PAA 2011 Submission – Draft

Outsider, Locals, and Insiders: Results from a Methodological Experiment on Data Collection

Alex Weinreb, University of Texas, Austin

Mariano Sana, Vanderbilt University

Guy Stecklov, Hebrew University

Abridged Abstract

It has been assumed since the early days of survey research that, in order to collect unbiased and valid data, interviewers must have no prior social relationship with respondents. We refer to this assumption as "the stranger-interviewer norm". In this paper we review social science literature related to this norm and describe an NIH-funded methodological experiment that we fielded in the Dominican Republic this past summer (May-July 2010) in order to examine its validity. This is the first-ever evaluation of the stranger-interviewer norm in an experimental context in any setting. Not surprisingly, preliminary results seem to be mixed, depending on the types of questions asked. A much larger array of results than those included in this initial document will be available to be presented at the time of the PAA meetings. We believe that the study's findings will have the potential to contribute to substantial improvements in data collection, especially in less developed countries.

Introduction

Over the last thirty years, survey research in less developed countries (LDCs) has been subjected to a series of "blanket condemnations" (Smith 1989:180) motivated, in large part, by two related factors: (i) validity-related concerns, that is, indications that survey data—and therefore analyses—are undermined by significant mismeasurement on key variables of interest; and (ii) the realization that there is a "dearth of theoretical studies upon which rational choices of methodology [can] be based" (Ross & Vaughan 1986:92). The response to these criticisms, many of which have also been directed at survey research in more developed societies (MDCs), has fallen into two distinct camps. One has focused on the development of statistical methods, applied after data are collected, that attempt to attenuate the effects of bias on analysis (Griliches 1986; Bollen 1989; Biemer & Stokes 1991). An altogether different response has been to demand that more attention be given "to the fundamentals of measurement" (Smith 1989:180), in particular, by identifying "precisely where surveys go wrong" (Axinn 1991:303). The result: survey researchers, particularly those working in non-western settings, have either attempted to improve data quality by incorporating techniques more commonly associated with qualitative research—thus the emergence of the "micro approach" (Caldwell, Reddy & Caldwell 1982), the "ethnosurvey" (Massey 1987), "microdemography" (Axinn, Fricke & Thornton 1991). Else, they have experimented with ways to eliminate interviewers—and therefore interviewer-related error—altogether: thus the emergence of Audio-Computer and Computer Assisted Self Interviewing (respectively, Audio-CASI & CASI) (Mensch et al. 2003).

Our research builds on this second response—that is, improving the quality of data before they land in the analyst's lap. Building on preliminary studies in Mexico and Kenya (Sana & Weinreb 2005; Sana & Weinreb 2008; Weinreb 2006) we evaluate whether one of the key weaknesses of contemporary data collection and data editing practices in LDCs is that they privilege "strangers" over "insiders." That is, although contemporary approaches to survey research in LDCs give a much more significant role to local researchers in the formulation of locally-valid survey instruments than they once did—rather than limiting them to more service-oriented roles as translators into appropriate local idioms—it remains the case that the actual data are collected by interviewers who have no direct and prior knowledge of respondents, their families, or their communities. We suggest that, contrary to this approach, there is intriguing evidence that more valid data could be generated if researchers were to rely more on "insiders" as data producers, at least in LDCs.

Our main aim in this paper is to report results of a methodological experiment that we fielded in the Dominican Republic (May-July, 2010) in order to systematically evaluate the insiderness hypothesis for the first time. Conducted in a single mid-sized town, the study involved drawing a sample of adult women respondents, then randomly assigning each of them to a female interviewer whose level of familiarity or "insiderness" with those respondents was known a priori. Each respondent was then administered a questionnaire on a variety of topics, including core questions used in international sociodemographic research (household structure, economic profile, remittances, contraceptive use,

children's vaccination history, religiosity, attitudes to stigmatized groups). They were also asked to fill-out a short self-administered questionnaire.

The analyses reported in this paper are focused primarily on evaluating response variation on this wide array of questions across levels of interviewer insiderness. We refer not only to comparing means and variances but also, on some questions, validating responses, either indirectly, by looking at the direction of response bias associated with a particular type of interview, or directly, by comparing it to some known value.

Note, too, that our data allow us to explore the mechanisms underlying any response effects. We collect an array of data on the interview setting itself – where the interview was conducted, repeated measure of third-party presence – that allow us to differentiate direct effects of insiderness on response from those that, for example, channel outsiders into different types of interview settings (more or less private).

Likewise, the data collection instruments do not only vary in terms of topic. Within each section they also intentionally vary in terms of level of sensitivity and level of cognitive effort required to provide the best possible answer. Each of these sources of variation allows us to address an important mechanism through which response effects occur, which in turn allows us evaluate the extent to which these mechanisms vary across different levels of interviewers' insiderness.

The paper proceeds in four sections. The first describes the theoretical underpinning of our study at greater length. We summarize relevant debates in related social sciences, and build on a large body of relevant social theory. Insiderness emerges from this section as a conceptually important, though empirically unexploited factor. In the second section we describe our data collection procedures – at some length – since the aims of the study require a complicated sampling strategy that integrates selection of local interviewers and respondents. The third section present preliminary results of our study – preliminary because the data were only collected recently (May-July, 2010), and we have only recently begun to analyze them. In the fourth and final section – not in this extended abstract at all – we look at the implications of these differential response patterns on analysis of relations among variables.

Section 1. Background

The literature on data collection techniques and response effects in MDCs is enormous, a highly specialized subfield that cuts across social science disciplines and a longstanding focus of research. A trickle of early papers (e.g., Rice 1929; Deming 1944) has given way to scholarly monographs which have drawn mainly on a series of data collection experiments (e.g., Hyman et al 1954; Sudman & Bradburn 1974; Dijkstra & van der Zouwen 1982; Tourangeau, Rips & Rasinski 2000).

