi 2020; Glennerster and Powers 2016; Teele 2014). This is fine as far as it goes, but discussions of the ethics of clinical research have moved beyond this, refining interpretations of the principles and applications found in *Belmont*, contesting these interpretations, and developing new concepts and principles to evaluate clinical research protocols. While clinical and social science research are different, the clinical research ethics literature may nonetheless yield valuable insights for discussions of the ethics of field experiments.

I explore this possibility in this chapter, mining the history of ethical reflection on clinical research to help identify the problems sponsors and researchers are likely to encounter in the course of their research, and to develop a set of concepts and principles they may use to resolve them. My focus is public policy experiments, that is, experiments designed to evaluate the efficacy of a policy intervention. I say very little about the ethics of other types of field experiments, for example electoral experiments, which are common in experimental political science. I focus on the ethical problems which have received the most attention in scholarly and public discussions regarding the use of policy experiments, including the fairness of random assignment, failures to secure informed consent, and the value of RCTs. I do not claim to offer a comprehensive framework for evaluating

policy experiments, but I do offer recommendations on these and other issues and identify several

additional issues deserving further ethical reflection in the conclusion. Like the analyses mentioned

above, I too appeal to the principles and applications outlined in *The Belmont Report*, but I also go

further, showing how recent work in clinical research ethics holds lessons for the ethical conduct of

policy experiments.

I begin with a discussion of the nature of public policy experiments, what they are, who

conducts them, and their purported value. I then consider what I take to be the central seven ethical

issues raised by these experiments and defend principles and recommendations for addressing them.

My focus throughout shall be policy experiments conducted or authorized by government agencies;

however, I shall also comment on how my recommendations apply to experiments conducted by

non-governmental organizations and academic researchers.

1. Public Policy Experiments

Public policy experiments involve the introduction of a policy intervention, or the alteration

of an existing policy intervention, as part of a systematic investigation aimed at generating

knowledge about the intervention's impact. Policy RCTs are experiments which use random

assignment to determine which individuals or groups are subject to the intervention. Policy

experiments may also involve the collection of data that would not otherwise be collected.[i]

As field experiments, policy experiments allow researchers to see the effects of an

intervention in the 'real world.' When experiments involve random assignment, they have the further

benefit of minimizing forms of bias that are more prominent in observational research (Shadish,

Cook, and Campbell 2002, 226-278). In particular, the random assignment of participants to intervention and control groups is often thought to minimize selection bias – i.e. systematic differences between those subject and not subject to the intervention that are potentially correlated with the outcomes of interest. With observational studies, researchers can never be certain whether outcomes are the result of the intervention, or of some distinctive characteristic(s) for which they have not controlled.

Government policy experiments are policy experiments that are conducted, authorized, or funded by a government agency or institution. Typically, government policy experiments involve government agencies or institutions partnering with academic researchers or firms to introduce and evaluate a novel intervention. For example, the Seattle-Denver Income Maintenance Experiment, was the joint effort of the U.S. Department of Health, Education, and Welfare, the states of Washington and Colorado, and the private research firms SRI International and Mathematica Policy Research (United States 1983). Non-government policy experiments, by contrast, are experiments in which government agencies are not involved as a partner or funder. They may instead involve non-governmental organizations such as philanthropic foundations working with researchers or research firms to introduce and evaluate an intervention. For example, researchers funded by Stichting Dioraphte, a private foundation, and the Stanford Center for Innovation in Global Health, recently partnered with Partners in Health, a non-governmental organization, to conduct an RCT in Malawi to test the impact of interventions designed to increase usage of chlorine to disinfect drinking water (Dupas et al. 2021). Households were randomly assigned to receive either coupons for free chlorine or some variation of a delivery method involving a community health worker supported by Partners in Health.

This distinction between government and non-government policy experiments is important since governments, non-governmental organizations such as research firms or philanthropic sponsors, and social scientists have different rights and obligations. While non-governmental organizations have duties of beneficence, governments have robust duties of justice to their residents, which may include duties to protect their liberty and security, protect and promote their health, provide a quality education, and ensure a decent quality of life. Legitimate governments also have a right to rule over their residents, that is, a right to enact law and policy. Private actors, by contrast, do not have these duties of justice nor do they have a right to rule. These differences in rights and obligations are relevant to the ethical conduct of policy experiments. As I discuss in greater detail below, because government agencies have robust duties of justice to residents, they must take greater care than non-government actors in ensuring that research participants are not subject to inferior interventions. Because non-government actors have no right to rule over participants, they must take greater care to secure the informed consent of participants.[ii]

I turn next to a discussion of seven central ethical issues raised by policy experiments and provide principles and recommendations to guide the design of studies. My focus is on government policy experiments, though I shall also comment on how the principles and recommendations I propose might differ for experiments carried out by non-government actors.

2. Social Value

In 2019, Abhijit Banerjee, Esther Duflo, and Michael Kremer were jointly awarded the Nobel Prize in Economic Sciences "for their experimental approach to alleviating global poverty (The Royal Swedish Academy of Sciences 2019)." RCTs are not without their critics, however,

despite the oft-heard claim that they constitute the 'gold standard' for establishing evidence of causation. First, Angus Deaton and Nancy Cartwright (2018, 10) rightly caution that the results of an RCT cannot be simply extrapolated or generalized to other contexts. The findings of an RCT in one context may be useful for the development of policy in another, but great care must be taken regarding their interpretation and application (Cartwright and Hardie 2012). Second, for experimental results to be truly useful, they must be taken up by policymakers in the design of public policy. This may not happen, either because the knowledge generated by experiments does not fill the "knowledge gaps" that policymakers need filling (Ravallion 2020, 71-76), because policymakers ignore the findings of an experiment, or because policymakers cancel an experiment before it is complete (McDowell and Ferdosi Forthcoming).

These concerns regarding the value of policy experiments are important, for they are expensive to conduct and often impose risks on participants and communities. Regarding the planning and design of individual experiments, a potentially useful concept for considering and adjudicating these concerns is *social value,* a widely recognized requirement for clinical research (Wenner Forthcoming; Emanuel, Wendler, and Grady 2000, 2703). Clinical research meets this requirement either by promising to generate clinically relevant generalizable knowledge – i.e. knowledge that could lead to improvements in clinical outcomes – or by promising to produce generalizable knowledge that, while upstream of any immediate application, is preliminary to the development of clinically relevant knowledge (Emanuel, Wendler, and Grady 2000, 2703).

There are several arguments in support of the social value requirement for clinical research, with some appealing to the need to justify the risks and expense of such research, and others appealing to the role it plays in securing health justice (see Wenner 2018; Wendler and Rid 2017; Emanuel, Wendler, and Grady 2000, 2703). For government policy research, the case for a social

value requirement is straightforward: such experiments are public projects, often employ public resources, and so must serve the public interest. Since governments have a fiduciary obligation to act in their residents' interests, they may only conduct, authorize, or fund policy experiments if they are reasonably expected to promote the public good (Glennerster and Powers 2016, 373).

The goals of policy and clinical research differ and so the content of the social value requirement for the former is different from that of the latter. The central purpose of public policy research is to generate policy relevant generalizable knowledge, that is, generalizable knowledge that is useful for the development and implementation of cost-effective policies for the realization of the goals of governments. Elsewhere, I introduce the concept of "target outcomes" to capture such goals (MacKay 2020, 323). Such outcomes "refer to specifications of the type and amount of goods or services governments have a duty to provide, and the outcomes they have a duty to realize (MacKay 2020, 323)." Target outcomes may include outcomes related to literacy and numeracy, college attendance, physical and/or mental health, crime and residents' physical security and liberty, income levels, food security, and subjective wellbeing, among others. The target outcomes that any government has a duty to realize is also likely to vary depending on cultural variation, resources, and political tradition.

Government policy research meets the social value requirement therefore if it promises to yield generalizable knowledge that is relevant to the development of cost-effective policies for the realization of target outcomes. Policy research would count as socially valuable then if it involved the evaluation of an untested intervention which promised to improve people's food security, literacy scores, subjective wellbeing, or mental health; if it tested a prediction of a promising social theory relevant to the development of affordable and effective policies; or if it promised to identify the ways in which people are treated unjustly, for example, in hiring or access to rental housing. It

would not count as socially valuable if it evaluated an intervention that could not realistically be implemented; if it tested a prediction of a social theory that is irrelevant to policymaking; if, for political reasons, it studied an issue that had already been resolved; or if it was unlikely to be completed, for example, due to a change in government.

The social value requirement for clinical research is widely understood as a threshold for permissible research (Wenner Forthcoming). Similarly, it is reasonable to suggest that public policy research need only be *sufficiently* socially valuable – not *maximally* socially valuable – to be permissible. Government agencies need only show, for example, that a proposed experiment in the sphere of education policy addresses a question relevant to the promotion of education-related target outcomes, not that it studies the most pressing question faced by policymakers concerned to realize such outcomes. As such, we can formulate the social value requirement for government policy experiments as follows:

> Social Value: Government policy experiments must be sufficiently socially valuable; they must be reasonably expected to yield generalizable knowledge that is relevant to the development of cost-effective policy interventions for the realization of target outcomes.

Prior to beginning an experiment, government agencies should explain to the public why they are confident the experiment will be completed, how the study is expected to contribute to existing knowledge (is not redundant), and how the study's findings are relevant to policymakers concerned to develop and implement cost-effective interventions for the realization of target outcomes.

