Undercoverage Rates and Undercoverage Bias in Traditional Housing Unit Listing

Stephanie Eckman
Institute for Employment Research
Nuremberg, Germany

Frauke Kreuter
Joint Program in Survey Methodology
University of Maryland, College Park
& Institute for Employment Research
Nuremberg, Germany

Published in: *Sociological Methods & Research* 42(3): 264-293

DOI: 10.1177/0049124113500477

Abstract

Many face-to-face surveys use field staff to create lists of housing units from which samples are selected. However, housing unit listing is vulnerable to errors of undercoverage: Some housing units are missed and have no chance to be selected. Such errors are not routinely measured and documented in survey reports. This study jointly investigates the rate of undercoverage, the correlates of undercoverage, and the bias in survey data due to undercoverage in listed housing unit frames. Working with the National Survey of Family Growth, we estimate an undercoverage rate for traditional listing efforts of 13.6 percent. We find that multi-unit status, rural areas, and map difficulties strongly correlate with undercoverage. We find significant bias in estimates of variables such as birth control use, pregnancies, and income. The results have important implications for users of data from surveys based on traditionally listed housing unit frames.

Acknowledgments:

In conducting this research, we received much help from staff members at the Survey Research Center at the University of Michigan, in particular Nicole Kirgis, Shonda Kruger-Ndiaye, Jim Lepkowski, James Wagner and Brady West. Thanks also to Dr. Mosher at the National Center for Health Statistics for allowing us to carry out this research with NSFG. For comments on earlier drafts of this paper, we thank Katharine G. Abraham, J. Michael Brick, Mick Couper, Colm O'Muircheartaigh, Jennifer Sinibaldi, and three reviewers. Support for this research was provided by the Centers for Disease Control and Prevention Grants for Public Health Dissertations, the Eunice Kennedy Shriver National Center for Child Health and Human Development grant R24-HD041041 to the Maryland Population Research Center, the Charles Cannell Fund in Survey Methodology at the Institute for Social Research of the University of Michigan, and the Rensis Likert Fund in Research in Survey Methodology at the Institute for Social Research. This article is based on the first author's dissertation, written at the Joint Program in Survey Methodology at the University of Maryland.

1 Introduction

Because of the importance of survey data to academic and public policy research, survey researchers have taken care to reduce nonresponse bias and measurement error. Coverage error, on the other hand, has received much less attention in large face-to-face surveys. Consider, as an example, the flagship survey for the study of fertility, sexual behavior and family formation in the U. S., the National Survey of Family Growth (NSFG). The NSFG team has conducted important research in recent years on responsive design techniques to reduce nonresponse bias without increasing costs (Groves and Heeringa, 2006; Axinn et al., 2011) and on reducing misreporting to sensitive questions (Couper et al., 2009; Peytchev et al., 2010). Coverage error has not been addressed, except indirectly by Joyner et al. (2012), who look at overall representativity in the NSFG, a term which subsumes nonresponse, undercoverage and measurement error. While the concept of representativity is insightful, teasing apart the effects of different error sources can help us understand the mechanisms of error at each point in the data collection process and thus assist in improving data quality.

Studies that explore undercoverage *rates* in face-to-face surveys often compute net coverage rates: the ratio of weighted survey totals to national control totals. Net coverage rates in these studies are as low as 90% and even lower among Hispanics, African-Americans and men (Botman, 1987; Hainer et al., 1988; Shapiro and Kostanich, 1988; Fay, 1989; Shapiro et al., 1993; Montaquila et al., 1996; Chromy et al., 1999; Horrigan et al., 1999; Judkins et al., 1999; Meier and Moore, 1999; Morton et al., 2006). These studies show that face-to-face surveys miss many people that they claim to represent. Such undercoverage can occur at two stages. In the first stage, the housing unit frame from which the sample is selected should contain all units that lie in the areas selected for the survey. In the second stage, each eligible member of the selected households should have a chance to be selected. For the latter stage, the literature on coverage of household members has revealed several mechanisms of undercoverage in the application of the household roster: those with tenuous attachments to households; those who do not contribute to household income; and those who wish to hide from authorities or generally try to avoid taking surveys tend to be undercovered (Valentine and Valentine, 1971; Hainer, 1987; Fein, 1990; Tourangeau et al., 1997; Martin, 2007; Tourangeau et al., 2012). In the first stage, however, the mechanisms of undercoverage of housing units are not yet well understood. For this reason, we focus in this paper on housing unit frame creation, specifically those created by field staff in a process called "listing." If frames created through listing miss some units that they should include, they suffer from undercoverage. The missed units have no chance to be selected, and the people who live there have no opportunity to take part in any survey using the listings. [1]

[1] Undercovered units could be covered via a missed housing unit procedure such as the half-open interval technique, though see Eckman and O'Muircheartaigh (2011) for evidence that this technique does not work as it should.

Just as nonresponse can lead to bias if the propensity to respond is related to variables collected in the survey (Groves, 2006), undercoverage can also lead to *bias* if the propensity to be included on the frame is related to survey variables. For example, if the field staff tend to miss small apartments, and the persons who live in these units are more likely to live alone, then estimates of household size could be biased due to errors in the frame. None of the studies mentioned above explores coverage bias in face-to-face surveys, but there is indirect evidence that implausible estimates of victimization rates from the National Crime Study (Martin, 1981; Cook, 1985) and relative labor force participation and school enrollment rates by race and sex from the Current Population Survey (Clogg et al., 1989) are due to coverage bias.

In most surveys, not only do we not have data about the cases that are undercovered, very often we do not even know how many such cases there are. This lack of information about the undercovered cases is the central challenge to studying coverage error. This paper takes a new approach to the study of coverage in listed housing unit frames. Working with professional field staff, we relisted segments in which NSFG interviews had already been carried out. We attempt to reveal patterns of undercoverage and estimate coverage bias by investigating whether completed cases, known to be valid housing units, were captured by the second listing. We answer the following three research questions:

1. What would the undercoverage rate in our sample segments have been, had only the traditional listing been done?
2. What do the patterns of undercoverage in these segments suggest about the mechanisms behind the listing errors?
3. How much bias would result in key NSFG demographic and fertility measures due to undercoverage on the housing unit frame?

2 Background

There are two commonly used methods to create housing unit frames in the field. In *traditional listing*, members of the field staff, called listers, create frames "from scratch." The listers are given maps which show streets, water features and local landmarks, and which indicate with shading the area to be listed. These selected areas, called segments, are one or more Census blocks, grouped together to meet a minimum population size and selected with probability proportional to population. Segments can be as small as one block in dense urban areas or as large as several square miles in sparsely settled rural areas. Using the maps as guides, the listers systematically travel around the selected area and record the address of every residential unit they see, on paper or on a computer (Harter et al., 2010). The *dependent listing* task is similar, but here listers are provided with an already existing list of addresses, which they update in the field, adding units that are missing and removing those that do not exist or are not residential. Traditional listing was for many years considered the highest quality method of frame construction, an improvement over quota and random walk procedures (Manheimer and Hyman, 1949; Boyd and Westfall, 1955). Dependent listing is rather new and can be less expensive then traditional listing (O'Muircheartaigh et al., 2002).

Regardless of the method, the goal of the listing process is to create a frame of all residential addresses that lie inside the selected area. These lists are returned to the central office and become a frame from which a sample of housing units is selected for interviewing. Interviewers then return to the areas and attempt to interview at the selected units.[2]

Previous research has illuminated an important mechanism of error in dependent listing: listers tend to become overly reliant on the existing list of addresses, and tend not to correct errors of inclusion and exclusion (Eckman and Kreuter, 2011). Less is known about the mechanisms of error in traditional listing, which is why we focus on this listing method here. Traditional listing is widely used in the U.S. and elsewhere. In 2009, the National Survey of Drug Use and Health listed more than 500,000 units via traditional listing (Morton et al., 2010, Table 3.4), and several countries participating in the European Social Survey also use traditional listing (Jowell and the Central Co-ordinating Team, 2003, 2005, 2007; Central Co-ordinating Team, 2010). Other surveys, such as the National Survey of Family Growth, the General Social Survey, and the Current Population Survey, use traditional listing in some parts of the country, and dependent listing in others (Groves et al., 2009; Harter et al., 2010; U.S. Census Bureau, 2006).

