

Conducting Ethical Economic Research

Complications from the Field

Harold Alderman

Jishnu Das

Vijayendra Rao

The World Bank
Development Research Group
Poverty and Inequality Team
May 2013



Abstract

This essay discusses practical issues confronted when conducting surveys as well as designing appropriate field trials. First, it looks at the challenge of ensuring transparency while maintaining confidentiality. Second, it explores the role of trust in light of asymmetric information held by the surveyor and by the respondents

as well as the latter's expectations as to what their participation will set in motion. The authors present case studies relevant to both of these issues. Finally, they discuss the role of ethical review from the perspective of research conducted through the World Bank.

This paper is a product of the Poverty and Inequality Team, Development Research Group. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://econ.worldbank.org>. The author may be contacted at vrao@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Conducting Ethical Economic Research: Complications from the Field

Harold Alderman, Jishnu Das, and Vijayendra Rao

Keywords: ethics, survey data collection; Institutional Review Boards

JEL Code: A13

Sector Board: POV, HNP

The authors wish to thank Gero Carletto, Jed Friedman, and Varun Gauri for shared insights on the topic.

Introduction

In 1993, anticipating the imminent majority government by the African National Congress, researchers in South Africa sought to establish a baseline of economic welfare in order to guide and assess the new government's progress. However, as the [then] Central Statistics Office was widely distrusted, the project was undertaken by a wide consortium of social scientists, health specialists, labor union representatives, and members of civil society convened by the Southern Africa Labour and Development Research Unit (SALDRU). To no surprise, the process of designing a survey by committee generated a number of lively debates. One, in particular, was in regards to documenting informed consent.

Many participants felt strongly that it was imperative that the survey teams not only inform the respondents of the research nature of the survey and the goal of confidentiality but also that the respondent had to indicate that he or she understood this with their signature. Others felt that in the prevailing political climate and with the limited literacy in rural areas this requirement would doom the project. Not only would a significant share of the sample decline to participate, as this group would be self-selected, the sample would no longer be representative of the population. Consensus, not voting, was a means for decision making among the team designing the survey, but casual observation suggested that the views of the social scientists differed from those in health fields, with economists offering an additional nuance, that answering the questions in the survey instrument was a form of revealed preference and thus implicit consent, a view that eventually prevailed. In the event, few respondents refused to be interviewed.

This incident serves as an introduction to some of the themes of this essay. It illustrates that many of the practical issues confronted in designing ethical field trials also need to be

considered in a wider range of data collection. It also hints that there are some glitches when translating the experience of research guidelines in clinical environments into field work undertaken by economic and other social sciences.¹ We group these issues into three broad categories, cautioning the reader that more often than not, we pose questions but have relatively little to say definitive about potential solutions.

Our first category considers the objectives of balancing transparency and confidentiality. Researchers familiar with the act of conducting a survey in a remote village are aware that there is seldom a single “respondent”. Even in the confines of his or her home, the survey respondent typically faces the entire family, with neighbors “dropping in” and a curious set of faces looking in from the window. How to maintain confidentiality in these settings, and what confidentiality *means* in a context where certain information is public—and therefore “OK” to share—but other information is not (and the researchers seldom knows which is which) is an ethically complicated question. The ethical ambiguity is heightened in cases where such confidential information implicates certain members of the family in behaviors that, at least in the eyes of the surveyor, fall outside the moral norms of society. We present a specific case study—surveys on domestic abuse— that takes on this issue.

The second category relates to trust between the surveyor and the surveyed and again goes to the issue of informed consent, and the sharing of information from the survey with the surveyed population. There are three fundamental issues that we address in this context. First, surveys are never isolated from the social history of the surveyed population. For instance, *any* survey on vaccinations among certain populations in India will have to tackle the surveyed populations’ perceptions, and in particular, whether they believe that the survey will lead to

¹ Stark (2012), however, reports that the question of obtaining signatures of informed consent – not of consent *per se* – was also controversial when the National Institute of Health was drafting its guidelines for research on human subjects.

follow-up action from the government, and if so, what these actions may be. In particular, the link between mandated government action on vaccination and previous experience with drastic population control policies has often led to a fundamental “trust deficit” that impinges and defines survey evidence to this date. See for instance Jegede (2007) on the Nigerian boycott of the Polio Vaccination Campaign or Hussain et al (2012) on continuing misinformation about polio vaccines in India leading to population resistance.² Further, the act of surveying in itself could change the surveyed populations’ behavior in the future; a hypothesis for which there is now some evidence. See Zwane et al. (2011) and Das and Leino (2011).

Second, what does informed consent really mean in a low literacy setting? Is consent really informed consent when the respondent does not fully understand the context of the research at hand? The broad notion of informed consent has, to our minds, an implicit social contract where the researcher believes that survey information will provide a better basis for decision making or deepen our understanding of the needs of the surveyed populations in the future. Actions stemming from the research will arguably help improve the lives of at least a subset of the surveyed population. But what should we do in cases where the surveyed population itself engages in practices that are not beneficial—and even harmful—to a broader population of individuals? We have in mind here, for instance, surveys of service providers, including teachers, doctors and licensing organizations. A number of studies in low-income countries have revealed that the behavior of such agents deviate sharply from that which professional ethics in their chosen profession would dictate and the neutrality of the survey team

² Hussain et al (2012) note for instance: *“Though most respondents supported the eradication program and vaccinated their children, many did not seem informed why the program had intensified the frequency of vaccination. Families described that when they asked the door-to-door vaccination teams why they visited them so often, they were usually not given an adequate response. Though the presence of medical interns helped, members of the vaccination team were observed sometimes providing dubious etiological explanations to the families: telling them that polio was “special” and needed a constant boost which other vaccines did not. One clinician who worked with routine immunization services explained that many patients did not understand why, whereas they vaccinated with BCG at fifteen days, DPT at one and a half months, and received a measles injection twice, they had to vaccinate for polio almost twelve times a year until they were about six years old.”*

in these circumstances is ethically indeterminate. Our second case-study presents the ethical dilemmas in this case using research on the behavior of health care providers. We illustrate these problems using on-going research on the interactions between health care providers and their patients in low-income settings, highlighting the difficult decisions that researchers face in such instances.

