

A Call for Structured Ethics Appendices in Social Science Papers

[Edward Asiedu](#)

University of Ghana

[Dean Karlan](#)

Northwestern University and IPR

[Monica Lambon-Quayefio](#)

University of Ghana

[Christopher Udry](#)

Northwestern University and IPR

Version: January 25, 2021

DRAFT

Please do not quote or distribute without permission.

Abstract

Ethics in social science experimentation and data collection are often discussed but rarely articulated in writing as part of research outputs. Although papers typically reference human subjects research approvals from relevant institutional review boards, most recognize that such boards are not comprehensive ethical assessments. The authors propose a structured ethics appendix to provide details on the following: policy equipoise, role of the researcher, potential harms to participants and nonparticipants, conflicts of interest, intellectual freedom, feedback to participants, and foreseeable misuse of research results. They discuss each of these, and some of the norms and challenging situations of each. They believe that discussing such issues explicitly in appendices of papers, even if briefly, will serve two purposes: more complete communication of ethics can improve discussions of papers and can clarify and improve the norms themselves.

We are grateful to Gharad Bryan, Shawn Cole, Andrew Foster, Christina Gravert, Steve Glazerman, Julian Jamison, Heather Lanthorn, Jeff Mosenkis, William Pariente, Kate Vyborny, and especially to Douglas Mackay, for comments and discussions.

Introduction

Social science researchers engaged in primary data collection often consider a range of ethical issues during the planning of their research, but they rarely discuss them in articles generated from the research. We believe that building explicit steps for considering and discussing ethical issues can lead to better research, better communications about research, and thus better impact of research as well. To facilitate this, we propose that authors include a structured ethics appendix in working papers and published online appendices, and we provide a framework below for guidance. Such information could also be used in grant applications.

Institutional review boards (IRBs) aim to protect research participants but ultimately examine a narrow set of ethics issues. We do not actually have “ethics review boards” that have a mandate over all ethical research issues. A broad range of ethical issues do not fall under the purview of an IRB but are important for primary data collection research and especially randomized controlled trials (RCTs). These issues, often left to self-regulation and peer review during grant applications and publications, deserve more thorough consideration for two broad reasons. First, active consideration and discussion of these matters can help form stronger and better norms and adherence to those norms. Second, social science writing style is often fairly terse on methods issues, in particular information relevant for an understanding of ethical issues. Many journals require an IRB approval number and nothing more. This leaves readers to fill in the unstated information with potentially incorrect assumptions. The short-form nature of some public discussions can then exacerbate misunderstandings.

We propose a “structured ethics appendix” for social science research engaged in primary data collection. This is not a checklist. Instead, it is a structured, but brief, set of nine questions which hopefully provide researchers a concise and consolidated platform to spark thoughtful consideration of ethical issues before a project begins and provide relevant information to readers of completed papers. We know of no specific recipe for “ethical approval” of a research project.

We created the Structured Ethics Appendix with the nine questions and brief explanations of each (see Appendix of this paper). Each of the nine questions of course merits deeper discussion. The topics are as follows: (1) Policy equipoise, (2) Role of researchers with respect to implementation, (3) Potential harms to participants or non-participants from the interventions or policies, (4) Potential harms to research participants or research staff from data collection (e.g., surveying, privacy, data management) or research protocols (e.g., random assignment), (5) Financial and reputational conflicts of interest, (6) Intellectual freedom, (7) Feedback to participants or communities, (8) Foreseeable misuse of research results, and (9) Other ethics issues to discuss.

1. Policy Equipoise and Scarcity

Is there policy equipoise? That is, is there uncertainty regarding participants’ net benefits from each arm of the study relative to the other arms and to the best possible policy to which participants could have access? If not, ethical randomization requires two conditions related to scarcity: (1) Was there scarcity, i.e., did the inclusion of multiple arms change the expected aggregate value of the programs delivered? (2) Do all ex-ante identifiable participants have equal moral or legal claims to the scarce programs?

Freedman (1987) argues that the therapeutic obligation of doctors generates a “clinical equipoise” ethical requirement for medical trials: the expert community must not have certainty that any arm in a trial is better therapeutically than any other arm.² For social science in particular, the word “certainty” can render such a requirement toothless: just note that even theories with strong empirical evidence will have some level of uncertainty regarding their applicability in a new context. Furthermore, in many cases one treatment arm clearly dominates another from the perspective of the participants, yet the better treatment arm is not viable as a policy for all, either due to scarcity or other practical or political issues. A study testing the returns to a large cash transfer is a perfect example. Yet evidence from the “better” treatment arm still serves an important social value, even if limited to the generation of abstract knowledge. This renders the concept of clinical equipoise wrong for social science experiments.

Is there a similar obligation for social scientists organizing or participating in RCTs? MacKay (2018; 2020) argues persuasively that a sufficient requirement for ethical randomization is “policy equipoise”, not “clinical equipoise”. We have adopted that view here, as well.

Policy equipoise builds on clinical equipoise but considers resource tradeoffs explicitly. Two arms of an RCT are in equipoise when there is genuine uncertainty about the efficacy of each arm in achieving the relevant outcomes of the study for all participants. Policy equipoise requires that all arms of the study be in equipoise with each other and with the best “proven, morally and practically attainable and sustainable” alternative policy for achieving improvements in the relevant outcomes of the study (MacKay 2018). Each of those words in quotes carries weight. “Proven” implies scientific consensus (and not, for example, adherence to any one methodology for creation of evidence). Clearly, this is not a strictly binary concept, but rather about how tight the priors are of relevant experts and stakeholders. “Morally ... attainable” requires that the alternative policy be consistent with individuals’ rights and liberties. “Practically attainable” means that the government or implementing agency has the resources to put the alternative policy into effect. “Sustainable” means that the government or implementing agency could maintain the policy, “given a just system of resource procurement and allocation”(MacKay 2018, 62).

The policy equipoise requirement is a sufficient condition for randomization, but it is quite strict. First, no participant in any arm of an RCT can be predicted to be better or worse off (in an ethically-relevant way) than he or she would be under any of the other arms of the study. Second, this equipoise must extend to the appropriately defined counterfactual policy, which constitutes whatever policy would have occurred in the absence of the research. Often this counterfactual policy is the current status quo, but that need not be the case. There may be an alternative policy that is moral, practically attainable and for which there is a consensus on its effectiveness (see the hypothetical Progressa case below). Or it could be that the current status quo is not itself sustainable. We emphasize a key distinction that we agree is important, made in MacKay (2018), that what matters is not what the actual counterfactual policy is, but rather what it could be.