The literature on response effects and data collection techniques in LDCs is very different. The earliest manuals, monographs and descriptions of fieldwork tended to draw much more heavily on anecdote and self-perpetuating lore than on formal experimental comparisons of different methods and modes (e.g., Back & Stycos 1959; Choldin et al 1967; Heisel 1968; Caldwell et al 1970; Bulmer & Warwick 1983). Later studies explored a limited set of mismeasurement problems posthoc (e.g., Chidambaram et al

1980; O'Muircheartaigh 1982; Thompson, Ali & Casterline 1982; Stone & Campbell 1984; Axinn 1991; Lozano Ascencio 1993, Dare & Cleland 1994; Massey & Parrado 1994; Becker et al 1995; Ghuman et al 2005; Bignami-van Assche et al 2003; Weinreb 2006; Weinreb & Sana 2008). Only recently, largely driven by AIDS-related research priorities, have researchers begun to employ experimental designs more typical of methodological evaluations in MDCs (e.g., Mensch et al 2003; Plummer et al 2004a; 2004b).

Although the literature on data collection in LDCs is highly variable—in both quality and approach—there is a single dominant theme throughout. Significant unexplained error and uncertainty affects reported behavior and attitudes across all substantive areas, including: contraceptive use and knowledge, sexual behavior, women's autonomy, measures of income and labor force participation, and migration-related behavior. Measurement error, in turn, threatens analysis in one of two ways: by (i) biasing univariate statistics; and (ii) undermining analysts' ability to correctly identify existing relationships both when mismeasurement is with the outcome variable as well as when it lies with the explanatory variable(s).

Responses to mismeasurement

The awareness of the degree of mismeasurement has generated strong skepticism about survey data from LDCs. One response has been the increased emphasis on, and attention to, methods such as instrumental variables procedures (Wansbeek and Meijer 2000; Greene 2002), latent variable models (Bollen 1989; Biemer and Stokes 1991), or multilevel models (Hox et al. 1991; O'Muircheartaigh and Campanelli 1999, Weinreb and Sana 2008) that aim to either "cleanse" and/or test the problem, thereby avoiding biased results. Yet, such "fixes" often bring costs, such as the loss of efficiency associated with the use of instrumental variables.

More pertinent to our approach have been the attempts to improve measurement within surveys. At least two of these efforts have incorporated techniques more commonly associated with qualitative research, yielding (i) a shift from fully structured and standardized questionnaires towards more conversational styles of interviewing (e.g., the "ethnosurvey"); and (ii) researcher's immersion in the field setting and flexibility in the design of survey tools (e.g., "microdemography"). A third response has involved eliminating interviewers—the middlemen who cause a considerable proportion of error—entirely. Thus the recent experiments with Audio-CASI (Mensch et al 2003), particularly on sensitive questions.

There are signs that each of these shifts in "interview modes" can generate more valid responses on sensitive questions (Plummer et al 2004a, 2004b), though this is not always the case (Mensch et al 2003). However, insofar as none of these new methods change the social relationship between the researcher and the respondent in any substantive way, there are also strong grounds for thinking that the returns to all such shifts will be limited. In particular, no shift in interview mode changes the fact that the researcher—including his/her agent, the interviewer—remains a "stranger" to the respondent. By this we mean that these two actors have no social relationship that predates specific interactional roles associated with their interview-focused interaction. More specific, a stranger-researcher has no

"entanglement in family and party interests" (Simmel 1908/1950:404), and no norms of reciprocity bind him/her to the respondent. Consequently, the respondent has no social leverage over the interviewer.

It is important to clarify our usage of the term "stranger" more explicitly. Although Simmel's foundational definition of "stranger" is inherently relational in that it addresses the existence or absence of a relationship between two individuals, underlying this relationality is the hint of a secondary dimension: interactants' physical location. Specifically, strangers are likely to be "outsiders"—by which we refer to someone who is not from a given area. Because of this, the literature which we review below sometimes confuses the term stranger with outsider. For both theoretical and empirical reasons we think it important not to but, rather, to conceive of them as independent, albeit frequently cross-cutting, dimensions. In other words, a "stranger" and an "insider" can be either a "local" or an "outsider." We return to this below.

Themes related to researchers, strangers, insiders and outsiders have long been the subject of intensive discussion in the literature on ethnography and other types of qualitative data collection (Sandoz 1930/1988; Nash 1963; Powdermaker 1966; Everhart 1977; Agar 1980; Jansen 1980; Cressey 1983; Mayoni 1983; Marcus 1992; Rogers 1999). That literature depicts researchers entering "as strangers into a small and culturally alien community" (Stocking 1983:7) with the aim of becoming familiar with "a variety of secrets which range from the assured truths of gossip to the collective myths of rumors" (Suttles 1972:36). It also defines the key to achieving that aim as an extended presence, since that is the only way to become embedded in local networks, and therefore known, trusted, and accountable. In other words, it is the key to attaining "the status of an intimate, an insider" (Agar 1980:59), which, from the perspective of data collection, is someone who is given the right to know things that are otherwise hidden by untrusting locals.

Notwithstanding the shifts towards techniques—and to some extent also epistemological positions—associated with qualitative research among survey researchers, the nature of this relationship between the researcher/interviewer and the respondent has, until Weinreb (2006), not been explored methodologically by survey researchers. In other words, a century of social research has yielded no estimates of the size and scope of insider versus stranger effects on data, no estimates of the size and scope of local versus outsider effects on data—again, we emphasize that the Simmelian stranger can be either a local or an outsider—and no systematic review of relevant theoretical or methodological issues concerning the collection of survey data by strangers or by outsiders in general. Even the large body of research on the effects on survey response of different types of "social distance" between interviewers and respondents includes little more than hints of these themes: (i) tangential asides in early methodological reviews, in which the possibility of stranger-effects has been acknowledged but then ignored empirically (e.g., Stycos 1960; Mauldin 1965; Sudman & Bradburn 1974), or (ii) more specific criticisms of survey research as practiced in the developing world (Heisel 1968; Stone & Campbell 1984). Instead, the extant literature has treated social distance as proximity/distance on "role-independent interviewer characteristics" (Sudman & Bradburn 1974)—that is, the extent to which interviewers and respondents share or differ on characteristics like gender, race, class, ethnicity. An interviewer's level of "insiderness," clearly, is completely different. Just as it can cut across the local/outsider spatial

dimension, it can cut across these ascribed social characteristics, potentially interacting with them (see Weinreb 2006 for empirical evidence) but also indexing a very different type of social relationship.