Our third category relates to review boards in low-income countries and the *absence* of such a board at the World Bank, an organization that routinely conducts more than 500 surveys in low-income countries every year. This scale and the fact that all of the authors of this essay have conducted surveys and trials as World Bank staff (two are still working in that capacity at the time of this writing) motivate the inclusion of this topic in the paper.

This essay will concentrate on practical issues of data collection and of communication with respondents and participants both in observational studies and in impact evaluations of interventions. This largely reflects the expertise of the authors and the large number of surveys that they have conducted over the previous three decades. Each of us has faced the problems that we highlight here in several instances. We have struggled with the answers, and in most cases, muddled through trying to keep the best interests of poor people around the world in view in making our decisions. Our hope here is that by recounting the issues of surveying populations in low-income countries, we do justice to the tough problems that researchers face and the need for a negotiated solution in each case. Many times, these problems (and their solutions) fall in the interstices of the human subjects courses that we take, and more often than not, these departures and a relative lack of guidance are part and parcel of surveying populations today. We recognize, of course, that there are many related ethical considerations in the choice of field studies to undertake and in the assignment of treatments and control groups in evaluation-based

designs. Some of the key issues in such studies are extensions of the data collection themes in this paper. Other issues, both in regard to the economic content of techniques employed and of the ethical challenges in setting up such empirical economic studies are well developed elsewhere (Deaton, 2010; Ravallion, 2012; Barrett and Carter, 2010; Glennerster, this volume) and will be secondary focuses here.

Balancing Transparency and Confidentiality

Participants in field trials and observational surveys have a recognized right that the information they provide will remain confidential. Field conditions, however, often put pressure on the best intentions of researchers. For example, it is difficult to arrange a one-on-one interview in a village setting. Moreover, by its very nature, it is impossible to do so in a focus group. A noted illustration of this challenge was reported by Nancy Scheper-Hughes who was run out of the village where she had conducted an award winning anthropological study after the study was perceived as overly negative and after observers deduced in which village she had worked and with whom she had interviewed despite her efforts to provide anonymity (Scheper-Hughes, 2000).³ This successful unraveling of attempted concealment of identities achieved long before the design of current search engines that make a concerted effort at re-identification likely to be successful.

While generally this is a minor concern for data collection it brings out the practical complexity of standard, though slightly conflicting, principles of research ethics beyond the voluntary nature of participation: the more or less categorical admonition to ‘do no harm’ and the somewhat more utilitarian concern that the risk to participants in research be vanishingly small

³ This incident was highlighted in an essay challenging the goal of anonymity on the grounds it is “usually impossible to ensure” and “often undesirable to try to do so” (Walford, 2005).

relative to the anticipated value of the research. But as illustrated in the following case study on research regarding domestic violence, this is not always the case. This case study comes from a survey conducted by one of the authors in 1992, which was one of the first attempts to collect quantitative data on domestic violence in the Indian sub-continent and, therefore, conducted without the advantage of the experience of other researchers (Rao 1997, 1998, Bloch and Rao 2002).

Case Study: Domestic Violence

When a man hits his wife it is, usually, a private act about which the wife feels great shame and is thus reluctant to reveal, particularly to an outsider. Thus, asking such questions in a survey raises various ethical concerns. Mindful of the sensitivity of the issue - and on the recommendation of various women's activists who had worked on the subject - in 1992 a team of social workers with specific experience in gender issues was hired to conduct field work. The team was strictly instructed to have all respondents sign an informed consent form and to ensure that interviews with women were conducted in private and behind closed doors at a time of day when men were away in the fields. They were also instructed to inform respondents that the survey was being conducted by the University of Michigan (where the principal investigator [PI] then worked).

The informed consent form had a general statement explaining that questions would be asked about the lives of women, that all responses would be kept confidential, and that respondents should feel free to stop the interview at any stage if they felt uncomfortable with the questions. The first ethical question raised by this is: Was this enough? Is documented "informed consent" - which followed the strict guidelines of the University of Michigan's

Institutional Review Board (IRB) – meaningful? In practice, over half the sample (consisting of married women of reproductive age) was illiterate. Consequently, proof of consent was obtained by thumbprint. Both signatures and thumbprints in the rural Indian context are associated with interactions with the state, which makes it very likely that the majority of respondents viewed the survey, regardless of the information they received from the interviewer, as associated with the government. Given this, almost all respondents likely viewed it as a precursor to some kind of benefit or intervention rather than as merely an interview. This raises two important questions: First, what is the meaning of “consent” when it is conveyed by signing a consent form in a context where the process of signing the form is very similar to interactions with the government? Second, with the consequent expectation of the possibility of assistance (which the interviewers know is not forthcoming) what is the ethical response to observations of egregious abuse?

In this particular case no one who was approached refused the survey and everyone who started the survey completed it. Many women who had suffered violence were hesitant to speak about it but did so with encouragement from the social worker-interviewers. In many cases this was a cathartic experience for the respondents who had never had the opportunity to share their trauma with a sympathetic stranger. We heard several heart-wrenching stories of wife-abuse that surprised us by the length of their duration and severity. In some instances the violence was current and ongoing, particularly with younger wives with alcoholic husbands or who were stuck in the middle of dowry disputes. But, at the end of the conversation, the survey team simply walked away without being obliged to intervene or to help in any way. A claim can perhaps be made - and is often made by researchers of domestic violence - that the very act of having women share their stories of trauma is healing, but the claim seems a little self-serving.

Most surveys in most developing countries with high levels of illiteracy face similar ethical concerns and survey researchers are acutely aware of these concerns. Often there is an expectation of help that underlies a respondent's consent to the survey. In cases where the survey is collecting relatively benign information, say on demographic characteristics, or consumption patterns, this is perhaps less of an issue, particularly since respondents have usually been subject to such survey questions in the past. But when the survey is trying to obtain information on particularly traumatic events, the question of what constitutes consent and the ethics of the obligation to intervene and help becomes particularly acute.

Collecting information on issues that are private, shameful, or imbedded within struggles for power in the surveyed population potentially violates an even more serious ethical principle than informed consent: the principle of "do no harm." Despite all our efforts to keep the interviews private, it was impossible to keep the nature of our questions a secret. When outsiders come to a village they are objects of curiosity, and people wonder what they are doing there. This can have serious consequences for respondents.