If policy equipoise is not satisfied, then randomization can be justified if no participant can be predicted to be worse off in any arm of the study than under the counterfactual policy, and if there is scarcity of

² There is a robust debate on the meaning and importance of clinical equipoise for ethical medical trials (see Gifford 2007; Chiong 2006). Is there a distinction in the role of doctors as researchers and as physicians providing treatment? Or do the ethical principles governing medical therapy apply to clinical research (Miller and Brody 2007)?

the resources required for the arms in which participants are better off. A treatment arm may be unambiguously the “best” for recipients but not viable as a systematic policy for all. The simplest example here, relevant in social science, would be large transfer programs. It is unethical to withhold this policy/treatment unless it is scarce (MacKay 2020). Budget constraints can make it impossible to reach everyone with a better costly intervention, in which case it may be permissible to allocate access randomly, if the second scarcity condition is also met, that is, if it is not known a priori who should receive that intervention. If all participants have equal claims to a scarce intervention, then randomization can be ethical. If some participants have stronger claims (for example, if there is local knowledge that they would benefit more), then these participants should have priority (Barrett and Carter 2010), rather than being randomized out of access.³

We highlight one further component of the definition of policy equipoise employed above: “achieving improvements in the relevant outcomes of the study”. An RCT could, for example, be focused on testing alternatives for reducing child malnutrition. One of the treatment arms could be a cash transfer. If we were to “know” that cash transfers have benefits beyond addressing child malnutrition (a reasonable conjecture, albeit one with uncertainty itself, certainly on the types and magnitudes of impacts), should the non-child malnutrition outcomes be considered when examining if any treatment arm is viably “better” than all of the others? To do so would require a full-blown multi-dimensional welfare analysis, weighing all ethically-relevant outcomes. We would argue that it is typically sufficient to constrain the comparison of treatment arms (and counterfactual policy) to the problem at hand (e.g., paths to improving the primary outcomes of the study). Realistically, anything that dominated in all dimensions with certainty would likely not be “practically attainable.” We do see merit in this discussion, though.

Randomization into different treatment arms, therefore, is unethical if two conditions are met. First there is not policy equipoise. And second, either the preferred arm does not require resources that are scarce or all participants do not have equal claim to those scarce resources. Some examples can clarify the requirements of policy equipoise and scarcity.

Karlan et al (2014) implements an RCT that includes an arm in which farmers are offered cash grants of approximately \$400, and also a control group. This RCT violates policy equipoise, because there is no genuine uncertainty among experts that receiving an offer of a cash grant is better for the recipient than not receiving such an offer. However, randomization remains ethical because of the two principles of scarcity. Providing cash grants to farmers predictably improves their welfare, so the cash grant arm is better than the counterfactual policy of no cash grants. However, there is no consensus that providing cash grants to farmers as a policy is practically attainable and sustainable, given the cost of that policy. This meets the first requirement for scarcity. Nor is there a consensus about which farm households have stronger claims over receipt of cash grants. This meets the second scarcity requirement, and therefore randomization of access to the grants is acceptable.

Similarly Glewwe et al. (2016) violates policy equipoise – there is no uncertainty that eyeglasses benefit those whose eyesight can be corrected by them. However, there is no consensus that eyeglass provision to school children is a cost-effective policy for improving school achievement. Therefore, it was not clear a priori that there was an alternative sustainable policy that dominates the control arm, and the first

³ The growing selective trials literature provides a promising path forward on capturing local information (Chassang, Padró I Miguel, and Snowberg 2012).

scarcity requirement is met. Randomization occurred at the township level, and there was no a priori knowledge of which township would benefit most from the intervention, satisfying the second scarcity requirement.

Policy equipoise would be violated if there is an expert consensus that an arm of a trial is dominated by another policy that is proven, attainable and sustainable. A hypothetical but realistic example of a violation of policy equipoise would occur if a government with sufficient resources for a national program agreed to withhold a Progresa-inspired conditional (on school enrollment) cash transfer from some households for a pure control group in an RCT (MacKay 2020, 351).⁴ In this hypothetical case, the first scarcity requirement also fails, because the government is presumed to have sufficient resources to make the CCT available to more households, so this randomization is unethical.

Consider a potential research project studying the effect of mobile money transfers on food security during a drought with a control group who receive nothing and a treatment group receiving the transfers. While policy equipoise is violated because the transfers can be expected to improve food security, the first scarcity requirement is met due to the budget constraint of the project. However, if the researchers know a priori through, say, remote sensing data, which households' farms were particularly affected by the drought, the second scarcity requirement may be violated. The most affected farmers should be prioritized for transfers, rather than randomized into treatment or control.⁵ Thus such a proposed randomization raises an ethical red flag.

2. Researcher Roles with Respect to Implementation

Are researchers “active” researchers, i.e. did the researchers have direct decision making power over whether and how to implement the program? If YES, what was the disclosure to participants and informed consent process for participation in the program? Providing IRB approval details may be sufficient but further clarification of any important issues should be discussed here. If NO, i.e., implementation was separate, explain the separation.

In social science, the role of the researcher is quite varied, and this has important implications for what may constitute ethical research. On one end of the spectrum, the researcher is merely the evaluator, with neither influence over nor responsibility for any of the interventions. In such an instance, understanding the ethics of the intervention may even be the motivation for the research. Then with more evidence on deleterious effects of the intervention, perhaps such policies can be altered or eliminated.⁶

On the other end of the spectrum, the researcher is the implementer (“active” in the jargon we put forward here). In such cases, the researcher for example secured funding, hired staff, decided the design of the intervention, and directly implemented it. In this case, it goes without saying that the researcher

⁴ Policy equipoise could be restored if the control group were replaced by an experimental arm expected to be as effective or better as the CCT.

⁵ Naturally “most affected” is undoubtedly mismeasured and not known with uncertainty; as such this could be done probabilistically, in which individuals' likelihood of assignment to treatment is a function of estimated prioritization. This maintains the ability to test impact throughout while still adhering to this ethical principle.

⁶ See Glennerster and Powers (2016) for a robust discussion of this distinction and its implications, and some further helpful examples.

is responsible for the ethics of the intervention itself. The researcher is changing the world not merely through the dissemination of results of the research but deliberately through the research process itself.

Even when the researcher is “active”, however, it may still be “ethical” to implement arms which raise valid ethical concerns. But certain criteria ought to be met. First, there must be genuine empirical uncertainty about the outcomes that raise the ethical concerns (and this may be the aim of the study). Second, the policy or intervention must be commonplace enough that assessing the intervention is of high social value. In these instances, the onus is on the researcher to establish “policy equipoise”, or explain why scarcity limits the extension of expected beneficial arms to the entire population as discussed in Section 1 and MacKay (2020).

Two recent examples in social science highlight the importance of this distinction, and also point to the complicated nature of defining the researcher’s role in some instances. In Bryan, Choi and Karlan (2020), researchers collaborated with a nonprofit organization in the Philippines which ran a four-month program that included various secular activities (e.g., savings groups, livelihood promotion, health education) alongside weekly meetings with a pastor that covered an evangelical Protestant curriculum. To understand the impact of the Protestant curriculum on individuals and their households, the nonprofit organization agreed to randomize across villages whether or not the Protestant curriculum was included in their program. The research did not lead to an increase or decrease in the quantity of sessions conducted, nor the number of people reached, by the program. Furthermore, there is no consensus on the impact of such programs, and they are commonplace in the world. Thus, we argue that there is social value in evaluating claims of impact or harm, and no ethical quandary for the researchers.

