Confirmation Bias in Housing Unit Listing

Stephanie Eckman
Frauke Kreuter
Joint Program in Survey Methodology
University of Maryland, College Park

February 12, 2010

Abstract

Field listing of housing units is an expensive and time-consuming stage of the survey process, and its error properties are not well understood. This paper uses an experimental repeated listing design to demonstrate the presence of confirmation bias in dependent listing. We find evidence that when provided with an initial listing to update in the field, listers can become too trusting of the list and tend not to add missing units or delete inappropriate units. This finding has implications not only for surveys that use dependent listing to create housing unit frames but also for studies of coverage of housing unit frames.[1]

[1]Many thanks to James Lepkowski, Ashley Bowers, Adam Schlecte, Andrew Hupp, Kat Donahue and Matt Jans for their assistance with this experiment, and to the listers for their participation. Thanks also to Roger Tourangeau, Carolina Casas-Cordero and Michael Lemay for comments on earlier drafts, and to Austin Nichols and Ben Jann for discussions of the analysis. We also acknowledge the helpful comments of the journal reviewers and the financial support for this work from the Rensis Likert Fund at the University of Michigan and the ETH in Zürich, Switzerland.

1 Introduction

Many area-probability surveys use field listing to construct a frame of housing units from which cases are selected. Errors in these frames can lead to bias in survey data, yet we know very little about the how listers do this work. While several studies have looked at coverage rates in housing unit listing (Manheimer and Hyman, 1949; Kish and Hess, 1958; Joncas, 1985; Hawkes, 1986; Jacobs, 1986; Childers, 1992; Barrett et al., 2002; O'Muircheartaigh et al., 2003; Pearson, 2003; Thompson and Turmelle, 2004; Turmelle et al., 2005; O'Muircheartaigh et al., 2006, 2007), to our knowledge no previous study has identified a mechanism by which errors of undercoverage and overcoverage are introduced.

There are two methods of housing unit listing. In *traditional listing* – also called scratch listing – listers are provided with maps of the selected area and travel around each selected block in the segment, recording the address or description of every housing unit (Kish, 1965; Survey Research Center, 1969, 1976). In *dependent listing* – also called update or enhanced listing – listers are provided with addresses believed to lie inside the selected area. Call these addresses the *input listing*: these might be from a previous listing of the area or from a geocoded address database. Listers travel around the segment and make corrections to the list to match what they find in the field. Dependent listing can be less expensive than traditional listing (O'Muircheartaigh et al., 2003), and some survey organizations, for example the Census Bureau and the Survey Research Center at the University of Michigan, use dependent listing regularly.

This paper examines a mechanism of error in dependent listing which we call *confirmation bias*: the tendency for listers to preserve the errors in the input listing. When listers are provided with an input list to update in the field, they may be likely to confirm that the input listing is correct. When a lister cannot locate a housing unit which is on the input listing, she may presume it is her own error rather than an error in the input list and choose not delete it, a *failure-to-delete* error. Conversely, a lister may not add units to the frame that are missing from the input list even though they exist in the field, a *failure-to-add* error. In this way errors in the initial listing, whether it comes from a commercial address database or a previous listing of the area, can be perpetuated on the final frame.

Small multi-unit buildings pose particular problems because listers have trouble determining how many units the building contains (Chakrabarty and Torres, 1992; Subcomittee on Survey Coverage, 1990). Buildings with many units tend to have just one entrance with a list of units on the mailboxes or near a phone. Smaller buildings often have a less organized presentation. The Census Bureau also recognizes the difficulty of listing small multi-unit buildings and has created a task force to address undercoverage and overcoverage in these buildings (2-19 units) for the 2010 Census (personal communication with Nathan Hillila, Labor and Crime Surveys Branch, U.S. Census Bureau). Given the special problems posed by small multi-unit buildings, we expect that confirmation bias will be more common with these units. For example, seeing a locked apartment building with two doors, four doorbells, five gas meters, and six mailboxes, a lister may decide that the four units included on the input listing are correct.