While this requirement is formulated as a threshold that experiments must satisfy, the concept of social value makes it in principle possible to compare proposed policy experiments in terms of their expected social value. Some policy experiments, for example, because they concern

policy spheres that are more central to residents' wellbeing, may be more socially valuable than others. The possibility of such comparative judgments is important for it opens the door to rational priority setting among policy experiments competing for scarce public funds (see Pierson and Millum 2018).

Must policy experiments carried out by non-governmental organizations also meet a social value requirement to be ethically permissible? The 'public good' argument presented above doesn't apply to such organizations since they are not using public resources. Similarly, some research ethicists are skeptical that the social value requirement applies to privately funded clinical research since private actors have much greater discretion over how they use their resources (Resnik 2018, 75; Wertheimer 2015, 303). But, while it may be challenging to argue that *all* non-governmental policy experiments are subject to a social value requirement, there are good arguments in support of the claim that *many* are. First, some experiments may impose high risks on participants and/or fail to secure participants' informed consent – e.g. studies involving children. For the imposition of risk in these studies to be justifiable, they must be socially valuable (Wendler and Rid 2017, 79-81). Second, non-profit agents may have duties to their home governments and funders to conduct socially valuable research. Non-profits often enjoy tax exempt status which is conditional on their using their resources in socially valuable ways and are also accountable to their donors to carry out their socially important missions (Pierson and Millum 2018, 12).

3. Favorable Risk-Benefit Ratio

In addition to concerns regarding the value of policy experiments, commentators have also raised concerns regarding the risks experiments impose on participants and the communities within

which they live (McDermott and Hatemi 2020). One of the epistemic benefits of policy experiments is that they are conducted in the *field*, but the flip side of this is potential negative spillover effects for communities.

A central ethical requirement of clinical research that holds also for government policy experiments is that studies have a favorable risk-benefit ratio: the potential benefits of the study should justify the risks (Rid and Wendler 2011; Emanuel, Wendler, and Grady 2000, 2705; The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research 1979, 8-9). This requirement is usually grounded in the consequentialist principle of beneficence which directs agents to maximize net benefits (The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research 1979, 5). However, with respect to government policy research in particular, this requirement is also derivable from the duty of governments to protect and promote their residents' interests. To fulfill this duty, government agencies should only conduct or authorize an experiment if the risks to participants and the broader society are justified by the potential benefits.

Emanuel, Wender, and Grady (2000, 2705) helpfully note that the favorable risk-benefit ratio requirement should be understood as three imperatives: (1) minimize risks to participants; (2) enhance benefits to participants; and (3) ensure that the potential benefits to participants and society are proportionate to the risks. (1) and (2) are not necessary to ensure that a study has a favorable risk-benefit ratio but follow from: (a) the consequentialist logic of the principle of beneficence, which requires decision-makers to choose the action with the greatest *net* benefit; and (b) the duty of governments to protect and promote residents' interests. For the purposes of evaluating public policy experiments however, these imperatives should be widened. While the clinical research ethics regulations and literature focus on risks to *participants* since it is participants who bear the principal

risks of novel clinical interventions (though see Shah et al. 2018), public policy interventions may affect both participants and the communities within which they live. Spillover risks to bystanders must therefore be considered in the evaluation of a study's risk-benefit ratio.

A risk is a possible harm; the severity of a risk is thus a combination of the degree of the harm and the probability that it may occur. In government policy experiments, there are four principal types of harms to consider when conducting a risk-benefit assessment. First, there are "intervention harms," that is, harms related to the delivery of the intervention that is the subject of the experiment (Phillips 2021, 279). These harms may include direct harms to participants should an experimental intervention prove to be worse than the counter-factual status quo (the policy to which participants would have been subject absent an experiment), and may include worse health, loss of housing, lower income, reduced subjective wellbeing, and worse educational outcomes, among others. These harms may also extend to members of participants' households, particularly children. In addition, where experiments evaluate the responses of employers, businesses, or civil servants, they may also occupy the attention and time of employees who would be otherwise engaged in other, valuable tasks (Phillips 2020, 280). Desposato (Forthcoming, 14) notes that while each engagement may be minor, the sum of such engagements may be large, constituting a sizeable "aggregate harm."

Intervention harms may also be indirect, wherein the introduction of the intervention leads to unforeseen negative consequences for participants and their broader communities (Phillips 2021, 280). For example, microcredit programs which aim to empower women by giving them access to capital may also increase women's risk of violence and victimization in patriarchal societies (Schuler, Hashemi, and Badal 1998). Similarly, the introduction of a large-scale alternative dispute resolution education campaign in Liberia led to higher rates of land dispute resolution and lower violence, but

also an increase in extrajudicial punishment (Blattman, Hartman, and Blair 2014). There is also evidence that randomized evaluations of cash transfer programs may harm people in the control arm by increasing the prices of important goods such as protein-rich foods (Filmer, Friedman, Kandpal, and Onishi 2018).

A second type of harm is 'data harms.' Researchers may need to collect data directly from individuals enrolled in an experiment and if this data is not kept confidential, its release could be harmful for participants. For example, the publication of data concerning an individuals' disease status, substance use, or sexual orientation, may lead to discrimination or stigmatization. Participants may also be harmed in the data collection process if it requires them to reveal facts about themselves that they find to be embarrassing or shameful, or whose revisitation is emotionally disturbing or traumatic.

A third type of harm is "exclusion harms" (Phillips 2021, 280). These are harms that result from the exclusion of participants from the treatment arm in an RCT or the exclusion of prospective participants from participation in the study. In many cases, participants who are assigned to the control arm in a study and so are not given access to a promising intervention, may experience feelings of jealousy, resentment, and anger which may in turn lead to actions – including the spread of misinformation and violence – that threaten community cohesion (Dionne, Harawa, and Honde 2016; MacPhail et al. 2013). People assigned to the control arm may also be harmed if they would have had access to a beneficial intervention if the experiment had not been conducted (Phillips 2021, 280).

A fourth type of harm is "outcome harms" (Phillips 2021, 281). These are harms related to the use, misuse, and dissemination of research findings. They may include the use of findings to

enact policies which make people worse off or the stigmatization of groups within society (Phillips 2021, 281; Warwick 1982, 114-115). For example, Gubler and Selway (2016, 177) worry that their research focus on caste prejudice among people from lower castes may lead people from higher castes to conclude that caste prejudice is only a phenomenon within lower caste communities. Admittedly, it may be very difficult for researchers to either predict or prevent outcome harms.

'Benefits' in risk-benefit assessments refer to the things of value which may be produced by research. Benefits contrast directly with harms so risk-benefit assessments examine the probability and magnitude of harms and benefits (The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research 1979, 8). The principal benefit of public policy research is policy relevant generalizable knowledge, that is, generalizable knowledge that is useful for the development and implementation of cost-effective policies for the realization of target outcomes. Risk-benefit assessments should also consider, however, potential direct benefits to participants and the communities in which they live. Experiments that evaluate interventions reasonably expected to be superior to the status quo – e.g. cash transfer programs – deliver benefits directly to participants.

Determining whether a policy experiment has a favorable risk-benefit ratio requires judgment for it involves the qualitative comparison of risks to participants and the value of policy relevant knowledge. In addition, as Trisha Phillips (2021, 281) makes clear, there is also a good deal of "normative ambiguity" regarding the classification of certain consequences of interventions. For example, suppose an experiment reveals serious corruption within a bureaucracy, resulting in the firing of several civil servants. The fired civil servants would seem to be harmed by the research, but should the probability of this harm really be considered a risk when conducting a risk-benefit assessment?

To resolve these problems, it is important to specify a baseline against which to evaluate changes in people's wellbeing (Phillips 2021, 281). One possibility is to employ a *descriptive* baseline. On this understanding, a person is harmed if they are worse off than they would otherwise be and benefitted if they are better off than they would otherwise be. This approach has implausible implications, however. It would imply that the above-mentioned possibility that corrupt civil servants might be fired should be counted as a *risk* of the research.

A second possibility is to employ a *normative* baseline (Phillips 2021, 281). On this approach, whether an impact on a person's wellbeing counts as a harm or benefit depends on considerations of how well off they *ought* to be. Employing such a baseline is demanding for it requires researchers to make normative judgments, but it yields more plausible implications in the above case. Since civil servants have no entitlement to perform their duties corruptly, the potential firing of civil servants for corruption should not be considered a risk of research.

Constructing a normative baseline may be challenging in some circumstances since people may reasonably disagree about the requirements of justice. But as will become clear later in this chapter – if it is not yet clear now – the consideration of the ethics of public policy research is shot through with the need to make difficult judgments of justice. To paraphrase John Rawls (1999, 3), justice is the first virtue of laws and policies just as truth is the first virtue of systems of thought. It is no surprise then that for researchers to conduct policy experiments ethically, it is necessary that they treat participants and non-participants justly, and so it is necessary for researchers to make potentially controversial judgments regarding what this entails.

With the above considerations in mind, we may now reformulate Emanuel, Wendler, and Grady (2000, 2705)'s favorable risk-benefit ratio requirement to provide guidance for government agencies planning policy experiments.

Favorable Risk-Benefit Ratio: Government policy experiments are permissible only if:

1. Risks to participants and affected bystanders are minimized;

2. Benefits to participants and society are enhanced; and

3. The potential benefits to participants and society justify the risks to participants and affected bystanders.

Favorable risk-benefit ratio is thus a necessary condition of government policy experiments, and as noted above, consideration of risks and benefits should employ a normative baseline.