In Kenya, researchers from universities as well as the World Bank collaborated with the Nairobi Water and Sewerage Company, a Kenyan utility company regulated by the government of Kenya. The researchers engaged with the utility company over several years, working broadly and collaboratively in order to increase access to water and sanitation services. Coville et al. (2020) reports on this collaboration, specifically on an experiment in which one of the treatment arms threatened tenants with having their water access shut off if the water bill was not paid by their landlord. While one could debate the ethics of that intervention arm as a policy, we would argue that if the researchers are not “active” then such a discussion is important and worthwhile but should be discussed as an ethical consideration for the utility company, not the researcher. For the researchers, in fact, the ethical issues may make this more important to study if this is a common policy that potentially infringes on important human rights (and in fact, cutting off utilities due to non-payment is a common policy around the world). Ultimately this example also highlights potential ambiguity on the binary categorization of “active.” If a researcher suggests a treatment arm, for example, does this make them responsible for its ethics, even if the researcher has no actual power over the decisions? Putting ethical onus on the researcher could be construed as ignoring (and thus disrespecting) the autonomy and agency of the local policymakers. We are not asserting how exactly to draw the line here, but argue that more transparency on this role would be fruitful for social science, so that norms can be better developed on this question.

Extreme examples, however, render this distinction meaningless. For example, research ethicists share a common consensus that medical data from experiments conducted on prisoners in Nazi concentration camps cannot now be used to answer medical research questions. But, even if one believes the Vietnam

War was unethical, using the lottery draft to study the impact of military service on later earnings is easily defensible as ethical. Perhaps one key distinction is the purpose of the original implementation, because of the incentives created for researchers. If the original implementation was being done for the sake of research (albeit unconscionable and unethical), then allowing future researchers to use such data creates perverse incentives to implement unethical research. However, when the original implementation was “natural” (e.g., the Vietnam lottery example above, or a government-run public lottery for secondary school scholarship (Angrist et al. 2001), or reservation of village leader slots for women (Chattopadhyay and Duflo 2004)), then we argue that use of such variation is valid ethically even if there are objections to the underlying policy.

3. Potential harms to participants or nonparticipants from the interventions or policies

Does the intervention, policy or product being studied pose potential harm to participants or non-participants? Related, are participants or likely affected non-participants particularly vulnerable? Also related, are participants’ access to future services or policies changed because of participation in the study? If yes to any of the above, what is being done to mitigate such risks?

We highlight two broad issues: First, despite best efforts IRBs are unable to fully oversee all aspects of potential harms to participants, and local knowledge in particular is invaluable and at the mercy of self-regulation; second, potential harms to nonparticipants is itself a topic of high interest to researchers on substantive levels, but thus also carries important ethical considerations.

Ethical guidelines for research on human subjects are primarily concerned with protecting study participant (Eyal et al. (2018)). Protocols typically expect researchers to highlight the benefits and risks of participation, confidentiality protocols, compensation, and instructions for withdrawing from the study at any stage. While in theory an IRB strives to consider all such issues, in reality IRBs, particularly ones from a different culture, are not always informed enough (due to omission by researchers or lack of knowledge of context and practices) to raise or adjudicate on them. Ultimately the responsibility of the researcher extends beyond mere IRB approval for such matters.

Cultural norms and nuances of the implementation may pose potential harms to participants that are beyond the viable purview of the IRB. For example, cultural norms or misinterpretations of implementation activities could lead participants to be perceived as being “superior” or suffer some form of stigma or discrimination. Such perceptions can boost or harm the reputation of participating households which may have implications on their social and perhaps economic lives within the community. Elected community leaders in participating communities may suffer abuse and intense competition from political rivals, as interventions are sometimes perceived as ways for elected local leaders to hold on to political power. In view of these potential indirect harms, when deemed plausible it is important for projects to integrate mechanisms to deal with such ethical concerns throughout the duration of the project.

While informed consent and IRBs are primarily concerned with the participants in a study, studies may also pose harm to nonparticipants. The National Bioethics Advisory Commission of the US asserts that regulatory oversight for research with human subjects extends beyond the protection of individual research participants to include the protection of social groups (Sharp and Foster 2002). The potential effects on non-participants become especially salient when researchers are actively engaged in interventions. It is, therefore, important to evaluate the scope and intensity of risks that are likely to

impact non-participants. In some cases, community-level consent may be appropriate (see section 4 for further discussion of community-level consent).

We highlight three examples in which nonparticipants are affected and discuss the ethical implications. Ashraf et al. (2014) conducts an experiment in Zambia on women's bargaining power and fertility. In one arm women were given access to contraceptives alone and in the other with their husbands. Since both arms of the trial have implications for a non-participating husband in terms of child-bearing outcomes, should there be household-level consent for both arms? If empowering women changes gender relations within households, does this compromise shared household values? If so, an active researcher should consider obtaining some household consent, one in which non-participants also indicate their support for the intervention or research. Or, instead, is a women's control of reproductive decisions considered a sacrosanct human right, and thus informed consent from her alone is sufficient? This issue of protecting non-participants of a study while maintaining the rights of research participants can be a tradeoff worth considering.

Second, suppose researchers are evaluating an education reform of teacher pay and training, with outcomes such as teacher absenteeism, teacher effort and student performance. Researchers would conduct this in collaboration (and thus consent) of key high-level stakeholders such as district school heads or head teachers. But since the intervention has implications for the quality of learning obtained by students, should consent be obtained from parents?

Third, incentivizing students to take part in an antiauthoritarian protest poses a direct risk to participants but also poses risks to non-participants who are also taking part in the protest (see Bursztyn et al. 2021). Thus, if students are rowdier in a protest than non-students, incentivizing more student participation in a protest increases the number of people in the protest and as such increases the risk for the non-student non-participants of the intervention. Inasmuch as it is important to minimize the risk for the study participants, it is in this case even more critical that the actions of participants do not increase the risk for non-participants.

A consideration of the risks and benefits for the group of non-participants is, therefore, encouraged, to help evaluate the ethics and plan any appropriate mitigation strategies. Naturally, however, it is easy to posit indirect effects, much harder to predict magnitude and likelihood of such effects. And, indeed, studying such indirect effects is itself a robust research agenda for many (e.g., Miguel and Kremer 2004). Thus, we do not argue that the mere posing of a possible indirect effect should render a study unethical. But we do suggest that such issues be considered, and as we learn more about the manifestation of such effects from ongoing research, we hope this space can become more evidence-based, learning when to expect and thus avoid harms to non-participants.

We acknowledge that these are critical issues that require further discussion and consensus building as a discipline on how to practically address them. As highlighted, research risks to communities and non-participants suggest that ethical protocols that focus on the protection of individual research participants will be insufficient (Sharp and Foster 2002). Ethical guidance for research in social science in general and RCTs specifically must be extended beyond the protection of individuals to include the protection of social groups i.e., family members, community members, etc. Group implications of identified individual-level harm can serve as a starting point to identify risk to non-participants and should be properly assessed and addressed in an ethics review process. It is also important to consider

working directly with community representatives to discuss informed consent and its implication and develop study methods that minimize potential group harms.