In summary, it is clear that a significant part of the edifice of survey research is built on an unvalidated idea: that using strangers, particularly those who come from some other place – as interviewers almost always do – is a superior data collection approach. We henceforth refer to this as the "stranger-interviewer norm."

Insiders and strangers in data collection

Past debates about strangers and data quality in the social science literature have universally focused on validity, not reliability. The underlying contention is that, as a general rule, the stranger-interviewer norm generates more valid data. We now summarize and critique this literature since it necessarily builds on one of two problematic ideas about the nature of (micro-level) social interaction: either people are "more likely to be reticent with those who are already known to them" (Bulmer 1993:215), or else an interviewer-respondent interaction is a distinct type of social interaction in which this rule operates, albeit with the assistance of role-specific interviewing techniques such as minimizing interactional biases and using confidentiality assurances.

i. Truthtelling and intimacy: The first of these ideas—that people are more open with strangers as a general rule—is inconsistent with the bulk of social theory, insofar as dominant paradigms about micro-level social action equate truthtelling with intimacy. The key assertion has remained largely unchanged since Tönnies (1961:109) argued that "one is his most 'informal' among brothers or among comrades; one 'lets himself go'," creating a social milieu which fosters truthtelling. Examples from other areas of social theory include: Goffman's dramaturgical theories, in which the most intimate and honest social contacts occur "backstage" with people we trust and with whom we have "anchored" relations (Goffman 1959, 1971:189); ethnomethodological approaches, which show that a larger range of mutually comprehensible indices are available to intimates than to strangers, meaning that speedy and accurate communication can rely on nonverbal signaling or on more concise verbal structures (Garfinkel 1967); neo-Durkheimian ritual interaction chains, in which the shared emotional focus from which interactional pleasures and social significance derive heighten in intimate interactions, enhancing motivation for open and trusting communication (Collins 1981; 2004); affect theory, in which an actual or prospective reciprocal exchange relationship fortifies a respondent's motivation to communicate a truth as a sign of their investment in the relationship (Lawler 2001); and the literature on trust conventions, in which those oriented towards reciprocal exchanges—more longstanding—are more reliable than those based on negotiated exchanges (Molm et al. 2000).

The idea that people are more open with strangers as a general rule is also inconsistent with the experimental and semiotic literature on micro-social interaction. For example, Self-Disclosure Theory suggests that individuals "tend to disclose more honestly and in greater detail to people with whom they feel emotionally comfortable" (Catania et al 1996:348; see also Dindia et al. 1997). Linguistic analysis shows that people can accomplish a given task more effectively with intimates because fewer interactional acknowledgments are necessary, and the relationship also allows for more interactional

advisements and interpretation than in an equivalent task-oriented relationship with strangers (Breault et al. 1987). Analysis also shows how rhythmic synchronization between interactants—in brain waves, gestures, and an array of vocal characteristics—is heightened in communication among intimates (reviewed in Collins 2004: 75-78). On the flipside, the relative discomfort in interactions with strangers has been shown through observation of facial expressions and automanipulations (Givens 1978) as well as, perhaps more objectively, through analyses of cardiovascular reactivity under varied experimental conditions (Christenfeld et al. 1997).

The literature on deception also tends to equate truthtelling and intimacy. It casts deception as a communicative ritual intended to structure micro-communities by defining social and moral boundaries (Simmel 1908/1950:315; Barnes 1994). This necessarily means differentiating insiders from strangers, as implied above in Suttles' (1972) reference to "secrets." In turn, this often entails making careful moral distinctions between lies told to strangers and lies told to insiders. The latter tend to be more morally problematic, and this distinction can be enshrined in lexicon, with lies to strangers, and sometimes outsiders in general, assigned a different name than lies told within one's immediate group (Barnes 1994).

ii. The effect of local structural arrangements on relationships with strangers: The assertion that people are more open with strangers as a general rule also implicitly disregards one of the basic axioms of social theory, the idea that group-level variation in micro-level social relations, which include interviewer-respondent relations, is affected by broad structural factors. Durkheim (1926/1964) and Douglas (1970), in two versions of a classical theoretical paradigm—other variants can be found in Tönnies (1961), Linton (1936), Levy (1972), and Giddens (1990)—differentiate between social groups along two structural continua: the levels of social density (i.e. the proximity of one member to another) and social diversity (i.e. internal differentiation). The combination of these two characteristics affects the composition of moral systems and, through these systems, interactional patterns. High density/low diversity groups have a relatively localized moral outlook that places high value on conformity and, crucially for social science research, also incorporates a tendency towards inwardly-oriented trust conventions—that is, a distrust of strangers—based primarily on reciprocal exchanges. The prevalence of such groups today tends to be higher in LDCs whereas the reverse tends to be true for MDCs.

An additional—though somewhat parallel—layer of structural effects is related to local political characteristics. That is, the extent to which trust conventions are oriented towards strangers or not is also a product of socioeconomic or political inclusion or disenfranchisement (e.g., the correspondence between ethnically-based patron-client networks in business and politics—see Horowitz 2000).

In short, the frequency, intensity, and value of interaction with strangers, as well as general attitudes to

strangers, appear to be a direct function of social density and diversity. Indeed, a primary index of modernity through different versions of these theories is the extent to which daily patterns of social interaction are stranger-oriented. These include both the sharing of general social space with strangers as well as the frequency of dealings with anonymous bureaucrats and workers, many of whom, like interviewers, ask questions and record answers for some unknown and possibly mistrusted authority.

iii. Distinctions between interpersonal and task-oriented interview-related norms: The secondary assertion underlying the stranger-interviewer norm is that unlike any other type of social interaction, there is greater truthtelling to strangers in interviewer-respondent interactions. This initially appears more robust. There is some evidence in the methodological literature that the warmth that is characteristic of more intimate interactions does not necessarily improve truthtelling since it may increase the respondent's susceptibility to a conformity or social desirability bias (see reviews in Fowler & Mangione 1990, Tourangeau et al 2000). This is consistent with situationally-oriented micro-social theory. For example, reporting a jarring truth to an insider may breach an unwritten interactional rule and, therefore, bring about an emotionally upsetting interaction ritual.