Must non-governmental policy experiments comply with this requirement to be ethical? I think so. While non-governmental organizations do not have the same fiduciary obligations as governments, they do have obligations of beneficence, understood here not as the utilitarian imperative to choose actions that maximize good outcomes, but rather as the less demanding obligation not to harm others without justification and to also (sometimes) promote the interests of others. Where non-governmental organizations decide to conduct a policy experiment, this duty of beneficence implies that they design the study in accordance with 1-3. The duty not to harm others without justification supports 1; the duty to promote the interests of others supports 2; and together these duties support 3.

4. Standard of Care

In a recent RCT in Nairobi, Kenya, researchers working for the World Bank partnered with the Nairobi City Water and Sewerage Company (NCWSC) to randomly assign households living in 'slums' to two interventions designed to increase payment rates for water services (Coville et al. 2021). Increased payment rates were necessary to allow the NCWSC to maintain the quality of its services for consumers. Controversially, one of the interventions involved serving official disconnection notices to non-paying customers followed by disconnections in cases of non-payment (Coville et al. 2021). Critics argue that it is unethical to randomly assign households to this intervention since they have a moral claim to water service, regardless of their ability or willingness to pay.

In the context of clinical research, the concept of 'standard of care' provides a useful resource for articulating this type of concern, identifying the level and type of clinical care to which people are entitled and so the 'floor' or 'baseline' of care participants should be guaranteed in a study. Critics of the Nairobi RCT can be understood to be arguing that the standard of care is access to water services regardless of households' ability or willingness to pay. The researchers acted wrongly, in their view, since some participants were assigned to an intervention expected to be inferior to the standard of care. Indeed, if we adopt a normative baseline for the classification of harms, critics can be understood to be arguing that the disconnection intervention harms participants. If residents are entitled to water, disconnection makes them worse off than they have a right to be.

The need to provide people with the standard of care treatment is often justified by reference to the fiduciary obligation of physicians to promote their patients' health related interests

(Angell 1997). While public policy researchers do not have such obligations to residents, governments do. As I discuss above, governments have obligations to realize target outcomes for their residents, and to realize these outcomes, they must implement policies and laws. Governments therefore have an obligation to provide residents with a 'standard of care' and so would act wrongly by authorizing or conducting a policy experiment in which residents are subject to a policy that is reasonably expected to be inferior with respect to the realization of target outcomes.

How might we characterize this standard of care? First, since governments have an obligation to realize target outcomes, they should implement policies that can be reasonably expected to realize those outcomes, that is, that are evidence-based. Second, since residents have rights which place constraints on the ways in which governments realize their goals, governments should implement policies that are consistent with people's rights. Finally, since governments must realize multiple target outcomes – e.g. outcomes related to personal security, education, health, etc. - for any one target outcome, they must implement a policy that is consistent with the realization of others. For example, governments must not implement an evidence-based education policy that is so expensive that they lack the resources necessary to provide basic healthcare to all. Trade-offs are necessary.

Governments should therefore implement policies that are (1) best proven, (2) consistent with people's rights, and (3) consistent with the realization of other target outcomes, given a reasonably just system of resource procurement and allocation. With respect to any particular target outcome, they should therefore implement the *best proven morally and practically attainable and sustainable policy*, or, *BPA* (best proven attainable) policy for short (MacKay 2020, 323-324).

The BPA policy will of course vary from regime to regime since governments and societies differ in terms of their financial resources, bureaucratic capacity and expertise, cultural and political traditions, regime types, and social capital, among other factors. In resource-poor countries, the BPA policy for a specific sphere may be no policy for the government may lack the resources to implement even the most basic evidence-based intervention. The concept of a BPA policy is thus a formal concept – analogous to standard of care – which researchers may use to identify the level of benefits or care they must provide to research participants.

Returning to the Nairobi RCT, the concept of a BPA policy provides a set of questions with which we might determine whether the RCT was indeed unethical. The key question is whether the NCWSC has the fiscal and bureaucratic capacities to sustainably implement a policy of universal provision of water services without disconnection in the case of non-payment. If it does, this policy is the BPA policy and the study is arguably unethical; if it does not, as the researchers argued (Coville et al. 2021), this policy is not the BPA policy, and the study did not necessarily deny participants access to level of services to which they have a moral claim.

A further complication raised by this case but also by others is that governments are not unitary agents but rather complex institutions, featuring legislative, executive, and judicial branches as well as administrative agencies. Policy experiments are often conducted or authorized by specific government agencies – e.g. the NCWSC – having a set amount of resources and limited authority regarding the policies they implement. As such, they may not have the authority or resources to implement the BPA policy – the best proven policy that the government as a whole has the authority, resources, and bureaucratic capacity to enact. Perhaps the Kenyan state can provide water and sewerage services to all residents without concern for people's ability or willingness to pay, but the NCWSC can not.

In cases where the government institution or agency conducting or authorizing a policy experiment has the authority and resources (or resource procurement power) to decide whether to implement the BPA policy or not, experiments are unethical if participants are subjected to a policy reasonably expected to be inferior to the BPA policy. But things are more complex when the institution or agency conducting or authorizing a policy experiment lacks this authority and resources – as in the case of the NCWSC. An ethical norm forbidding agencies such as the NCWSC from authorizing or conducting policy experiments that subject residents to policies reasonably expected to be worse than the BPA policy may have the consequence of preventing socially valuable experiments – e.g. an experiment evaluating a promising intervention that, while inferior to the BPA policy, is potentially superior to the status quo.

In my view, such experiments are ethically permissible, provided the agency authorizing or conducting the study does not have the authority and resources to implement the BPA policy. The principal obligation of government agencies is to implement the most effective policies which they have the resources and authority to implement. Government agencies should not conduct, authorize, or fund policy experiments in which residents are subject to a policy reasonably expected to be inferior to this policy; but, forbidding their involvement with policy experiments under which people are subjected to policies inferior to the BPA policy would prevent socially valuable public policy research in which no one is made worse off than the agency in question has a duty to make them. I therefore suggest the following 'standard of care' requirement which accounts for the complexity of government institutions:

> *Standard of Care*: Government agency or institution A may assign participants to an
> intervention only if it is not reasonably expected to be inferior with respect to the realization

of target outcomes than the policy A has a duty to implement, that is, the most effective

policy which it has the authority and resources to sustainably implement.

The implications of this requirement for the Nairobi RCT are not clear without careful

consideration of the relevant empirical facts. But, if, as seems likely, the NCWSC has no duty to

provide water services to all regardless of ability or willingness to pay, there is a good case to be

made that this study satisfies this requirement.

This requirement does not apply to non-governmental agents conducting policy experiments

since it is grounded in the government's duty to realize target outcomes for its residents. Non-

governmental agents do have duties of beneficence though, and these duties have implications for

the interventions they offer to participants. Since these agents should promote participants' interests,

they should not offer participants an intervention that is reasonably expected to leave them worse

off than they would otherwise be. It may be permissible, therefore, for NGOs working in unjust

countries to evaluate promising interventions that, while reasonably expected to be inferior to the

BPA policy, are not ex ante inferior to the status quo. For example, many residents of the U.S.

currently lack health insurance because several states have refused to expand Medicaid. It is arguably

permissible for an NGO to pilot a form of catastrophic health insurance that while inferior to the

BPA policy – Medicaid – leaves participants ex ante no worse off than they would otherwise be.

5. Fair Randomization

Many policy experiments involve the random assignment of participants to an intervention.

While proponents of RCTs often suggest that it is fair to allocate access to a promising intervention

by lottery (in addition to being epistemically ideal), community members, among others, sometimes question this, wondering why the intervention should not instead be given to those who need it the most (Rayzberg 2019; Kombe et al. 2019). While RCTs may be socially valuable since they promise high-quality evidence regarding the causal efficacy of an intervention, access to interventions must be allocated *fairly* if a study is to be ethical.

RCTs are commonly used in clinical research and so there is a sizeable literature on the ethics of random assignment. Elsewhere, I make use of this work to outline a set of conditions under which random assignment is permissible (MacKay 2020). Appealing to Peter Stone's (2011) insight that lotteries are permissible when decision-makers are in a state of indeterminacy among the options before them, I argue that random assignment is permissible if: (1) government researchers occupy a state of indeterminacy regarding whether the intervention arm is superior to the control in realizing the relevant target outcomes; or (2) they occupy a state of indeterminacy regarding who should have access to a superior intervention. I discuss each of these conditions in turn.

The first type of indeterminacy is when researchers have no sufficiently strong reasons to think that either arm of the RCT is superior in terms of realizing the relevant target outcomes. In the simplest type of case, there is reasonable disagreement within the social science community regarding whether intervention A or intervention B is more effective in realizing target outcome X. There may also be more complex cases where there are multiple target outcomes at stake but where there is reasonable disagreement regarding whether A or B promises the best 'bundle' of outcomes. For example, researchers may be interested in a particular intervention for young children because it promises to improve both health and educational outcomes.

When researchers are indeterminate among the different arms of an RCT in this way, they are 'equally poised' between them and so occupy a state of 'equipoise.' Just as 'clinical' equipoise is a prominent norm in biomedical research (London Forthcoming), several scholars have proposed 'policy equipoise' as a condition under which random assignment in policy experiments is permissible (MacKay 2020; Abramowicz and Szafarz 2020; MacKay 2018; Petticrew et al. 2013; Baele 2013).