4. Potential Harms from Data Collection or Research Protocols

Are data collection and/or research procedures adherent to privacy, confidentiality, risk-management, and informed consent protocols with regard to human subjects? Are they respectful of community norms, e.g., community consent not merely individual consent, when appropriate?⁷ Are there potential harms to research staff from conducting the data collection that are beyond “normal” risks?

The fundamental principle of ethical conduct of human subjects’ research centers around minimizing the risk of harm and respecting agency (i.e., securing appropriate informed consent). While Section 3 was focused on how interventions themselves create such issues, here we focus on the data collection and random assignment protocols. Respect for participants is the guiding principle.

Informed Consent: Data Collection

Genuine consent is present only when participants have the capacity to fully understand and process the information that is being shared (Hewlett (1996)). Processes for obtaining informed consent differ widely in practice, and are often at the mercy of a staff’s adherence to protocols. Field staff may treat consent as a mere formality where the main task is to get the participant’s signature or thumb print on a form, thus leading field workers to hurriedly read through the prepared briefs without fully engaging participants. In a culture or setting where the educated, well-dressed and elderly folk are associated with perceived power or influence; participants, especially the vulnerable, may feel compelled to continue interviews merely due to fear of being denied access to an important service or of being misjudged as hiding something. This may be exacerbated by differing age and education level of the interviewer. Electronic data collection offers paths to oversight; e.g., survey managers can examine time stamps on the informed consent process to make sure they are not being rushed through in mere seconds. However, not all interaction can be monitored and thus staff training is critical. When a gift is provided, even if minor such as a bar of soap or a tin of milk, such appropriate compensation may have the unintended consequence of distracting from the importance of the informed consent information.

Enumerators are often encouraged to probe during their interviews to obtain more accurate responses. While the skill of probing is useful, it has the potential of creating discomfort to the participant if questions are considered to be intrusive and lead to discussions which may expose the vulnerabilities of the participant and affect their self-esteem. There is a fine line between probing and harassment.

Moreover, although consent forms provide participants with contact details of the researchers or the principal investigators of the project, they may be less likely to contact the researchers or field supervisors when they experience any violations of their privacy for reasons such as lack of funds for

⁷ Example of sub-questions to consider as part of the broad question: Are there any risks that could ensue because of the data collection process or storage, e.g. discomfort to being asked certain questions or breach of confidentiality? If so, what are the mitigation strategies? Are there costs to the participant for the data collection process, such as their time, and if so, what is the strategy or rationale for offsetting this cost? Because these are all issues covered by most IRB processes, a sufficient explanation for a “yes” response may be to provide the IRB approval numbers for all IRBs that have approved the project. However, if there are particular issues that are important to discuss, please do so here.

telecommunication charges, access to telecommunication services or perhaps perceived language or status barrier.

Ultimately the burden is on the enumerator (and the research team with respect to training the enumerators) to engage respectfully with participants through the informed consent process.

Informed Consent: Randomization

Informed consent on the randomization process itself is often not discussed. In medical trials, partly because the medical professional has a “standard of care” therapeutics to which each patient has a clear right, informed consent on the randomization process is essential to having agency over one’s own health decisions. However, in social science, the intervention is often implemented independently from the researchers (see Section 3), and “informed consent” is effectively implied via voluntary participation in an intervention (e.g., if a lender randomly offers people different loans, their decision to borrow is not forced by the lender nor by the researcher, and consent is implied by their decision to apply and/or sign a contract for a loan). Furthermore, in social science scarcity often limits the total number of individuals that can receive a certain intervention, and the randomization is being done not for research purposes but rather to allocate a scarce resource fairly and with minimal risk of corruption (e.g., see Angrist et al. 2001).

Hawthorne and John Henry effects are often discussed as threats to internal validity, but they each generate ethical issues as well (and not merely by reducing the social value of the research). A Hawthorne effect is generated when individuals, because they know they are in a treatment group, behave differently. A John Henry effect is the analog for a control group. A prime concern here that raises ethical issues is if there is an “unlucky stigma” among households or individuals leading them to harbor negative feelings or even psychological effects of being the unlucky ones that did not receive an intervention (for control group participants). The analog for a treatment group is potentially more nuanced. If, for example, a treatment participant receives a cash transfer, observing jealousy, demands for sharing, or other such effects, are not necessarily a risk to internal validity, but rather perhaps exactly the kind of changes in behavior that the researcher (and policymaker) seek to learn more about. The question for ethics is whether the randomization process itself generated such effects over and beyond what the transfer itself generates. This is difficult to know, empirically, but is a risk nonetheless that may be more relevant in some settings. When such a risk is deemed more likely, we suggest it be discussed.

Informed Consent: Individual vs Community

Perhaps one of the most topical issues in extending RCTs to developing countries is the conflict in eliciting both individual and group level consent, or only individual-level consent. Social norms in many developing countries, particularly in sub-Saharan Africa, mandate that visitors to communities first meet with chiefs and community leaders, explain their reasons for visiting the community, and in some cases provide customary gifts, such as presenting kola nuts or drinks. This level of community consent is sometimes important and useful at the barest minimum to ensure the security of field enumerators and researchers. This also may be necessary etiquette for the sake of future researchers, so that future enumerators are not greeted with mistrust. Community consent also could be critical for managing local security risks that may arise (and due to timing, be far outside the purview of the IRB process).

Aside from issues relating to obtaining individual versus community consent, there are other secondary level ethical concerns that projects ought to consider in the process of obtaining community or individual level consent. In some settings research projects may inadvertently contribute to creating an expectant culture of participants or facilitate the corrupt behavior of local community and political leaders. For example, community leaders whose assistance may be required during the community entry or obtaining community consent for an intervention may take advantage of the project to advance their own agenda or illegally take bribes from community members in exchange of providing some service that the project may be providing for free.

Local IRBs can prove particularly useful for these issues, as they undoubtedly will have better information about the expectation and norms in the local context. But as with many issues, no IRB is sufficient.

Privacy, Confidentiality and Sensitive Information

Once data are collected, the responsibility of ensuring confidentiality of the information gathered rests on the researcher. Researchers are expected to not betray the trust of participants with respect to data management, storage, and protection of privacy. Failure to ensure the confidentiality of the participants may have serious consequences. For example, sensitive information on the health status, particularly sexually transmitted diseases such as HIV, or sexual behavior of a participant can cause potential harm to participants if such information is not properly guarded. Similarly, financial data could be used by a financial institution in ways contrary to a client's self-interest. Confidentiality is, therefore, key to safeguarding and reducing the risk of potential stigmatization of participants, especially for the vulnerable population. In some cases, the potential harm of a lax data management system may extend beyond the individual participant to his or her extended family.

For sensitive questions (e.g., questions about sex, health, politics, religion, etc.), discussing the specific survey methods may be helpful both for explaining the adherence to ethical protocols but also as guidance for future researchers looking to mimic or improve upon such methods.

