Published in final edited form as:

Popul Stud (Camb). 2016 ; 70(1): 73–92. doi:10.1080/00324728.2016.1139740.

A test of the stranger-interviewer norm in the Dominican Republic

Mariano Sana,

Department of Sociology, Vanderbilt University

Guy Stecklov, and

Department of Sociology and Anthropology, Hebrew University of Jerusalem

Alexander A. Weinreb

Department of Sociology and Population Research Center, University of Texas at Austin

Abstract

We offer the first empirical test of the “stranger-interviewer norm,” which states that in social, demographic, and health surveys, interviewers should be strangers—not personally familiar with respondents. We use data from an experimental survey fielded in the Dominican Republic, which featured three types of interviewers: from out of town (outsiders), local but unknown to the respondent (local-stranger), and local with a prior relationship with the respondent (insiders). We were able to validate answers to up to 18 questions per respondent, mainly by checking official documents in their possession. Contrary to expectations derived from the stranger-interviewer norm, respondents were more reluctant to show the documents needed for validation when the interviewer was an outsider. Furthermore, and again at odds with the stranger-interviewer norm, we found no difference in accuracy by type of interviewer. Our results have considerable implications for survey data collection approaches in less developed and non-Western settings.

Keywords

surveys; interviewers; interviewer effects; interpersonal relations; data collection; data quality; field workers; developing countries

Do survey interviewers really have to be strangers?

The collection of survey data to produce socioeconomic, demographic, and health statistics in less developed countries (LDCs) has become an enormous enterprise over the past several decades. While the statistical principles that underlie survey research are universal, survey implementation is often constrained by context-specific factors, in particular those related to culture, language, politics, and field costs.

Recruitment of interviewers, a central aspect of survey implementation, is affected by these factors. In almost all field implementations of the Demographic and Health Surveys (DHS), for example, interviewers and respondents are matched on gender, and in most there is an attempt to match them on mother tongue. One key aspect concerning the choice of interviewers, however, has long been considered independent of context- or respondent-

specific considerations. Elsewhere, we have labeled it the “stranger-interviewer norm” (Weinreb et al. 2011)—a reference to the data collection strategy that researchers employ interviewers who have no prior relationship with respondents.

The stranger-interviewer norm is so deeply ingrained in the survey ethos that it appears beyond question. However, it is not a statistical principle but a methodological mandate based on little more than precedent. The central goal of this paper is to test its implied claim that stranger-interviewers collect more valid data than the alternative—interviewers with some degree of prior familiarity with the respondents. We report results from a methodological experiment explicitly designed to test the stranger-interviewer norm in a town in the Dominican Republic countryside. We make no claims with regards to the suitability of the stranger-interviewer norm in more developed countries. Instead, based on our experimental results, we argue that an uncritical acceptance of this norm in less developed settings may be unwarranted.

The stranger-interviewer norm

The longstanding dominant view on the relationship between interviewer and respondent is that no relationship should predate the interview, even if interviewers and respondents may be matched along certain demographic, or “role-independent,” characteristics such as gender or ethnicity (Sudman and Bradburn 1974; Schaeffer et al 2010). This no-relationship paradigm is driven by assumptions about response behavior, especially the idea that a pre-existing relationship will increase risks of response bias. Consequently, the standard survey interviewing practice is to match respondents with interviewers they do not know. This practice is so entrenched in the survey ethos that methodological texts on data collection and survey interviewing manuals rarely bother to make the point explicit. When they do, they leave little room for doubt: “Interviewers must approach strangers” is the explicit instruction given in the Demographic and Health Surveys field-training manual (Macro International Inc. 2007, p. 6).

Some aspects of the interviewer-respondent relationship, including matching and mismatching on gender, race, or other attributes, have a long history of empirical scrutiny (Hyman et al 1954; Sudman and Bradburn 1974; Fowler and Mangione 1990). The notion that interviewers who are strangers collect better (i.e. unbiased and more accurate) data than interviewers who are not strangers is not among these regularly examined topics. It is a commonly held idea, but it has never been empirically tested. Consider, for example, “the stranger on the train” phenomenon. It describes how “people are highly disclosive and honest about their feelings, attitudes, and behaviors” (Hancock 2008, p. 159) when facing a stranger unlikely to be seen again. Thibaut and Kelley (1959) provide a theoretical rationale in support of this claim: people are guarded toward strangers because they feel constrained by uncertainty about the future of the dyadic relationship with the stranger. With a passing stranger that constraint disappears as future interaction is unlikely.

This claim that passing strangers collect better data has an illustrious origin in classical sociology (Simmel [1908]1950). Early research on sensitive topics, such as the Kinsey study (Kinsey et al. 1948, 1953) and its small-scale replication in Britain (England 1949), persuaded survey researchers that the claim was true, since people did sometimes disclose

the most intimate information to strangers (Maccoby and Maccoby 1954, p. 463). Yet, all early survey research did, and ongoing survey research continues to do, is demonstrate that it is *feasible* to collect personal information from respondents using stranger-interviewers. Whether such disclosures are more honest than those given to non-strangers is not really known. That comparison has never been made in any robust or systematic way.

Contrasting methodological principles associated with other forms of data collection suggest that there is a case for evaluating the stranger-interviewer norm. Ethnography, in particular, completely inverts the norm's underlying assumption. An ethnographers' explicit goal is to develop strong interpersonal relationships with informants since this is seen as the most effective way to collect high quality data (Agar 1980, p. 59; Stocking 1983, p. 7). Thus, Agar, for example, recommends leaving more intimate topics for later stages of fieldwork, once a social relationship is established.

Lying somewhere between the standard stranger-interviewer approach in survey research and the insider-approach in ethnography is community-based participatory research (CBPR). In CBPR, interviewer recruitment draws from a pool of people who have preexisting relationships to the community, although they may not know specific respondents or their families (Parrado et al. 2005; Axinn and Pearce 2006; Marcelli et al 2009a, 2009b; Sana and Conway 2012). The recruitment of community-based interviewers in CBPR is premised on the idea that reducing social distance between interviewer and respondent improves the validity of survey data by making the respondent feel that the interviewer is less of a stranger. While not directly known, the interviewer is at least local and potentially *knowable* through some network, a claim that cannot be equally applied to outsiders.

More generally, both social theory and empirical work on social interaction suggest that truth-telling, intimacy and trust are more likely to happen between parties who know each other, while deception is more likely to be directed at strangers (Garfinkel 1967; Goffman 1971; Barnes 1994), including deception of researchers, a type of "sucker bias" (Agar 1980, p. 61; Barley 1983, p. 89; Stone and Campbell 1984; Bulmer 1993; Weinreb 2006). Although researchers have documented considerable overlap between general interaction and survey interaction (e.g., Maynard and Schaeffer 2002; Schaeffer and Presser 2003), a survey interview may be more immune to these dynamics when it is perceived as something different from an ordinary conversation. This is arguably the case in developed societies, where a survey interview is akin to other types of instrumental, interrogative, and highly regulated interactions involving strangers such as lawyers, doctors, therapists, teachers, or the police. In less developed settings, in the absence of a strong legal system or tradition that can protect their privacy, promises of confidentiality are less meaningful and reassuring to respondents or, alternatively, they are only as credible as the messengers who deliver them (Agar 1980; Stone and Campbell 1984; Weinreb 2006). It follows that the stranger-interviewer norm should be especially subject to empirical scrutiny in less developed societies.

The stranger-interviewer norm guides fieldwork practices in less developed countries as much as it does in more developed ones, though interviewer recruitment strategies vary considerably along the local/outsider continuum. Some of the largest and most established

endeavors, such as the Demographic and Health Surveys (DHS) or Living Standards Measurement Studies (LSMS), tend to employ experienced fieldworkers associated with national statistical offices and native to the countries, or regions within countries, where the surveys are implemented. On the other hand, data collection within the INDEPTH network and its associated health and demographic surveillance system (HDSS) field sites in Africa, Asia and Oceania, relies on local interviewers. So do researchers fielding smaller-scale studies in spatially concentrated settings (e.g., the Malawi/Kenya Diffusion and Ideational Change Project, the Tsogolo la Thanzi study, or the Likoma Network Study). This substantially reduces field costs and gives survey organizers direct access to local informants.