On the other hand, methodological experiments and fieldworkers' accounts have increasingly pointed to the difficulty in neatly distinguishing interpersonal norms from task-oriented interview-related norms, implying that the boundaries between these two types of interpersonal communication are not clearly marked. Rather, the standardized survey interview, as Schaeffer and Presser (2003:65) have recently concluded, "is a distinct genre of interaction with unique rules, but it shares many features with ordinary interaction because social and conversational norms as well as processes of comprehension, memory, and the like are imported into the interview from the situations in which they were learned and practiced." An emerging body of research in MDCs, for example, has applied a range of micro-social approaches to the architecture of survey interviews and survey interactions (e.g., a collection of papers in Maynard, Houtkoop-Steenstra, Schaeffer & van der Zouwen 2002). One of the key aims of this research has been to substitute traditional stimulus-equivalence methods with an awareness of the "situational oscillations between formal rule-following and tacit practices that interviewers enact with survey respondents" (Maynard and Schaeffer 2002:4). This new approach also builds on a more longstanding methodological literature that compares the validity of data collected across more and less standardized interviewing formats. Much of that literature has found that response validity can be improved through the use of various interpersonal techniques such as "contingent feedback" methods (Cannell, Miller and Oksenberg 1981) and more flexible styles of interviewing than the fully impersonal standardized techniques (Cicourel 1974; Dijkstra 1987; Suchman and Jordan 1990; Tourangeau et al 2000).

As yet, no systematic research of this type has been conducted in LDCs. But the effect of higher density, lower diversity characteristics, and in particular the low exposure to researchers and other types of strangers who ask questions, implies that the correspondence between interpersonal and interview-related norms will tend to be higher yet in LDCs. So too do field researchers' accounts. The anthropological literature, in particular, is replete with accounts of culturally sanctioned, and even culturally prescribed, deception of stranger-researchers in poorer countries, some with comical, and others quite serious results (e.g., Dentan 1970:92, Agar 1980:61, Barley 1983:89, Stone & Campbell 1984:32, Anderson 1986:342-3). Bulmer (1993) refers to these as types of "sucker-bias," a phenomenon that is specific to strangers (since having mastered local cultural codes, insiders know how to avoid being suckered). Agar (1980) and Stone and Campbell (1984) are particularly significant since they report matching prior survey reports from given respondents in rural India and Nepal with post-enumeration reports gathered in informal interviews in communities in which they were embedded. In both cases,

respondents reported telling untruths to the survey-interviewers because the latter were strangers whose motives were unknown and therefore suspect.

iv. Confidentiality assurances & response anonymity: One of the principal ways in which researchers/interviewers protect their right to ask questions is by giving confidentiality assurances and various levels of response anonymity, or complete anonymity in self-administered questionnaires (SAQs). There is, however, no extant literature on how effective these assurances and promises are in non-Western settings. For example, all 113 of the studies analyzed in Singer et al's (1995) meta-analysis of the utility of confidentiality assurances were conducted in MDCs. In poorer societies, in contrast, it has been asserted that "[t]he anonymity of survey research is generally incomprehensible" (Kroeger 1983:467) and that, therefore, emphasizing the confidentiality of responses in these contexts "may not be necessary or even desirable" (Hershfield et al 1993:244). An extreme case of the undesirable effects of a confidentiality assurance is reported by Back and Stycos (1959:8) in their discussion of their survey of Jamaican fertility. They describe how in the exploratory stage of the project, interviewers "nearly lost their lives…" after confidentiality assurances sparked deep concerns about witchcraft, forcing the researchers to abandon the confidentiality assurance in the main survey. Similar stories appear in field reports from other early LDC surveys: Bangladeshi interviewers were considered "godless women heralding the end of the world" (Choldin, Kahn & Ara 1967:251); in a Burundian survey, Tutsi interviewers were "shortened"—that is, killed—by their untrusting Hutu respondents (Michael Bracher 2005: personal communication). In some of our own fieldwork in Mexico, Puerto Rico, the Dominican Republic, Kenya and Malawi, fieldworkers reported levels of mistrust that confidentiality assurances did little to reduce.

These cases are directly relevant to insider-stranger debates since they suggest that the credibility of a confidentiality assurance—and by extension, a promise of response anonymity—is only as strong as the respondent's confidence in the promise-giver, in this case the interviewer and/or the organization that she or he represents. Absent that confidence, and no matter how vociferous the promises are, confidentiality assurances may do little to negate people's fear of telling the "wrong people" about an undesirable activity or attitude. It is noteworthy that these concerns have their parallels in the response effects literature in MDCs. Aquilino (1994), for example, demonstrates more pronounced response variation between face-to-face, telephone, and self-administered questionnaires among African-Americans, and connects this empirically to trust in government and related institutions.

v. Insiders, strangers and data reliability: No prior empirical research has been conducted on the relationship between insiders, strangers and reliability. In addition, there is no parallel theoretical literature on this relationship either. Finally, inferring the most likely direction of the relationship between insiderness and reliability also doesn't help too much since an equally plausible case can be made for stranger-interviewers generating both more and less reliable data.. For example, we would expect stranger-interviewers to generate more reliable data than insiders where there is greater homogeneity in attitudes to strangers over insiders. This may well be the case in an LDC setting characterized by high density/low diversity (B.2.2) since two randomly selected stranger-interviews would always be mistrusted, whereas two randomly selected insider-interviews would only sometimes be mistrusted. On the other hand, the "motivational" aspect that underlies respondents' behavior may

flip this reliability issue. That is, greater familiarity with insiders may lead respondents to invest more effort in answering their questions relative to those of strangers, resulting in lower reliability of data collected by strangers. Again, a priori we have no way of judging which of these effects is likely to be stronger.