Potential Harms to Field Staff

Ironically, risks to field staff are not a consideration for most (if not all) IRBs, since they are not research participants. Yet clearly field staff are often unsung heroes of the research effort; without high quality data, a randomized controlled trial is akin to an elaborate birthday party without guests.⁸ But field work does create risks to staff. Risks range from "normal" such as road accidents or verbal abuse to extreme such as sexual harassment, political violence or pandemics. Pandemics of course pose even further risks to communities, in which traveling staff may unintentionally worsen the spread of a virus particularly in situations like COVID-19 where carriers are often asymptomatic.

5. Financial and reputational conflicts of interest

Do any of the researchers have financial conflicts of interest with regard to the results of the research? Do any of the researchers have potential reputational conflicts of interest?

⁸ Analogy courtesy of Christina Gravert, with the follow-on explanation: "Tons of work to set up and then makes you cry out of disappointment." An alternative and subtler analogy, courtesy of Andrew Foster: "A compass in Antarctica".

Financial conflicts of interest reporting policies are similar but distinct across universities, multilateral organizations, and research conduits such as the National Bureau of Economic Research. It is important to report the ex-ante possibility of a conflict of interest, so that the research community can judge for itself whether the interpretation of the empirical results was biased in favor of the possibly competing interest. These rules become vaguer when the income source and the data source differ but are from the same industry. If earning consulting income from Bank A, and writing a paper with data from Bank B, does the Bank A consulting income need to be reported in the research about Bank B? Ultimately researcher judgement is required to make such decisions; the goal of the structured appendix is to report potentially important issues without being a burden.

Reputational conflicts of interest are not addressed by any university, multilateral or research organization rules that we have seen. But for many researchers, idea promotion is a vital self-interest (whether due to altruism, reputation, or future financial remuneration). While money is an interest that is easily defined, and typically traceable via contracts and payment records, reputational self-interest is person-specific, and difficult to define and trace. We thus define a reputational conflict of interest with respect to a particular paper quite broadly: when prior writing or advocacy could be contradicted by specific results in the new paper, and such contradiction would pose reputational risks to the author. Taken to an extreme, such disclosures could be exhausting (and thus then ignored). We do not recommend that. A key illuminating question could be whether the scholar has a public presence regarding a specific prior finding or scholarly theory. We provide two examples from our own research.

First, for Karlan, his paper “Tying Odysseus to the Mast” (Ashraf, Karlan, and Yin 2006) evaluates the take-up and impact of a commitment savings product in the Philippines. He subsequently published a paper on commitment contracts for smoking cessation. Both find generally positive impacts from the commitment devices. Subsequent to the above paper, Karlan started a for-profit company stickK.com that helps individuals write commitment contract to change some future behavior (typically weight loss, exercise or smoking). This company has been written about frequently in the media. Karlan owned equity when he co-founded stickK.com, but has subsequently transferred all of his equity to charitable structures and he receives no financial remuneration from stickK.com. As such he no longer has a financial conflict of interest that is reportable per standard university rules. But he does have a reputational conflict of interest on the topic of commitment contracts, as his work and stickK are often cited in the media as an example of successful “nudges”. To guard against such risks, he can coauthor with others who have no such reputational risks and for prospective studies and he can use preregistration and pre-analysis plans. While each assuages the issue, neither is dispositive. A disclosure would be appropriate for Karlan on future work related to committed contracts that states “Karlan has a public presence related to his co-founding and prior ownership of stickK.com, a company dedicated to using incentives to help individuals change future behavior.”

For Udry, his paper “Gender, agricultural production, and the theory of the household” (Udry 1996), finds that households do not achieve pareto efficiency (yet much empirical work using household data starts with such an assumption and never looks back). If Udry were to have a new paper in which he finds pareto efficiency, this would conflict with his prior finding but would not constitute in our opinion a reputational conflict of interest. Even though his earlier work is commonly cited for testing for efficiency within the household, Udry has no non-academic reputation that revolves around that finding (and also, for the record, the other three of us are certain Udry would be quite happy to find such

evidence, and then would undoubtedly try to concoct some neoclassical economic theory as to why that context yielded such a result, but other contexts did not).

Lastly, as we alluded to above, registration of RCTs and pre-analysis plans both can serve to mitigate reputation conflicts of interest issues because they both tie the hands of the research to some extent. Neither of course is dispositive: the choice of specification and the choice of outcomes (timing, measurement approach, and selection of proxies) could all be biased. And naturally the way results are interpreted (even with refereeing and editing at journals) can be biased. Thus, reporting such conflicts, and what steps if any were taken to mitigate, could be useful.

6. Intellectual freedom

Were there any contractual limitations on the ability of the researchers to report the results of the study? If so, what were those restrictions, and who were they from?

While reporting of conflicts of interest makes potential biases transparent, parties with vested interest may have other means of controlling research output and thus creating bias. Specifically, researchers and collaborators (here we define collaborators as either implementing parties, sources of data, or funders) often sign contracts regarding funding, intervention plans, or data rights. If such contracts infringe on the intellectual freedom of the researcher, this poses harm to the credibility of the research. Failure to disclose such conditions could reduce the social value of the research, constituting an ethical transgression. As such, we have included an explicit question on this in the appendix.

The most egregious infraction is the simplest: a contract which provides an external party (funder or implementer) unilateral power to prevent publication of the results. This should be unacceptable to any researcher interested in putting forth credible research, but at a minimum if such were the rights of the partner this ought to be disclosed in any and all publications.

We consider four restrictions to be benign but still worthy of reporting for the sake of completeness: (1) permission required to report the name of the collaborating institution, (2) a comment commitment: collaborating institution can provide comments on the research output and the researchers agree to consider these comments to the best of their judgement (but are not committed to incorporating them), (3) a timing commitment: the collaborating institution has a right to see the results before others, with a specified period of time, and (4) proprietary intellectual property: the collaborating institution has developed a proprietary technology (perhaps an algorithm, for example) that is relevant for the research, but can only be described broadly without revealing its inner workings (naturally such restrictions limit the scientific value of a publication, but that is not an ethical consideration; rather, that is for editors and referees to decide on scholarly merit). We do not assert this to be a complete list of all acceptable restrictions, but we are unaware of any others that we consider acceptable.

While researcher independence is critical for combatting nefarious actors (e.g., see Goldacre 2013) , it is also critical for protection from biased pressure.

A few details regarding researcher independence are important to note:

- Timing matters: Intellectual freedom must be granted in advance. A clause in an agreement which states that this will be decided later is a farce to the concept of researcher independence.⁹
- Partial set of coauthors: If a subset of the coauthors on an academic paper are conflicted or lack independence (e.g., most commonly, because they are employed by the funder or the implementing organization), then this section of the appendix ought to make clear that the independent researchers had full rights to publish on their own without the conflicted coauthor.

7. Feedback to participants or communities

Is there a plan for providing feedback on research results to participants or communities? If yes, what is the plan? If not, why not?

Informed consent ultimately is about respect for the agency of an individual as well as their privacy and property rights to their information. But often attention to this focuses on the informed consent process. Recently more attention, but still too little, has been placed on engaging with research participants, local government, and communities after the research is complete.