Hiring local interviewers in this way opens the door to the often neglected possibility, especially when the survey setting is more spatially concentrated and networks are dense, that interviewers find themselves collecting data from people they know, or whose families they know. How frequently this occurs in general is unknown, since records of interviewer-respondent familiarity are not normally kept. An exception comes from a survey in Kenya where the field manager recorded relationships between interviewers and respondents, and found that 23 percent of interviews involved interviewers and respondents who knew each other prior to the interview. Subsequent analysis suggested that both response rates and data quality were higher where respondents were interviewed by insiders than strangers (Weinreb 2006).

A contrasting perspective on local (and insider) data collection was found in a population-based HIV surveillance program in South Africa (Reynolds et al. 2013). Some subjects reported hesitation to having a blood sample drawn when requested by a fieldworker they knew personally, especially if the fieldworker was younger than the respondent. However, while most respondents reportedly wished for fieldworkers to be strangers from other localities, there were those who favored locals. This conflict raises two important issues to be considered below: the sensitivity of the questions asked and the relative social status of the interviewer.

Research question

The competing DHS, CBPR and HDSS approaches point to the epistemological tension that we evaluate empirically in this paper. If survey researchers think that respondents will provide more valid data where they feel that the interviewer is not a stranger but a member of their own community, and hence at least potentially *knowable*, it is worth questioning what level of interpersonal familiarity between interviewer and respondent is the most desirable in terms of data quality.

Prior to the research presented here (and in related recent pieces—Weinreb et al. 2011; Rodriguez et al. 2015; Stecklov et al. forthcoming), the effects of interviewer-respondent familiarity on data quality have only been examined statistically using non-experimental data from Kenya (Weinreb 2006). In contrast, we use survey data gathered in a carefully designed experiment fielded in the Dominican Republic. Our key research question is: do outsider interviewers collect more accurate survey data than either interviewers who are familiar with the respondent or other locals who are not directly familiar with the

respondent? Thus, our study focuses on an overlooked factor to contribute to the vast literature on interviewer effects, which has largely emphasized race, ethnicity, and gender (Davis et al. 2010), or considered all interviewer attributes at once as indexed by interviewer identification variables (Barnighausen et al. 2011). To answer our question, we document the effect on response validity when we randomize interviewer assignments to respondents based on their pre-interview level of familiarity. We limit our analysis to a relatively small number of survey responses that we can directly validate. These pertain to questions primarily addressing administrative data frequently used in demographic, health, and development-related research. While our focus here is on response validity, elsewhere we have examined the effects of interviewer-respondent familiarity on responses to questions that we could not validate, including item non-response rates and response patterns, social desirability bias and other identifiable differences by type of interviewer (Weinreb et al. 2011), response differences to sensitive questions in a self-administered portion of our survey instrument (Rodriguez et al. 2015), and response variation on questions about fertility and contraceptive use (Stecklov et al. forthcoming).

Setting and experimental design

We fielded our experiment in a small urban setting in the Dominican Republic (DR), which offered two principal advantages. First, the rapid pace of urbanization worldwide implies that data collection efforts in less developed countries are progressively shifting from rural to small urban areas such as our survey setting. Second, the DR is a prime example of a country in transition, a bridge between the developed and the developing world. Its Human Development Index of 0.702 in 2012 ranked it 96th out of 186 countries, very near the top within the group of countries labeled as “Medium Human Development” (United Nations 2013). In comparison to the rural setting in Kenya that provided the data examined by Weinreb (2006), our Dominican setting augments the case for generalizability to many other mid-ranked development settings.

We collected our data in a town that we refer to as San Benito, as part of the NIH-funded Interviewer-Respondent Familiarity Project (IRFP). San Benito is a mid-size municipal seat, about 4 hours away by car from the capital city of Santo Domingo, with social and economic indicators that approach the average for the DR. Our interviewer recruitment and training process was very selective. We trained 32 locals out of a pool of 64 applicants, together with eight experienced interviewers from Santo Domingo. At the end of training, we evaluated all applicants on the interviewing protocols and dismissed ten of them, resulting in a final team of 24 local interviewers and six from Santo Domingo. Some of the local interviewers had experience in collecting data from respondents, although not for social surveys. All thirty interviewers were female. Since all respondents were also female (see below), this helped to avoid adding another layer of effects associated with interviewer-respondent gender interactions. These gender effects are worth exploring with proper randomization of interviewer assignments by gender, but this would have required a much larger sample size and far greater resources that we had at our disposal.

There were multiple differences between local and outsider (i.e. Santo Domingo) interviewers. Locals were younger, with a mean age of 24.04 (ranging from 18 to 37), while

the outsiders' mean age was 33.50 (ranging from 28 to 46). Five of the 24 locals were married and ten were mothers, while five of the six outsiders were both. All interviewers had completed high school, but 15 out of 24 locals listed "student" as their current occupation, while only one outsider did. Given the overwhelming importance of skin color in the stratification system of the DR (see for example Sidanius et al. 2001), we asked each interviewer to report her skin tone, measured on her forearm, using a palette designed by the Project on Ethnicity and Race in Latin America (PERLA) at Princeton University (Flores and Telles 2012). The palette ranges from 1 (lightest) to 11 (darkest). The mean on this skin tone variable was approximately similar: 5.50 for locals and 5.83 for outsiders, both ranging from 4 to 7. In short, in addition to their status as locals or outsiders, on which the two groups are fully distinct, local and outsider interviewers differed in age, marital and parental status, and current occupation, while they were not substantially different in skin tone. Outsiders also had more interviewing experience, though our strict recruitment and training process aimed at producing two groups with comparable interviewing skills. Overall, it is reasonably to wonder whether measured differences between data collected by locals and outsiders stem from their status as local/outside or something else. In the statistical analyses below we control for two of these characteristics (age and skin tone), returning to the larger question in the discussion section.

Eligibility for the survey was set to women age 20-50, Dominican nationals and residents of San Benito. As the first step of the sampling process, the 24 local interviewers were asked to list all the women between the ages of 20 and 50, excluding fellow interviewers, who lived in the town and they knew well or were familiar with, in addition to being able to identify their houses. We refer to this process as "claiming" houses (or respondents). Combining the 24 lists, and after accounting for duplications, produced a total of 635 local women "claimed" for the survey. The field team had detailed census maps of San Benito (1:900 scale), secured from the Dominican census office, which included all properties. The next step was to identify the houses of the claimed women on the San Benito map, a process that took five days during which small teams composed of both researchers and local interviewers canvassed the town marking claimed households on copies of the town map.

The end product of this process was the splitting of the town, for survey purposes, into two strata: the claimed stratum with all households identified by local interviewers and marked on the map as a result of the process outlined above, and the unclaimed stratum with all other properties on the map. A total of 5,469 properties were listed on the map: 635 in the claimed stratum and 4,834 in the unclaimed stratum. Based on power calculations we decided to interview everyone in the claimed stratum and a random sample from the unclaimed stratum. In statistical analyses presented below, we control for stratum or conduct separate analyses for the two strata, although the differences in variable means between respondents in both strata were generally small and nonsignificant.

Our approach enabled us to carry out three types of interviews, defined as follows—*Insider* interviews, where interviewer and respondent knew each other prior to the survey interview, *Local-stranger* interviews, where the interviewer was local but there was no prior familiarity between the two parties, and *Outsider* interviews, conducted by the interviewers from Santo Domingo, who knew no one in town.

It follows that an insider interview can only happen in the claimed stratum, since it results from matching a local interviewer with a respondent she listed during the claiming process. However, not all interviews in the claimed stratum are insider interviews because, within each stratum, we randomized the assignment of interviewers to avoid any possible bias related to the composition of each stratum. In the claimed stratum, we aimed for equal numbers of interviews of the three types. Thus, we first selected a third of these respondents randomly and assigned them the interviewers who had listed them—if someone had been claimed by more than one interviewer, then we selected one of those interviewers randomly. Afterward, we divided the rest of the claimed stratum respondents, randomly and equally, between outsiders and local interviewers (excluding, in each case, the local interviewer/s that had listed the respondent) so as to produce the other two types of interviews. In the unclaimed stratum, the procedure was simpler: we randomized interviewer assignments, without any constraints.