Because governments have a duty to ensure people are subject to BPA policies however, mere policy equipoise is not sufficient to render random assignment permissible. Instead, researchers must be equally poised among all arms of the RCT and the BPA policy. Because people are entitled to the level of target outcomes provided by the BPA policy, it would be wrong to randomize people to an intervention that is reasonably expected to be inferior. I formulate this sufficient condition for permissible randomization as follows (MacKay 2020, 329):

> Government agencies may randomly assign participants to different policy interventions if they are in a state of genuine equipoise regarding all arms of the study and the BPA policy.

Despite the centrality of the idea of equality to equipoise, most commentators reject the idea that the expert community must be equally divided on, or have equally strong reasons in favor of, the interventions in question. Such an interpretation of equipoise would be unworkable since it would be rarely met and easily disturbed. Instead, equipoise holds when the expert social scientific community is collectively uncertain about whether one intervention is more effective than another in realizing the relevant target outcomes, for example, when a minority of experts have good reasons to think A superior to B, even though the majority disagrees (London Forthcoming). As Asiedu, Karlan, Lambon-Quayefio, and Udry (2021, 2-3) put it, equipoise is satisfied when there is

"meaningful uncertainty" among experts, that is, "a degree of uncertainty deemed important and likely enough as to make reasonable and informed stakeholders disagree on the optimal policy." One task for future work is to provide a framework that researchers may use to determine when equipoise holds and when it has been disturbed.

While it is sometimes supposed that clinical equipoise is a necessary condition of permissible randomization in clinical research, policy equipoise need not be satisfied for randomization to be permissible (MacKay 2020). Randomization is also sometimes permissible when one arm of an RCT is preferable with respect to the realization of target outcomes, but the intervention is scarce. Where researchers are indeterminate regarding *who* should have access to the scarce intervention, it may be fair to employ a lottery (MacKay 2020, 333).

Commentators often appeal to this line of argument to justify randomization in cases where one arm of the study is (or is perceived to be) superior. For example, discussing the need to "avoid the potential of creating bad feeling due to the randomization," Abhijit V. Banerjee and Esther Duflo (2014, 101) write:

> 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 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.

While Banerjee and Duflo present this line of argument in an arguably cynical way – i.e. as a way to placate people complaining (perhaps rightfully) about the use of randomization – there is nonetheless something to it.

There are at least two types of cases where government agencies may have good reason to evaluate a scarce intervention using random assignment. First, a government agency may decide that a particular policy is the BPA policy but lack the financial and/or administrative capacity to cover all eligible people at the same time. It may wish to introduce randomization into the policy rollout to gather further evidence regarding its effectiveness. For example, beginning in 1998, the Mexican federal government conducted an RCT in its roll out of *Progresa*, a conditional cash transfer program, to determine its effectiveness more precisely (Schultz 2004). Second, the BPA policy may be scarce for the foreseeable future and the government agency responsible for implementing it may similarly wish to use random assignment to gather further evidence or experiment with policy 'tweaks.' The "Moving to Opportunity" experiment is a good example of this type of RCT, wherein the U.S. Department of Housing and Urban Development randomized eligible households to a standard Section 8 housing voucher, an experimental Section 8 housing voucher, and no voucher (Chetty, Hendren, and Katz 2016). Since the U.S. Congress does not provide sufficient funding to provide Section housing vouchers to all eligible households, it was deemed acceptable to randomize some households to a voucher and others to no voucher.

Banerjee and Duflo are correct to note that random assignment may be permissible in these types of cases, but their line of argument needs further development and qualification. First, the scarcity must be genuine – i.e. not due to injustice or lack of political will – and must be sufficient to allow a well-designed RCT to be conducted. People should not be denied access to the BPA policy for any period for the purposes of conducting an RCT (MacKay 2020, 345). Second, since lotteries

are a permissible way to allocate access to a scarce good if people have *equally strong claims* to the

good (Stone 2011, 52-53), eligible people must indeed have equally strong claims to the scarce BPA

policy (MacKay 2020, 345). This condition may be satisfied less often than researchers think, for if

some people have stronger claims to the policy than others, for example, on the grounds of need,

and it is possible to identify these individuals, it is reasonable to think that these people should be

targeted for access to the policy (Barrett and Carter 2014, 66).

Finally, the randomization procedure must recognize the claims of all people who are eligible

for the BPA policy (MacKay 2020, 345). For example, suppose there is only enough resources to

cover 25% of the eligible population with the BPA policy. Researchers would devalue people's

claims to the policy if they, for feasibility reasons, carried out an RCT in half the country, thus giving

people in this area a 50% chance of coverage, and people outside a 0% change of coverage. Instead,

the randomization procedure should grant all eligible people a 25% chance of coverage. Putting

these points together, there is thus an additional sufficient condition for permissible randomization

(MacKay 2020, 345-346):

> If the BPA policy is a scarce policy, government agencies may randomly assign participants
> to either the BPA policy or some alternative intervention or non-intervention (which may be
> inferior) if:

>> 1. No participant is subject to an intervention reasonably expected to be
>>   inferior to the non-BPA status quo policy;

>> 2. Participants have equally strong claims to be subject to the BPA policy; and

3. No person's claim to the BPA policy is devalued for the purposes of conducting the RCT.

These are not the only two conditions under which randomization is permissible. Elsewhere I formulates additional conditions to cover other types of cases, including conditions to cover non-ideal cases in which governments fail to implement the BPA policies, and I do not claim that even this expanded list is exhaustive (MacKay 2020). While too brief, I hope this discussion outlines the central ethical considerations relevant to the question of random assignment.

Governments have stronger duties to people than non-governmental organizations and so where it is permissible for government agencies to employ random assignment, it is no doubt permissible for non-governmental agencies to do so as well. However, I also think non-governmental agents may use random assignment under less stringent conditions.

First, because non-governmental agents have no duty to provide people with the BPA policy, they may permissibly randomize participants to interventions expected to be inferior to this policy. The only constraint, as discussed above in the previous section, is that people are not assigned to policies reasonably expected to make them worse off than they would be otherwise.

Second, while non-governmental agents have duties to treat to people equally, this duty is far weaker than that of governments. Governments have duties to ensure that residents receive the equal protection of law and policies. Where residents have equally strong claims to a scarce intervention therefore, governments must take special care to ensure that these claims are not devalued for arbitrary reasons. By contrast, people do not have claims of justice on the resources and programs of non-governmental agents. It is widely understood to be permissible for NGOs to offer programs and benefits to some people within a jurisdiction and not others. While non-

governmental agents have a duty of beneficence and so a duty to promote the interests of others, they also have discretion regarding whom and how they decide to help.

6. Fair Subject Selection

In both high-income and low-income countries, government policy experiments tend to enroll people who are recipients of social safety net programs and so politically and economically disadvantaged. Several scholars have suggested that this disproportionate enrolment of vulnerable people raises questions of justice (Teele 2014, 133; Blustein 2005, 838). As Angus Deaton (2020, 43) puts it, even "in the US, nearly all RCTs on the welfare system are RCTs done *by* better-heeled, better-educated, and paler people *on* lower income, less-educated and darker people."

To determine if and when this disproportionate enrollment disadvantaged people in government policy experiments is defensible, it is helpful to turn to fair subject selection, a widely recognized requirement of ethical clinical research (Emanuel, Wendler, and Grady 2000, 2704; National Commission for the Protection of Human Subjects of Clinical and Behavioral Research 1979, 9). The selection of participants has implications for the distribution of the benefits and burdens of clinical research – including burdens associated with participation – and so researchers must recruit and select participants with an eye towards ensuring that these benefits and burdens are distributed fairly (MacKay and Saylor 2020, 6). This requirement applies to public policy research too, for decisions regarding where to carry out an experiment and with whom influences the distribution of these benefits and burdens. I suggest here that fair subject selection has three key dimensions for public policy research.

First, the chief benefit of public policy research is generalizable policy relevant knowledge. The selection of participants is directly relevant to the distribution of this benefit since the selection of participants for a study has implications for the generalizability of any resulting knowledge. The knowledge generated by an experiment will be most valuable for those populations most similar to the participants, and so the selection of participants has implications for which groups in society benefit from the research (MacKay and Saylor 2020, 6). Since governments have a duty to produce knowledge that is relevant to the realization of target outcomes for residents, study sites should be chosen, and participants recruited and selected, with an eye towards ensuring that any resulting knowledge will be generalizable to those diverse populations that are the object of these duties. A chief dimension of fair subject selection is thus 'fair inclusion:'

> Fair Inclusion: The selection of research participants must be sufficiently inclusive to ensure that the research in question fairly benefits members of society (MacKay and Saylor 2020, 7).

To comply with this requirement, researchers must choose study sites and select participants in ways that are inclusive of distinct groups. The population enrolled in research should reflect the diversity of the population which is likely to be subject to any subsequent intervention. As Rachel Glennerster and Shawn Powers (2016, 373) point out, this can be achieved either by conducting RCTs on "large representative samples" or studying the same policy in different contexts.

For example, the Canadian province of Ontario recently initiated a Basic Income Pilot Project, randomly assigning participants to either a basic income intervention or the status quo social safety net policy (Government of Ontario). Because any subsequent basic income policy would apply to *all* low-income Ontarians aged 18-64, researchers conducted the pilot in multiple

communities (including mid-size, urban, rural, and urban/rural mixed communities) to ensure the pilot would be "representative of Ontario's population," and matched the inclusion criteria for the study with the likely eligibility criteria of the experimental intervention (Government of Ontario).