Studies of coverage of traditionally listed housing unit frames have found net coverage rates from 80% to more than 99% (Manheimer and Hyman, 1949; Kish and Hess, 1958; Hawkes, 1986; Childers, 1992; Barrett et al., 2002; Pearson, 2003; O'Muircheartaigh et al., 2006, 2007). The neighborhood characteristics associated with undercoverage are low median household income (Manheimer and Hyman, 1949; O'Muircheartaigh et al., 2007) and rural areas (O'Muircheartaigh et al., 2007). At the housing unit level, vacant units, trailers, and those in multi-unit buildings, particularly small buildings with two to nine units, are vulnerable to undercoverage (U. S. Census Bureau, 1993; Childers, 1992; Barrett et al., 2002). Lister characteristics, and interactions between lister and segment characteristics, have been hypothesized to affect frame quality. However, the only study to look at lister characteristics found no significant relationships between lister experience and errors of undercoverage and overcoverage (Pearson, 2003).

Like all coverage studies, these previous studies of coverage in housing unit frames must make assumptions about the gold standard frame against which they compare the frame under investigation. Kish and Hess (1958) calculate net coverage

[2] Listing should not be confused with random route or address random techniques that rely on interviewers walking a serpentine route, selecting housing units as they go, which are used in several European countries and in developing countries (Fink, 1963; Schnell, 2008). The listing task differs from these techniques due to the clear separation in time between the frame creation and sampling tasks.

rates relative to an estimate of the true frame size. This approach is quite sensitive to the accuracy of the estimates and furthermore cannot determine whether the *correct* housing units are on the frame, only that the correct *number* of units are on the frame. Other studies assume that a second listing, by a more senior lister or via dependent listing (or both), is more accurate (Childers, 1993; Barrett et al., 2002; O'Muircheartaigh et al., 2002; Pearson, 2003; Thompson and Turmelle, 2004; O'Muircheartaigh et al., 2006, 2007). However, we are not aware of any empirical evidence that justifies this assumption, and it may be the case that junior listers make fewer errors than do senior listers, as junior staff may have more regular and recent listing experience. The use of dependent listing to create a gold standard frame can also be problematic, given the susceptibility of that listing method to confirmation bias (Eckman and Kreuter, 2011). In addition to their reliance on problematic gold standard assumptions, none of these studies estimates coverage bias due to errors in the listed frame.

To understand what drives the errors in housing unit listing, we first consider the listing task in detail. A good lister should circle a block several times, and look carefully at each building, to determine how many residential units it contains. This behavior may attract negative attention from residents or even the police. Listers work largely on their own, and may be the only member of the project staff for hundreds of miles. When they encounter difficult or unclear listing situations, they are not able to discuss them with another lister or supervisor, except by phone. In addition, their work is usually not fully reviewed in the field by another project member, due to the high costs of traveling other listers to the area.[3]

The rational actor theory holds that individuals weigh costs and benefits when making decisions (Coleman, 1994). Similar perspectives have been used to study survey nonresponse and measurement error (Philipson and Lawless, 1997; Kennickell, 2000, 2003; Singer, 2011), and thus rational actor theory may be useful in understanding the mechanisms of error in housing unit listing as well. In the context of listing, this perspective suggests that listers working alone in unfamiliar areas, without direct contact with a supervisor, may weigh the benefits of closely investigating every structure for hidden housing units against the costs of doing so in terms of time spent away from home, personal safety, and so on. When a map does not match the situation on the ground, they may guess at the boundaries of the selected area. When a building or complex is gated, they may estimate the number of the housing units inside instead of trying to gain access and count the units. Listers may further compromise quality for comfort and convenience by driving rather than walking. Such behaviors make the listing task easier for the lister, but can also lead to errors of undercoverage.

To test for the pattern of errors in traditional listing suggested by the rational actor theory, and to estimate coverage bias in one of our most important demographic

[3]To our knowledge, only the U. S. Census Bureau routinely performs in-field relisting as a check on the accuracy of their listed frames. Every year, senior field staff do a dependent check of two blocks listed by every lister and provide feedback on errors. See Pearson (2003) for more details.

surveys, we conducted a study of listing error in conjunction with the National Survey of Family Growth.

3 Data

In 2009 we worked with the National Survey of Family Growth (NSFG) to relist a sample of segments. The second listing was always done via traditional listing, and thus we can estimate undercoverage in traditional listing by calculating how many units found by the first listing were also covered by the second. We can also use the subset of those cases in the first listing which completed the interview to calculate coverage bias due to traditional listing. This design, involving known housing units about which we have survey data, is unique in the coverage literature, and permits exploration of both undercoverage rates and undercoverage bias.

Cycle 7 of the NSFG was a national area-probability study conducted from 2006 to 2010 by the Division of Vital Statistics at the National Center for Health Statistics to study fertility behavior. The initial sample was selected in four stages. The sampling units at the first stage were counties or groups of counties. At the second stage, segments were selected within these counties. After listing, housing units within the segments were sampled, and interviewers attempted to screen each household to determine if any residents were eligible (15-44 years old). In the last stage, an eligible household member was selected for the interview. A fifth stage was used towards the end of data collection to subsample nonresponding cases for extra interviewer effort (Lepkowski et al., 2010).

Data collection for NSFG Cycle 7 was carried out by the Survey Research Center at the University of Michigan. The long data collection period was broken into quarters lasting three months each. In each quarter, a nationally representative sample of segments was in the field.

All interviewers were also listers: in every quarter they interviewed cases in their active segments and listed segments which would become active the next quarter. As they listed, NSFG listers recorded housing unit addresses into a tablet computer. The software ensured that listers parsed addresses into fields (street number, street name, apartment designator) and provided a drop-down menu of street names preloaded in the central office to minimize spelling errors and standardize abbreviations. Listers also recorded observations of the neighborhood on the computer: languages spoken, safety and accessibility concerns, and whether they drove or walked while listing.[4] The first listing of each segment was conducted by the interviewer assigned by the project, who used dependent listing whenever the address database maintained by the vendor provided more than six addresses in the selected segment, and traditional listing otherwise. This listing took place from January to

[4]These observations are standard for Survey Research Center surveys and are used by interviewers to plan field work strategies.

March, 2009 after which cases were selected and sent to the field for screening and interviewing in the next quarter (April - June 2009).[5]

For the coverage study presented here, a stratified sample of 49 segments from the 104 segments fielded in quarter 12 of Cycle 7 was selected for a second listing. Segments with characteristics found in previous studies to be related to undercoverage in traditional listing, such as many multi-unit dwellings and trailers, or a high percentage of households in poverty, were placed into the one stratum, and all remaining segments were placed in the other. All segments (n=40) in the first stratum were selected, and nine out of 64 in the second stratum. The 49 selected segments were a nationally-representative sample.

For the purposes of this study, the selected segments were each listed a second time by a different NSFG lister, who used the same software and made independent neighborhood observations.[6] This second listing was always via traditional listing. The second listing took place from April to June of 2009 and was always within four months of the first. Due to the delay between the first and second listings, some real change might have take place in the housing unit stock in these areas. Any true loss of housing units in the selected segments (through, for example, demolition, fire, merging with other units) would appear in our analyses as undercoverage. However, fewer than 0.03% of all housing units are destroyed in these manners each year, so we believe the effect of such loss on our estimates is small (Eggers and Moumen, 2011).

Our analyses required matching the frames created by the two listings together: the Appendix provides details on the matching procedures. Housing units in three very rural segments could not be matched, because both listers used descriptions rather than addresses to identify units and seemed to follow different paths as they listed. The housing units in these segments represent about 3.4 million units nationally. Because these segments could not be matched, they are not used in our analyses.

In the 46 segments that we analyze, sixteen interviewers performed the first listings for the NSFG study, and 11 the second listings done for this study. Seven listers participated in both listing exercises, but never listed the same area twice. Most listed three segments in the second listing, but one listed nine, and one only one. The listers who did the second listing were not informed about the experiment, rather they were told that the listing was done to test new features of the listing software.

From the interviewer questionnaire, we know that those who performed the second listing were slightly more experienced but otherwise quite similar to those who

[5]While approaching the selected cases, NSFG interviewers also conduct a half-open interval missed unit procedure which is meant to give any units undercovered in the initial listing a chance of selection. In the segments studied here, interviewers did not find any missed units, though they did find some in other segments in Cycle 7.

[6]This investigation does not explore the inter-rater reliability of the neighborhood observations made by the two listers. See Casas-Cordero et al. (2013) for such an analysis.

did the first (Table 1). By design, all NSFG interviewers, and thus all listers, were female.