High response rates were achieved in both strata: 87.9% among claimed households and 86.8% among the unclaimed (estimates calculated using standard AAPOR Outcome Rate Calculator). In the claimed stratum, insiders and local-strangers obtained higher response rates (88.0% and 89.5% respectively) than outsiders (85.7%), while in the unclaimed stratum, outsiders obtained higher response rates than local interviewers (88.5% vs. 86.0%). However, none of these differences was statistically significant, as shown by t-tests for the difference between proportions. The final sample size for completed interviews was 1,207.

Survey instrument and validation data

The survey instrument primarily consisted of questions frequently asked in social, demographic and health surveys fielded in LDCs. The topics included household demographic data, characteristics of the residence, wealth and assets, inflows and outflows of money and goods, family planning, detailed information on small children, religious beliefs and activities, attitudinal questions, and a set of sensitive questions collected by means of a separate, self-administered questionnaire.

Two definitions are in order. First, we define *validation* as the ability to confirm with certainty or near certainty whether the respondent provided a truthful or untruthful answer, regardless of the respondent's intentions. Therefore, validation occurs when, in addition to the respondent's answer, we had a trustworthy external source (as described below) that we could use to check the accuracy of the respondent's answer. Second, validation can have two outcomes: *match* or *mismatch*, depending on whether the answer provided by the respondent matches the trustworthy source or not.

We included in the survey instrument 16 items subject to external validation, in addition to two questions described below that were, by design, automatically validated. In order to validate the 16 items, we queried information that we knew could be checked against official documents that respondents were supposed to possess: a national ID card (or *cédula*) that includes information on a person's date and place of birth, in addition to skin color; and vaccination cards (for small children). Questions were posed during the interview that targeted information available in the *cédula* or vaccination card, enabling us to request these official documents later in the interview and validate answers. Interviewers were trained to

justify their request for documentation in terms of the researchers' desire to test respondent recall—a deception, approved by the overseeing Institutional Review Board, necessary to preserve the research from inevitable contamination were the true goal of the study apparent to the local population. A question on respondent's blood type could sometimes, but not always, be validated with the *cédula*—some list it and some do not. When their blood type was not listed on the *cédula*, or the *cédula* was not presented, respondents were asked if they possessed any other kind of documentation that could verify it. Likewise, birth weight could be validated with alternative records when it was not listed on the vaccination card.

Our interviewers were able to examine the *cédula* of 831 of our 1,207 respondents, while open refusal to show the document was recorded in only 45 cases. Various other reasons cited for the unavailability of the *cédula* included it being at another location, lost or misplaced, or the respondent never having had one. Validation of data on small children included only those between 18 months and their fifth birthday. There were 309 respondents who had at least one child in this age range, 36 of whom had two, and none with more than two.

Of the 16 survey items subject to external validation, four described the respondent's own characteristics, six provided information on one of the children of eligible age (if any), and the remaining six provided equivalent information on the second child of eligible age (if any). In addition to these sixteen items, two other items were subject to automatic, rather than external, validation. The survey included a series of questions that asked the respondent whether she had heard of specific “famous” people. While the list did include seven highly recognizable names (Michael Jackson, Fidel Castro, Hillary Clinton, and four well-known Dominican personalities) it also included two Dominican-sounding names of fictitious persons (plausible names but certainly not famous) enumerated in fifth and last place, so as to mix them with the real famous people. The respondent's answer was validated as a match if she had not heard of them and a mismatch if she reported she had. All items subject to validation, together with the validation documentation used, are listed in Table 1. Note that eligibility for validation varies across items.

Analytical strategy

Our goal is to first examine whether interviewer-respondent familiarity—indexed by insider, local-stranger, and outsider interviewers affects our ability to validate responses. Subsequently, conditional on the feasibility of validation, we test whether familiarity affects the accuracy of responses. We conduct two sets of analyses. In the first set the unit of analysis is the respondent. In the second set it is the respondent-item, or items within respondents. For this second analytical approach, the data set was expanded so that, for each respondent, there are as many observations as items made eligible for validation by the respondent's cooperation in providing the necessary documentation. Appendix Table A1 illustrates the two alternative layouts of the data set.

The first set of analyses, with the respondent as the unit of analysis, includes two measures of validation. First, we consider whether a respondent was selected into the validation analysis at all, which is the case whenever we were able to validate at least one item for that respondent. The second one is the percent of validated items per respondent, out of the total

eligible for validation. Our measure for the accuracy of responses is the percent of matches among those items that we were able to validate. After presenting descriptive statistics, we show results from regression models predicting selection into validation, proportion validated, and proportion of matches. The standard errors in these models are adjusted for clustering of respondents within interviewers.

In the second set of analyses, with the expanded data set and respondent-item as the unit of analysis, all our measures are dichotomies. Validation was measured as a dummy variable with value 1 if we were able to validate the item, and accuracy of response was measured as a dummy variable with value 1 if the response was a match. Again, we present descriptive statistics and regression analyses. The latter consist of probit models on the likelihood of validation and the likelihood of matching, with standard errors adjusted for clustering of items within respondents.

When the analytical focus is on matches, we present our analyses on five different subsets of items. These subsets include all 18 items, only the two items corresponding to fictitious people, all items excluding the fictitious people (that is, 16 items), respondent-only items (up to four items), and child-only items (up to 12 items). When the focus is on validation, only the last three subsets apply, because the questions on fictitious people are validated automatically.

Table 2 shows the breakdown of our sample size by stratum and type of interview. The top panel (1,207 cases) is based on the respondent as the unit of analysis while the bottom panel (9,279 cases) is based on the respondent-item as the unit of analysis.

Validation and validated matches

Table 3 presents the percent of respondents selected into validation (that is, those for whom we could validate at least one item), means for the percent of items validated (among those eligible for validation) across respondents, and means for the percent of matches (among validated items) across respondents. The figures are broken down by stratum and by subset of items. As stated above, validation means for all 18 questions and for the two fictitious people questions are not shown because everyone was automatically validated on the latter. We also provide results from t-tests for the comparison of the means for insiders and local-strangers with outsiders, adjusted for clustering of respondents within interviewers.

When it comes to validation, these first results show, contrary to the expectations of the stranger-interviewer norm, a positive effect of personal familiarity between interviewer and respondent. Except for the child-only subset of questions in the unclaimed stratum, the percent of respondents selected into validation and the means of percent validated are always higher for insiders and local-strangers than for outsiders. In the claimed stratum and for the subsets of 16 items and respondent-only items, these differences are always statistically significant. A look at the results on percent of matches shows familiarity to have a significant positive effect when all 18 items are considered, something that appears to result from a strong effect of familiarity on the two fictitious people questions. Once the fictitious people questions are put aside, familiarity has no effect on the percent of matches, except for child-only items in the unclaimed stratum.

Table 4 shows analogous information but with the respondent-item as the unit of analysis. Qualitatively, results are remarkably similar to those where the respondent is the unit of analysis. Since the stranger-interviewer norm predicts that respondents will be more open and honest to strangers, and in particular to the passing stranger (the one never to be seen again, or the outsider) these first descriptive results call the norm into question.

Regression analyses

Considering the respondent as the unit of analysis first, Table 5 presents regression coefficients from a probit model predicting selection of respondents into validation, and OLS models predicting the proportion validated and the proportion matched. The table presents the coefficients for the type of interview dummies, while full model results are shown in tables A2, A3 and A4 in the appendix. All these models control for the following covariates: stratum (claimed or unclaimed), whether the respondent lived in the eastern side of San Benito (which proved empirically relevant), number of children ever born, respondent's age, educational level, two measures of wealth, and two interviewer attributes—age and skin tone. The measures of wealth consist of an additive index of appliances or amenities (refrigerator, stereo equipment, washing machine, air conditioning, computer, cable TV, Internet, and microwave oven) and a dummy variable with value 1 if the household owned both the family residence and an additional property, which could be agricultural land, a house, an empty lot, etc.