Summary

Insiders in data collection: In terms of data collection, there appear to be two key problems arguing that stranger-interviewers should be the fieldworkers of choice across groups with different structural and cultural characteristics. First, the basic proposition that a given respondent is more likely to provide more valid data to stranger-interviewers draws on assumptions about micro-social relations which are at best inconsistent with, and at worst, antithetical to a large body of social theory. Second, the successful suspension of normal interactional rules in a research interview cannot automatically be relied upon since the assumption of normative distinctiveness between standard interpersonal norms and task-oriented interview-related norms is weaker than has traditionally been assumed. In short, there is weak a priori support for the stranger-interviewer norm, that is, for the assertion that, as a general rule, people are more open with strangers or stranger-interviewers across all types of settings. Put another way, it may be true that "Surveys often venture into areas that people do not ordinarily discuss with strangers" (Tourangeau, Rips and Rasinski 2000:14), and that strangers have received "the most surprising openness" (Simmel 1908/1950:404) over a century of formal data collection. But this in itself tells us nothing about the validity or reliability of the information they have received. There are good theoretical reasons for supposing that stranger-interviewers working in groups with more socially diverse and less socially dense structural characteristics would tend to collect more valid and more reliable data than insiders. But there are equally good theoretical reasons for supposing that the situation in more socially dense and less socially diverse groups is, in general, the reverse: that in these types of groups, insider-interviewers would tend to collect more valid and/or reliable information than strangers.

Similarly, although it is reasonable to expect some reticence in respondents' willingness to disclose controversial opinions or disreputable facts to insiders—whether stemming from fear of gossip, subordination to local status hierarchies, or conformity with local cultural norms—there is no extant empirical evidence that either local or outsider stranger-interviewers represent a more attractive and trustworthy alternative. There is, of course, some indication that either of these types of stranger-researchers in general can become more trusted over time. But to the best of our knowledge this literature refers solely to ethnographic researchers who gradually embed themselves in local networks (e.g., Dentan 1970, Agar 1980, Barley 1983). It does not refer to survey interviewers who, unknown and uninvited, knock on people's doors and request information, however politely they do so, and with however much emphasis on the confidentiality of the respondent's answers. Finally, moving beyond this general pattern of insider-stranger variation, we might also expect the degree to which insiderness matters to be affected by a question's particular attributes. One of these is surely the sensitivity of the information being sought. For example, stranger-interviewers may do no better or worse where the requested information pertains to other people or to characteristics of the community, but much worse where it is personal information. Similarly, there might also be insider-stranger variation on questions

which do not appear sensitive but require some motivation and effort to answer well, such as those that demand the scanning and sorting of memories. As noted below our research design will allow us to explore insider-stranger effects across different types of questions.

Section 2. Data

Our key aim is to gauge the impact of insiderness on two fundamental aspects of data quality: validity and reliability. In order to do this with confidence our research design has two crucial elements. First, it randomizes insider and stranger interviewers across respondents. Second, it includes a large number of questions across a range of important outcomes whose responses we can validate. Before describing these two elements—and a number of other concerns that arise from them—in greater detail, we justify our country selection.

The Dominican Republic (D.R.) offered an ideal setting for carrying out our research for both substantive and practical reasons. The former are related to the D.R.'s mid-level development characteristics. Specifically, it is precisely its middle-income status—its 9 million inhabitants have a GNI per capita of 7,150 USD, life expectancy at birth of 68 years, and are 64 percent urban—that makes it an attractive location for our research. This is because, as elsewhere in Latin America and the Caribbean, considerable sub-national variation lies beneath these averages, allowing us to identify a single research site in which there is both relatively low and higher levels of social density and social diversity—e.g., relatively wealthy, westernized middle class, and relatively traditional rural poor. In other words, we expect Dominicans' attitudes to a survey interview to be, on average, more like those of their American counterparts than of those of the rural Africans who provided us with our preliminary results.

However, we also expect to find more variation in underlying attitudes to surveys and strangers that we can leverage analytically – something less obviously available in our earlier focus on rural African settings. The larger epistemological benefit of this is that if insider/stranger effects emerge in our D.R. settings our case for generalizability will be stronger.

Insiders, strangers, and outsiders

As shown in Table 1, interviewer–respondent matches are classified along the two dimensions which we briefly distinguished above: (i) the degree of familiarity describes the extent to which the interviewer knows the respondent, or someone in the respondent's immediate family, as per the classical Simmelian archetype; and (ii) the degree of locality describes the extent to which the interviewer is from the community and, therefore, can be understood to be “one of us.”

These two dimensions enable us to distinguish four theoretically distinct categories, three of which are incorporated into our research strategy (not surprisingly, the fourth, non-locals who know the respondent or their household, did not exist in our data). Two of the three are prototypical categories and representative of the theoretical distinctions described above: (i) "insiders" are defined as both “local” and “familiar”; and (ii) “outsider-strangers” are both “non-local” and “unfamiliar” with the

respondent and the community, and therefore represent the classical type of survey interviewer. Finally, the third category, "local-strangers", captures an in-between status, those that are from the community but have no personal relationship with the respondent or their family. As noted earlier, this third category is less embedded in classical theory than the first two. That said, it is an increasingly visible type of interviewer selection strategy in LDC settings, in particular in projects with a participatory research approach. Moreover, its inclusion helps us to identify which of the axes of difference matters more—location or familiarity.

Table 1. Type of interviewer, by familiarity and residential proximity to the respondent

		Familiar with the respondent or respondent's family	
		Yes	*No*
Lives locally	*Yes*	Insider	Local-stranger
	No	--	Outsider

Sample population

Our respondent sample is composed of women aged 20-50 in a single administrative zone (mostly a town with some surrounding rural areas) with mid-ranking development characteristics (for the D.R.). Here we refer to the town as San Fernando. It is a town with a population of between 20-30,000 people, and was selected after a fact-finding trip to more than 10 towns of equivalent size in different areas of the country. We concluded that it appeared to provide the most suitable setting for the study since it was relatively developed, with a slightly higher than average educational attainment, but it was also relatively isolated.