Table 1: Characteristics of Participating Listers

	1st Listing	2nd Listing
Number of listers	16	11
African-American	13%	27%
Speaks Spanish	56%	55%
Has second job	38%	45%
Interviewing experience, in years	5.5	6.8
Range, in years	0-12	3-12

To gain insight into the listing task, we conducted 30 to 60 minute individual debriefings with seven of the listers who did the second listing, after both listing tasks were completed, but before analyzing the data. Many of the listing difficulties that came up in our discussions echo the findings of the studies discussed above. Listers have difficulties with small multi-unit buildings: they often cannot tell how many units a building contains, and will count doorbells, mailboxes or utility meters. Several listers mentioned hidden multi-units, buildings that appear to be single family homes but, upon closer inspection during interviewing, contain an additional unit. Rural segments are also difficult, they said, because they tend to cover a large area, the homes can be set quite far back from the road, and many of the streets are unnamed and the houses unnumbered.

Every lister mentioned the quality of her listing maps as a concern, saying that they were hard to read or out-of-date. The maps provided to the listers were based on the Census Bureau's Topologically Integrated Geographic Encoding and Referencing (TIGER) data, which were out-of-date and inaccurate in some areas at the time of the listing for this study, which was prior to the update for the 2010 Census (Zandbergen, 2009). To quantify the error in the listing maps, we compared them to Google's online maps. We coded whether the selected area on Google Maps contained additional streets not shown on the listing map. Twelve segments had such missing interior streets, which may be an indicator of recent growth and construction in the segments. Eight segments had other map errors: mislabeled streets or segment boundaries that were a different shape than those shown online. Such errors make the selected area difficult to identify and could affect coverage.

A map issue that all listers mentioned as troublesome is non-visible boundaries, which are block edges that do not correspond to streets or other visible landmarks: they are often political boundaries such as town or county limits. While non-visible boundaries are not map errors, they can cause problems if listers do not know where they should start and stop listing in such areas. Seventeen segments in this study had non-visible boundaries.

Driving would seem to create problems finding hidden units. A lister in a car cannot see the units well, especially from the driver's side of the car (listers always list the units on their right). The listers explained that they drive because many rural segments are so large, they cannot cover them in a reasonable amount of time except by car. Others mentioned that in single family neighborhoods, where they feel rather certain there are no hidden units, driving is simply easier and faster. In this study, 31 of the 46 segments were listed while the lister herself was driving. All but one of the eight rural segments were listed while driving, and more than half of the urban segments.[7]

4 Methods

To address the first research question concerning the undercoverage rate in the traditional listing, we use the 46 matched segments and two different case bases. The first case base is all 9,059 housing units from the first listing, and the undercoverage rate is the weighted percent of these cases that were also captured in the second listing. However, among these units may be some that were overcovered by the first listing and thus should not have been listed by the second lister. Such errors in the first listing will introduce positive bias into our estimate of the undercoverage rate. To address this shortcoming, we use the second case base, the 1,970 units from the first listing that were selected for the survey and found to have been properly listed by the interviewer.[8] We can be reasonable certain that these housing units existed at the time of data collection and thus that they should have been listed by the second lister. The first approach uses a larger case base and has more external validity, while the second has a higher claim to being a gold standard. Together the two methods give us a good sense of undercoverage in traditional listing.

A logistic model of listing propensity explores the housing unit, segment and lister characteristics that correlate with the probability that a unit was included on the traditionally listed frame. The model allows us to answer the second research question and to test the hypotheses about undercoverage developed above. The dependent variable in model is an indicator of whether the second listing included that unit (1) or not (0). The independent variables are: housing unit characteristics captured by the interviewers during data collection; area characteristics recorded during the second listing, coded from maps, or merged in from Census data; and lister characteristics from the interviewer questionnaire. Positive estimated coefficients on these characteristics indicate those that make a unit *more likely* to be listed and negative

[7] Urbanicity was operationalized as the percent of all housing units in the block group which contains the segment that are in blocks flagged as rural in the 2000 Census. A block group is a Census Bureau geography made up of one or more blocks, smaller than a tract or county.

[8] 1,994 cases were selected for the NSFG screener from our 46 segments. During field work, interviewers flagged 24 cases as improperly listed (either non-residential or outside the boundaries of the segment). These units should not have been listed by either lister and thus are not counted as undercovered.

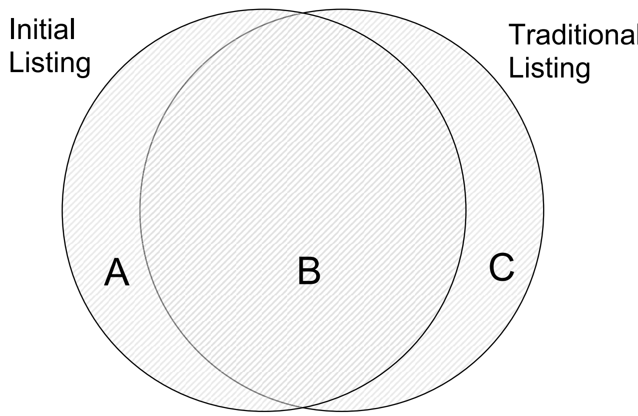


Figure 1: Venn Diagram of Two Listed Housing Unit Frames

coefficients those that make a unit *less likely* to be listed. The model uses the second case base (n=1,970).

The two listings of the completed cases, together with the collected NSFG survey data, also allow us to estimate undercoverage bias and answer the third research question. To help explain the bias estimation methods we use, Figure 1 shows a Venn diagram of the frames created in the two listings. The housing units listed in both the initial listing and the traditional listing are in set B. Some units were listed in the first listing and not the second: these are in set A. There are also units that were listed in the second listing and not the first: set C.[9]

The ideal measure of bias would be the difference between the mean calculated from the units in the traditional listing ($B \bigcup C$) and the mean on all cases ($A \bigcup B \bigcup C$). If we let $\bar{Y}_{BC}$ be the first mean and $\bar{Y}_{ABC}$ the second, the bias is:

$$\text{bias}(\bar{Y}_{\text{trad}}) = \bar{Y}_{BC} - \bar{Y}_{ABC} \tag{1}$$

However, we do not have survey data for any of the cases in C. We have data only for a sample of cases from the first listing ($A \bigcup B$). For this reason, we must make additional assumptions when calculating bias. We use two different approaches to estimating the ideal bias in Equation 1. Each relies upon an assumption that is unlikely to be true in practice, but taken together they likely bound the true bias we are after.[10]

In the first approach, we compare the estimates from all completed cases with the estimates from those cases which were also covered by the second lister. That is, we compare the estimates from the completed cases in $A \bigcup B$ to those only in B. Let $\bar{y}_{AB}$ be the estimate of the mean of a given variable on all completed cases in these

[9]Note that the Venn diagram refers to the full frames created by the two listers, not only to the selected cases. The selected cases are a sample from $A \bigcup B$.

[10]The bias estimates developed here reflect the hypothetical risk of bias in data collected by NSFG, and are not estimates of actual bias in the released NSFG Cycle 7 data.

segments. Let $\bar{y}_B$ be the same mean calculated on only those cases which were also included in the traditional (second) listing. (The lower case y here indicates that we do not have data about all cases in A and B, but only for a sample of cases.) Then the estimate of relative bias from the first method is:

$$\text{rel}\hat{\text{b}}\text{ias}_1(\bar{y}_{\text{trad}}) = \frac{\bar{y}_B - \bar{y}_{AB}}{\bar{y}_{AB}} \tag{2}$$

We use relative bias rather than absolute so that we can compare across variables in the results section.

This approach assumes that the first listing has no undercoverage, that is, that any units in the second listing but not the first (set C) are nonresidential, outside the segment, or otherwise not eligible for the NSFG interview. This assumption is likely to be too strong, but we do note that the first listing was carefully reviewed by the NSFG staff, while the second was not. The first listing was also done via dependent listing in 39 of the 46 segments. While dependent listing is not without its own problems, in NSFG it is based on the Postal Service's address database which often contains the hard-to-see units that listers can miss (O'Muircheartaigh et al., 2002).