The regression results basically reinforce the findings of the descriptive statistics, consistent with the randomization of interviewer assignment embedded in the research design. The probit model shows that familiarity has a positive and significant effect on the selection of respondents into validation when considering either the whole batch of 16 questions under analysis or the respondent-only questions. When considering the OLS model predicting proportion validated and the subset of 16 questions, the effects of familiarity are weaker, as evidenced by only a marginally significant coefficient for local-stranger—yet, they are not negative, as the stranger-interviewer norm would suggest. In contrast, results for the subset of respondent-only questions are strong, with the coefficients for insider and local-stranger positive and significant. Finally, the models predicting the proportion of matches that include the two fictitious people questions, either as part of all 18 items or by themselves, show a positive and significant effect of familiarity. Mirroring the descriptive statistics, when the two fictitious people questions are removed and we consider 16 questions, respondent-only questions, or child-only questions, the effect of type of interview on the accuracy of responses vanishes.

With respondent-item as the unit of analysis, we then ran probit models to separately predict the likelihoods of validation and matching. We used the same covariates, adding controls for type of question—whether a fictitious people question, a respondent item, a child1 question, or a child2 question. Results for the type of interview dummies are shown in Table 6—full regression results are in tables A5 and A6. Again, familiarity either has a positive effect or no effect at all, but never a negative effect as presumed by the stranger-interviewer norm. The models for likelihood of validation show a positive and significant effect of local-stranger when we use the subset of 16 questions, and of both insider and local-stranger when

we focus on respondent-only questions. The models predicting likelihood of matching show a positive and significant effect of insider interviews when using the full batch of 18 questions, and a strong positive effect of both types of local interviewers when the model is restricted to the two questions on fictitious people.

Our findings so far are clear: respondents are more prone to produce the documentation requested by interviewers when the interviewer is local, regardless of their degree of familiarity with them, as opposed to an outsider. In other words, insiders and local-strangers seem to be more effective at either generating the respondent's trust to show these documents, or inspiring her to invest the time and effort needed to find them. Local interviewers also have an advantage over outsiders when it comes to the accuracy of responses, but our results indicate this to be entirely due to two specific questions in our batch of 18. When these are removed from the analysis, differences in accuracy by type of interview disappear.

In an attempt to shed more light into the apparent advantage of local interviewers when it comes to our ability to validate, we examined validation by question using the expanded data set with respondent-item as the unit of analysis. Table 7 shows the percent of validated cases for each question, by type of interview. Because of the low number of cases of respondents with a second small child, we did not distinguish by child order on the six questions corresponding to children and merged them, so that we now have 10 questions instead of 16. These statistics suggest that the advantage of local interviewers stems from the respondent-only questions, consistent with the results shown above, where interviewer-respondent familiarity had a positive and significant effect when the sample was reduced to respondent-only questions.

The advantage of local interviewers may be further elucidated by examining differences in validation likelihood for different items. We therefore estimated a single additional model that predicts the likelihood of validation, where we replaced the variables for type of question with a dummy variable for each of the 10 questions plus interaction terms for each of these dummies with type of interview. Results from this model are shown in the appendix in Table A7. Table A8, in turn, shows all the resulting Wald tests for the null hypotheses of no difference in the effect between insiders/local-strangers and outsiders on the likelihood of validation given each question. In short, the tests show that both insiders and local-stranger interviewers were significantly more successful at securing validation on three out of four items concerning information on the respondent. The children variables show no differences, with one exception: local-stranger interviewers were significantly more successful than outsiders at validating the question on child's birth weight.

Discussion and conclusion

Our findings, while statistically clear and robust, are constrained by two factors: differences between local and outsider interviewers, and the nature of the questions we were able to validate.

Local and outsider interviewers differed in important ways, as noted earlier. Outsiders were older, more likely to be married and to have children, and more likely to have jobs rather

than being full-time students. These attributes, in addition to being from the country's capital (disclosed in the process of obtaining respondent's consent) and their more professional demeanor as experienced interviewers, are all markers of social status. Thus, our findings might be interpreted as indicating the disadvantage of higher status interviewers, at least when the survey is fielded in a provincial mid-size town, as opposed to the disadvantage of outsidership. Because we included age and skin tone in our models—determinants of status in the Dominican Republic—we can formally claim that the effect of familiarity is independent of these factors.

A more important argument concerning this alternative interpretation is linked to the general goal of our research as it applies to the real world of survey interviewing in LDCs. In that real world, experienced interviewers, as compared to available locals, are usually outsiders with higher social status across many dimensions. A researcher intending to field a survey in a setting like the one we studied will typically find what we found: a large supply of reasonably educated locals who are young, enthusiastic, need income, want the job, and are different in multiple ways from the professional interviewers that the same researcher would have to recruit in the big city. Thus, the ideal lab-like experiment, where locals and outsiders differ only in terms of their status as locals or outsiders, may not only be nearly impossible but actually pointless, since a researcher hiring locals will be getting the entire package of differences that distinguish locals and outsiders interviewers. The design in our experiment reflected this built-in difference. We feel that our theoretical framework can account for all these factors, insofar as the typical attributes of professional interviewers enhance their outsidership.

Because the research presented here focuses on validation, we are limited to mostly non-sensitive administrative types of data that we can validate with certainty. This makes it difficult to generalize our findings to other types of questions. It is possible that the effects of familiarity depend on the nature of the questions asked. For example, using the IRFP data, Rodriguez et al (2015) show no significant differences in responses and item non-response to the most sensitive questions in our survey instrument, which include number of lifetime and recent sex partners, abortions, and drug use, among others. However, as dictated by the IRB overseeing our study, these questions were placed in a self-administered part of our questionnaire, which may have diluted the significance of familiarity—though respondents had to still trust that the interviewer would not open the sealed envelope with their responses once out of sight. In addition, the responses could not be validated. Other research papers using IRFP data tackle similar insider-, local and outsider-interviewer effects on a larger array of questions, but because direct external validation is impossible in those other analyses as well, the analytical tools are restricted to exploring differences in responses with considerations of the likely direction of social desirability bias (see Weinreb et al. 2011, Rodriguez et al. 2015, and Stecklov et al. forthcoming). As far as direct validation is concerned, we cannot go further than what we show in the present article, but all the pieces cited above empirically question and weaken the stranger-interviewer norm.

These constraints notwithstanding, our findings on validation provide a remarkably consistent rejection of the stranger-interviewer norm, which would predict that familiarity between interviewers and respondents weakens validity. In our 18 questions, with two

exceptions, interviewer-respondent familiarity has *no effect* on the accuracy of survey responses. The exceptions, the two questions on fictitious people, provide an even clearer negation of the stranger-interviewer norm: respondents were less truthful when interviewed by outsiders. One explanation for the difference between these two exceptions and the other 16 questions is that the former did not request information on the respondent or the respondent's small children. Instead, they focused on respondent knowledge. Misreporting knowledge of these fictitious people is a type of social desirability bias: avoiding embarrassment by appearing to be more cosmopolitan and knowledgeable than the respondent really is. Either way, the effects of these feelings on data quality were more notable when the questions were asked by the interviewer idealized by the stranger-interviewer norm: an outsider who is never to be seen again.

Our analyses also demonstrate that our ability to validate in the first place is itself dependent on interviewer-respondent familiarity. This means that respondents were more likely to produce the documentation needed to undertake validation when the interviewer was local (either a stranger or known to the respondent) than when the interviewer was an outsider. This is another check mark against the stranger-interviewer norm: respondents in this setting were not more open with complete strangers, even those unlikely to be seen again. Below we speculate on mechanisms that seem to give an edge to local or known interviewers, but in any event, familiarity with place, as in the case of a local-stranger interviewer, or interpersonal familiarity, as in the case of an insider interviewer, appears to induce respondents to cooperate.