Providing feedback to research participants and communities embodies the principles of respect for persons, beneficence, and justice, which are widely recognized to shape the conduct of research with human subjects (Fernandez, Kodish, and Weijer 2003). Operationalizing the principle of respect throughout the research, from inception to completion is key to this principle.

Providing feedback from the research to participants and communities can further this aspiration of respect to participants and communities. Providing such feedback reduces the likelihood that the participant may feel exploited by the researcher. Public awareness of research results also may be a stepping stone towards research uptake to improve the lives of poor communities. Providing feedback also improves the willingness of communities to participate in future research. Note that in many cases, key gatekeepers (community leaders) may have been part of community-level consent (see above), and as such the feedback may be appropriate to such gatekeepers (when not viable to all participants).

Despite the ideal of providing such feedback, there are several practical challenges that often renders this aspiration unwise. We highlight three: budget, finding people, literacy/knowledge necessary to understand the research.

Budget

Follow-up visits to all communities can be costly and in some cases these costs can outweigh the benefits of directly providing feedback. It is therefore important for researchers, right from the beginning of the project to consult communities, stakeholders and partners on their preferred mode for providing feedback, and subsequently budget accordingly. However, with cellphone data and internet access expanding rapidly, communicating results of the research to participants could be viable with

⁹ In our collective experiences, we have only encountered such a request once (specifically, the assertion was “to decide researcher independence later, after results are in”). The individual demanding this condition led a large, corporate “philanthropy” effort, and subsequently left their post at the corporation to take a high-ranking position with the Trump administration.

respect to the budget, e.g., posting of information on a website and sending individuals text messages with links to the url, or sending pre-recorded video or audio via text message.

Finding people

As benign as it can be, finding research participants after a study can be challenging. Participants may move from the communities within which the study took place, and any aggressive attempts to reach such people could involve non-trivial research cost. Before and during the project, researchers must build the infrastructure or logistics for retaining contact with research participants after the conclusion of the research

Literacy/knowledge necessary to understand the research

Depending on the context and the research, study participants may struggle to understand research findings. Poor understanding of research findings can generate emotional distress for research participants and affect community involvement in future studies. Therefore, providing feedback to participants should not be done solely for the purpose of ticking a box, but with genuine engagement. Furthermore, any study with nuanced results may get misused intentionally or unintentionally. Researchers must examine the context of their study so that proper measures can be taken to enhance understanding of their research results. We propose that the offer of feedback to participants be included as part of the informed consent processes for social science research. As also argued in the medical literature, this process of providing feedback must be designed to provide clear opportunity for the participant to decline receipt of results (see Rigby and Fernandez 2005).

For policy-oriented research, it may be prudent and developmentally effective to provide feedback at a much higher level than at the individual level. Policy prescriptions often are made at the community, district, or national level. Researchers should consider the level at which to provide feedback to be relevant for policy.

Providing feedback to participants may itself affect behavior. Before implementing such a process, the possible adverse consequences of providing feedback should be considered. Given the complexity of some research, a risk is that feedback is conveyed but then misconstrued, leading to detrimental actions. This concern reinforces the importance of careful communication and thoughtful consideration of the way the results may be interpreted. And, of course, this is a question for research itself, does provision of such feedback change later behavior?

8. Foreseeable misuse of research results

Is there a foreseeable and plausible risk that the results of the research will be misused and/or deliberately misinterpreted by interested parties to the detriment of other interested parties?

In research settings characterized by strong imbalances of power between interested parties, there may be foreseeable, plausible risks that a powerful party might be able to use findings from the research in ways that will harm participants or non-participants. Research might reveal a vulnerability of a sub-population that can be exploited for the gain of a more powerful party. If this is the case, the researcher has an ethical obligation to mitigate these risks.

An example might be an innovation in microcredit that research shows is overall profitable and has positive average effect on borrower income. However, consider two dimensions of heterogeneity that

could raise ethical concerns about misuse of these results. First, suppose that the microcredit innovation has a significant detrimental effect on a vulnerable minority of borrowers. Second, suppose that while the innovation leads to an increase in borrower income, it has other psychological or social costs to borrowers that outweigh the income gain for at least some participants. In either case, as discussed sections 2, 3 and 4, the researcher has an obligation to mitigate such risks for participants in the research. In addition, in this example, there is a foreseeable and plausible risk that the research results could be misused by this or other microcredit organizations to expand the use of this innovation without costly mitigation measures to protect the vulnerable minority (in the first case), or to address the psychological or social costs of the innovation (in the second case). This observation calls for the research to take steps to address the possibility of such misuse of the research results.

In a context of a repressive or non-democratic government, additional care is required for the ethical conduct of some policy research. As discussed in previous sections, the researcher must consider the potential for direct harm to participants or non-participants from the research itself. In addition, the researcher should consider the possibility that the results of the research could be misused by a repressive government. Location-specific research to improve digital identification services in such a context, for example, raises the possibility that the results could be used for repressive purposes and researchers should consider this possibility when considering and designing their work.

We use the phrase “foreseeable and plausible” to limit the scope of this question. Research findings can be misused in many ways that should not be considered to be the responsibility of the researcher (companies making use of the discovery of a novel behavioral pattern to sell unneeded products; the police force of a repressive government using techniques discovered in research elsewhere on making bureaucracy function more efficiently). The ethical responsibility of the researcher is to consider harmful uses of their work that are predictable and reasonably tied to the research context.

9. Conclusion

While much has been written on ethics of research, we perceive there to be a large empirical research gap. Broadly we see two categories of empirical research needed: documenting contextual factors that render a particular ethical concern critical or negligible, and learning field research methods to improve adherence to stated intentions.

Examples of the first category include the following: Under what circumstances does random assignment (whether to treatment or control) generate stigma that harms participants? Does promising the control group a later service change their behavior now, during the observation period of the study?

Examples of the second category include the following: How much of an informed consent is understood by participants, and what wording and content changes lead to higher or lower informed consent comprehension rates (e.g., see Wade et al. 2009; Mills et al. 2014)? What surveyor methods and staff trainings are most effective for reducing harm from data collection, e.g. from adverse emotional reactions to sensitive questions? When does conducting surveys change later behavior, i.e., a mere measurement effect (e.g., see Zwane et al. 2011).

We hope that the proposed structured ethics appendix will accompany papers or grant applications, and that it will spark further consideration of ethical issues in field research. We have aimed for brevity, understanding that anything too burdensome will not be adopted. In a public comment period for this

appendix, we received several useful suggestions to include questions and topics we had not considered adding. We adopted several, but not all. For example, some suggested *ex post* discussion of actual harms. We decided not to include this because those are for the most part covered by IRB protocols, and we also believe the ethical decisions ought to be adjudicated blind to results; otherwise “no harm happened” could be misconstrued as evidence of an ethical setup. We also did not extend the reporting of potential harms to instead be a reporting of potential welfare change. Although we admire the aspiration, we fear that such an inclusion would inevitably require considerable work and untenable assumptions in order to assert. Furthermore, loss aversion in both law and ethics is commonplace. Lastly, we did not include a call to discuss the social value of the research with respect to potential policy. Although this is an important consideration and articulating the policy implications is often useful for motivating research and then improving the research-to-policy nexus, we believe this would be too burdensome if addressed more robustly than merely regurgitating points made in the introduction of most papers. We did not receive any suggestions for items to remove, although a common suggestion is to make this meaningful yet short in order to increase participation and impact.