Given the project's aims, our sampling strategy begins with the selection of interviewers. We advertized open interviewer positions in the San Fernando area and, after a test and informal interview with each candidate, selected 30 trainees. This was in addition to the 8 outsider interviewers (professionals) that we had brought with us from Santo Domingo, the country's capital city. All these trainees – both the locals and outsiders – were then trained in the same room for a period of 5 days. Training included not only the content of the questionnaire, but also technical aspects related to how to conduct an interview. After a final test at the end of training, 24 local and 6 outsider trainees were selected as project interviewers.

Sampling could now begin. We asked each of the local interviewers to free-list the names and addresses of all women (in the sample age group) who they knew well or very well. Over a period of 3 days, we then took each interviewer around town in order to find and mark each of these households on detailed census maps that we had acquired from the DR Central Bureau of Statistics (1:900 scale). These maps list every building in the town and immediate surrounds, with special markings for number of stories, mixed-use business and residential, and buildings under construction. The maps were 3-years old and, we found, highly accurate in general. Where we found missing houses, whether claimed or not, we hand-drew them onto the hard copy of the maps, following the original map-drawing protocol.

These marked households represent our "claimed" stratum. Each such household contains someone that at least one of our local interviewers knows well or very well. Subsequently, the named respondents in these claimed household were randomly assigned to each of the three types of interviewers – the insiders, local strangers, and outsiders – albeit with a slight weighting towards insiders. Where the same household had been claimed by more than one local interviewer, it was randomly assigned to one of those claimants.

This randomization from within an identical subsample allows us to make direct comparisons between mean responses to different types of interviewers within the claimed stratum. In particular, it allows us to attribute any observed differences in mean response value across the three types of interviewer-respondent match to the difference in interviewers' insiderness.

The second stratum is quite different. Since it seemed likely, *a priori*, that sampled women in the claimed households were not a representative subsample of women on the whole, we also specified an "unclaimed" stratum. Specifically, a code representing every unmarked household on the map – that is, every household that had not been claimed by a local interviewer – was entered into a spreadsheet. A simple random sample of these unclaimed households was drawn and then randomly assigned to one of our 30 interviewers. Our analyses of data from this unclaimed stratum focus on a comparison between locals – none of whom know the respondents – and outsiders. This is a more realistic representation of the range of current data collection practices in LDC surveys. It also allows us to more readily generalize from our claimed sample to the town's population of adult women as a whole. But it inevitably cannot address the core issue of interviewer-respondent familiarity that our claimed stratum addresses.

The number of completed interviews by these two strata – claimed and unclaimed – and by type of interviewer within each strata, is presented in Table 2. Also presented are the overall response rates and refusal rates associated with each type and strata. Overall, 1,204 interviews were conducted, 509 within the claimed sample, and 695 within the unclaimed. Within the claimed strata we see signs of slight weighting towards insider interviewers. 224 of the claimed women were assigned to an insider interviewer. Of these, 192 were successfully interviewed, representing an 85.7% response rate. The response rate associated with other locals is similar (86.1), and that associated with outsiders a few percentage points lower (82.1%). [Note that each household was visited up to 4 times, each visit on a separate day, but we haven't yet numerically coded the "reason not completed." Based on anecdotal reports, however, the main reason for non-completion appears to have been the lack of an eligible respondent (i.e., a women in the requisite age range). Only once we have coded these data can we will be able to calculate refusal rates by interviewer type.]

Table 2. Number of sampled households and completed interviews, and overall response and refusal rates, by strata and type of interviewer

	Sampled	Complete Interviews	Overall response rate %	Refusal rate %
Claimed strata				
Insiders	224	192	85.7	
Other locals	208	179	86.1	

Outsiders	168	138	82.1
Subtotal	600	509	84.8
Unclaimed strata			
Locals	754	479	63.5
Outsiders	362	216	59.7
Subtotal	1116	695	62.3
TOTAL	1716	1204	70.2

Notes: One of the main reasons for lack of completion was the lack of an eligible woman in the household. We are currently coding the reason for non-completion.

Field procedures

Each supervisor – who we'd trained separately in Santo Domingo prior to moving to San Fernando – was responsible for a particular sector of the town. Interviewers moved between sectors, and local interviewers alternated between insider and outsider roles – depending on whether they were interviewing one of their own claimed households, a household claimed by someone else, or a household in the unclaimed stratum. In other words, we ensured that local interviewers didn't conduct all their insider-interviews first, and only then move on to other households – or vice-versa – since this would potentially confound an insider effect with an interviewer's changing skill levels.

Instrument

Our questionnaire was organized into distinct sections, many of which included questions frequently used in international sociodemographic research (e.g., from DHS instruments). For example, it included questions on household structure (a standard household roster), household economic profile (ownership of durables and property), sources of income, remittances received and sent, fertility and contraceptive use, children's vaccination history, religious affiliation and religiosity, and attitudes to stigmatized groups. Respondents were also asked to fill-out a short self-administered questionnaire with highly sensitive questions. Finally, a number of other pieces of data were collected on the interview itself: where in the house it was conducted; if there were any interruptions, and if so, for how long; repeated measures of other people being within earshot of the interview (and if so, their rough age group); and the number of folds respondents made in the SAQ before putting it in the envelope.

Each such question and section was designed to allow us to do a few things. At the simplest level we simply wanted to examine the reliability of frequently used questions in the social sciences across different types of interviewer-respondent matches.