After completing the survey interviews the team of social workers started conducting focus group discussions with the female respondents. One of these was held in the home of an older woman whom the survey team had befriended. A group of about 8 women had gathered for the discussion at about 11 in the morning. Given the dense clustering of homes in the village we could not keep the location of the group discussion a secret. About half an hour into the discussion a man burst in to the room (the door was unlocked) and furiously walked over to one of the women and hit her in the head screaming "get back and do your work you ----, how dare you ignore your children. You should be at home cooking, not talking to these people." He then grabbed her right ear and pulled her out the room. We knew from our interview with her that the

woman was a victim of domestic violence. Clearly our interview had made her husband angry enough that he had come back home from the fields to punish his wife publically and make the point that talking to us would have adverse consequences. The very act of conducting the interviews had raised the risk of violence against our respondents.

Could this have been prevented? It is not clear how. The survey team was experienced with women's issues. The sample size was small (less than 200 women) so the PI was able to personally supervise the survey. All existing protocols on informed consent had been followed. Every effort was made to keep the interviews (especially the individual interviews) confidential. But, the very act of asking questions of this nature is fraught with risk for the respondent. And asking about traumatic events makes concerns about the nature of informed consent in low literacy populations even more acute.

Such concerns are common to any effort to collect information on domestic violence and other issues that are private, involve some shame, and are embedded within power relationships. Cognizant of these concerns, and also of the importance of collecting representative data on domestic violence, the World Health Organization (2001) has developed a set of ethical guidelines which they follow in collecting their influential data on domestic violence in several developing countries⁴. There has been an explosion of research on the topic and a vast expansion in the availability of data on domestic violence over the last decade. The

⁴ The guidelines are: 1) The safety of respondents and the research team is paramount, and should guide all project decisions. 2) Prevalence studies need to be methodologically sound and to build upon current research experience about how to minimize the under-reporting of violence. 3) Protecting confidentiality is essential to ensure both women's safety and data quality. 4) All research team members should be carefully selected and receive specialized training and on-going support. 5) The study design must include actions aimed at reducing any possible distress caused to the participants by the research. 6) Fieldworkers should be trained to refer women requesting assistance to available local services and sources of support. Where few resources exist, it may be necessary for the study to create short term support mechanisms. 7) Researchers and donors have an ethical obligation to help ensure that their findings are properly interpreted and used to advance policy and intervention development. 8) Violence questions should only be incorporated into surveys designed for other purposes when ethical and methodological requirements can be met.

Demographic and Health Surveys, which have collected nationally representative data on domestic violence for a variety of countries, state that they follow a modified version of the WHO guidelines (Kishor and Johnson, 2004). However, given the extremely large samples involved (ranging from 2,000 to 90,000), it seems fair to ask if such guidelines can be effectively implemented at reasonable cost.

What Does Privacy Mean?

Beyond the difficulty of conducting private interviews illustrated with the challenging example of studying domestic violence there are more general issues of privacy to consider. For example, once data is collected the researcher often has limited say in regards to the subsequent dissemination of the data. Indeed, while the ethical rationale for privacy is readily apparent, there are also ethical considerations that encourage wide data access. For instance, students, particularly in low income countries, generally have limited access to research grants and other funds that could be used to collect data that will further their research. From this perspective broad access is less extractive than exclusive rights to use a data set granted to a single research team, often from outside the countries where the data has been generated. Moreover, as field research is not as easily replicated as are clinical studies, public posting of data provides a check against inadvertent research errors or excessive data manipulation. Thus, it is common policy that funders of research require that the data collected as part of the project be made publically available within a specified period.⁵

⁵ The research department at the World Bank has such a policy. Additionally, the World Bank strongly encourages governmental statistical offices to make data collected using instruments it has designed, such as the Living Standards Measurement Survey, publically available. It often provides a draft letter of data protocol for the country statistical director as part of technical assistance. However, the ultimate decision on data access to such surveys remains within the statistical office and is governed by local laws and the preference of the government.

Implementing data access, however, poses a few practical issues. While individual identifiers are generally stripped out of data that are publically posted, community identifiers often are not. For example, global positioning system [GPS] coordinates are frequently used in data analysis and are valuable for steps such as determining the distance from the household to a school or clinic. Since, in principle, GPS data can be used to locate a household, it is common practice to either use a shifter to disguise the original site or to convert GPS information to commonly used distances and then to remove the raw indicator. Still, a determined researcher can likely use location data to extract information as to the likely households that participated in the survey and undermine the guarantee of anonymity. Fortunately, however, such steps increase the required effort and thus plausibly make the exertion not worth undertaking.

There are also times when the research actually may want to retain the identity of the household in the data set. Not only is a baseline-resurvey approach a standard research tool, panels are increasing used to identify long term impacts of programs (see, for example, Maluccio et al. 2009) or the dynamics of investments in children (see, Helmers and Patnam, 2011). To reduce attrition in such panel data sets, researchers commonly ask for names of individuals who will be in a position to help locate the initial respondents (and, not infrequently, their children) in the future. As the team conducting follow up research may be quite distinct from the original research team – and, indeed, the initial survey team may not have even conceived of the topic of the follow up study and thus would not have discussed the subject of the new research with the respondent – the validity of the initial confidentiality agreement ultimately rests on the chain of trust between the respondent, the initial researchers, the repository of the data, and the subsequent users of the archived data.

Trust and Communication: Informed Consent and a Two-Way Exchange of Information

Trust is also at the core of the concern for deception in laboratory experiments conducted by economists. Jamison, Karlan, and Schechter (2008) note that deception is generally proscribed in protocols for laboratory experiments in economics in contrast to, say, psychology and offer guarded evidence that this reflects, in part, the potential costs of lost trust. They claim, further, that this particular concern has little overlap with purview of ethical review boards.

Does the observation that participant pools at universities may be influenced by deception have any relevance for field work? Mistrust, particularly of distant governments and of strangers is ubiquitous and ignoring the influence of the various social constructs that have evolved to reinforce social cohesion is likely to result in misleading research. This point was made apparent to one of the authors of this essay [Alderman] in a futile attempt to elicit time preference in rural Gambia using standard instruments in a survey pretest; the virtually infinite discount rate implicit in the respondents' preference for small immediate payments over larger future payments said much about trust but little about inter-temporal utility.