In addition to having implications for site selection and inclusion criteria, fair inclusion also has implications for study recruitment. Researchers should cast a wide net and take proactive steps to ensure that the resultant set of participants is sufficiently inclusive. Ontario's experiment faced challenges in this regard; due to privacy restrictions regarding the sharing of tax data and the fact that many low-income Ontarians do not file tax returns, researchers faced serious obstacles in the recruitment of a representative sample (Mason 2018).

A more direct way that the selection of participants influences the distribution of the benefits of policy research is when participation is reasonably expected to be net beneficial for participants (MacKay and Saylor 2020, 6). Where participation is a benefit, the selection of participants directly allocates it and so must be carried out in a fair manner. Participation may be ex ante net beneficial in cases where the intervention under study is expected to be superior to the status quo policy for the realization of target outcomes. For example, in the Ontario Basic Income Pilot Project, the basic income intervention was far more generous than the status quo policy, offering a maximum cash transfer of $16,989/year for individuals compared to $8,472 (Government of Ontario). Ex ante, participants therefore had a 50% chance of being assigned to a superior intervention.

In cases where participation is ex ante net beneficial therefore, researchers must ensure that prospective participants have a 'fair opportunity' to participate:

Fair Opportunity: Prospective participants must be granted a fair opportunity to participate in public policy research that is expected to be net beneficial (MacKay and Saylor 2020, 9).

This dimension of fair subject selection forbids researchers from selecting study sites or designing the inclusion and exclusion criteria in ways that exclude prospective participants for arbitrary reasons and requires them to cast a wide net when recruiting participants, for example, through the placement of advertisements.

Participation may not always be ex ante net beneficial for participants, however. Even in cases where participants are not subject to a policy that is reasonably expected to be worse than the status quo, participation may involve harmful and/or time-consuming data collection procedures such as lengthy surveys, focus groups, or interviews requiring participants to recall traumatic events. In these cases, participation may be ex ante net burdensome and researchers must ensure that the burden of participation is shared fairly:

Fair Burden Sharing: The burdens of participation in public policy research must be shared fairly (MacKay and Saylor 2020, 8).

According to The Belmont Report, this dimension of fair subject selection requires that researchers should not "select only 'undesirable' persons for risky research" (National Commission for the Protection of Human Subjects of Clinical and Behavioral Research 1979, 9) and that participation in burdensome research should be allocated based on prospective participants' ability to bear the burden in question. The above-mentioned concern that politically and economically disadvantaged people are disproportionately represented in policy experiments can be expressed as the concern that the burdens of participation in policy research are not being fairly shared within a population. I comment further on this concern below.

Fair subject selection is thus a multidimensional requirement. To select participants fairly, researchers must comply with the principles of fair inclusion, fair opportunity, and fair burden sharing. Unfortunately, these principles may yield conflicting imperative, requiring researchers to make tradeoffs. For example, fair inclusion may conflict with fair burden sharing, requiring that researchers include socio-economically disadvantaged people who are less able to bear the burdens of research. It may also conflict with fair opportunity, requiring that researchers target prospective participants with specific backgrounds to ensure that the participant pool is sufficiently inclusive.

With Katherine Saylor, I argue elsewhere that where these conflicts arise, fair inclusion should take precedence (MacKay and Saylor 2020). The central goal of research is to develop socially valuable knowledge that fairly benefits the public and fair inclusions facilitates the realization of this goal. We formulate a decision-procedure researchers may follow to ensure participants are selected fairly:

1. Design inclusion and exclusion criteria to answer the scientific question in a way that fairly benefits members of society (fair inclusion).

2. Among potential participants meeting inclusion criteria and not meeting exclusion criteria, set and meet goals for enrollment of potential participants to ensure research fairly benefits members of society (fair inclusion).

3. Consistent with step 2, ensure all prospective participants satisfying inclusion and exclusion criteria have a fair opportunity to participate, and refrain from targeting disadvantaged prospective participants for research which is ex ante net burdensome (fair opportunity and fair burden sharing).

Step 3 is applicable to studies for which it is difficult to say whether it is either ex ante net beneficial or ex ante net burdensome for all participants. The chief concerns with the principles of fair opportunity and fair burden sharing are that the benefit or burden of participation be widely shared among eligible participants and that disadvantaged participants not be targeted. Step 3 is worded to incorporate these concerns.

Does the above analysis and decision-procedure apply to non-governmental policy experiments? For the most part, yes. Non-governmental agents too have a duty to treat prospective participants equally and not to inflict harm without justification. They are therefore bound by the principles of fair opportunity and fair burden sharing. To the extent that they are bound by the social value requirement (see above), they should also comply with fair inclusion. The above decision-procedure also offers non-governmental agents a defensible way to work through conflicts among these principles.

The one difference for non-governmental agents conducting policy experiments is that they don't have the same robust duties of distributive justice possessed by governments. It may be permissible, therefore, for non-governmental organization to be concerned with solving a policy problem for a highly specific population. As such, it may be permissible for NGOs to interpret the fair inclusion requirement in a narrower fashion than governments. For example, in the context of educational policy, governments have a responsibility to realize target outcomes for all students and so should fund a portfolio of research that is responsive to this responsibility and that includes a diverse range of students. By contrast, an NGO may be devoted to improving educational outcomes for undocumented children, for example, and so may permissibly conduct pilot projects that are more limited in terms of the student populations included in the research.

7. Informed Consent

Many public policy experiments do not secure the informed consent of their participants, either because doing so is not feasible or would introduce certain forms of bias, for example, selection bias and Hawthorne effects. Informed consent by participants or a surrogate decision-maker – e.g. in the case of research with children – is almost always necessary for clinical research, and so some scholars argue that policy experiments which do not secure participants' informed consent are morally concerning, if not deeply unethical (Hoffman 2020a; Hoffman 2020b; Barrett and Carter 2014, 64-65; Teele 2014, 124-128; Baele 2013). I suggest here that things are more complex. Sometimes informed consent is ethically necessary for participation and/or data collection; but there are also circumstances where it is not.

According to *The Belmont Report*, people give informed consent to participate in a study if and only if: (1) researchers disclose the materially relevant information regarding the study to participants; (2) participants comprehend this information; and (3) participants voluntarily agree to participate, free of coercion and undue influence (The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research 1979, 6-7). The requirement of informed consent is grounded in the principle of respect for persons. Joseph Millum and Danielle Bromwich (Forthcoming, 1) helpfully interpret this principle as a form of recognition respect for rights-holders: we respect persons when we recognize their moral status as rights-holders. For people who are competent, this means respecting their autonomy rights, that is, their rights to exercise self-determination (Millum and Bromwich Forthcoming). For people who are not competent, this means respecting their rights not to be harmed.

Because competent people have autonomy rights, researchers must often secure people's valid informed consent prior to enrolling them in research. Clinical studies often involve the ingestion of pharmaceutical interventions and blood draws, and since people's autonomy rights include a right to control what happens to their body, informed consent is necessary to make these otherwise wrongful actions rightful. For noncompetent persons, respect for persons entails that their rights not to be harmed be respected, and that consent be secured by a surrogate decision-maker.

But, if the principal reason for securing informed consent is to respect people's autonomy rights, it is not necessary in cases where research does not infringe such rights (MacKay and Chakrabarti 2019; Gelinas, Wertheimer, and Miller 2016). This matters for government policy experiments. As I argue below, the subjection of people to a policy intervention as part of a study need not infringe people's autonomy rights. Moreover, while certain forms of data collection do infringe people's autonomy rights and so require informed consent, not all forms of data collection do so – e.g. the collection of administrative data.

People's autonomy rights impose limits on governments' legitimate sphere of policymaking, carving out a sphere of action in which individuals are sovereign. Governments should not therefore pass laws that impinge people's basic freedoms – e.g. freedom of religion, liberty of the person, freedom of expression, etc. But legitimate governments have a 'right to rule' and so are themselves sovereign over certain spheres of action (MacKay and Chakrabarti 2019, 191). Political philosophers disagree on exactly how to demarcate these spheres of sovereignty, and some argue that no state can be legitimate (Peter 2017), but within liberal democracies, governments are widely recognized to have the constitutional authority to rule over many areas, including welfare, health, and education,

among others. When legitimate governments implement laws and policies within the spheres of action within which they are sovereign, they do not infringe people's autonomy rights.

Recall the Ontario Basic Income Pilot. Researchers secured the informed consent of participants for this study but had the Legislative Assembly of Ontario simply decided to enact the policy, it would not have needed Ontarians' informed consent to do so. The Legislative Assembly has a right to rule over welfare policy, and just as it has the right to implement a basic income policy permanently without the informed consent of Ontarians, it arguably does not need their informed consent to permissibly evaluate this intervention by means of an experiment.

One might object that citizens of liberal democracies do have autonomy rights over policymaking since they have democratic rights. But consider that these rights are different from the autonomy rights people possess. Democratic rights include the right to vote, the right to run for political office, and the right to participate in public discussions. Whereas autonomy rights grant individuals sovereignty over the use of their body and mind, democratic rights only grant people the right to participate in the process of collective will-formation. Residents do not therefore have a right to control the government's actions, such that it could not pass a law or carry out a public project without securing the consent of *all* residents prior to enacting a law or carrying out a public project (MacKay and Chakrabarti 2019, 193). Governments do not therefore violate residents' autonomy rights when they exercise their right to rule without securing the consent of all residents, nor do they violate residents' democratic rights, provided they comply with the appropriate procedures.