A second aim was to examine the validity of different types of data. In most cases, we can only validate the data indirectly, that is, by inferring from the likely direction of the response bias which type of interviewer generates the truer responses. In some cases, however, we are able to validate responses directly, having collected other data to which we can directly compare the respondents'. This is true, for

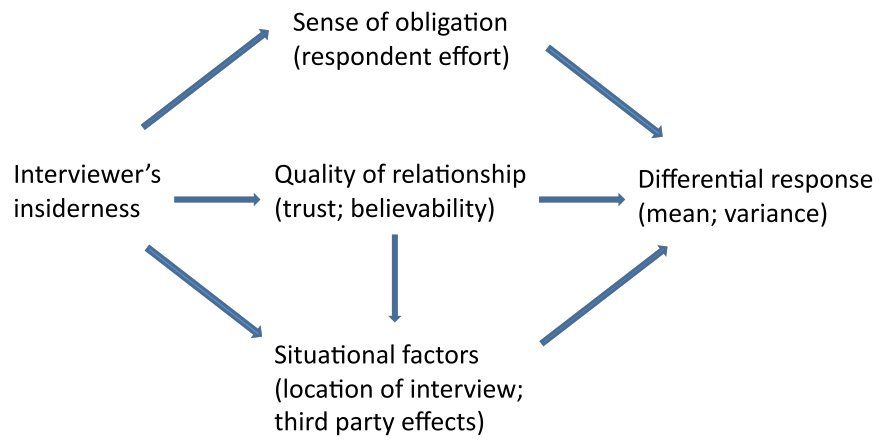
example, with reported child vaccination data, data that appears on the respondent's ID card (*cedula*) , reported distances from the respondent's house to various locations in town, and certain types of reported public assistance (which have a set value).

A third aim of the instrument was to allow us to differentiate three mechanisms through which any observed insider-stranger-outsider effects on data occur. These mechanisms are described in Figure 1. We begin with the central pathway linking a given interviewer-respondent match to differential responses through "quality of relationship." This is the main focus of the theoretical literature reviewed above. It deals with how levels of insiderness or familiarity affect trust and believability. We have argued that, under certain conditions, greater familiarity is thought to be associated with greater trust, which has implications for respondents' willingness to share certain personal pieces of information.

A second mechanism – the upper pathway in Figure 1 – deals with patterns of obligation. Building on ideas associated with social exchange theory – both Mauss' foundational ideas and those formalized by Blau and colleagues – this mechanism is focused on the amount of effort that individuals are willing to invest in answering cognitively challenging questions (e.g., those that require them to recall people's exact ages or exact dates or add together amounts of money from various sources). Not surprisingly, where a respondent knows an interviewer, she will more likely feel under some obligation to provide more exact, better quality data.

The third mechanism is quite different. This is the lower pathway in Figure 1. It describes the extent to which an interviewer's insiderness affects the way that she is treated in the respondent's home. For example, a given level of familiarity may affect where in the house the interview takes place – an insider may invited into the inner sanctums, whereas strangers are restricted to more formal and public settings. This can have an independent effect on response patterns, even if the primary cause remains level of insiderness. A similar argument can be made about third party effects, that is, the extent to which the interview is conducted in private, out of earshot of other household members.

Figure 1. Mechanisms through which a given level of insiderness in an interviewer-respondent match affects differential response.



Section 3. Analysis

Since we only collected the data this past summer (May-July, 2010), and recently received the entered data, here we present only some preliminary results from the claimed stratum, that is, the primary sample through which we can differentiate response effects across insider/local stranger/outsider interviewers. Underlying all our analyses in this section is the assumption that, because of the systematic randomization of interviewers to respondents, all observed differences in response values across types of interviewers within a given stratum are a product of differences in the type of interviewer.

We begin by looking at systematic differences in interview context and then move to response values.

Situational factors

Table 3 presents predicted values of five discrete dimensions of interview setting across the types of interviewers, all derived from questions filled in by the interviewer at the end of the interview. The order in Table 3 loosely follows the chronology of an interview. Thus, we see that whereas 91% of respondents interviewed by insiders were judged to have given their consent easily, the same is true for 88% of other locals, and only 80% of outsiders (the last of these differences being statistically significant at the p=003 level).

Table 3 about here

As expected – described in relation to Figure 1 – insiderness also affects where in the house an interview is conducted, and the differences between the three types of interviewers grows as the house size increases. Specifically, whereas insiders are invited to conduct their interviews in the inner sanctum of homes (living rooms, bedrooms and kitchens) in 65% of cases (houses with at least 2 rooms) and 60% of cases (houses with >5 rooms), the same is true of only 51% and 31% of outsider interviewers, respectively (all effects highly significant). Other local interviewers fall between these two extremes, albeit closer to the insiders.

This difference in location and quality of interaction is also reflected in the minutes lost to interruption. Here, too, we see large and statistically significant differences across types of interviewers. It is close to 5 minutes in the case of insiders, under 4 for local strangers, and 2 minutes for outsiders. We have no information on the causes of the interruption – they could equally be family members entering and greeting their acquaintance, or the respondent leaving for a few minutes in order to take care of pressing business. But either way they point to a more relaxed interactional environment.

Similar results can also be seen in relation to interviewers' final assessments: the extent to which the respondent's general cooperation appeared to be good throughout the interview; and did the interviewer ever get the sense that they were lying. Here, too, insiders were most likely to report good

cooperation throughout, and the least amount of lying. Outsiders were at the opposite end of the extreme. And local strangers, again, were in the middle.

Overall, then, we see substantial differences in situational factors across levels of interviewer insiderness. It is not clear, however, how these affect actual data patterns. As a first step to addressing that question we describe some baseline patterns of response variation across the three levels of insiderness. We begin with the variables related to respondent effort.

Response effects related to respondent effort

Three main parts of the questionnaire allow us to look at respondent effort. The first is the household roster. Following normal practice, we asked people to list all the people in their household. We then followed up with a battery of questions about each of these, including age. We assumed that both a larger reported household size and more age heaping of household members would serve as an indicator of effort. A second area is where we asked questions about the respondent's official status on a number of parameters and then asked her to bring her official ID card. Not bringing the ID, and claiming to not have it or not be able to find it would be an indicator of lack of effort. A third indicator is less an area of the questionnaire than a particular respondent behavior: the greater propensity to report "don't know" to questions that involve some effort (e.g., without the assistance of a child's vaccination card, trying to remember which vaccinations that child has had).