Research ethics is ultimately self-regulated, and our goal is to create a norm of discussing these issues explicitly, both to strengthen and improve the norms and research practices and also to improve the discussion of these issues. To do so we believe that all cases ought to be discussed, not just the ones that raise questions. We hope that a norm of inclusion of a structured ethics appendix can help the research community advance.

References

- Angrist, Joshua, Eric Bettinger, Erik Bloom, Elizabeth King, and Michael Kremer. 2001. "Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment." Working Paper 8343. National Bureau of Economic Research. <http://www.nber.org/papers/w8343>.
- Ashraf, Nava, Erica Field, and Jean Lee. 2014. "Household Bargaining and Excess Fertility: An Experimental Study in Zambia." *American Economic Review* 104 (7): 2210–37. <https://doi.org/10.1257/aer.104.7.2210>.
- Ashraf, Nava, Dean Karlan, and Wesley Yin. 2006. "Tying Odysseus to the Mast: Evidence from a Commitment Savings Product in the Philippines." *Quarterly Journal of Economics* 121 (2): 673–97.
- Barrett, Christopher B., and Michael R. Carter. 2010. "The Power and Pitfalls of Experiments in Development Economics: Some Non-random Reflections." *Applied Economic Perspectives and Policy* 32 (4): 515–48. <https://doi.org/10.1093/aep/32.4.515>.
- Bryan, Gharad, James J Choi, and Dean Karlan. 2020. "Randomizing Religion: The Impact of Protestant Evangelism on Economic Outcomes." *The Quarterly Journal of Economics*, June, qjaa023. <https://doi.org/10.1093/qje/qjaa023>.
- Bursztyjn, Leonardo, Davide Cantoni, David Y. Yang, Noam Yuchtman, and Y. Jane Zhang. 2021. "Persistent Political Engagement: Social Interactions and the Dynamics of Protest Movements." *American Economic Review: Insights*. <https://doi.org/10.1257/aeri.20200261>.
- Chassang, Sylvain, Gerard Padró I Miguel, and Erik Snowberg. 2012. "Selective Trials: A Principal-Agent Approach to Randomized Controlled Experiments." *American Economic Review* 102 (4): 1279–1309. <https://doi.org/10.1257/aer.102.4.1279>.
- Chattopadhyay, Raghavendra, and Esther Duflo. 2004. "Women as Policy Makers: Evidence from a Randomized Policy Experiment in India." *Econometrica* 72 (5): 1409–43. <https://doi.org/10.1111/j.1468-0262.2004.00539.x>.
- Chiong, Winston. 2006. "The Real Problem with Equipoise." *The American Journal of Bioethics* 6 (4): 37–47. <https://doi.org/10.1080/15265160600755565>.
- Coville, Aidan, Sebastian Galiani, Paul J. Gertler, and Susumu Yoshida. 2020. "Enforcing Payment for Water and Sanitation Services in Nairobi's Slums." SSRN Scholarly Paper ID 3658865. Rochester, NY: Social Science Research Network. <https://papers.ssrn.com/abstract=3658865>.
- Eyal, Nir, Marc Lipsitch, Till Bärnighausen, and Dan Wikler. 2018. "Opinion: Risk to Study Nonparticipants: A Procedural Approach." *Proceedings of the National Academy of Sciences* 115 (32): 8051–53. <https://doi.org/10.1073/pnas.1810920115>.
- Fernandez, Conrad V., Eric Kodish, and Charles Weijer. 2003. "Informing Study Participants of Research Results: An Ethical Imperative." *IRB* 25 (3): 12–19.
- Freedman, Benjamin. 1987. "Equipoise and the Ethics of Clinical Research." *New England Journal of Medicine* 317 (3): 141–45. <https://doi.org/10.1056/NEJM198707163170304>.

Gifford, Fred. 2007. "So-Called 'Clinical Equipoise' and the Argument from Design." *Journal of Medicine and Philosophy* 32 (2): 135–50. <https://doi.org/10.1080/03605310701255743>.

Glennester, 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 F. DeMartino and Deirdre McCloskey, 366–401. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199766635.013.017>.

Glewwe, Paul, Albert Park, and Meng Zhao. 2016. "A Better Vision for Development: Eyeglasses and Academic Performance in Rural Primary Schools in China." *Journal of Development Economics* 122 (September): 170–82. <https://doi.org/10.1016/j.jdeveco.2016.05.007>.

Goldacre, Ben. 2013. *Bad Pharma: How Drug Companies Mislead Doctors and Harm Patients*. First American edition. New York: Faber and Faber, Inc., an affiliate of Farrar, Straus and Giroux.

Hewlett, Sarah E. 1996. "Is Consent to Participate in Research Voluntary?" *Arthritis & Rheumatism* 9 (5): 400–404. [https://doi.org/10.1002/1529-0131\(199610\)9:5<400::AID-ANR1790090509>3.0.CO;2-I](https://doi.org/10.1002/1529-0131(199610)9:5<400::AID-ANR1790090509>3.0.CO;2-I).

Karlan, Dean, Robert Osei, Isaac Osei-Akoto, and Christopher Udry. 2014. "Agricultural Decisions after Relaxing Credit and Risk Constraints." *The Quarterly Journal of Economics* 129 (2): 597–652. <https://doi.org/10.1093/qje/qju002>.

MacKay, Douglas. 2018. "The Ethics of Public Policy RCTs: The Principle of Policy Equipoise." *Bioethics* 32 (1): 59–67. <https://doi.org/10.1111/bioe.12403>.

———. 2020. "Government Policy Experiments and the Ethics of Randomization." *Philosophy & Public Affairs* 48 (4): 319–52. <https://doi.org/10.1111/papa.12174>.

Miguel, Edward, and Michael Kremer. 2004. "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities." *Econometrica* 72 (1): 159–217.

Miller, Franklin G., and Howard Brody. 2007. "Clinical Equipoise and the Incoherence of Research Ethics." *The Journal of Medicine and Philosophy* 32 (2): 151–65. <https://doi.org/10.1080/03605310701255750>.

Mills, Nicola, Jane M Blazeby, Freddie C Hamdy, David E Neal, Bruce Campbell, Caroline Wilson, Sangeetha Paramasivan, and Jenny L Donovan. 2014. "Training Recruiters to Randomized Trials to Facilitate Recruitment and Informed Consent by Exploring Patients' Treatment Preferences." *Trials* 15 (1): 323. <https://doi.org/10.1186/1745-6215-15-323>.

Rigby, Heather, and Conrad V. Fernandez. 2005. "Providing Research Results to Study Participants: Support versus Practice of Researchers Presenting at the American Society of Hematology Annual Meeting." *Blood* 106 (4): 1199–1202. <https://doi.org/10.1182/blood-2005-02-0556>.