Now, the fact that legitimate governments have a right to rule over certain policy spheres does not entail that government agencies need not secure people's consent when conducting or

authorizing an experiment concerning these spheres. First, governments are not unitary agents and different institutions that may conduct or authorize research – e.g. legislative assemblies and administrative agencies – have differing rights to rule. The authority of the U.S. Congress is sufficiently wider than that of the Department of Housing and Urban Development. If a government agency wishes to introduce a novel policy intervention as part of an RCT or pilot project without seeking people's consent, it must in fact have the right to rule over the sphere of policy in question. If it does not, the experiment may infringe people's autonomy rights and may also infringe their democratic rights – e.g. if the agency in question is exceeding its constitutional authority. There may therefore be cases where it is permissible for a legislative body to conduct or authorize a policy experiment without people's consent but not an administrative agency.

Second, even if a government agency has the right to introduce an experimental policy intervention, it may not be possible to collect the data it requires to evaluate the intervention without violating people's autonomy rights. If the agency needs to conduct surveys or interviews to evaluate the intervention, it will need to secure people's informed consent to do so. The mere fact that governments need not secure people's informed consent to permissibly subject them to a policy intervention need not imply that informed consent is not necessary for data collection.

Finally, while respect for autonomy rights is the principal reason to seek informed consent, it is not the only one. Neal Dickert et al. (2017, 4) argue that informed consent secures six purposes in addition to respecting people's autonomy rights, namely, "providing transparency," "promoting concordance with participants' values," "protecting participants' welfare interests," "promoting trust," "satisfying regulatory requirements," and "promoting the integrity of research and researchers." Government agencies do not necessarily act wrongly by failing to realize these purposes via informed consent, but the fact that informed consent processes promote values other

than respect for autonomy rights is a reason not to abandon them absent justification. Researchers should have good reasons not to seek informed consent, for example, because doing so is not feasible, would be too expensive, or threaten the scientific validity of the study through the introduction of various forms of bias.

Pulling these threads together, I suggest that government agencies conducting or authorizing policy research need not secure people's consent when the following conditions are satisfied:

1. The government institution conducting or authorizing the research possesses a right to rule over the spheres of policy targeted by the research;

2. Data collection does not involve the infringement of participants' autonomy rights; and

3. There is a strong justification for not obtaining participants' informed consent (MacKay and Chakrabarti 2019, 196).

While research that satisfies 1-2 does not infringe people's autonomy rights, Luke Gelinas, Alan Wertheimer, and Franklin G. Miller (2016) argue convincingly that research without consent is also sometimes permissible when the research involves the infringement of people's autonomy rights. Since rights are not absolute, there may be cases where the infringement is minor enough, and the social value of the research large enough, that it may be permissible to conduct research without consent. Government policy research without consent is thus also permissible when:

A. There is a strong justification for not obtaining participants' informed consent;

B. The infringement of participants' autonomy rights is minor; and

C. The expected social value of the research is sufficient to outweigh the infringement of participants' rights (Gelinas, Wertheimer, and Miller 2016, 36-37).

Conditions 1-3 do not apply to non-governmental agents since they have no right to rule over people. NGOs conducting a pilot project may therefore need to secure people's consent for data collection and to subject them to an intervention in the first place. Conditions A-C may sometimes license non-governmental agents to conduct research without informed consent, but great care should be taken in interpreting these conditions to avoid infringing people's autonomy rights without justification.

8. Community Engagement

As I note above, a central concern with government policy experiments is that they involve powerful institutions experimenting on politically and economically disadvantaged individuals, often without their consent or input. One might worry that my analysis thus far fails to address this concern and even provides researchers with the means to justify the continuation of this problematic practice. After all, I argue above that informed consent is not necessary in many cases, and that the disproportionate enrollment of disadvantaged people is justified when their inclusion is necessitated by the problem under study. In this last section, I defend a requirement for policy experiments that does address these concerns.

'Community engagement' is an emerging norm for clinical and public health research (Anderson and Spellecy Forthcoming; Pratt and de Vries 2018). It involves researchers collaborating with research stakeholders – e.g. patients, participants, community members, healthcare providers,

policymakers, and/or local organizations – to design and/or carry out various aspects of the research (Anderson and Spellecy Forthcoming). Community engagement exists along a spectrum, ranging from outreach on one end, through consultation, collaboration, and shared leadership on the other (Anderson and Spellecy Forthcoming; Pratt and de Vries 2018, 459-460). Stakeholders may serve on advisory boards or provide consultation on specific aspects of a study, or they may be involved in "study-specific tasks such as participant recruitment, focus group facilitation, or leading informal events such as town halls and community forums (Anderson and Spellecy Forthcoming)."

There are four prominent arguments in favor of community engagement for clinical and public health research. The first two arguments are instrumental ones. First, community engagement promotes socially valuable research by helping researchers design and conduct studies that address real problems, identify promising interventions, yield valid data, and generate trust in downstream implementation. Prospective participants may have unique lived experience regarding the problems they face and interventions that may be useful and involving them and community members in study design and data analysis may improve recruitment, retention, the validity and reliability of data, and trust in the research results (Anderson and Spellecy Forthcoming; Dickert and Sugarman 2005, 1125; Emanuel, Wendler, Killen, and Grady 2004). Second, community engagement leads to better protection of research participants, for community members may be better able to identify individual and community risks and help improve the quality of the informed consent process (Anderson and Spellecy Forthcoming; King et al. 2014, 2; Dickert and Sugarman 2005, 1124; Emanuel, Wendler, Killen, and Grady 2004).

The remaining two arguments in support of community engagement are non-instrumental. First, community engagement is necessary, particularly with respect to public health research, to respect people's democratic rights (Pratt and de Vries 2018, 457; King et al. 2014, 3; Dickert and

Sugarman 2005, 1125). People have rights to participate in and/or influence decision-making processes that affect their interests and community engagement processes, involving public outreach and consultation, are an important way to respect these rights.

Community engagement is also sometimes necessary, second, to respect the collective rights of communities. Where researchers wish to conduct research with people from a different political or cultural community, community engagement is necessary to respect the community's collective right to determine what happens within its jurisdiction and to its people (Emanuel, Wendler, Killen, and Grady 2004; Weijer 1999). For example, researchers from high-income countries wishing to conduct research in low-income countries should consult with and seek the consent of representatives of these countries (Emanuel, Wendler, Killen, and Grady 2004). Similarly, researchers wishing to conduct research with indigenous communities have a similar obligation to consult with and seek the consent of these communities to respect their rights to self-determination (Brunger, Russell, and Wall Forthcoming).

Do these four arguments hold for government policy research? Since the first two arguments are instrumental ones, whether they hold or not depends on complex empirical factors. Nonetheless, it is plausible to think that community engagement processes would enhance the social value of policy research and promote the protection of participants (Evans 2021, 3; Ouma 2020; Davis 2020; Kombe et al. 2019; Barrett and Carter 2014, 64-65). The latter two arguments provide strong, non-contingent support for a community engagement requirement. First, governments have a duty to respect and fulfill the democratic rights of their residents and so agencies, when planning and conducting policy experiments, should inform affected residents of the planned research and provide participants and affected residents with a forum within which they may voice their concerns. Since governments have a right to rule, agencies need not seek the *consent* of residents likely to be

affected by a proposed experiment and it is not desirable that they do so, for there are significant

drawbacks to direct citizen control over government agencies (Heath 2020, 72-78). But agencies do

have a duty to inform residents of the proposed research, hear residents' concerns, and respond to

them (McDermott and Hatemi 2020, 30019). Similarly, in the U.S., federal agencies provide citizens

with a process by which they may comment on proposed rules and regulations.

Second, where government researchers are planning research with members of a distinct

political or cultural community, it is imperative that they respect any collective rights of these

groups. For example, aid agencies from high-income countries conducting or authorizing research in

low-income countries should work collaboratively with representatives of these communities,

respecting the rights of the latter to make important decisions regarding the problems to be studied

and the interventions to be tried.

These two arguments support a community engagement requirement for government policy

research. The content of this requirement flows from the need to respect people's democratic rights

and the self-government rights of communities. At minimum, it therefore requires that agencies and

researchers: (1) provide materially relevant information to affected individuals and provide them

with an opportunity for meaningful feedback; and where community rights are in play, (2) secure

consent from the authorized community representatives regarding the design and conduct of the

research. The instrumental arguments discussed above, moreover, suggest that more robust

community engagement practices, though not ethically required in all cases, may nonetheless be

effective ways to protect participants and communities and enhance the social value of the research.

Are non-governmental agents also bound by a community engagement requirement? Such

agents certainly have a duty to respect the collective rights of communities, where these rights are

relevant. But it's not clear that such agents have the same duties to respect and fulfill the democratic rights of residents, and so that NGOs, for example, have a robust duty to inform and consult with all community members potentially affected by a proposed study. This need not be terribly worrying, however, since as I discuss above such agents face much stricter requirements regarding informed consent than government agencies.