The absence of trust undermines the ability to implement protocols for informed consent. While researchers must provide the respondents a background on the purpose of the data collection there is no guarantee that even when this is clearly communicated the respondents will take this at face value. As many respondents will have little familiarity with research it is not surprising that they will try to understand the surveyors' motive in terms closer to their own experience and concerns. This natural tendency fuels speculation – often with rather creative proposals. An example of fairly benign attempts to place research in the framework of local concerns, comes from one of the author's [Alderman] study of allocation of time between farmers' own fields and landlords' as part of his thesis during Indira Gandhi's emergency in

1976. At the time, a local teacher reported that his neighbors believed that this work was preparation for the return of the Raj; noteworthy, this was perceived locally as a worthy endeavor. More worrisome, however, such speculation may be a fertile area for manipulation, as tragically noted in regards to some thwarted interventions such as polio inoculations in Nigeria (Jedege 2007). While outright refusal to participate in research is relatively rare, at least in rural contexts, strategic response can be influenced by the perceived motives of the survey team, for example, in regards to collecting information believed to be useful for taxation. Thus, informed consent, with or without written documentation, again reduces to trust.

But to whom do the researchers owe their primary responsibility? This is at the heart of ethical research but is not without its ambiguities. We discuss here, in some detail, the particular case of poor service provision by doctors in low-income countries, a topic on which one of the authors has led survey design for the last decade.

Case Study: Health Care Providers in Low-Income Countries

In early 2003, researchers at the World Bank started collecting unique information on teachers and health care providers to help guide key policy decisions, ranging from regulatory concerns - should unqualified health care providers be allowed to practice - to specific investments for public providers. The initial surveys focused on the availability of staff and resources in “frontline” clinics, documenting both high levels of absence and a leaky funding pipeline from ministries to schools or clinics.

Using these facts as a starting point, in a series of papers Das and Hammer (2005, 2007b) – henceforth, DH - ask questions about health care provider quality and prices, an agenda squarely within the discipline of the organization of firms and industries. They soon discovered

that there were no data and no established method for assessing provider quality in these settings. Many different measures had been proposed—ranging from the availability of medicines to the quality of infrastructure or to patient satisfaction—but these data had seldom been validated against other measures of the quality of medical advice and theoretically, they held little promise. The availability of medicines—a widely used measure of health care *quality*—for instance, is high in a commercial chain pharmaceutical in the U.S., but few individuals would go to one if they had chest pains; at most *free* medications was a measure of subsidies rather than the quality of care.

Over 3 years, DH developed protocols that would measure both the technical *competence* of providers (that is, how much they knew about medical conditions) and their *practice* (that is, what they do when an actual patient comes to them). Measurements of competence were based on medical *vignettes* (tests based on hypothetical patients given to doctors) and fell within standard protocols for surveys of firms, an activity with a clear precedent and well defined notions of “informed consent”. Measurements of practice, on the other hand, required them to observe doctors with their real patients and here things became trickier.

To consider why, think of the kinds of clinics they were visiting. Since their sample was selected from a list of all providers in select neighborhoods, they were sitting with providers ranging in expertise from those with a degree from the United Kingdom to those who had no medical education at all. In many cases, “clinics” were small one-room affairs facing the main road. The doctor sat behind desk and one or two benches for the patients spilled over into the street. Patients sat on the bench, and the doctor saw the one who was sitting closest to him. As they document, the average interaction lasted 3 minutes, the doctor asked 3 questions, performed 1 exam and gave 3 medicines. In many cases, these medicines would be taken from a bottle that

contained a large number of pills (bought from one of the “wholesale” markets for medicines in Delhi), put in a mortar, crushed together and dispensed in a small piece of paper. There was no hint of privacy, and none was expected: Every patient observed the full doctor-patient interaction, and, only in cases where a woman’s examination required her to undress, did the doctor use a second room, separated from the main clinic by a curtain. The time spent with patients was low on average, and among the busiest doctors, each interaction lasted 30 seconds. On the data front, the key challenge that the surveyors had to meet was maintaining similar data quality irrespective of the doctor’s practice. From the viewpoint of the analysis down the line, data on clinical practice that emerged from crowded clinics where interactions are short and doctors may be “seeing” multiple patients at the same time had to be as accurate as data from a clean and well-kept clinic where patients were seen one-by-one in private.

In one notable example, the two researchers and a surveyor approached a doctor’s clinic in a low/middle-income locality. On seeing them approach, the “doctor” promptly got up and disappeared before the researchers could reach the clinic, leaving behind 3 patients—one at the table, one lying down on a bench with an oxygen mask and a third waiting behind. The researchers then sat in a tea-stall for 30 minutes, until the doctor re-entered the clinic. They then split up and approached the clinic from different directions and managed to talk to the provider. The provider was forthcoming, and immediately gave consent to observing interactions between him and his customers—as long as the researchers went to his *real* job, which was selling bolts and other industrial material in a hardware shop! The medical practice, he insisted, was a side-line: *“I only do this for charity, because these poor people don’t have others to go to. Who knows, maybe one day soon, I will decide to close this up and go somewhere else. So what is the point of sitting here with me? Come with me to my real shop, and you can sit there all you*

want.” Notably, the provider’s clinic had been in operation for more than a decade and there were close to 60 other medical clinics in the same locality, with similar prices at his level of quality. Ultimately, it took us four visits to this provider to convince him of the anonymity of the survey as well as the benefits of allowing us to collect these data within the context of a broader democratic debate on health care.