In the second approach to estimating bias, we assume that both listings suffer from undercoverage and that they undercover the same types of housing units. That is, set C does contain proper housing units, and the units in C are just like those in A. (Analysis of the full set of cases listed in the first and second listings, for the few available housing unit characteristics, supports this assumption.) We can then estimate $\bar{Y}_{ABC}$, the mean on all cases in both the listings, with $\bar{y}_{ABA}$, double weighting the cases in A to make up for the missing data about the cases in C. We can also estimate $\bar{Y}_{BC}$, the mean on all cases included in the second listing, with $\bar{y}_{BA}$. Then the estimate of relative bias from the second method is:

$$\text{rel}\hat{\text{b}}\text{ias}_2(\bar{y}_{\text{trad}}) = \frac{\bar{y}_{BA} - \bar{y}_{ABA}}{\bar{y}_{ABA}} \tag{3}$$

The assumption in method 2 that the two listings undercover the same type of units is probably overly conservative, just as the assumption in method 1 that the traditional listing had no undercoverage was overly liberal. Together these two methods should bound the true relative bias that we cannot estimate directly.

All estimates of bias use the 678 completed interviews in our selected segments. Because not all segments contained a completed interview, the total number of segments in these analyses is 44.

Testing the significance of the relative bias estimates is not straightforward, as relative bias is a ratio of two estimated quantities. Instead, we test whether the bias itself, the numerators in Equations 2 and 3, are significantly different from zero, using regression models. The dependent variable in each regression is one of the 30 variables for which we have estimated bias. The sole predictor variable in these models is the undercoverage indicator appropriate for the method. Where the estimated coefficient on the indicator variable is significantly different than zero in a standard t-test, we conclude that the bias due to undercoverage in traditional listing is significant. Because these tests for undercoverage bias are exploratory, we do not adjust for multiple comparisons (Bender and Lange, 2001).

Estimation of all coverage rates, the logistic regression model, biases and the associated regressions testing for significance in the bias estimates are weighted to account for all stages of selection into the NSFG sample and for the selection of segments into our listing study. The weights do not permit inference to the national population, due to the missing contributions from the three unmatched segments. However, the weights do adjust for the diverse selection probabilities in NSFG and in our study. At the segment-level, the weights are not equal due to the probability proportional to size selection of the first and second stage clusters: rural segments with low populations are selected with very low probabilities, and thus have large weights, and vice versa. Housing units were also not selected with equal probability in the NSFG sampling scheme (Lepkowski et al., 2010).

NSFG suffers from nonresponse at both the screener and main interview stages. The overall response rate in Cycle 7 was 77% (Martinez et al., 2012). We do not adjust for nonresponse in any of our analyses. The first two analyses, estimation of the undercoverage rate and the listing propensity model, are not affected by nonresponse because the case bases we use include both responding and nonresponding housing units. Nonresponse could affect our third analysis, the estimates of undercoverage bias, but only if the nonrespondents in sets A, B and C of the Venn diagram (Figure 1) are different than the respondents in the same set. That is, we worry about nonresponse bias in our coverage estimates only if we believe there is an interaction between the mechanisms of undercoverage and nonresponse. There are two ways that such an interaction could arise. First, listers who know they will later be interviewing a sample of the cases that they list might purposefully undercover cases that look like they will be nonrespondents, in an attempt to keep their response rates high: such a connection has been demonstrated in a few studies (see, for example, Tourangeau et al., 2012). However, this phenomenon is not likely in this study because the listers who did the second listing knew that they would not be interviewing in those areas. Second, there could be a joint mechanism that makes some households less likely to be listed and also less likely to respond: our study cannot detect such a relationship.

Our bias analyses do not use the NSFG nonresponse adjusted weights because the adjustment includes a post-stratification step. Post-stratification is intended to reduce both nonresponse and coverage error, and thus analyses which used these weights would not capture the coverage bias we are interested in.

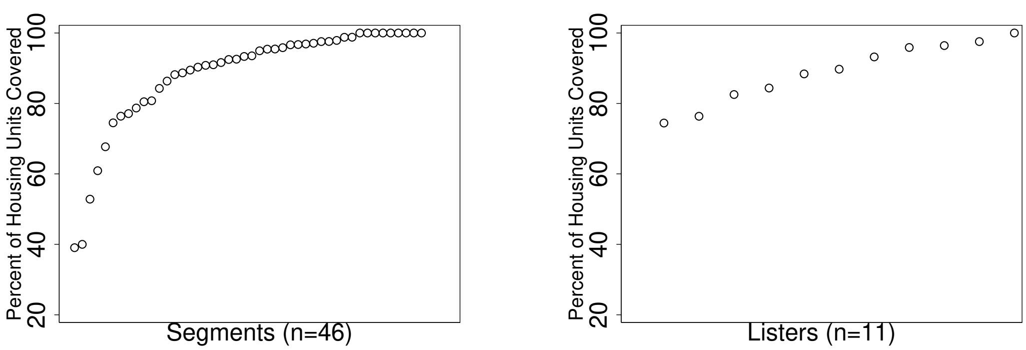
All standard error estimates and significance tests given below adjust for the clustering of the cases into segments, because the units in each segment are similar (Kreuter and Valliant, 2007; Heeringa et al., 2010). All of our analyses use the `svy` commands in Stata 12 to implement the Taylor series linear approximation variance estimation method.

5 Results

The two matched listings, the listing propensity model, and the bias calculations allow us to address the three research questions set out in the introduction. We first report the undercoverage rate for the traditionally listed housing unit frame and then explore the housing unit, area and lister characteristics that correlate with undercoverage, to understand the mechanisms of undercoverage. Finally we give estimates of coverage bias in 30 means calculated from the NSFG survey data.

Undercoverage Rates

Overall, the (weighted) coverage rate of the second listing among all cases included on the first listing is 86.4% (standard error 4.5 percentage points), an undercoverage rate of 13.6%. The top panel of Figure 2 shows the diversity of coverage rates across the 46 segments, from 39.1% to 100%. There are nine segments with coverage rates equal to 100%: all housing units in those segments were listed by the second listers, using traditional listing. The bottom panel of Figure 2 shows that the variation across the eleven listers is much smaller. All listers covered 74% or more of the housing units we examined, and only one lister had a coverage rate of 100%. In the second case base, those which were selected and were proper listings, the coverage rate is 91.3% (standard error 2.8 percentage points). Although our sample of segments is not quite representative, as described above, these two estimates of undercoverage are strong evidence that undercoverage occurs in traditional listing.



(a) Coverage Rate by Segments (b) Coverage Rate by Listers

Figure 2: Coverage Rates, by Segments and Listers, in Percent

Models of Listing Propensity

While the overall coverage rate for traditional listing is high, some housing units and segments are particularly vulnerable to undercoverage. The logistic model of listing propensity described above helps us better understand the undercoverage we see in

Figure 2 and test hypotheses about the mechanisms of lister error. The dependent variable is the logit of the probability that the second lister included a unit, and the independent variables are housing unit, segment and lister characteristics. The case base is the 1,970 housing units listed by the first lister and flagged as appropriate listings during data collection. Table 2 contains estimated beta coefficients and standard errors.

Table 2: Traditional Listing Coverage Propensity Models

	Range	Mean	$\hat{\beta}$	(SE)
Housing Unit Characteristics				
Single Family	0,1	75.9%	reference	
Unit in small multi-unit building 1-9 units	0,1	11.6%	-2.315*	(0.592)
Unit in large multi-unit building > 10 units	0,1	12.5%	0.00160	(0.755)
Unit is mobile home	0,1	2.4%	-2.342*	(0.942)
Unit was vacant at time of screening	0,1	10.2%	-1.121*	(0.485)
Segment & Lister Characteristics				
Pct. of HHs in block group in rural blocks^{a}	0 – 1	7.8%^{b}	-2.170*	(0.917)
Lister drove herself, in urban segment	0,1	52.2%^{b}	0.323	(0.602)
Lister reported safety concerns	0,1	13.0%^{b}	-0.822	(0.567)
Lister speaks language of residents	0,1	82.6%^{b}	0.00946	(0.622)
Lister reported access issues	0,1	19.6%^{b}	-0.340	(0.450)
Pct. of HHs with income less than \$25,000^{a}	0.01 – 0.70	27.9%^{b}	0.548	(1.412)
Years of interviewing experience (log)	1.10 – 2.48	1.77^{c}	-0.296	(0.827)
Map missing interior streets	0,1	26.0%^{b}	0.696	(0.456)
Map had other errors	0,1	17.3%^{b}	-2.403*	(0.603)
Block has non-visible boundary	0,1	37.0%^{b}	-1.012^{+}	(0.574)
N			1970	
Goodness of Fit F statistic (9,37)			72088.3*	
Area Under the ROC curve			0.818	

Model also contains a constant term which is not displayed

$^{+}$ $p < 0.10$, * $p < 0.05$

^{a} Source: Census 2000

^{b} Mean reported at segment level, n=46

^{c} Mean reported at lister level, n=11

As expected from previous studies, units in small multi-unit buildings are less likely to be listed than single family homes ($\hat{\beta} = -2.3, p < 0.01$) and units in large multi-unit buildings ($p < 0.01$). Also in line with previous studies are the findings of lower listing propensities for mobile homes ($\hat{\beta} = -2.3, p < 0.05$) and vacant units ($\hat{\beta} = -1.1, p < 0.05$).[11] The more rural an area is, the lower the listing propensity of

[11] Vacancy was determined by the NSFG listers at the time of data collections, which occurred during the same three month period as the traditional listing we investigate in the model.

the housing units ($\hat{\beta} = -2.2, p < 0.05$), which is also consistent with other research.