A closer look at the questions that mattered most provides some insight on the mechanisms at work. In three of four questions referring to information on the respondent, validation was more likely to happen when the interviewer was local. This was not a straightforward outcome of producing the *cédula*, since the respondent's blood type (one of those three questions) could be validated with other documentation and is not always specified on the *cédula*. When considering the child-specific questions, we found local-stranger interviewers to be significantly more likely than outsiders to be able to validate the child's birth weight. Insider interviewers also outperformed outsiders, but the difference felt just short of statistical significance—given the magnitude of the difference shown in Table 7, we believe this was a consequence of small sample size. While all the other child variables could only be validated with vaccination cards (and therefore the appearance of the card would allow for blanket validation of the other five questions on children), birth weight could be validated with a variety of hospital records but was not always specified on the vaccination card. Another way to look at this is to note that, when producing their children's vaccination cards, respondents were as cooperative with outsiders as they were with locals. Our interviewers were trained to assist respondents with basic information on their children's recommended vaccination schedules, and they delivered a brochure on the topic. In other words, when offered a potential benefit to their own children, respondents lowered their guard against outsiders. Otherwise, they were more willing to invest effort to produce the requested documentation, or more willing to trust the interviewer with such documentation, when the interviewer was someone they knew, or otherwise a fellow townspeople. Outsiders were at a disadvantage here.

Another possibility is that respondents were less likely to produce documentation when they knew they had lied, or perhaps guessed, when asked for the relevant information. There is no way to know this, but if it was indeed the case, our results (since those interviewed by outsiders were less likely to enable validation) indirectly suggest a higher likelihood of misreporting information to the outsider interviewers. Whatever the reason for the familiarity effect on our ability to validate, the appearance of those records was a good thing, and they were more likely to appear when the interviewer was an insider, or a local-stranger, than an outsider.

Again, when records are available, we find no evidence that variation in interviewer-respondent familiarity affects data accuracy. In other words, in terms of the accuracy of data on the types of variables used in this analysis, there is a tie between the three different types of interviewers, each of whom indexes a discrete level of interviewer-respondent familiarity. This runs counter to the stranger-interviewer norm and provides support to researchers who choose to employ local interviewers, irrespective of their level of familiarity with respondents.

We began this paper by noting that, despite the widespread acceptance of the stranger-interviewer norm, it has never been subject to rigorous empirical testing, and there are no robust empirical foundations to support it. Our results, based on data from the first experimental evaluation of the stranger-interviewer norm, show that there are now sufficient empirical grounds for questioning the assumptions associated with that norm. Until methodological research demonstrates otherwise, stranger-interviewers need not be the default choice in survey data collection in developing country contexts.

Acknowledgments

This research was funded by NIH Grant 1R21HD054731-01A1 to the authors. We are grateful for the comments and suggestions provided by three anonymous reviewers and an anonymous statistical consultant.

Appendix

Table A1

Data structure with alternative units of analysis. Example with fictitious data. Interviewer-responder familiarity project, Dominican Republic, 2010.

Panel A: Respondent as the unit of analysis				
Respondent ID	Selected into validation	Proportion validated	Proportion matched	Rest of variables
1	1	0.67	1.00	Interview type, stratum, East San Benito, respondent characteristics, interviewer characteristics
2	1	0.50	0.00	
3	1	1.00	0.50	
4	0	0.00		

A respondent is “Selected into validation” if at least one eligible item could be validated. “Proportion validated” is subject to eligibility (see Table 1). In this example, we validated two out of three eligible items for Respondent 1 (see Panel B below), leading to a proportion validated of 0.67. The “Proportion matched” is restricted to the validated items, so that it is missing when there are no validated items (Respondent 4).

In the analysis on the subset of questions on children only, “Selected into validation” would be missing if the respondent had no minor children, making her not eligible for validation. Accordingly, “Proportion validated” and “Proportion matched” would be missing as well.

✓ We define validation as the ability to confirm with certainty or near certainty whether a response was accurate, by comparison to a trustworthy external source (see text). Validation can have two outcomes: match or mismatch, depending on whether the answer provided by the respondent matches the external source or not.

Table A1

Data structure with alternative units of analysis. Example with fictitious data. Interviewer-responder familiarity project, Dominican Republic, 2010.

Panel B: Respondent-item as the unit of analysis				
Respondent ID	Item	Validated	Match	Rest of variables
1	R birth place	1	1	Interview type, stratum, East San Benito, respondent characteristics, interviewer characteristics
1	FP Medina	1	1	
1	Child 1 TB vacc	0	.	
1	Child 2 bweight	.	.	
2	R birth place	0	0	
2	FP Medina	1	0	
2	Child 1 TB vacc	.	.	
2	Child 2 bweight	.	.	
3	R birth place	.	.	
3	FP Medina	1	1	
3	Child 1 TB vacc	1	0	
3	Child 2 bweight	.	.	

For simplicity, only four items are considered in this example: the respondent's birth place, her response to the question on one of the fictitious people, whether Child 1 received the TB vaccine, and birth weight of Child 2. "Validated" is 1 if the item could be validated, and a missing value indicates non-eligibility. In this example, no respondent has a second minor child, Respondent 2 has no minor child at all, and Respondent 3 reported to never have had a cedula, making her birth place ineligible for validation, as per Table 1. Respondent 2, on the other hand, said she does have a cédula, but did not show it to the interviewer.

We define validation as the ability to confirm with certainty or near certainty whether a response was accurate, by comparison to a trustworthy external source (see text). Validation can have two outcomes: match or mismatch, depending on whether the answer provided by the respondent matches the external source or not.

Table A2

Probit models predicting likelihood of selection of respondents into validation. Respondent as unit of analysis. Interviewer-respondent familiarity project, Dominican Republic, 2010.

Predictors	16 questions	Respondent-only	Child-only
Type of interview (ref.outsider)			
Insider	0.438 **	0.438 *	-0.161
Local-stranger	0.308 *	0.365 **	-0.001
Claimed stratum	0.052	0.058	0.078
East San Benito	0.241 *	0.235 *	0.092
Children ever born	0.029	0.013	-0.174
Age	-0.01	-0.004	0.011
High school diploma	0.065	0.079	-0.071
Household appliances	-0.072 *	-0.074 *	-0.029
Multiple properties	0.23 *	0.162 ^	-0.09
Interviewer skin color	0.128 *	0.121 *	0.11
Interviewer age	0.004	0.006	0.001
Constant	0.319	0.088	-0.354
N	1200	1200	309
Log pseudolikelihood	-521.7777	-541.93747	-209.39
Pseudo R-squared	0.0321	0.0288	0.0218

^
p<0.10,
*
p<0.05,
**
p<0.01

Table A3

OLS models predicting proportion validated. Respondent as unit of analysis. Interviewer-respondent familiarity project, Dominican Republic, 2010.

Predictors	16 questions	Respondent-only	Child-only
Type of interview (ref.outsider)			
Insider	0.063	0.089 *	-0.031
Local-stranger	0.068 ^	0.082 *	0.003
Claimed stratum	0.026	0.016	0.034
East San Benito	0.043 *	0.055 *	-0.025
Children ever born	-0.026 **	-0.008	-0.069 **
Age	0.007 **	0.004 *	0.006
High school diploma	0.036	0.058 *	-0.037
Household appliances	-0.009	-0.011	-0.008
Multiple properties	0.049 *	0.042 *	-0.008
Interviewer skin color	0.03 *	0.023 *	0.036
Interviewer age	-0.001	0.000	0.000
Constant	0.181	0.285 *	0.195

Predictors	16 questions	Respondent-only	Child-only
N	1200	1200	309
Pseudo R-squared	0.0495	0.0355	0.0431

[^]
 p<0.10,
 *
 p<0.05,
 **
 p<0.01

Table A4

OLS models predicting proportion of matches. Respondent as unit of analysis. Interviewer-respondent familiarity project, Dominican Republic, 2010.