On the ethical front, several issues were raised. To begin with, what would “privacy” mean in this context? There was a concern that by observing interactions we could be violating the privacy of the doctor-patient interaction. Yet, in a context where such privacy was violated by the very nature of the practice and was expected neither by the doctor nor the patient, it appeared to us that further observation by a silent observer would not be detrimental to the patient or the provider as long as informed consent was obtained. And here the researchers ran into further problems. Technically, informed consent would require them to approach both the patient and the provider. That makes sense when there is a separate waiting room and patients can be “enlisted” prior to the actual interaction, neither of which held in this context. In the pilot phase, explaining the informed consent form to the patient in every case took an average of 5-6 minutes and DH could not do this prior to the actual interaction; in many cases, this turned out to be a distraction to the doctor and to other patients. DH started asking patients what they thought of the informed consent, and they reported it to be a bother at the time of the interaction; the suggestion from patients and doctors instead was for us to take the consent from the doctor and elicit the patient’s feedback only if (a) the patient specifically requested information on the observer or (b) the doctor wanted to examine the patient in private (again, mostly female patients). Both patients and doctors agreed that this would be a far more efficient way of proceeding, as it would retain the sanctity of the interaction and cause no disruptions to the

practice. DH's eventual decision was to take informed consent from the provider (*"We will be sitting with you for the duration of your practice today and taking some notes. During this time, we will observe your interactions with patients, but at no time will we ask questions from you about these interactions. If you, or the patient, feel that you would rather not have the observer present for a particular interaction, please let us know and we will leave the room"*) and, in the case of private examinations, DH decided to omit the observation entirely (for survey purposes, this happened in less than 0.5% of cases).

The researchers' final ethical dilemma was the identity of the observer. At the beginning of the study, DH thought that they could use trained nurses and/or doctors for the observation, but the ethics of this were difficult to handle. Specifically, in many cases (more on this below), the doctor diagnosed the patient wrongly, gave the wrong medicines and may have engaged in behavior detrimental to the patient (such as using an unsterilized injection). Doctors as observers now faced a dual dilemma that could not be resolved. On the one hand, they were required in their role as doctors to inform on the interaction and provide better care to the patient. On the other, they could not break the sanctity of the provider-patient relationship *unless* the observed provider actively solicited a second opinion. Eventually, DH had to give up on this idea and use non-medical personnel who had no medical judgment regarding the quality of the interaction to conduct the observations.

The results from this study were startling in what they disclosed. DH demonstrated empirically the very low levels of effort that providers exerted in their clinical interactions. They were also able to show that effort was a key component of medical care quality—public sector providers were more *competent* than unqualified private sector providers, but because of abysmally low effort levels (an average interaction of 90 seconds) actually provided medical

advice that was *worse* than that provided by unqualified providers. The results opened up a substantial debate on the practices of the medical sector and the importance of incentives in the profession, a debate that eventually led to an incentive-based design in India's quest for universal health coverage.

Clearly, the researchers muddled through. DH made several decisions that reflected both their own prior beliefs that the study was critical to understanding the nature of health care delivery in India and to provide information for better policy that would eventually help the poor. Each of the 3 key decisions: "violating" privacy, informed consent only from providers and not using medical personnel could rightly be criticized using strict definitions from human resource subjects courses. Yet, when DH asked the fundamental question "how could this study be done with full privacy, informed consent from every patient and medical personnel as observers", people who were commenting could offer no alternative; they even heard the view that "such data should not be collected because all the ethical protocols cannot be strictly followed." Yet, DH were not harming either the patient or the doctor, the data that they collected played (and continue to play) an important role in improving the lives of the poor, and subjects in the research (both doctors and patients) agreed with the study protocols. And, in many cases, patients actively welcomed our research on medical practices. It appears to us that we have little guidance in such situations.

Addressing deeper questions regarding the validity of diagnoses obtained by patients and differences between doctors in the treatment of patients raises further issues pertaining to the ethics of research on survey delivery. Specifically, the DH studies using interactions with real patients were still subject to biases in estimations of practice-quality across doctors due to unobserved case and patient selection. For instance, it could be argued that lower process quality

among public doctors reflected a different patient or sickness profile. Furthermore, the DH method, when combined with their inability to use medical personnel in the observation stage, did not provide any information on the accuracy of diagnoses; the researchers suspected that the accuracy of diagnosis was low, but could not validate their suspicions with data.

Therefore, in 2010, Das and others (2012) deployed *standardized patients*—surveyors recruited from the local community and extensively trained to portray standardized presentations of limited cases—in a sample of health care providers in Delhi and the rural regions of an Indian state. Such standardized patients (SPs) are widely recognized as the “gold standard” in measurements of quality, but the use of SPs in large samples of randomly selected providers, many of them with low medical training, had not been previously attempted.

In Delhi, the study sought informed consent from all providers (where the providers knew that a standardized patient would be sent, but did not know the date) and follow-up surveys showed (A) a single case of adverse interaction between the provider and the patient (out of 230 interactions) and (B) less than a 1 percent detection rate for the SPs. On the basis of this pilot, the study was rolled out in the rural district *without* informed consent from providers, who were all part of a larger study (with informed consent) similar to that in the DH Delhi studies. The reasons for the lack of informed consent stemmed primarily from the remoteness of the sampled populations. In rural villages, most patients were known to the doctor and although unknown patients often presented to the providers, the combination of announcing the SP study *and* a remote rural setting would have raised suspicions, invalidating the study design. The study was cleared through Institutional Review Boards at Harvard University and Innovations for Poverty Action. For the purposes of the discussion, we note that the study demonstrated, for the first time, the very poor diagnoses rates in Indian health care (the correct diagnosis of unstable angina

was provided to 12 percent of patients) and the complete lack of correlation between the quality of care and the availability of equipment and medicines or patient-load. Further, the study also confirmed the earlier DH results that quality in the public sector was *worse* than among completely untrained private sector providers. See Das et al (2012).

Despite the official clearance from IRBs, again, there are tricky issues that arise in the use of informed consent and the nature of the study. Had the researchers been forced to inform providers of the arrival of SPs, detection rates would have been high in the rural context and the study could not have been completed. In the course of the study, careful safeguards ensured that all data were anonymous *and* that no provider even knew that they had participated in such a study. Therefore, it is difficult to argue that there was any harm to either patients or providers through the course of the study, but a strict interpretation of informed consent “at all costs” was not possible in this context.