Table 4 about here

Table 4 presents the results of basic analyses across these questions. They are mixed. Relative household size is smallest amongst outsiders, and the likelihood of age heaping is lowest, but these effects are not statistically different from those associated with either insiders or local strangers. In contrast, the other measures of effort – which all come later in the questionnaire – are more productive. Outsider interviewers are significantly less likely to have been shown a *cedula* by respondents. They are also more likely to have received "don't know" answers to questions about how many doses of the tetanus or polio vaccine has your child received. In fact, these latter effects are enormous. Because we followed DHS protocol and only asked about children aged 1 to 30 months, our sample size dropped to less than 90 for the claimed stratum, substantially weakening our statistical power. Notwithstanding this reduction, whereas 11.5 of women interviewed by insiders claim not to know how many tetanus or polio vaccines their child has received, 31 and 37% respectively of women interviewed by outsiders claimed not to know. These substantial effects are statistically significant at the .09 and .02 levels.

Response effects related to trust

As noted above, response effects related to trust are the main focus of the theoretical literature reviewed above. Indicators of trust appear throughout the questionnaire. They are geared toward questions that demand that respondent share sensitive pieces of information.

Results from a preliminary analysis of these variables are presented in Table 5. We look at variation across a range of indicators, in most cases combining related questions into a single scale in order to both increase the content validity of each indicator, and also to increase available variance on the measure.

Table 5 about here

Models 1 and 2 deal look at interviewer-related variation on measures of wealth. The first measure is a scale of within-house durables, largely electronics and furniture. There are no differences here between insiders and outsiders. Since interviewers could often get a sense of the extent to which respondents were lying on these questions – they may have been sitting within sight of many of the items that they were asking about – we are not surprised by this lack of variation (though it is somewhat curious that local strangers emerge as the least wealthy (borderline significant).

Model 2 is more appealing. It looks at variation on other types of property that anyone in the household owns and that, overall, would be much harder for the interviewer to see – a cell-phone, bicycle, motorbike, car, or van. This makes it easier for respondents to claim whatever they want. And indeed, here we see significantly higher reports of wealth to insiders than to outsiders ($p<.001$), with locals returning to their intermediate position between the two extremes.

Model 3 looks at an additive scale of popular religious activities that people were asked to report participation in. These are a common source of social desirability bias. Our results suggest that there is lower reporting of such activities to outsiders than to insiders ($p=.010$).

Model 4 looks at a combined scale built using questions on attitudes to homosexuals, Haitians and prostitutes. In all cases, higher numbers reflect more stigmatizing attitudes. Here, too, outsider interviewers receive marginally higher reports (significant at the $p=.010$ level).

Finally, model 5 looks at a simple dichotomous outcome: does a respondent claim to be using a method of family planning. Here we find an extremely large and statistically powerful effect. Whereas only 44% of women interviewed by an insider claim to be using a method of contraception, 73% of those interviewed by outsiders claim to be doing so. Locals, with 48%, are indistinguishable from insiders.

Planned analyses:

- More systematic/across a larger range of variables
- More explicit discussion of apparent direction of bias

Next section: Summary of differential variance across levels of insiderness

Final section: Discussing implications of the research for using data to model behavior in LDC or mid-tier settings.

Table 3. Predicted differences in interview setting and interviewers' evaluation of respondents' general cooperation, by level of interviewer's insiderness

Model #		*Insiders*	*Locals*	*p[z]*	*Outsiders*	*p[z]*	*N*
1	**Interviewer assesses that it was easy to get the respondent's consent**	0.910	0.876	0.309	0.800	0.003	500
2	**Interview was conducted in the living room, kitchen or bedroom**						
	>2 rooms	0.648	0.638	0.839	0.508	0.011	494
	>3 rooms	0.644	0.648	0.942	0.501	0.012	463
	>4 rooms	0.613	0.596	0.784	0.433	0.006	357
	>5 rooms	0.596	0.517	0.362	0.311	0.003	181
3	**Amount of minutes lost to interruptions**	4.45	3.48	0.088	2.05	0.001	402
4	**Interviewer assesses that respondent's cooperation was good throughout the interview**	0.872	0.788	0.044	0.629	0.001	499
5	**Interviewer thinks that the respondent did not lie once**	0.272	0.152	0.008	0.099	0.001	486

Notes: Model 3 was estimated in OLS; all others are probits.

Table 4. Differences in respondents' effort, by level of interviewer's insiderness

Model #		*Insiders*	*Locals*	*p[t]*	*Outsiders*	*p[t]*	*N*
1	**Household size**	4.240	4.252	0.934	4.031	0.188	509
2	**Mean number of HH members with age ending on a 0 or 5**	0.208	0.217	0.682	0.224	0.487	509
3	***Cedula* (National ID card) not shown to the respondent**	0.79	0.80	0.769	0.66	0.008	430
4	**Responds "Don't know" to question: how many doses of tetanus vaccine has your child received?**	0.115	0.193	0.469	0.307	0.090	83
5	**Responds "Don't know" to question: how many doses of antipolio vaccine has your child received?**	0.115	0.148	0.767	0.370	0.023	80

Notes: Models 1 and 2 were estimated as OLS, models 3 - 5 as probits.

Table 5. Reporting sensitive information, by level of interviewer's insiderness

Model #		*Insiders*	*Locals*	*p[t]*	*Outsiders*	*p[t]*	*N*
1	**Ownership of household durables (0-13 scale)**	7.760	7.237	0.065	7.960	0.511	497
2	**Ownership of other property (0-1)**	0.479	0.459	0.324	0.406	0.001	509
3	**Involvement in religious activities (0-5)**	1.077	1.135	0.689	0.674	0.010	498
4	**Stigma scale - less socially desirable attitudes (3 vars)**	22.30	23.89	0.264	26.25	0.010	506
5	**Currently using a method of contraception**	0.44	0.48	0.452	0.73	0.001	441

Notes: Model 5 was estimated as a probit; all others as OLS.