To capture the relationship between driving and listing propensity, we include in the model an indicator variable for the 24 urban segments where the lister drove while she listed. Although we had expected that driving in these segments would be associated with undercoverage, we see no significant effect. However, driving was not randomly assigned in this study. It could be that listers chose to drive while listing only in those areas where the listing task was least challenging, for example, in segments with few hidden units. Safety concerns, speaking the language of the residents and access concerns, such as gated communities or doorman-controlled buildings, also have no significant relationship with listing propensity. These findings are in contrast to the ideas we developed above that listers would make errors of undercoverage when faced with challenging areas.

The coefficients on two of the three variables coded from the maps are significant. While missing streets on the listing maps do not seem to be related to undercoverage, other map errors, such as incorrectly labeled streets and segments with the wrong shape, are associated with lower listing propensities for the units in those segments ($\hat{\beta} = -2.4, p < 0.01$). Units in segments with nonvisible block boundaries are also vulnerable to undercoverage ($\hat{\beta} = -1.0, p < 0.09$).

At the bottom of Table 2, the goodness-of-fit test shows that this model fits the data well. The area under the ROC curve measure also indicates a very good fit (0.818): Hosmer and Lemeshow (2000, p. 162) state that a value between 80% and 90% from this test indicates excellent discrimination.

The findings are largely consistent with previous work: lower listing rates for units in small multi-unit buildings, mobile homes and vacant units, and those in rural areas. The model does not support our hypotheses that listers take shortcuts which lead to undercoverage when faced with dangerous areas or access barriers. The model does however indicate that the maps we provide listers with are not of high enough quality to allow them to do their jobs well. Nonvisible boundaries also present a challenge to coverage.

Coverage Bias

The results in the previous two sections consider the patterns of undercoverage. However, the more important question in terms of overall data quality in NSFG, and the many other studies which use traditional housing unit listing, is bias due to undercoverage. If the undercovered units are no different from the covered units, in terms of the variables measured by NSFG, then we need not worry about the effect of thirteen percent undercoverage in traditionally listed frames on survey estimates. We test for bias in NSFG variables using the two methods outlined above (see Equations 2 and 3).

Figure 3 presents estimates of relative bias in the means of three sets of variables:

ten about reproductive behavior, a core module in the NSFG questionnaire; ten items on health and sexual behavior; and ten demographic and financial items. More than 350 of our 678 cases responded to each of these questions. (Not all respondents answer all questions due to questionnaire skip patterns and item nonresponse.) For each variable, the figure contains two points: the circle is the relative bias under the first approach, and the triangle the relative bias under the second approach. In every case, the estimates using Method 2 are slightly smaller than those from Method 1 (that is, attenuated towards zero). Table 3 gives the relative bias estimates alongside the p-values from the significance tests of the regression coefficients (see the Methods section for an explanation of these tests). Most of the discussion below focuses on the Method 1 results.

The first section of the figure contains estimates of 10 items on the NSFG survey related to family formation and fertility. In the first row, we see that the mean number of induced abortions captured in the survey would be 9.3% too low, relative to the mean from all completed cases in these segments, if we used only the traditional listing to make our estimates. That is, we would have missed more than nine percent of all abortions if we had done only the traditional listing. People living in the housing units missed by the traditional listers on average have more induced abortions than those living in the units covered by the traditional listers. In the fifth row of this section of the table, we see that the estimate of the proportion of people who have used birth control would be 3.5% too high if the housing units uncovered by the second listing were not included in the responding sample. This result means that the people living in housing units that were covered in the first listing but missed by the traditional listers are less likely to use birth control than those who live in units covered by both listings: the traditional listing disproportionately undercovers households that do not use birth control. Similarly, other measures of family size and growth would be too high if we used only cases in the traditional listing to make estimates. The proportion ever using birth control pills would be 4.0% too high, the number of children under 13 in the household would be 5.8% too high, and the number of live births 7.0% too high. Other related variables have positive bias as well, as can be seen in the figure. The sign of the bias that we detect in these variables is likely related to undercoverage of units in small multi-unit buildings by the traditional listers. These homes may be smaller than single family homes and the families inside of them smaller too. Thus the undercoverage we detected in the models above can lead to bias in these important NSFG variables.

The second section of the figure shows the bias in ten health and sexual behavior items. These estimates are all close to zero, indicating only small differences between the cases covered via the two listing methods. The third section of the graph in Figure 3 contains estimates of relative bias in ten demographic and financial items on the NSFG questionnaire. The largest relative bias (in absolute value) is in this section. The estimate of the percent Jewish would be 22.8% too low relative to the full sample of completed cases. This estimate would be smaller by one-fifth if the cases missed by the traditional lister were dropped from the dataset. It is not clear why this variable is so strongly affected by the undercoverage in traditional listing.

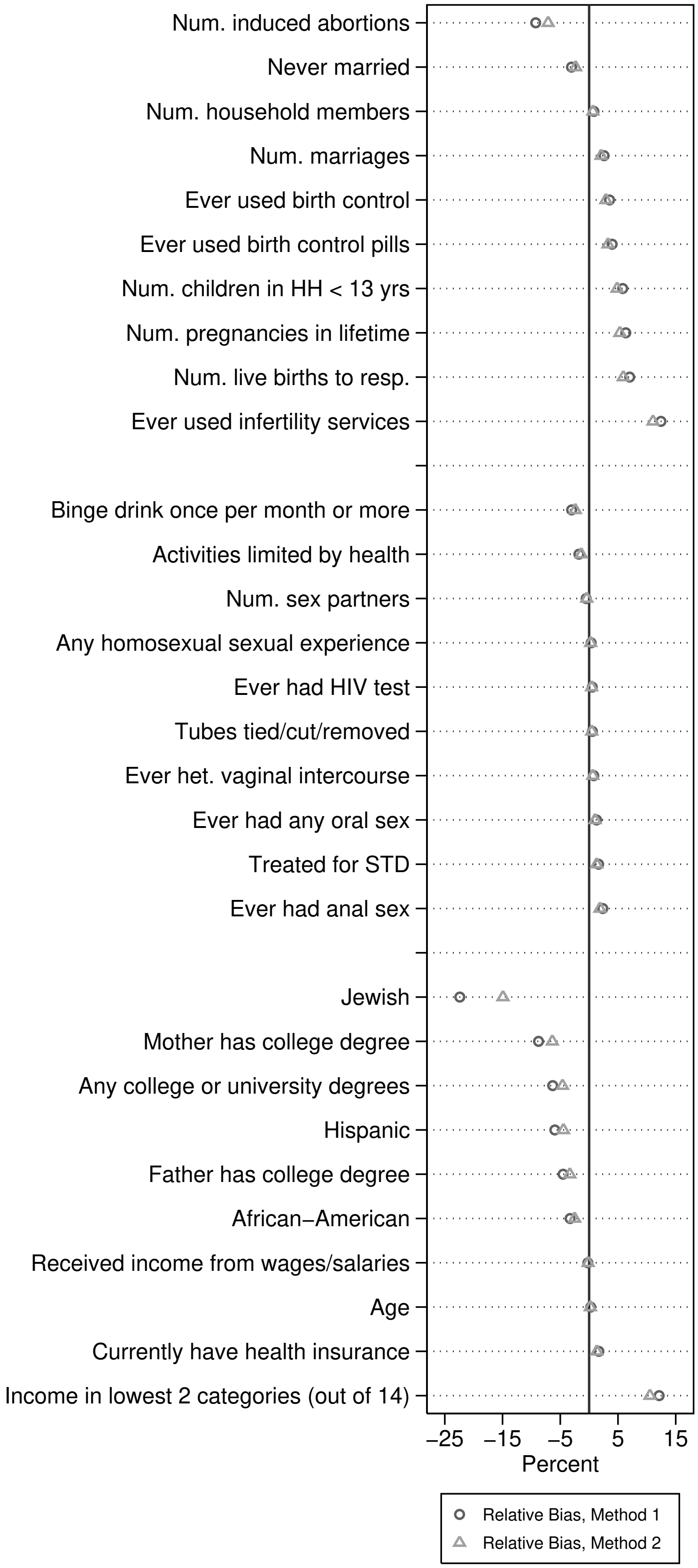


Figure 3: Estimates of Relative Bias in Selected Variables from Two Methods, by Topic (from top to bottom: Family and Fertility, Health and Sexuality, Demographic and Financial)