This issue relates to a wider debate over the use of stealth surveys by anonymous observers sent to ascertain the quality of services.⁶ Other examples from the literature pertain to the use of false shoppers or patients who observe the activities of a doctor, teacher, or other agent but who do not seek their consent or inform them directly of the purpose of the visit. In such situations a researcher is presumably seeking an assessment of average quality of services and not that of a specific individual. Thus, the information is expected to remain anonymous and not to be used to punish the individual or seek damages. In partial contrast, employers may use stealth clients to monitor performance usually with the employee informed that they may be observed in this manner. Researchers are sometimes advised to use a similar strategy of informing potential subjects (participants seem an inappropriate label in this circumstance) that

⁶ See a blog by Jed Friedman on this subject, <http://blogs.worldbank.org/impactevaluations/sometimes-it-is-ethical-to-lie-to-your-study-subjects>

they may be observed without indicating who will be studied or when this might occur. However, it is not clear that there are always practical means to communicate this to the service providers (we return to the logistics of communication below). Moreover, this may lead to strangers, often easy to recognize in a remote rural setting, receiving higher quality services than usual. Thus, the clearest ethical guideline for such types of research pertains not to the rights of individuals being observed, but the interests of the wider community which these individuals are expected to serve, since any such stealth visits should be designed in a manner that does not crowd out or otherwise reduce access for the general population.

Further Discussion on Communications with Subjects beyond Informed Consent

Survey respondents or trial participants are not only expected to be informed of the purpose of the study, they are generally expected to be informed of the results. Chambers (2001), among others, makes the point that surveys take the time of respondents and to a degree interviews are, therefore, extractive, even if the consent to be interviewed is voluntary.⁷ Communication of the results as they pertain to the community then can be viewed as a minimal compensation for participation, although even the process of an interview serves a degree of communication since the interview may raise awareness of health or other concerns even in the absence of a direct intervention. The advice to share results is, however, more practical for focus groups than for broad surveys. The statistical power of surveys is determined by the interplay of the number clusters or communities included in the sample and the number of respondents in each cluster. Transport costs – which are often inputs into software used for survey design – are a key factor

⁷ However, this observation does not, by itself, imply that such data collection is unethical; taxes are also extractive yet viewed by most economists as justifiable to fund public goods even if they entail various distortions which economists seek to minimize. The value of the tax on the respondent's time, then, is a function of the public value of research.

in determining these two elements of sampling. These costs will also be a factor in the ability of a research team to hold community level meetings to disseminate results. However, this obstacle is reduced in the case of panel data collection in which repeated visits to the community provide opportunities to share observations. Still, there is a challenge to communicate research results from a survey in which the community is a cluster in a larger sample when the results are not designed to be representative at the cluster level.

Rather than - or in addition to - sharing results, respondents can be compensated more directly with cash grants or small gifts in kind. The former, however, are hard to track and may end up with members of the field team instead of the households. Gifts may be in the form of items useful for the households, such as soap, a flashlight, or daris (light blankets commonly used in India and Pakistan) or they may be a collective good for the community such as a repair to a school roof. While such gifts are modest relative to the cost of data collection, they may be significant relative to the daily wage of poor households and, thus, increase the incentive for the household to seek to provide the information that they believe the survey team would like to receive. This may be due to the fact that a gift is offered as much as it is to the value of the gift *per se* (Ariely, 2008). Whether this compensation biases the data collected is an area for research.

Even in the absence of direct payments, survey design can lessen the burden on respondents both by paring down the size of the survey instrument (a practical consideration for quality as well) and by scheduling the data collection in seasons and times of day that least interfere with work, although it is difficult to envision what time fulfils that recommendation in regards to a child care provider.

Generally, individuals who have a threatening but treatable ailment that is found in the course of a clinical trial are removed from that trial and given appropriate therapy.⁸ This stands in partial contrast to communication with respondents in observational studies, for example in regards to whether they or their children are malnourished or anemic. Arguably there are few curative (rather than preventative) therapies for stunting, although there clearly are recommended practices to address the heightened risk of mortality attendant to severe acute malnutrition as indicated by low weight for height (wasting) or low upper arm circumference. The former measure of weight for height requires comparison to international reference tables and is often not calculated in the field, although upper arm circumference cutoffs for acute malnutrition can be assessed in the field with the data routinely collected in many household surveys.

Protocols for Demographic and Health Surveys advise direct advice on anemia for individuals deemed at risk and referrals to clinics for severe cases⁹. However, neither acute malnutrition nor anemia can be treated in a single visit and in many contexts knowledge of the status is ineffective without the presence of a health system that facilitates regular interaction between health workers and the individual at risk.

Communication at the cluster level. One of the authors of this essay submitted a randomized cluster-based trial of deworming that analyzed data from over 30,000 children derived from registers at child health days in Uganda. The paper was initially rejected primarily because it was not a double blind study. A letter was sent to the journal suggesting that scale trials should have different methodologies than clinical trials. The journal's 'hanging committee' – their

⁸ This may set up an incentive for willful ignorance in which a researcher avoids ascertaining information such as parasite loads or anemia before the final stages of a trial.

⁹ Some techniques for assessing anemia such as Hemocue© require additional analysis not generally conducted in the field.

choice of title - agreed and the paper was eventually published (Alderman et al. 2006). Do we need a similar modification of agreed ethical guidelines for large scale and cluster based studies? This is not just a question for economic programs; the issue has been mooted in various health journals, sometimes in an expected benefit, expected risk framework (Osrin et al. 2009; Winkins et al. 1997). Nor is the issue of how scale influences informed consent raised only with controlled trials. Scale affects consent even in observational studies since a community leader is often the initial contact before individuals are approached for their consent. Whether the community leader represents the full interests of the population is often doubtful. In the SALDRU data collection mentioned above, much of the non-response from the initial sample was due to a landowner refusing access to his tenants or hired laborers.

In the case of a cluster-based trial, often a local official or community leader provides consent for the cluster. Indeed, in most observational studies, as well as many impact evaluations, the central government's approval is considered binding on the community. Local officials are often relatively powerless to decline involvement in a survey undertaken when the central government is a partner in the study or when participation in an impact evaluation has been approved at a higher level. Similarly, households may feel compelled to follow the lead of the community leader.

In some cases individuals themselves can opt out by not taking up the service offered. This take-up provides important practical information and the difference between the intention to treat and the treatment on those taking up the treatment is an important aspect of the analysis. But in many cases all members of the cluster share some opportunities and risks. As communities are generally heterogeneous in their needs, the consent of one individual, even if this person was chosen by consent of the community, does not cover the risks for all individuals

affected by the intervention. In principle, a researcher seeking consent can stratify by gender, caste ethnicity etc. and seek consent from a representative of each strata. In practice, this is rare.