Conclusion

My aim in this chapter has been to use the clinical research ethics literature as a resource to provide guidance to researchers and sponsors on the ethical problems they are likely to encounter when planning and conducting policy experiments. I have focused on a set of issues that have received a good deal of attention in scholarly and public discussions of the ethics of policy experiments, formulating guidance for issues of social value, risk-benefit assessments, standard of care, fair randomization, fair subject selection, informed consent, and community engagement.

Unfortunately, there is still much to discuss. First, while I have covered the conditions under which informed consent is and is not necessary, guidance is also necessary on the conditions that must be satisfied for consent to be morally transformative. For example, must participants fully comprehend all aspects of the research for their consent to be valid? Second, what is the scope of researchers' responsibility when they partner with government agencies? Relatedly, should they offer their expertise in cases where aspects of the research design or policy – over which they have little control – are unethical (see Humphreys 2015; and Asiedu et al. 2021)? Third, should participants be paid? If so, what level of pay is appropriate and for what reasons should they be compensated? Is it permissible to pay participants in-kind – e.g. with training programs, stationary, or food stuffs –

rather than with cash? Fourth, what conditions must be satisfied to ensure that researchers and research sponsors do not exploit participants and local communities? Is it enough to pay participants, or must research be responsive to the needs and priorities of local communities and/or offer benefits to them? Finally, while I have commented on the ethical obligations of both governments and private organizations, my analysis is weighted towards the former. More work is needed to carefully think through the ethical obligations of private organizations such as NGOs and outline their implications for the design and conduct of policy experiments.

As with many of the topics discussed above, the clinical ethics literature may be helpful in addressing these remaining questions (see Millum and Bromwich 2021; Gelinas et al. 2018; Shah, Wolitz, and Emanuel 2013; El Setouhy at al. 2004). While my chapter does not address these additional issues, I hope it will introduce several new concepts and considerations to discussions of the ethics of policy experiments and spark a broader conversation.

Bibliography

Abramowicz, Michel, and Ariane Szafarz. 2020. "Ethics of RCTs: Should Economists Care about Equipoise?" In *Randomized Control Trials in the Field of Development: A Critical Perspective*, edited by Florent Bédécarrats, Isabelle Guérin, and François Roubaud, 280-292. New York: Oxford University Press.

Angell, Marcia. 1997. "The Ethics of Clinical Research in the Third World." *The New England Journal of Medicine* 337, no. 12: 847-849.

Asiedu, Edward, Dean Karlan, Monica P. Lambon-Quayefio, and Christopher R. Udry. 2021. "A

Call for Structured Ethics Appendices in Social Science Papers." *PNAS* 118, no. 29: e2024570118.

Baele, Stéphane J. 2013. "The Ethics of New Development Economics: Is the Experimental

Approach to Development Economics Morally Wrong?" *The Journal of Philosophical Economics* VII, no.

1: 2-42.

Baldassarri, Delia, and Maria Abascal. 2017. "Field Experiments Across the Social Sciences." *Annual*

Review of Sociology 43, no. 1: 41-73.

Banerjee, Abhijit V., and Esther Duflo. 2014. "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, 78-114. New Haven: Yale University Press.

Barrett, Christopher B., and Michael R. Carter. 2014. "A Retreat from Radical Skepticism:

Rebalancing Theory, Observational Data, and Randomization in 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, 58-77. New Haven: Yale University Press.

Blattman, Christopher, Alexandra C. Hartman, and Robert A. Blair. 2014. "How to Promote Order

and Property Rights under Weak Rule of Law? An Experiment in Changing Dispute Resolution

Behavior through Community Education." *American Political Science Review* 108, no. 1: 100-120.

Blustein, Jan. 2005. "Toward a More Public Discussion of the Ethics of Federal Social Program

Evaluation." *Journal of Policy Analysis and Management* 24, no. 4: 824-846.

Brunger, Fern, Todd Russell, and Darlene Wall. Forthcoming. "Decolonizing Research Ethics: Best Practices for the Ethical Conduct of Research Involving Indigenous Peoples." In *The Oxford Handbook of Research Ethics*. edited by Ana Iltis and Douglas MacKay. New York: Oxford University Press.

Chetty, Raj, Nathaniel Hendren, and Lawrence F. Katz. 2016. "The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment." *American Economic Review* 106, no 4: 855-902.

Coville, Aidan, Sebastian Galiani, Paul Gertler, and Susuma Yoshida. 2021. "Financing Municipal Water and Sanitation Services in Nairobi's Informal Settlements." Policy Research Working Paper, no. WPS 9725, Impact Evaluation Series. Washington, D.C. : World Bank Group.

Crouch, Robert A. and John D. Arras. 1998. "AZT Trials and Tribulations." *The Hastings Center Report* 28, no. 6: 26-34.

Davis, Justine M. 2020. "Manipulating Africa? Perspectives on the Experimental Method in the Study of African Politics." *African Affairs* 119, no. 476: 452-467.

Deaton, Angus. 2020. "Introduction: Randomization in the Tropics Revisited, a Theme and Eleven Variations." In *Randomized Control Trials in the Field of Development: A Critical Perspective*, edited by Florent Bédécarrats, Isabelle Guérin, and François Roubaud, 29-46. New York: Oxford University Press.

Deaton, Angus, and Nancy Cartwright. 2018. "Understanding and Misunderstanding Randomized Controlled Trials." *Social Science & Medicine* 210: 2-21.

Desposato, Scott. Forthcoming. "The Ethical Challenges of Political Science Field Experiments." In *The Oxford Handbook of Research Ethics*. edited by Ana Iltis and Douglas MacKay. New York: Oxford University Press.

de Souza Leão, Luciana, and Gil Eyal. 2019. "The Rise of Randomized Controlled Trials (RCTs) in International Development in Historical Perspective." *Theory and Society* 48, no. 3: 383-418.

Dickert, Neal, and Jeremy Sugarman. 2005. "Ethical Goals of Community Consultation in Research." *American Journal of Public Health* 95, no. 7: 1123-1127.

Dupas, Pascaline, Basimenya Nhlema, Zach Wagner, Aaron Wolf and Emily Wroe. 2021. "Expanding Access to Clean Water for the Rural Poor: Experimental Evidence from Malawi." *Poverty Action Lab Working Paper*, April 2021.

El Setouhy, Maged, Tsiri Agbenyega, Francis Anto, Christine Alexandra Clerk, Kwadwo A. Koram, Michael English, Rashid Juma, Catherine Molyneux, Norbert Peshu, Newton Kumwenda, Joseph Mfutso-Bengu, Malcolm Molyneux, Terrie Taylor, Doumbia Aissata Diarra, Saïbou Maiga, Mamadou Sylla, Dione Youssouf, Catherine Olufunke Falade, Segun Gbadegesin, Reidar Lie, Ferdinand Mugusi, David Ngassapa, Julius Ecuru, Ambrose Talisuna, Ezekiel Emanuel, Christine Grady, Elizabeth Higgs, Christopher Plowe, Jeremy Sugarman, and David Wendler. 2004. "Moral Standards for Research in Developing Countries from 'Reasonable Availability' to 'Fair Benefits.'" *Hastings Center Report* 34, no. 3 (2004): 17-27.

Emanuel, Ezekiel J., David Wendler, Jack Killen, and Christine Grady. 2004. "What Makes Clinical Research in Developing Countries Ethical? The Benchmarks of Ethical Research." *The Journal of Infectious Diseases* 189, no. 5: 930-937.

Evans, David K. 2021. "Towards Improved and More Transparent Ethics in Randomised Controlled Trials in Development Social Science." CGD Working Paper 565. Washington, DC: Center for Global Development.

Filmer, Deon P., Jed Friedman, Eeshani Kandpal, and Junko Onishi. 2018. "General Equilibrium Effects of Targeted Cash Transfers: Nutrition Impacts on Non-Beneficiary Children." World Bank Policy, Working Paper No. WPS8377.

Final Report of the Seattle-Denver Income Maintenance Experiment (Washington, D.C.: U.S. Government Printing Office, 1983)

Gelinas, Luke, Alan Wertheimer, and Franklin G. Miller. 2016. "When and Why Is Research Without Consent Permissible?" *Hastings Center Report* 46, no. 2: 35-43.

Gelinas, Luke, Emily A. Largent, I. Glenn Cohen, Susan Kornetsky, Barbara E. Bierer, and Holly Fernandez Lynch. 2018. "A Framework for Ethical Payment to Research Participants." *The New England Journal of Medicine* 378, no. 8: 766-771.

Glennerster, Rachel, and Shawn Powers. 2016. "Balancing Risk and Benefit: Ethical Tradeoffs in Running Randomized Evaluations." In *The Oxford Handbook of Professional Economic Ethics*, edited by George DeMartino and Deirdre McCloskey, 367-401. New York: Oxford University Press.

Government of Ontario. "Ontario Basic Income Pilot." https://www.ontario.ca/page/ontario-basic-income-pilot.

Greenberg, David, and Mark Shroder. 2004. *The Digest of Social Experiments*, Third Edition. Washington, DC: The Urban Institute Press.

Gubler, Joshua R., and Joel S. Selway. 2016. "Considering the Political Consequences of

Comparative Politics Experiments." In *Ethics and Experiments: Problems and Solutions for Social Scientists

and Policy Professionals*, edited by Scott Desposato, 171-182. New York: Routledge.

Gueron, Judith M., and Howard Rolston. 2013. *Fighting for Reliable Evidence*. New. York: Russell Sage

Foundation.

Heath, Joseph. 2020. *The Machinery of Government: Public Administration and the Liberal State*. New York:

Oxford University Press.