Table 3: Relative Bias Estimates from Two Methods

			Method 1		Method 2	
Category	Variable	N	Rel. Bias[a]	p-value[b]	Rel. Bias[a]	p-value[b]
Family/Fertility						
	Num. induced abortions	351	-9.25	0.168	-7.09	0.171
	Never married	678	-3.05	0.114	-2.33	0.105
	Num. household members	678	.836	0.188	.658	0.177
	Num. marriages	678	2.6	0.144	2.08	0.141
	Ever used birth control	381	3.55	0.021	2.9	0.011
	Ever used birth control pills	381	3.99	0.012	3.28	0.007
	Num. children in HH < 13 yrs	381	5.82	0.128	4.85	0.117
	Num. pregnancies in lifetime	381	6.36	0.052	5.33	0.038
	Num. live births to respondent	381	7.03	0.045	5.92	0.034
	Ever used infertility services	585	12.4	0.000	11.1	0.000
Health/Sexuality						
	Binge drink once per month or more	515	-3.03	0.193	-2.39	0.191
	Activities limited by health	678	-1.76	0.448	-1.36	0.448
	Num. sex partners	538	-.51	0.442	-.401	0.442
	Any homosexual sexual experience	676	.352	0.481	.277	0.481
	Ever had HIV test	678	.55	0.297	.432	0.293
	Tubes tied/cut/removed	381	.565	0.476	.452	0.476
	Ever het. vaginal intercourse	677	.814	0.182	.641	0.168
	Ever had any oral sex	678	1.3	0.172	1.02	0.161
	Treated for STD	677	1.64	0.407	1.3	0.407
	Ever had anal sex	674	2.37	0.235	1.89	0.228
Demographic/Financial						
	Jewish	678	-22.4	0.004	-14.9	0.001
	Mother has college degree	668	-8.75	0.043	-6.38	0.030
	Any college or university degrees	366	-6.31	0.014	-4.6	0.008
	Hispanic	678	-5.94	0.202	-4.44	0.211
	Father has college degree	604	-4.5	0.149	-3.34	0.136
	African-American	650	-3.29	0.148	-2.5	0.151
	Received income from wages/salaries	665	-.216	0.429	-.169	0.428
	Age	678	.273	0.269	.214	0.271
	Currently have health insurance	678	1.69	0.221	1.34	0.211
	Income in lowest 2 categories (out of 14)	595	12.1	0.001	10.5	0.000

[a] in percent

[b] p-values do not belong to the rel. bias estimates but to absolute bias estimates, as explained in Methods section

Method 1 assumes no undercoverage in the first listing

Method 2 assumes undercovered units in the first listing are just like those undercovered in the second

The last row shows a 12.1% overestimate of the number of households in the lowest two income categories. The traditional listing contains too many people in these low income categories when compared to the full sample of completed cases. Listers using traditional listing tend to miss the units containing families in the higher income categories, which may be related to the undercoverage of rural areas.

Several of these bias estimates are significantly different from zero. In Method 1, the three largest effects (in absolute value) are significant at the one percent level: ever used infertility services, income in the lowest two categories, and percent Jewish. Additionally, ever used birth control, ever used birth control pills, number of live births, mother has college degree, and respondent has college degree are significant at the five percent level and number of lifetime pregnancies is significant at the ten percent level. In Method 2, the same variables show significant relative biases, though the p-values are generally smaller (see Table 3). The agreement between our two methods of calculating bias is a sign that our estimates are robust.

6 Discussion

This study has demonstrated that traditional listing suffers from substantial undercoverage. More than 13% of all housing units that should have been captured by traditional listing were undercovered in this study. This undercoverage appears to be related to several housing unit and segment characteristics. Vacant homes, mobile homes, units in small multi-unit buildings, and those in segments with non-visible boundaries and in rural areas are undercovered. Map errors also negatively impact coverage. The patterns of undercoverage that we find in this study are correlated with variables collected on the survey, particularly those related to pregnancies, births, contraception use, household income and education, among the most important variables collected on the NSFG survey. Hence, undercoverage should be a concern of researchers using traditional listing methods to create housing unit frames. The techniques used in this investigation could be implemented in additional studies, to explore whether the findings presented here generalize to other topics, and expand our knowledge of undercoverage further.

We recognize the shortcomings of this study. Due to matching difficulties, the estimates presented above are not nationally representative: they exclude the most rural areas of the U.S. We have also not considered inter-lister variability in the listing process; see Eckman (2013) for a discussion. In addition, we calculate bias under two different assumptions: the first that the initial listing contained no undercoverage; the second that the undercovered units in the first listing are just like those undercovered in the traditional listing, in terms of the variables collected in NSFG. Neither of these assumptions is true in practice, but together they likely bound the true estimate. It is also possible that there are other valid housing units in these segments that were undercovered in both listings: this assertion was not tested here.[12]

[12] A third listing of these segments, done as part of the same study but not used in this paper,

Although we are not able to estimate the ideal bias and must make additional assumptions, our estimates of coverage bias are the first that we are aware of to look specifically at the contribution of housing unit listing to coverage bias.

In actual practice using the NSFG data, the effects of coverage bias are likely to be smaller than we have estimated here, for several reasons. First, NSFG did not use traditional listing, except when the vendor was able to supply only a few addresses in a segment. Second, the released Cycle 7 NSFG data contains weights which are poststratified to age, sex, race and ethnicity totals for the U.S. population (Lepkowski et al., 2010). These are not the variables found in the propensity models or in the bias calculations to be correlated with coverage propensity, but it is possible that these adjustments remove some of the coverage bias we find from actual estimates calculated with NSFG data. Researchers who are particularly concerned about the effects of housing unit undercoverage on their estimates, for example, those who wish to analyze the fertility behavior of the U.S. Jewish population, may wish to perform additional poststratification adjustments by variables related to housing unit coverage, such as building size and urbanicity.

Some of the problems that we find in traditional listing could perhaps be corrected. In preparation for the 2010 Census, the Census Bureau improved its map data, which forms the basis for the commercial map data used by many survey organizations to produce listing maps (Zandbergen, 2009). Thus we hope some of the map errors we detected will not be a problem in the future. In the meantime, those who manage listers may wish to compare the listing maps with online resources, as we did, and discuss with the lister how to handle any discrepancies. To address the non-visible block boundaries, we suggest that blocks should be combined in ways that eliminate such boundaries whenever possible (O'Muircheartaigh and English, 2011).

To improve the coverage of units in small multi-unit buildings, we can imagine a new version of the missed housing unit procedure conducted by interviewers as they are out collecting data (Kish, 1965, 341-342). When a unit in a small multi-unit building is selected, interviewers often gain access to the building or speak with a resident, and can thus verify the number of units in the building. When the interviewer finds more units than the lister included, the additional units could be added to the frame and given a chance of selection. This technique should reduce undercoverage in multi-unit buildings, but see Eckman and O'Muircheartaigh (2011) for evidence that interviewers do not carry out a similar missed housing unit procedure successfully.

Traditional listing is often used when dependent listing is not possible due to the absence of a prior frame to serve as the input list. NSFG Cycle 7 used traditional listing when its vendor is not able to provide at least six addresses in a segment, and the Current Population Survey's four-frame design used a similar criterion (U.S. Census Bureau, 2006; Lepkowski et al., 2010). In practice such a decision rule often means

suggests that there are additional housing units in these segments that were not captured in either listing.

using traditional listing in rural areas where city-style addresses are not available from the postal delivery database (Eckman and English, 2012). But as we have seen, it is in just these rural areas that traditional listing performs poorly. A better understanding of the unique challenges of rural listing is needed to avoid undercoverage in these areas.