In a similar manner, as some interventions change power relations in a community – either as a direct goal or as an indirect outcome – refusal for participation may perpetuate inequality and thus be unethical on a criterion rather different than that which guides clinical trials. Imagine, for example, excluding a community from an evaluation of an improvement in access to education in the American South in the 1950s because the local mayor declined or dropping an experiment to mediate inheritance claims of widows in Sub-Saharan Africa on the view of a single male dignitary. At the very least, the results of the study of those communities that did not opt out will be biased. More generally, the option of having a representative decide on participation in cluster based trials is clearer in medical trials where power relations and distributions of resources are not central.¹⁰

Review boards in cross-country research. Ethical review boards, often multi-disciplinary, have been standard for a sufficiently long time for their overall rules of operation to be well worked out to the satisfaction of legal advisors as well as ethicists (Stark, 2012). There are even a number of companies that provide ethical review services on a commercial basis. Many health journals require that submissions indicate the nature of the review process that the study utilized, although this is not commonly required by economic journals even in the case of trials, never mind in the case of observational studies.¹¹ The World Bank generally follows the lead of the WHO as to ethical guidelines and has a link to these guidelines for applicants to funds approved

¹⁰ However, this approach of using a representative to prove consent in cluster trials is generally advocated for low risk interventions.

¹¹ Economic journals are also less likely to solicit letters criticizing research design or even challenging the published results.

by the research board. However, it does not have either an internal board dealing specifically with ethical research on human subjects nor a policy on the utilization of external boards.

Discussions among staff as to the need for such a policy have stressed, among other things, the reputational risk for the Bank. Yet, it seems to us that a review board that simply mimics boards in universities and the private sector would be a missed opportunity for the World Bank in an area where it has decades of field experience in a large number of countries. Three particularities of the World Bank make it fundamentally different from other research institutions.

First, unlike most American and European universities, the bulk of the research that the World Bank engages in is in low-income countries, rather than on the United States and Europe. As we have sought to demonstrate through this paper, the legal framework, the experience of field-work and the organization of communities, households and firms is very different in each of the countries in which we have worked. Consequently, a bio-medical model of ethics largely developed in the U.S. (for instance, all researchers in the U.S. are required to clear a human subjects course designed by the National Institution of Health) has serious limitations when applied to much of the research undertaken at the World Bank. As one example, there is a criterion that you can't with-hold known efficacious treatments from people. But what if you know that it is efficacious on one dimension but not on another? This is may be the case with deworming---we know the effect on health, but not on education for instance.

Second, much of the research at the World Bank is jointly undertaken with its “operational” arm in projects that are run by the Governments of the concerned countries. Every operational project is subject to a number of “safeguards” designed to protect the vulnerable and ensure that environmental concerns are adequately addressed. Every such project also generates

a large amount of data, ranging from surveys of beneficiaries and households to bespoke monitoring and information systems that track resources and aid. An open question is whether projects that also have a research arm—perhaps asking the simple question of whether project resources reached the intended beneficiaries—should be subject to further IRB clearance when the project itself is not.

Third, it has also been pointed out that the presence of an internal review board would hardly suffice when a researcher conducts a study in one of its 188 member countries and the internal board's view is not in accord with local boards. Therefore, although having an internal board might facilitate reviews by host country panels, this would not supersede the need for local clearance. On the other hand, not every country has a board that is functional and diligent. Thus, local review may also not be sufficient. Whether one should privilege the views of the local board of the World Bank's internal board in these cases remains an open question. One option is that the union of the recommendations of both a local board (if any) and that of an external board might be prudent.

By highlighting these three issues we do not, by any means, wish to argue for World Bank exceptionalism when it comes to ensuring that research is carried out in an ethical manner consistent with the goals of the institution. Quite the opposite. The WB should be equally ethical in its research, but given the vast experience it has accumulated, it should actually be a leader in addressing the issues that arise in the contexts it works in. We feel that the World Bank is actually in a unique position to lead in the creation of an institutional review board that can move beyond the Eurocentric bio-medical model that is currently in use. The process of creating such a review board could bring up many questions that researchers working in low-income countries

face, and over time, these issues could be addressed by a panel of ethical experts backed by the considerable experience of researchers in the institution.

For instance: Is a different process of review needed for observational data collection as opposed to reviews for research that involve trials or treatments to a group of participants? Or, how should reviews that vet questionnaires and are often cumbersome since, in principal, redesign after pretesting requires additional clearance, be handled? Various types of data collection involving focus groups or other qualitative techniques are inherently open ended and thus cannot easily be preapproved. How should the review process take this into account? These are all difficult questions and they add to the considerable ethical complexities that we have all faced in the field and tried to document here. At this stage, we feel that the Bank should start from the beginning by first documenting the kinds of issues that have arisen in both Bank projects and Bank research, perhaps by looking at a random set of publications and projects within the institution. Such a step would enable the creation of a database of ethical issues that we feel have been insufficiently addressed by the current IRB process—not surprising given the very different contexts that researchers in low-income countries work in. Our hope is that the conversations around such a process would generate valuable insights into how to proceed—not just so that World Bank research is “covered” by an ethical approval, but so that that institution can become a global leader in thinking of ethical research in resource poor settings.

Conclusion

This essay presents no objections to or modifications of standard guidelines and protocols for research ethics. We do, however, argue that these protocols are tools, the application of which in the field relies on trust between the research and the community, among other things. That is,

the guidelines are imperfect substitutes for trust; the guidelines can assist a researcher in handling the burden of that trust but they cannot eliminate ambiguities.

Among the ambiguities is the question: To whom is the researcher obligated? This question needs to be confronted, for example, in cases of conflicts of interests within a household or within a community – conflicts that seldom arise between participants in medical trials. They occur quite clearly in the situation of domestic violence illustrated above, but also in cluster-based studies where individuals are asked to speak for the larger community. They also arise between principles and their agents or between service providers and their clients.

Economists generally seek Pareto improvements where actions harm no one and help some. No harm, of course, also lies at the very core of research ethics. Economists know that Pareto improvements are rare outside a classroom discussion. Fortunately, no harm is the norm in research. But as we attempt to illustrate here, researchers in the field do have to be acutely aware that there are times when they need to go beyond existing protocols in order to judge whose welfare are they obligated to protect.