Predictors	18 questions	Fictitious people	16 questions	Respondent-only	Child-only
Type of interview (ref.outsider)					
Insider	0.060 *	0.128 **	-0.021	-0.026	-0.009
Local-stranger	0.018	0.098 **	-0.026	-0.034	0.031
Claimed stratum	0.004	0.014	0.030 ^	0.033 *	-0.009
East San Benito	-0.017	0.000	-0.026 *	-0.036 *	0.038
Children ever born	-0.023 **	-0.020 *	-0.029 **	-0.030 **	0.011
Age	0.001	0.001	0.000	-0.001	0.002
High school diploma	0.090 **	0.057 *	0.102 **	0.110 **	0.049
Household appliances	0.003	0.001	0.012 *	0.013 *	-0.008
Multiple properties	-0.012	-0.007	0.001	0.003	-0.010
Interviewer skin color	-0.016	-0.018	-0.005	-0.009	0.015
Interviewer age	-0.003 ^	-0.003	-0.002	-0.002	-0.002
Constant	0.902 **	0.914 **	0.832 **	0.875 **	0.623 **
N	1200	1200	1001	989	159
Pseudo R-squared	0.117	0.082	0.142	0.157	0.049

^

p<0.10,

* p<0.05,

** p<0.01

Table A5

Probit models predicting likelihood of validation. Respondent-item as unit of analysis.
Interviewer-respondent familiarity project, Dominican Republic, 2010.

Predictors	16 questions	Respondent-only	Child-only
Type of interview (ref.outsider)			
Insider	0.131	0.227 *	−0.085
Local-stranger	0.153 *	0.197 **	0.045
Claimed stratum	0.047	0.027	0.101
East San Benito	0.091	0.141 *	−0.042
Children ever born	−0.045 *	−0.019	−0.18 **
Age	0.009 **	0.009 *	0.013
High school diploma	0.088	0.148 *	−0.122
Household appliances	−0.023	−0.028 ^	−0.011
Multiple properties	0.093	0.122 ^	0.026
Interviewer skin color	0.073 **	0.055 *	0.124 ^
Interviewer age	−0.002	−0.001	−0.002
Child 1 validation	−0.514 **		
Child 2 validation	−1.101 **		
Constant	−0.45 ^	−0.485 *	−0.864
N	6840	4770	2070
Log pseudolikelihood	−4503.231	−3174.997	−1319.553
Pseudo R-squared	0.048	0.011	0.023

^
p<0.10,
*
p<0.05,
**
p<0.01

Table A6

Probit models predicting likelihood of matching. Respondent-item as unit of analysis. Interviewer-respondent familiarity project, Dominican Republic, 2010.

Predictors	18 questions	Fictitious people	16 questions	Respondent-only	Child-only
Type of interview (ref.outsider)					
Insider	0.159 *	0.519 **	-0.074	-0.114	0.048
Local-stranger	0.095	0.35 **	-0.073	-0.138	0.162
Claimed stratum	0.063	0.054	0.084	0.123 ^	-0.066
East San Benito	-0.062	0.004	-0.092 ^	-0.116 ^	0.031
Children ever born	-0.071 **	-0.076 **	-0.068 **	-0.081 **	0.039
Age	-0.001	0.003	-0.004	-0.004	0.002
High school diploma	0.265 **	0.226 **	0.292 **	0.333 **	0.159
Household appliances	0.024 ^	0.001	0.04 **	0.051 **	-0.027
Multiple properties	-0.008	-0.015	-0.013	0.01	-0.034
Interviewer skin color	-0.033	-0.064 ^	-0.014	-0.03	0.072
Interviewer age	-0.008 *	-0.011 ^	-0.006	-0.007	-0.004
Fictitious people validation	0.186 **				
Child 1 validation	-0.169 **		-0.203 **		
Child 2 validation	0.233		0.219		
Constant	1.025 **	1.256 **	1.027 **	1.146 **	0.111
N	6000	2397	3603	2861	742
Log pseudolikelihood	-2982.239	-1061.998	-1892.894	-1451.0228	-426.6436
Pseudo R-squared	0.037	0.054	0.035	0.05	0.009

^
p<0.10,

*
p<0.05,

**
p<0.01

Table A7

Probit model predicting likelihood of validation. Respondent-item as unit of analysis.
Interviewer-respondent familiarity project, Dominican Republic, 2010.

Predictors	Coefficient	p-value
Type of interview (ref.outsider)		
Insider	0.297	0.025
Local-stranger	0.307	0.004
Claimed stratum	0.056	0.428
East San Benito	0.093	0.154
Children ever born	-0.060	0.008
Age	0.011	0.006
High school diploma	0.089	0.191
Household appliances	-0.026	0.172
Multiple properties	0.112	0.126
Interviewer skin color	0.086	0.006
Interviewer age	-0.001	0.807
Validation items (ref. R's blood type)		
R's birth place	1.421	0.000
R's birth date	1.440	0.000
R's skin color	0.480	0.000
Child's birth date	0.736	0.000
Child's TB vaccine	0.759	0.000
Child's tetanus vaccine	0.290	0.049
Child's polio vaccine	0.264	0.073
Child's measles vaccine	0.714	0.000
Child's birth weight	-1.069	0.000
Interactions		
R's birth place × insider	0.011	0.940
R's birth place × stranger	-0.094	0.414
R's birth date × insider	-0.105	0.469
R's birth date × stranger	-0.183	0.103
R's skin color × insider	-0.061	0.664
R's skin color × stranger	-0.011	0.918
Child's birth date × insider	-0.430	0.072
Child's birth date × stranger	-0.347	0.048
Child's TB vaccine × insider	-0.453	0.059
Child's TB vaccine × stranger	-0.369	0.036
Child's tetanus vaccine × insider	-0.277	0.266
Child's tetanus vaccine × stranger	-0.318	0.089
Child's polio vaccine × insider	-0.303	0.224
Child's polio vaccine × stranger	-0.276	0.138

Predictors	Coefficient	p-value
Child's measles vaccine × insider	−0.408	0.088
Child's measles vaccine × stranger	−0.353	0.045
Child's birth weight × insider	0.264	0.437
Child's birth weight × stranger	0.605	0.025
Constant	−1.417	0.000

Number of observations = 6840.

Pseudo R-squared = 0.1613.

Log pseudolikelihood = −3968.361.

Table A8

Wald tests for the difference between insider/local-stranger and outsider, for each specific question, in coefficients from a Probit model predicting likelihood of validation. Interviewer-responder familiarity project, Dominican Republic, 2010.

Question	Test	Difference ¹	Chi-sq	p-value
R's blood type	Insider v Outsider	0.297	5.02	0.0250 *
	Local-stranger v Outsider	0.307	8.39	0.0038 **
R's birth place	Insider v Outsider	0.308	4.50	0.0338 *
	Local-stranger v Outsider	0.213	3.84	0.0500 ^
R's birth date	Insider v Outsider	0.192	1.80	0.1802
	Local-stranger v Outsider	0.124	1.30	0.2537
R's skin color	Insider v Outsider	0.236	3.39	0.0657 ^
	Local-stranger v Outsider	0.296	8.39	0.0038 **
Child's birth date	Insider v Outsider	−0.133	0.37	0.5424
	Local-stranger v Outsider	−0.040	0.06	0.8054
Child's TB vaccine	Insider v Outsider	−0.156	0.51	0.4771
	Local-stranger v Outsider	−0.062	0.14	0.7034
Child's tetanus vaccine	Insider v Outsider	0.020	0.01	0.9298
	Local-stranger v Outsider	−0.010	0.00	0.9509
Child's polio vaccine	Insider v Outsider	−0.006	0.00	0.9783
	Local-stranger v Outsider	0.032	0.04	0.8516
Child's measles vaccine	Insider v Outsider	−0.111	0.26	0.6128
	Local-stranger v Outsider	−0.046	0.08	0.7796
Child's birth weight	Insider v Outsider	0.561	3.11	0.0778
	Local-stranger v Outsider	0.912	12.76	0.0004 **

¹For R's blood type, coefficient for main effect. For all other questions, sum of main effect and the corresponding interaction term. Full model results provided in Table A7.