Sharp, Richard R, and Morris W Foster. 2002. "Community Involvement in the Ethical Review of Genetic Research: Lessons from American Indian and Alaska Native Populations." *Environmental Health Perspectives* 110 (suppl 2): 145–48. <https://doi.org/10.1289/ehp.02110s2145>.

Udry, Christopher. 1996. "Gender, Agricultural Production, and the Theory of the Household." *Journal of Political Economy* 104 (5): 1010–46. <https://doi.org/10.1086/262050>.

Wade, Julia, Jenny L. Donovan, J. Athene Lane, David E. Neal, and Freddie C. Hamdy. 2009. "It's Not Just What You Say, It's Also How You Say It: Opening the 'Black Box' of Informed Consent Appointments in Randomised Controlled Trials." *Social Science & Medicine* 68 (11): 2018–28. <https://doi.org/10.1016/j.socscimed.2009.02.023>.

Zwane, A. P., J. Zinman, E. Van Dusen, W. Pariente, C. Null, E. Miguel, M. Kremer, et al. 2011. "Being Surveyed Can Change Later Behavior and Related Parameter Estimates." *Proceedings of the National Academy of Sciences* 108 (5): 1821–26. <https://doi.org/10.1073/pnas.1000776108>.

Appendix

Structured Ethics Appendix Questions: Brief Explanations

1. Policy Equipoise

Is there policy equipoise? That is, is there uncertainty regarding participants' net benefits from each arm of the study relative to the other arms and to the best possible policy to which participants could have access? If not, ethical randomization requires two conditions related to scarcity: (1) Was there scarcity, i.e., did the inclusion of multiple arms change the expected aggregate value of the programs delivered? (2) Do all ex-ante identifiable participants have equal moral or legal claims to the scarce programs?

If there is no reasonable expectation that one arm of the study produces more benefits to participants than any other arm or than the best possible alternative policy, then randomization is ethically unproblematic. If not, then excluding some participants from the superior treatment arm can only be justified by scarcity. Scarcity conditions are two-fold: (1) resources are not sufficient, given constraints, to include all participants in the superior treatment arm; (2) no ex-ante identifiable participants are excluded from the superior arm and have a greater claim to those resources than any participant assigned to the superior arm. See MacKay (2018, 2020) for more complete discussions of policy equipoise.

2. Role of researchers with respect to implementation

Are researchers "active" researchers, i.e. did the researchers have direct decision making power over whether and how to implement the program? If YES, what was the disclosure to participants and informed consent process for participation in the program? Providing IRB approval details may be sufficient but further clarification of any important issues should be discussed here. If NO, i.e., implementation was separate, explain the separation.

A researcher should be considered "active" if, for example, the implementing staff are employed by an institution at which the PI is employed, and the staff report either directly or indirectly to the PI at this institution with regard to this project. Or if researchers control funding for implementation, or have direct decision-making power over key implementation decisions.

Some key factors that help illuminate whether the researchers are "active" or not (here "researchers" are defined as the PIs and the staff that report directly or indirectly to the PIs): Did researchers directly provide any of the interventions, or parts thereof, to participants? Did researchers interact directly with participants and implicitly endorse one or more of the interventions?

3. Potential harms to participants or nonparticipants from the interventions or policies

Does the intervention, policy or product being studied pose potential harm to participants or non-participants? Related, are participants or likely affected non-participants particularly vulnerable? Also related, are participants' access to future services or policies changed because of participation in the study? If yes to any of the above, what is being done to mitigate such risks?

It may be important to consider whether the researchers are "active" (see above) or not for this discussion. If the researchers are "active", then they are responsible for the potential harms, and thus a robust discussion is appropriate. If the researchers are not "active", then while they may not be responsible for potential harms, a discussion of this would be appropriate here.

There will almost always be some potential harms, if nothing else because of complementary investments such as time that participants in an intervention necessarily redirect from one activity to another.

Quantifying these risks and complementary investments may be difficult ex-ante, but a discussion of what they are here would help the reader assess their likely importance relative to the potential benefits of the tested intervention. Also note that measuring any harms ex-post may be the exact reason for the study, particularly when the intervention is common.

If risks to nonparticipants exist, discuss the mechanisms through which the risk arises from the study and provide an estimate of the magnitude of the risk and the probability of harm.

4. Potential harms to research participants or research staff from data collection (e.g., surveying, privacy, data management) or research protocols (e.g., random assignment)

Are data collection and/or research procedures adherent to privacy, confidentiality, risk-management, and informed consent protocols with regard to human subjects? Are they respectful of community norms, e.g., community consent not merely individual consent, when appropriate? Are there potential harms to research staff from conducting the data collection that are beyond “normal” risks?

Example of sub-questions to consider as part of the broad question: Are there any risks that could ensue because of the data collection process or storage, e.g. discomfort to being asked certain questions or breach of confidentiality? If so, what are the mitigation strategies? Are there costs to the participant for the data collection process, such as their time, and if so, what is the strategy or rationale for offsetting this cost?

Because these are all issues covered by most IRB processes, a sufficient explanation for a “yes” response may be to provide the IRB approval numbers for all IRBs that have approved the project. However, if there are particular issues that are important to discuss, please do so here.

Harms to research staff could include, e.g., exposure to political violence, exposure to unusual levels of a communicable disease, mistrust due to lack of perceived lack of community consent, or emotional wellbeing from surveying about difficult subject matters. This would not include, e.g., traffic accidents.

5. Financial and reputational conflicts of interest

Do any of the researchers have financial conflicts of interest with regard to the results of the research? Do any of the researchers have potential reputational conflicts of interest?

We define financial conflicts of interest as that used by the researcher’s institutional (e.g., their university) guidelines. We define a reputational conflict of interest as one in which prior writing or advocacy could be contradicted by specific results pursued in this study, and such contradiction would pose reputational risks to the author.

6. Intellectual freedom

Were there any contractual limitations on the ability of the researchers to report the results of the study? If so, what were those restrictions, and who were they from?

This could include, for example, approval of release of the paper and restrictions on data release, but does not include things such as a “comment period” during which interested parties have a right to review and provide comments prior to release but not to control the outputs of the study.

7. Feedback to participants or communities

Is there a plan for providing feedback on research results to participants or communities? If yes, what is the plan? If not, why not?

Engaging in post-study feedback is a way of acknowledging the agency of participants and communities, and is thus a desired practice. However, it may be impractical due to costs, timing, challenges communicating the results, or potential harms if such communication may itself change behavior in undesirable ways.

8. Foreseeable misuse of research results

Is there a foreseeable and plausible risk that the results of the research will be misused and/or deliberately misinterpreted by interested parties to the detriment of other interested parties? If yes, please explain any efforts to mitigate such risk.

In settings with strong imbalances of power between interested parties, there may be foreseeable risks that a powerful party could use deliberately selected research findings to their advantage and to the harm of participants or non-participants, including for general public policy. For example, if the research might reveal the vulnerability of some that can be exploited for the gain of the more powerful party, what steps does the researcher plan to mitigate this risk?

9. Other Ethics Issues to Discuss

Are there any other issues to discuss?