References

- Alderman Harold, Joseph Konde-Lule, Isaac Sebuliba, Donald Bundy, and Andrew Hall. Increased Weight Gain in Preschool Children due to Mass Albendazole Treatment Given During “Child Health Days” in Uganda: A Cluster Randomized Controlled Trial. 2006. British Medical Journal, 333:122-126.
- Ariely, Dan. 2008. Predictably Irrational. The Hidden forces that Shape our Decisions. New York: HarperCollins.
- Barrett, Christopher and Michael Carter. 2010. The Power and Pitfalls of Experiments in Development Economics: Some Non-random Reflections. Applied Economic Perspectives and Policy 32(4): 515-548.
- Bloch, Francis and Vijayendra Rao, “Terror as a Bargaining Instrument: A Case Study of Dowry Violence in Rural India,” American Economic Review, Vol 92, #4, September 2002, Pp: 1029-1043
- Chambers, Robert. Qualitative Approaches: Self-Criticism and What Can be Gained from Quantitative Approaches. In Kanbur, Ravi (ed.). 2001. Q-squared : Combining Qualitative and Quantitative Methods in Poverty Appraisal. New Delhi: Permanent Black Press.
- Das, Jishnu and Jessica Leino. 2011. “Lessons From an Experimental Campaign for Evaluating the RSBY.” Economic and Political Weekly. August 6th 2011, Volume XLVI (32).
- Das, Jishnu and Jeffrey Hammer. 2005. “Which Doctor: Combining Vignettes and Item Response to Measure Doctor Quality.” Journal of Development Economics 78: 348-83
- Das, Jishnu, Jeffrey Hammer and Kenneth Leonard. 2008. “The Quality of Medical Advice in Low-Income Countries.” Journal of Economic Perspectives, 22(2), pp. 93-114.
- Das, Jishnu and Jeffrey Hammer. 2007a. “Location, location, location: Residence, Wealth and the Quality of Medical Care in Delhi, India.” Health Affairs 26, No. 3: w338-w351.
- Das, Jishnu and Jeffrey Hammer. 2007b. “Money for Nothing: The Dire Straits of Medical Practice in India.”, Journal of Development Economics, 83(1): 1-36
- Das, Jishnu, Alaka Holla, Veena Das, Manoj Mohanan, Diana Tabak and Brian Chan. 2012. “The Quality of Medical Care in Clinics: Evidence from a Standardized Patient Study in a Low-Income Setting”. Health Affairs, Vol. 31 (12): 2274-2784.

- Deaton, Angus. 2010. Instruments, Randomization, and Learning about Development. Journal of Economic Literature. 48(2): 424-455.
- Helmers, Christian and Manasa Patnam. 2011. The Formation and Evolution of Childhood Skill Acquisition: Evidence from India Journal of Development Economics. 95 (2): 252-266.
- Hussain RS, McGarvey ST, Shahab T, Fruzzetti LM (2012) Fatigue and Fear with Shifting Polio Eradication Strategies in India: A Study of Social Resistance to Vaccination. PLoS ONE 7(9): e46274. doi:10.1371/journal.pone.0046274.
- Jegede AS (2007) What Led to the Nigerian Boycott of the Polio Vaccination Campaign? PLoS Med 4(3): e73. doi:10.1371/journal.pmed.0040073.
- Jamison, Julian, Dean Karlan and Laura Schechter. 2008. To Deceive or Not to Deceive: The Effect of Deception on Behavior in Future Laboratory Experiments. Journal of Economic Behavior and Organization. 68: 477-488.
- Kishor, Sunita and Kiersten Johnson, 2004. *Profiling Domestic Violence: A Multi-Country Study*, Calverton, Maryland: ORC Macro.
- Maluccio, JA, Hoddinott, J, Behrman JR, Martorell R. Quisumbing, AR, Stein, AD. The Impact of Improving Nutrition During Early Childhood on Education Among Guatemalan Adults. Economic Journal, April 2009, v. 119, iss. 537, pp. 734-63.
- Osrin, David Kishwar Azad, Armida Fernandez' Dharma S Manandhar' Charles W Mwansambo, Prasanta Tripathy' and Anthony M Costello. 2009. Ethical challenges in cluster randomized controlled trials: experiences from public health interventions in Africa and Asia. Bulletin of the World Health Organization. 87(10): 772-779
- Rao, Vijayendra, "Wife Beating in Rural South India: A Qualitative and Econometric Analysis," Social Science and Medicine , Vol. 44, # 8, Pp. 1169-1180, April 1997.
- Rao, Vijayendra, "Wife-Abuse, Its Causes and Its Impact on Intra-Household Resource Allocation in Rural Karnataka: A "Participatory Econometric" Analysis, in Gender, Population, and Development, M.Krishnaraj, R.Sudarshan, A.Sharif (ed.), Oxford University Press, 1998.
- Ravallion, Martin. Fighting Poverty One Experiment at a Time: Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty: Review Essay. Journal of Economic Literature. 2012. Journal of Economic Literature, 50(1): 103-14.
- [Scheper-Hughes](#), Nancy. 2000. Ire in Ireland. Ethnography 1(1): 117-140.
- Stark, Laura. 2012. Behind Closed Doors: IRBs and the Making of Ethical Research. Chicago: University of Chicago Press.
- Walford, Geoffrey. 2005. Research ethical guidelines and anonymity. International Journal of Research & Method in Education. 28(1): 83-93.

Winkens, R. A. G., J.A. Knottnerus, A.D.M. Kester, R.P.T.M. Grol, and P. Pop. 1997. Fitting a routine health-care activity into a randomized trial: An experiment possible without informed consent? Journal of Clinical Epidemiology 50(4): 435-439.

World Health Organization, 2001. *Putting Women First: Ethical and Safety Recommendations for Research on Domestic Violence Against Women*, Department of Gender and Women's Health, Family and Community Health, World Health Organization, Geneva.

Zwane, Alix Peterson, Jonathan Zinman, Eric Van Dusen, William Pariente, Clair Null, Edward Miguel, Michael Kremer, Dean S. Karlan, Richard Hornbeck, Xavier Giné, Esther Duflo, Florencia Devoto, Bruno Crepon, and Abhijit Banerjee. 2011. "Being surveyed can change later behavior and related parameter estimates." PNAS 2011 108 (5) 1821-1826.