References

- Agar, Michael. *The Professional Stranger: An Informal Introduction to Ethnography*. Academic Press; New York: 1980.
- Axinn, William, Pearce, Lisa D. *Mixed Method Data Collection Strategies*. Cambridge University Press; New York: 2006.
- Barley, Nigel. *The Innocent Anthropologist: Notes from a Mud Hut*. Penguin Books; London, England: 1983.
- Barnes, John A. *A Pack of Lies: Towards a Sociology of Lying*. Cambridge University Press; Cambridge, England: 1994.
- Barnighausen, Till, Bor, Jacob, Wandira-Kazibwe, Speciosa, Canning, David. Correcting HIV prevalence estimates for survey non-participation using Heckman-type selection models. *Epidemiology*. 2011; 22(1):27–35. [PubMed: 21150352]
- Bulmer, Martin. Interviewing and Field Organization. In: Martin, Bulmer, Warwick, Donald P., editors. *Social Research in Developing Countries: Surveys and Censuses in the Third World*. 2nd. UCL Press; London, England: 1993. p. 205-217.
- Davis RE, Couper MP, Janz NK, Caldwell CH, Resnicow K. Interviewer Effects in Public Health Surveys. *Health Education Research*. 2010; 25:14–26. [PubMed: 19762354]
- England LR. Little Kinsey: An Outline of Sex Attitudes in Britain. *Public Opinion Quarterly*. 1949; 13(4):587–600.
- Flores, René, Telles, Edward. Social Stratification in Mexico: Disentangling Color, Ethnicity, and Class. *American Sociological Review*. 2012; 77(3):486–94. [PubMed: 25382861]
- Fowler, Floyd J., Mangione, Thomas W. *Standardized Survey Interviewing: Minimizing Interviewer-Related Error*. Sage Publications; Newbury Park, CA: 1990.
- Garfinkel, Harold. *Studies in Ethnomethodology*. Prentice-Hall; Englewood Cliffs, NJ: 1967.
- Goffman, Erving. *Relations in Public: Microstudies of the Public Order*. Basic Books; New York: 1971.
- Hancock, Jeffrey T., Conrad, Frederick G., Schober, Michael F. Envisioning the survey interview of the future. Hoboken, NJ; John Wiley & Sons, Inc.: 2008. Disclosure and Deception in Tomorrow's Survey Interview: The Role of Information Technology; p. 179-194.
- Hyman, Herbert H., Cobb, William J., Feldman, Jacob J., Hart, Clyde W., Stember, Charles Herbert. *Interviewing in Social Research*. University of Chicago Press; Chicago: 1954.
- Kinsey, Alfred C., Pomeroy, Wardell B., Martin, Clyde E. *Sexual Behavior in the Human Male*. Saunders; Philadelphia: W.B: 1948.
- Kinsey, Alfred C., Pomeroy, Wardell B., Martin, Clyde E., Gebhard, Paul H. *Sexual Behavior in the Human Female*. Saunders; Philadelphia: W.B: 1953.
- Maccoby, Eleanor E., Maccoby, Nathan. The Interview: A Tool of Social Science. In: Lindzey, Gardner, editor. *Handbook of Social Psychology*. Addison-Wesley Publishing Company; Cambridge, MA: 1954. p. 449-487.
- Macro International Inc. *AIDS Indicator Survey: Guidelines for Interviewer Training*. Macro International Inc.; Calverton, Maryland, USA: 2007. MEASURE DHS
- Marcelli, Enrico, Holmes, Louisa, Estella, David, da Rocha, Fausto, Granberry, Phillip, Buxton, Orfeu. (In)Visible (Im)Migrants: The Health and Socioeconomic Integration of Brazilians in Metropolitan Boston. Center for Behavioral and Community Health Studies, San Diego State University; San Diego, CA: 2009a.
- Marcelli, Enrico, Holmes, Louisa, Troncoso, Magalis, Granberry, Phillip, Buxton, Orfeu. Permanently Temporary?: The Health and Socioeconomic Integration of Dominicans in Metropolitan Boston. Center for Behavioral and Community Health Studies, San Diego State University; San Diego, CA: 2009b.

- Maynard, Douglas W., Schaeffer, Nora Cate, Maynard, DW., Houtkoop-Steenstra, H., Schaeffer, NC., van der Zouwen, H. Standardization and Tacit Knowledge: Interaction and Practice in the Survey Interview. Wiley Interscience; New York: 2002. Standardization and Its Discontents; p. 3-45.
- Parrado, Emilio A., McQuiston, Chris, Flippen, Chenoa A. Participatory survey research: Integrating community collaboration and quantitative methods for the study of gender and HIV risks among Hispanic migrants. *Sociological Methods & Research*. 2005; 34(2):204–239.
- Reynolds, Lindsey, Cousins, Thomas, Newell, Marie-Louise, Imrie, John. The social dynamics of consent and refusal in HIV surveillance in rural South Africa. *Social Science and Medicine*. 2013; 77:118–125. [PubMed: 23219165]
- Rodriguez, Leslie, Sana, Mariano, Sisk, Blake. Self-Administered Questions and Interviewer-Respondent Familiarity. *Field Methods*. 2015; 27(2):163–181.
- Sana, Mariano, Conway, Becky, Gold, Steven J., Nawyn, Stephanie J. The International Handbook of Migration Studies. Routledge; 2012. Surveys and Ethnosurveys; p. 484-493.
- Schaeffer, Nora Cate, Presser, Stanley. The Science of Asking Questions. *Annual Review of Sociology*. 2003; 29:65–88.
- Schaeffer, Nora Cate, Dykema, Jennifer, Maynard, Douglas W., Marsden, Peter V., Wright, James D. Handbook of Survey Research. Emerald Group Publishing; 2010. Interviewers and interviewing; p. 437-470.
- Sidanius, Jim, Peña, Yesilernis, Sawyer, Mark. Inclusionary Discrimination: Pigmentocracy and Patriotism in the Dominican Republic. *Political Psychology*. 2001; 22(4):827–51.
- Simmel, Georg. The Stranger. In: Wolf, Kurt H., editor. *The Sociology of Georg Simmel*. Free Press; New York: 1908/1950. p. 402-408.
- Stecklov, Guy, Weinreb, Alexander A., Sana, Mariano. Forthcoming. Reporting family planning use to strangers: A survey-based experiment from the Dominican Republic. Forthcoming in PLOS ONE.
- Stocking, George W., Stocking, George W. *Observers Observed: Essays on Ethnographic Fieldwork*. University of Wisconsin Press; Madison, WI: 1983. History of Anthropology: Whence/Whither; p. 3-12.
- Stone, Linda, Campbell, J. Gabriel The Use and Misuse of Surveys in International Development: An Experiment from Nepal. *Human Organization*. 1984; 43:27–37.
- Sudman, Seymour, Bradburn, Norman M. *Response Effects in Surveys: A Review and Synthesis*. Aldine Publishing Company; Chicago: 1974.
- Thibaut, John W., Kelley, Harold H. *The Social Psychology of Groups*. John Wiley & Sons, Inc.; New York: 1959.
- United Nations. Human Development Report 2013. The Rise of the Global South: Human Progress in a Diverse World. United Nations Development Program; 2013.
- Weinreb, Alexander A. The Limitations of Stranger-Interviewers in Rural Kenya. *American Sociological Review*. 2006; 71:1014–1039.
- Weinreb, Alexander A., Sana, Mariano, Stecklov, Guy. Outsiders, Locals, and Insiders: Results from a Methodological Experiment on Data Collection; Paper presented at the Annual Meeting of the Population Association of America; Washington D.C.. 2011; March 31 - April 2;

Table 1

Validation items. Interviewer-responder familiarity project, Dominican Republic, 2010.

Item	Validation document	Eligibility
Respondent		
Birth date	<i>cédula</i>	R with <i>cédula</i> ²
Birth place (town and municipio)	<i>cédula</i>	R with <i>cédula</i> ²
“Official” skin color	<i>cédula</i>	R with <i>cédula</i> ²
Blood type	any available	All Rs
First small child		
Birth date	vaccination card	R with a small child
Birth weight	any available ¹	R with a small child
TB vaccine (yes/no)	vaccination card	R with a small child
No. doses of tetanus vaccine	vaccination card	R with a small child
No. doses of polio vaccine	vaccination card	R with a small child
Measles vaccine (yes/no)	vaccination card	R with a small child
Second small child		
Birth date	vaccination card	R with a second small child
Birth weight	any available ¹	R with a second small child
TB vaccine (yes/no)	vaccination card	R with a second small child
No. doses of tetanus vaccine	vaccination card	R with a second small child
No. doses of polio vaccine	vaccination card	R with a second small child
Measles vaccine (yes/no)	vaccination card	R with a second small child
Fictitious people		
Medina	validated by design	All respondents
Comodoro	validated by design	All respondents

Source: Interviewer-Respondent Familiarity Project, DR 2010 data.