Appendix: Matching Addresses in Two Listings

The analyses presented in this paper required matching the two frames of listed housing units. The quality of this matching work greatly affects the quality of the results; both false matches and false non-matches could cause errors in our analyses. This appendix explains the matching procedures in detail. All matching was done only within segments. Our goal was to match addresses that would lead to the same housing unit being selected. However, we did not allow any many-to-one matches: each address on a frame could have one and only one match on another frame.

Although there have been substantial advancements recently in probabilistic matching algorithms (Herzog et al., 2007), these sophisticated techniques are not needed here. The segments in this study are quite small: the average number of housing units per segment is less than 200. Thus manual review of all listed units within each segment was possible. Additionally, the probabilistic matching routines are very good at resolving spelling errors in street names, but the listing software used in this study provides a drop-down menu of street names, which minimizes spelling differences.

The software collects addresses in three parts: house number, street and apartment number. We used a SAS program to parse the street variable into four fields:

- Pre-direction: Street direction, when it precedes the street name (e.g. `N`, `E`, `NW`)
- Street Name: e.g. `Main`, `37th`, `Martin Luther King`
- Street Type: e.g. `Ave`, `St`, `Dr`, `Circle`
- Post-direction: Street direction, when it follows the street name

The parser routine also standardizes the address parts to improve matching (for example, `Drive` becomes `Dr`, `North` becomes `N`).

The first listing contained 9,059 listed units and the second 8,988. We first used SAS programs to identify perfect matches across all of the address variables in both frames. This first matching pass found 6,516 pairs.

Subsequent passes through the unmatched addresses relaxed the matching criteria. For example, the second pass would match `1495 Beard Ave`[13] to `1495 S Beard Ave`. The final automatic matching pass would match `659 Wayne Dr` to `659 Wayne Rd`. Note that only cases which did not match in earlier passes went on to the later, more permissive, passes. That is, only if both `659 Wayne Rd` and `659 Wayne Dr` had no better partners would they be matched together. We carefully reviewed all matches from the later passes to ensure that they seemed reasonable. Table 4 shows

[13] All example addresses are fictitious.

the matching criteria at each pass as well as the number of matches found. All passes required matching house numbers, street names and apartment numbers.[14] Altogether, these SAS routines identified 6,874 matching pairs of addresses.

Table 4: Automatic Matches found, by Pass

Field	Pass 1	Pass 2	Pass 3	Pass 4	Manual Review
Segment	X	X	X	X	
House number	X	X	X	X	
Pre-direction	X				
Street Name	X	X	X	X	
Street Type	X	X	X		
Post-direction	X	X			
Apartment	X	X	X	X	
Matches Found	6,516	19	151	188	1,135

In the final step, we reviewed the addresses and descriptions of all remaining unmatched addresses.[15] We created spreadsheets of all addresses in each segment, both matched and unmatched, sorted in street, house number and apartment order. This step identified matches of five types:

Inconsistent Apartment Numbers Listers sometimes used different designators to refer to apartments in a building: for example `A, B, C; 1, 2, 3; Front, Rear; 1st floor, 2nd floor, 3rd floor; 101, 102, 103`. When the two frames agreed on the number of units in a building, we matched all the units in the order implied by the unit designators. In those cases where one list contained more units at an address than the other list, we left some units unmatched.

Sometimes one lister thought a structure was a single family home and the other saw more than one unit at the address. Because interviewers are trained to approach the first unit when a selected single-family case turns out to be a multi-unit building, we matched the single family unit to the first unit and left the other units unmatched.

Different Street Names Some roads have two different names, such as Main St and State Road 95. When the two frames had the same house numbers, but different street names, and when online research indicated the two streets likely coincided, we matched these addresses. In a few other situations, one of the listers edited the street name to correct a misspelling.

[14]Units where the house number was missing or was some variant of "No #" were matched via manual matching.

[15]Listers are instructed to provide descriptions of the housing units whenever the address does not include a house number or when it might be unclear which unit is meant.

Different House Numbers When two units appeared to be the same except for small differences in the house number, and there were no or very few other unmatched units with the same street name, we matched these units. In these situations, we felt that the interviewer would probably come to the same conclusion in the field if the case were selected. For example, if an interviewer cannot find `12195 Willow Rd`, but does see `12159 Willow Rd`, she is likely to interview at the second address.

No House Number When a house number is not available, listers are taught to write "No #" in the house number field and include a description that uniquely identifies the housing unit. We were able to match these units when they were the only unmatched cases on a street or when the descriptions made it clear that they referred to the same unit.

Manual matching led to 1,135 additional matches, 60% due to inconsistent apartment numbers.

The total number of matches between the first and second listed frames across all passes is 8,009. That is, 88.4% (unweighted) of the housing units listed by the first lister in this study also appeared in the second listing, and 89.1% of those in the second appeared in the first. We carefully reviewed all matches. While we recognize that there will likely be some false matches and false nonmatches, we have taken all possible steps to minimize them.

References

Axinn, W. G., C. F. Link, and R. M. Groves (2011). Responsive Survey Design, Demographic Data Collection, and Models of Demographic Behavior. *Demography 48*(3), 1127–1149.

Barrett, D. F., M. Beaghen, D. Smith, and J. Burcham (2002). Census 2000 Housing Unit Coverage Study. In *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 146–151.

Bender, R. and S. Lange (2001). Adjusting for Multiple Testing – When and How? *Journal of Clinical Epidemiology 54*, 343–349.

Botman, S. L. (1987). Coverage Implications of a Poststratification Adjustment on Estimates from a Household Survey. In *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 395–400.

Boyd, H. W. and R. Westfall (1955). Interviewers as a Source of Error in Surveys. *Journal of Marketing 19*(4), 311–324.

Casas-Cordero, C., F. Kreuter, Y. Wang, and S. Babey (2013). Assessing the Measurement Error Properties of Interviewer Observations of Neighbourhood Characteristics. *Journal of the Royal Statistical Society, Series A 176*(1), 227–249.

Central Co-ordinating Team (2010). European Social Survey Round 4 2008/2009. Final Activity Report ESS4e03.0, City University London.

Childers, D. R. (1992). The 1990 Housing Unit Coverage Study. In *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 506–511.

Childers, D. R. (1993). Coverage of Housing in the 1990 Decennial Census. In *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 635–640.

Chromy, J. R., K. R. Bowman, C. J. Crump, L. E. Packer, and M. A. Penne (1999). Population Coverage in the National Household Survey on Drug Abuse. In *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 576–580.

Clogg, C., M. Massagli, and S. Eliason (1989). Population Undercount and Social Science Research. *Social Indicators Research 21*(6), 559–598.

Coleman, J. S. (1994). A Rational Choice Perspective on Economic Sociology. In N. J. Smelser and R. Swedberg (Eds.), *The Handbook of Economic Sociology*, pp. 166–180. Russel Sage Foundation.

Cook, P. (1985). The Case of the Missing Victims: Gunshot Woundings in the National Crime Survey. *Journal of Quantitative Criminology 1*(1), 91–102.

Couper, M. P., R. Tourangeau, and T. Marvin (2009). Taking the Audio Out of Audio-CASI. *Public Opinion Quarterly 73*(2), 281–303.

Eckman, S. (2013). Do Different Listers Make the Same Housing Unit Frame? Variability in Housing Unit Listing. *Journal of Official Statistics 29*(2), 249–259.

Eckman, S. and N. English (2012). Creating Housing Unit Frames from Address Databases: Geocoding Precision and Net Coverage Rates. *Field Methods 24*(4), 399–408.

Eckman, S. and F. Kreuter (2011). Confirmation Bias in Housing Unit Listing. *Public Opinion Quarterly 75*(2), 139–150.

Eckman, S. and C. O'Muircheartaigh (2011). Performance of the Half–Open Interval Missed Housing Unit Procedure. *Survey Research Methods 5*(3), 125–131.

Eggers, F. J. and F. Moumen (2011). American Housing Survey Components of Inventory Change: 2007-2009. Technical report, U.S. Department of Housing & Urban Development Office of Policy Development & Research.

Fay, R. E. (1989). An Analysis of Within Household Undercoverage in the Current Population Survey. In *Proceedings of the Bureau of the Census Annual Research Conference*, pp. 156–175.

Fein, D. J. (1990). Racial and Ethnic Differences in U.S. Census Omission Rates. *Demography 27*(2), 285–302.