Hoffman, Nimi. 2020a. "Involuntary Experiments in Former Colonies: The Case for a

Moratorium." *World Development*. 127, March: 1-3.

Hoffman, Nimi. 2020b. "Involuntary Social Experimentation: Revisiting the Case for a

Moratorium." *CODESRIA Bulletin*, no. 1: 7-14.

Humphreys, Macartan. 2015. "Reflections on the Ethics of Social Experimentation." *Journal of

Globalization and Development* 6, no. 1: 87-112.

Karlan, Dean, and Jacob Appel. 2011. *More Than Good Intentions: How a New Economics is Helping to

Solve Global Poverty*. New York: Dutton.

King, Katherine F., Pamela Kolopack, Maria W. Merritt, and James V. Lavery. 2014. "Community

Engagement and the Human Infrastructure of Global Health Research." *BMC Medical Ethics* 15, no.

84: 1-6.

Kombe, Maureen Mupeta, Joseph Mumba Zulu, Charles Michelo and Ingvild F. Sandøy. 2019. "Community Perspectives on Randomisation and Fairness in a Cluster Randomised Controlled Trial in Zambia." *BMC Medical Ethics* 20, no. 99.

Leigh, Andrew. 2018. *Randomistas: How Radical Researchers are Changing Our World*. New Haven: Yale University Press.

Levitt, Steven D., and John A. List. 2009. "Field Experiments in Economics: The Past, the Present, and the Future." *European Economics Review* 53, no. 1: 1-18.

London, Alex John. Forthcoming. "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.

Lurie, Peter, and Sidney M. Wolfe. 1997. "Unethical Trials of Interventions to Reduce Perinatal Transmission of the Human Immunodeficiency Virus in Developing Countries," *The New England Journal of Medicine* 337, no. 12: 853-855.

MacKay, Douglas. 2015. "Standard of Care, Institutional Obligations, and Distributive Justice." *Bioethics* 29, no. 4: 262-273.

MacKay, Douglas. 2018. "The Ethics of Public Policy RCTs: The Principle of Policy Equipoise." *Bioethics* 32, no. 1: 59-67.

MacKay, Douglas. 2020. "Government Policy Experiments and the Ethics of Randomization." *Philosophy & Public Affairs* 48, no. 4: 319-352.

MacKay, Douglas, and Averi Chakrabarti. 2019. "Government Policy Experiments and Informed Consent." *Public Health Ethics* 12, no. 2: 188-201.

MacKay, Douglas, and Katherine Witte Saylor. 2020. "Four Faces of Fair Subject Selection." *The American Journal of Bioethics* 20, no. 2: 5-19.

Macklin, Ruth. 2004. *Double Standards in Medical Research in Developing Countries.* Cambridge: Cambridge University Press.

MacPhail, Catherine, Michelle Adato, Kathleen Kahn, Amanda Selin, Rhian Twine, Samson Khoza, Molly Rosenberg, Nadia Nguyen, Elizabeth Becker, and Audrey Pettifor. 2013. "Acceptability and Feasibility of Cash Transfers for HIV Prevention Among Adolescent South African Women." *AIDS and Behavior* 17, no. 7: 2301-2312.

Mason, Gregory C. 2018. "Ontario's Scrapped Basic-Income Pilot Project Was Actually Deeply Flawed." *Macleans.* August 10. https://www.macleans.ca/opinion/ontarios-scrapped-basic-income-pilot-project-was-actually-deeply-flawed/.

McDowell, Tom, and Mohammad Ferdosi. Forthcoming. "The Impacts of the Ontario Basic Income Pilot: A Comparative Analysis of the Findings from the Hamilton Region." *Basic Income Studies.*

Millum, Joseph, and Danielle Bromwich. 2021. "Informed Consent: What Must Be Disclosed and What Must Be Understood?" *The American Journal of Bioethics* 21, no. 5: 46-58.

Millum, Joseph, and Danielle Bromwich. Forthcoming. "Respect for Persons." In *The Oxford Handbook of Research Ethics,* edited by Ana Iltis and Douglas MacKay. New York: Oxford University Press.

Ogden, Timothy N. 2016. "Introduction." in Experimental Conversations: Perspectives on Randomized Trials in Development Economics, edited by Timothy N. Ogden, ix-xxxii. Cambridge, MA: The MIT Press.

Ouma, Marion. 2020. "Trust, Legitimacy and Community Perceptions on Randomisation of Cash Transfers." *CODESRIA Bulletin*, no. 1: 25-29.

Peter, Fabienne 2017. "Political Legitimacy." *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, https://plato.stanford.edu/archives/sum2017/entries/legitimacy/.

Petticrew, M., M. McKee, K. Lock., J. Green., and G. Phillips. 2013. "In Search of Social Equipoise." *BMJ* 346: f4016.

Phillips, Trisha. 2021. "Ethics of Field Experiments." *Annual Review of Political Science* 24: 277-300.

Pierson, Leah, and Joseph Millum. 2018. "Health Research Priority Setting: The Duties of Individual Funders." *The American Journal of Bioethics* 18, no. 11: 6-17.

Ravallion, Martin. 2020. "Should Randomistas (Continue to) Rule?" In *Randomized Control Trials in the Field of Development: A Critical Perspective*, edited by Florent Bédécarrats, Isabelle Guérin, and François Roubaud, 47-78. New York: Oxford University Press.

Rawls, John. 1999. *A Theory of Justice*. Revised Edition. Cambridge, MA: Harvard University Press, 1999.

Rayzberg, Margarita S. 2019. "Fairness in the Field: The Ethics of Resource Allocation in Randomized Controlled Field Experiments." *Science, Technology, & Human Values* 44, no. 3: 371-398.

Rid, Annette, and David Wendler. 2011. "A Framework for Risk-Benefit Evaluations in Biomedical Research." *Kennedy Institute of Ethics Journal* 21, no. 2: 141-179.

Schuler, Sidney Ruth, Syed M. Hashemi, and Shamsul Huda Badal. 1998. "Men's Violence Against Women in Rural Bangladesh: Undermined or Exacerbated by Microcredit Programmes?" *Development in Practice* 8, no. 2: 148-157.

Schultz, T. Paul. 2004. "School Subsidies for the Poor: Evaluating the Mexican Progresa Poverty Program." *Journal of Development Economics* 74, no. 1: 199-250.

Shah, S. K., J. Kimmelman, A. D. Lyerly, H. F. Lynch, F. G. Miller, R. Palacios, C. A. Pardo, and C. Zorrilla. 2018. "Bystander Risk, Social Value, and Ethics of Human Research." *Science* 360, no. 6385: 158–59.

Shah, Seema, Rebecca Wolitz, and Ezekiel Emanuel. 2013. "Refocusing the Responsiveness Requirement." *Bioethics* 27, no. 3: 151-159.

Shadish, William R., Thomas D. Cook, and Donald T. Campbell. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference.* Boston: Houghton Mifflin.

Stone, Peter. 2011. *The Luck of the Draw: The Role of Lotteries in Decision Making.* New York: Oxford University Press.

Teele, Dawn Langan. 2014. "Reflections on the Ethics of Field Experiments." In *Field Experiments and Their Critics: Essays on the Uses and Abuses of Experimentation in the Social Sciences*, edited by Dawn Langan Teele, 115-140. New Haven: Yale University Press.

The National Commission for the Protection of Human Subjects of Clinical and Behavioral Research. 1979. "The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research." Washington, DC: Department of Health, Education, and Welfare.

The Royal Swedish Academy of Sciences. 2019. "The Prize in Economic Sciences." https://www.nobelprize.org/prizes/economic-sciences/2019/press-release/

United States. 1983. *Overview of the Seattle-Denver Income Maintenance Experiment Final Report*. Washington, D.C.: Office of Income Security Policy, Office of the Assistant Secretary for Planning and Evaluation, U.S. Dept. Of Health and Human Services.

Weijer, Charles. 1999. "Protecting Communities in Research: Philosophical and Pragmatic Challenges." *Cambridge Quarterly of Healthcare Ethics* 8, no. 4: 501-513.

Wendler, David, and Annette Rid. 2017. "In Defense of a Social Value Requirement for Clinical Research." *Bioethics* 31, no. 2: 77-86.

Wenner, Danielle M. 2018. "The Social Value Requirement in Research: From the Transactional to the Basic Structure Model of Stakeholder Obligations." *The Hastings Center Report* 48, no. 6: 25-32.

Wenner, Danielle M. Forthcoming. "The Social Value of Research and Health Research Priority-Setting." In *The Oxford Handbook of Research Ethics*. edited by Ana Iltis and Douglas MacKay. New York: Oxford University Press.

 Wertheimer, Alan. 2015. "The Social Value Requirement Reconsidered." *Bioethics* 29, no. 5: 301-308.

Whitfield, Gregory. 2019. "TRENDS: Toward a Separate Ethics of Political Field Experiments." *Political Research Quarterly* 72, no. 3: 527-538.

---

[i] My definition of a policy RCT is very similar to David Greenberg and Mark Shroder's (2004, 4) definition of a social experiment, according to which a social experiment involves: (1) random assignment, (2) a policy intervention, (3) follow-up data collection, and (4) evaluation of the intervention's effectiveness.

[ii] For further discussion of the relevance of the institutional identity of the research sponsor to the ethics of research see Pierson and Millum 2018; Ogden 2016, xx-xxi; MacKay 2015.