¹ Sometimes, but not always, recorded on the vaccination card.² These items are eligible for validation if the respondent indicated that she had been issued a *cédula*, irrespective of whether she produced the *cédula* when asked to show it.

Table 2

Sample size by stratum and type of interview, by unit of analysis. Interviewer-responder familiarity project, Dominican Republic, 2010.

Stratum	Type of interview		Outsider	Total
	Insider	Local-stranger		
<i>Respondents</i>				
Unclaimed	0	455	216	671
Claimed	219	179	138	536
Total	219	634	354	1207
<i>Respondent-items</i>				
Unclaimed	0	3537	1701	5238
Claimed	1635	1326	1080	4041
Total	1635	4863	2781	9279

Source: As for Table 1.

Table 3

Percent of respondents selected into validation, percent of responses validated, and percent of matches among validated responses, by type of interview and stratum. Respondent as unit of analysis. Interviewer-respondent familiarity project, Dominican Republic, 2010.

Subset of questions	Unclaimed		Claimed		
	Local-stranger	Outsider	Insider	Local-stranger	Outsider
<i>Percent of respondents selected into validation¹</i>					
16 questions	82.9	81.5	87.7 **	87.7 **	74.6
<i>N</i>	455	216	219	179	138
Respondent-only	82.2	80.1	86.3 **	87.7 **	72.5
<i>N</i>	455	216	219	179	138
Child-only	47.5	58.1	47.1	64.1 ^	46.2
<i>N</i>	118	62	51	39	39
<i>Mean of percent validated across respondents²</i>					
16 questions	54.8	53.4	59.9 **	63.8 **	47.1
Respondent-only	59.2	56.5	63.7 **	66.2 **	50.0
Child-only	32.1	41.3	36.0	49.2 ^	31.0
<i>Mean of percent of matches across respondents</i>					
18 questions	79.2 ^	73.6	85.2 **	81.4 ^	75.3
<i>N</i>	453	216	219	179	138
Fictitious people	84.9 *	70.4	90.0 **	86.9 **	73.6
<i>N</i>	453	216	219	179	138
16 questions	74.4	74.1	80.5	78.7	80.9
<i>N</i>	377	176	192	157	103
Respondent-only	74.7	75.1	81.6	79.8	82.2
<i>N</i>	374	173	189	157	100
Child-only	78.8 *	70.9	72.6	74.5	75.2
<i>N</i>	56	36	24	25	18

Significance symbols refer to tests for the comparison of coefficients (outsider is the reference) from OLS regressions with no controls and adjustment of clustering of respondents within interviewers.

Source: As for Table 1.

¹Percent of respondents with at least one validated answer (i.e. those used for the calculation of percent of matches).

²Same sample sizes as the top panel.

[^]p<0.10,

^{*}p<0.05,

^{**}p<0.01

Table 4

Percent validated and percent of matches among validated observations, by type of interview and stratum. Respondent-item as unit of analysis. Interviewer-responder familiarity project, Dominican Republic, 2010.

Subset of questions	Unclaimed		Claimed		
	Local-stranger	Outsider	Insider	Local-stranger	Outsider
<i>Percent of validated observations</i>					
16 questions	51.0	51.5	56.1 **	62.1 **	43.5
<i>N</i>	2627	1269	1197	968	804
Respondent-only	59.7	57.5	63.9 **	66.2 **	50.0
<i>N</i>	1805	849	873	716	552
Child-only	31.9	39.5	34.9	50.4 *	29.4
<i>N</i>	822	420	324	252	252
<i>Percent of matches among validated observations</i>					
18 questions	78.7 **	72.6	83.6 **	81.7 ^	77.1
<i>N</i>	2244	1085	1109	959	625
Fictitious people	84.9 **	70.3	90.0 **	86.9 **	73.8
<i>N</i>	905	431	438	358	275
16 questions	74.6	74.2	79.4	78.5	79.7
<i>N</i>	1339	654	671	601	350
Respondent-only	74.0	75.6	81.2	79.8	82.3
<i>N</i>	1077	488	558	474	276
Child-only	77.1 ^	69.9	70.8	74.0	70.3
<i>N</i>	262	166	113	127	74

Significance symbols refer to tests for the comparison of coefficients (outsider is the reference) from OLS regressions with no controls and adjustment of clustering of respondents-items within respondents.

Source: As for Table 1.

^
p<0.10,

*
p<0.05,

**
p<0.01

Table 5

Regression coefficients for type of interview dummies from a probit model predicting selection of respondents into validation, and OLS models predicting proportion validated and proportion matched. Respondent as unit of analysis. Interviewer-responder familiarity project, Dominican Republic, 2010.

Subset of questions	Probit for selection into validation			OLS for proportion validated			OLS for proportion matched		
	Insider ^I	Local-stranger ^I	Insider ^I	Local-stranger ^I	Insider ^I	Local-stranger ^I	Insider ^I	Local-stranger ^I	Local-stranger ^I
18 questions	n/a	n/a	n/a	n/a	0.060 *		0.060 *		0.018
Fictitious people	n/a	n/a	n/a	n/a	0.128 **		0.128 **		0.098 **
16 questions	0.438 **	0.308 *	0.063	0.068 ^	-0.021		-0.021		-0.026
Respondent-only	0.438 *	0.365 **	0.089 *	0.082 *	-0.026		-0.026		-0.034
Child-only	-0.161	-0.001	-0.031	0.003	-0.009		-0.009		0.031

All standard errors adjusted for clustering within interviewers.

Source: As for Table 1.

^I Reference category is outsider. All models control for stratum (claimed vs. unclaimed), location in East San Benito, respondent's fertility, age, and high school diploma, household appliances, ownership of multiple properties, and interviewer's age and skin tone. See the appendix tables for complete model outputs.

^ p<0.10,

* p<0.05,

** p<0.01

Table 6

Regression coefficients for type of interview dummies from probit models predicting validation and matching. Respondent-item as unit of analysis. Interviewer-respondent familiarity project, Dominican Republic, 2010.

Subset of questions	Probit for validation		Probit for matching	
	Insider ^I	Local-stranger ^I	Insider ^I	Local-stranger ^I
18 questions	n/a	n/a	0.159 *	0.095
fictitious people	n/a	n/a	0.519 **	0.350 **
16 questions	0.131	0.153 *	-0.074	-0.073
respondent-only	0.227 *	0.197 **	-0.114	-0.138
child-only	-0.085	0.045	0.048	0.162

All standard errors adjusted for clustering within interviewers.

Source: As for Table 1.

^I Reference category is outsider. All models control for stratum (claimed vs. unclaimed), location in East San Benito, respondent's fertility, age, and high school diploma, household appliances, ownership of multiple properties, and interviewer's age and skin tone. See the appendix tables for complete model outputs.

[^] p<0.10,

* p<0.05,

** p<0.01

Table 7

Percent validated by question and type of interview. Respondent-item as the unit of analysis. Interviewer-respondent familiarity project, Dominican Republic, 2010.

Question	Insider	Local-stranger	Outside r
R's blood type	35.62 (219)	33.75 (634)	24.29 (354)
R's birth place	85.32 (218)	81.40 (629)	76.22 (349)
R's birth date	83.03 (218)	79.49 (629)	76.79 (349)
R's skin color	51.83 (218)	51.67 (629)	41.26 (349)
Child's birth date	44.44 (54)	46.37 (179)	49.11 (112)
Child's TB vaccine	44.44 (54)	46.37 (179)	50.00 (112)
Child's tetanus vaccine	33.33 (54)	30.73 (179)	32.14 (112)
Child's polio vaccine	31.48 (54)	31.28 (179)	31.25 (112)
Child's measles vaccine	44.44 (54)	45.25 (179)	48.21 (112)
Child's birth weight	11.11 (54)	17.32 (179)	3.57 (112)

Source: As for Table 1.