Fink, R. (1963). Interviewer Training and Supervision in a Survey of Laos. *International Social Science Journal 15*(1), 21–34.

Groves, Robert M., M. W. D., J. Lepkowski, and N. G. Kirgis (2009). Planning and Development of the Continuous National Survey of Family Growth. *Vital Health Statistics 1*(48).

Groves, R. M. (2006). Nonresponse Rates and Nonresponse Bias in Household Surveys. *Public Opinion Quarterly 70*(5), 646–675.

Groves, R. M. and S. G. Heeringa (2006). Responsive Design for Household Surveys: Tools for Actively Controlling Survey Errors and Costs. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 169*(3), 439–457.

Hainer, P. (1987). A Brief and Qualitative Anthropological Study Exploring the Reasons for Census Coverage Error Among Low Income Black Households. In *Report Prepared under Contract with the Census Bureau*.

Hainer, P., C. Hines, E. Martin, and G. M. Shapiro (1988). Research on Improving Coverage in Household Surveys. In *Proceedings of the Fourth Annual Census Bureau Research Conference*, pp. 513–539. U.S. Census Bureau.

Harter, R., S. Eckman, N. English, and C. O'Muircheartaigh (2010). Applied Sampling for Large-Scale Multi-Stage Area Probability Designs. In P. V. Marsden and J. Wright (Eds.), *Handbook of Survey Research* (Second ed.)., pp. 169–197. Emerald.

Hawkes, W. (1986). Census Data Quality: A User's View. *Journal of Official Statistics 2*(4), 531–544.

Heeringa, S. G., B. T. West, and P. A. Berglund (2010). *Applied Survey Data Analysis*. Chapman & Hall/CRC.

Herzog, T. N., F. J. Scheuren, and W. E. Winkler (2007). *Data Quality and Record Linkage Techniques*. New York: Springer.

Horrigan, M., W. Moore, S. Pedlow, and K. Wolter (1999). Undercoverage in a Large National Screening Survey for Youths. In *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 570–575.

Hosmer and Lemeshow (2000). *Applied Logistic Regression* (2nd ed.). New York: Wiley-Interscience.

Jowell, R. and the Central Co-ordinating Team (2003). European Social Survey 2002/2003. Technical Report ESS1e06.1, City University.

Jowell, R. and the Central Co-ordinating Team (2005). European Social Survey Round 2 2004/2005. Technical Report ESS2e03.1, City University.

Jowell, R. and the Central Co-ordinating Team (2007). European Social Survey Round 3 2006/2007. Technical Report ESS3e03.2, City University.

Joyner, K., H. Peters, K. Hynes, A. Sikora, J. Taber, and M. Rendall (2012). The Quality of Male Fertility Data in Major U.S. Surveys. *Demography 49*, 101–124.

Judkins, D., R. DiGaetano, A. Chu, and G. Shapiro (1999). Coverage in Screening Surveys at Westat. In *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 581–586.

Kennickell, A. B. (2000). Asymmetric Information, Interviewer Behavior, and Unit Nonresponse. In *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 238–243.

Kennickell, A. B. (2003). Reordering the Darkness: Application of Effort and Unit Nonresponse in the Survey of Consumer Finances. In *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 2119–2126.

Kish, L. (1965). *Survey Sampling*. John Wiley and Sons.

Kish, L. and I. Hess (1958). On Noncoverage of Sample Dwellings. *Journal of the American Statistical Association 53*(282), 509–524.

Kreuter, F. and R. Valliant (2007). A Survey on Survey Statistics: What is Done and Can be Done in Stata. *Stata Journal 7*(1), 1–21.

Lepkowski, J. M., W. D. Mosher, K. Davis, R. M. Groves, and J. V. Hoewyk (2010). The 2006–2010 National Survey of Family Growth: Sample Design and Analysis of a Continuous Survey. *Vital Health Statistics 2*(150).

Manheimer, D. and H. Hyman (1949). Interviewer Performance in Area Sampling. *Public Opinion Quarterly 13*(1), 83–92.

Martin, E. (1981). A Twist on the Heisenberg Principle: Or, How Crime Affects its Measurement. *Social Indicators Research 9*(2), 197–223.

Martin, E. (2007). Strength of Attachment: Survey Coverage of People with Tenuous Ties to Residences. *Demography 44*, 427–440.

Martinez, G., K. Daniels, and A. Chandra (2012). Fertility of Men and Woman Aged 15-44 Years in the United States: National Survey of Family Growth, 2006–2010. *Vital Health Statistics 1*(51).

Meier, F. and T. Moore (1999). The Effect of Screening on Coverage in the National Health Interview Study. In *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 566–569.

Montaquila, J. M., L. Mohadjer, J. Waksberg, and M. Khare (1996). A Detailed Look at Coverage in the Third National Health and Nutrition Examination Survey. In *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 532–537.

Morton, K. B., S. R. Hunter, J. R. Chromy, and P. C. Martin (2006). Population Coverage in the National Survey of Drug Use and Health. In *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 3441–3446.

Morton, K. B., P. C. Martin, J. R. Chromy, M. Foster, and E. L. Hirsch (2010). 2009 National Survey on Drug Use and Health Sample Design Report. Technical report, Substance Abuse and Mental Health Services Administration, Rockville, Maryland.

O'Muircheartaigh, C. A., S. A. Eckman, and C. Weiss (2002). Traditional and Enhanced Field Listing for Probability Sampling. In *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 2563–2567.

O'Muircheartaigh, C. A., E. M. English, and S. A. Eckman (2007). Predicting the Relative Quality of Alternative Sampling Frames. In *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 551–574.

O'Muircheartaigh, C. A., E. M. English, S. A. Eckman, H. Upchurch, E. Garcia, and J. Lepkowski (2006). Validating a Sampling Revolution: Benchmarking Address Lists against Traditional Listing. In *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 4189–4196.

O'Muircheartaigh, C. and N. English (2011). Eliminating Invisible Boundary Problems in Area Samples, with an Application to the National Children's Study (NCS). In *Presentation at the American Association for Public Opinion Research Conference*.

Pearson, J. M. (2003). Quality and Coverage of Listings for Area Sampling. In *Proceedings of the Section on Government Statistics, American Statistical Association*.

Peytchev, A., E. Peytcheva, and R. M. Groves (2010). Measurement Error, Unit Nonresponse, and Self-Reports of Abortion Experiences. *Public Opinion Quarterly 74*(2), 319–327.

Philipson, T. and T. Lawless (1997). Multiple-Output Agency Incentives in Data Production: Experimental Evidence. *European Economic Review 41*(3-5), 961 – 970.

Schnell, R. (2008). Avoiding Problems of Traditional Sampling Strategies for Household Surveys in Germany: Some New Suggestions. Technical report, Deutsches Institut fÃ¼r Wirtschaftsforschung.

Shapiro, G., G. Diffendal, and D. Cantor (1993). Survey Undercoverage: Major Causes and New Estimates of Magnitude. In *Paper delivered at the US Bureau of the Census Annual Research Conference, Washington, DC*.

Shapiro, G. M. and D. Kostanich (1988). High Response Error and Poor Coverage are Severely Hurting the Value of Household Survey Data. In *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 443–448.

Singer, E. (2011). Toward a Benefit-Cost Theory of Survey Participation: Evidence, Further Tests, and Implications. *Journal of Official Statistics 27*(2), 379–392.

Thompson, G. and C. Turmelle (2004). Classification of Address Register Coverage Rates: A Field Study. In *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 4477–4484.

Tourangeau, R., F. Kreuter, and S. Eckman (2012). Motivated Underreporting in Screening Interviews. *Public Opinion Quarterly 76*(3), 453–469.

Tourangeau, R., G. Shapiro, A. Kearney, and L. Ernst (1997). Who Lives Here? Survey Undercoverage and Household Roster Questions. *Journal of Official Statistics 13*(1), 1–18.

U. S. Census Bureau (1993). Programs to Improve Coverage in the 1990 Census. Technical report. 1990 CPH-E-3.

U.S. Census Bureau (2006). Technical Paper 66: Design and Methodology, Current Population Survey. Technical report.

Valentine, C. and B. Valentine (1971). Missing Men: A Comparative Methodological Study of Underenumeration and Related Problems. *Report prepared under contract for the US Census Bureau. Washington, DC 23*.

Zandbergen, P. A. (2009). Geocoding Quality and Implications for Spatial Analysis. *Geography Compass 3*(2), 647–680.