Confirmation bias occurs in other stages of the survey process. For example, when a second coder reviews the coding of open-ended occupation responses and can see the value assigned by the first, she is more likely to assign the same code than if she did an independent second coding. (Biemer and Lyberg, 2003). In dependent interviewing, when a second interviewer can see the response recorded by the first, the two interviews tend to capture similar responses (O'Muircheartaigh, 2004; Lynn and Sala, 2006).[2] In this study we investigate whether a similar phenomenon occurs in dependent listing.

2 Design

This study introduces known errors into the input listing and examines whether they persist after field updating. We implemented an experiment in the Practicum course of the University of Michigan's Program in Survey Methodology. Students in this course carry out all stages of a survey: they develop and pretest a questionnaire, list housing units, conduct in-person and telephone interviews, and analyze data. In the spring of 2009, they completed two listings of 14 segments, seven in Ann Arbor, MI and seven in Ypsilanti, MI.

The segments were groups of Census blocks which contained at least 30 housing units, according to 2000 Census counts. We chose segments that contained features associated with listing errors: we included one trailer park in each city[3] and purposefully selected areas containing many small multi-unit buildings. Each of the selected segments was listed twice, by two different listers. Working separately, one lister used traditional listing, and the other used dependent listing. Listers were randomly assigned to segments and to listing methods. Each lister listed one Ann Arbor segment and one Ypsilanti segment, using traditional listing in one and dependent in the other. Overall there were 13 listers for the 14 segments.[4] The listers read the standard SRC lister manual and received instruction both in class and in the field from the same staff members who train listers for large SRC surveys. The total training time was approximately three hours.

The listing maps were the standard maps used in all listing projects conducted by the Survey Research Center (SRC) at the University of Michigan; they included marks indicating where to begin listing and what direction to travel around the block. The input to the dependent listing was from a geocoded address database provided by Marketing Systems Group. The addresses appeared on the listing sheets in the order that they were expected to occur in the field.

To test the confirmation bias hypothesis, we manipulated the input listing in each of the 14 selected segments. These manipulations were not done randomly, though they were designed to cover all possible and reasonable scenarios. We deleted entire multi-unit buildings, the last unit in a multi-unit building, an entire side of the block, the only unit on a block,

[2]We use the term confirmation bias in the same sense that Biemer and Lyberg use it. Psychologists use the term in a different way: see Klayman (1995) for a summary of the uses in that field.

[3]See Childers (1992) on the difficulties of covering housing units in trailer parks in the 1990 Census.

[4]Two listers worked together to list four different segments; we were told that these students did all of their assignments together.

units flagged as vacant on the delivery file, and units with nothing remarkable about them. The added units were, for example, an even-numbered unit between two others, additional units in multi-unit buildings, units across the street from the selected block, and additional units at the end of a street. We added a total of 24 cases to the input listing, 12 in multi-unit buildings and 12 single unit buildings. We deleted 58 cases, 14 in multi-unit structures and 44 single-unit. We deleted at least one unit from each segment and added units to 12 of the 14 segments.[5] After manipulation, the input listing contained 756 addresses. The listers were not informed about the manipulation of the input listing or about the double-listing design.

3 Analysis Methods

If the dependent listers are subject to failure-to-add confirmation bias, we would expect them to tend not to add the units we deleted. To test for this tendency, we use the traditional listers as a control and calculate a difference-in-differences estimate of the effect of the treatment, the suppression of the housing units from the input listing. The difference-in-differences technique is commonly used with panel data to derive treatment effects from non-randomized designs: each case serves as its own control and the difference in the change from period 1 to period 2 between those who did and did not receive the treatment is the average treatment effect (Angrist and Pischke, 2009, pp. 221–247).

Let L_{unm}^{dep} be the fraction of unmanipulated cases on the input list that were listed by the dependent lister. L_{del}^{trad} is the fraction of cases deleted from the input list that were listed by the traditional lister. $L_{del}^{dep} - L_{del}^{trad}$ captures the difference in the listing propensities of the deleted units between those listers who were and were not subject to the manipulation, which is part of the effect we are interested in, but it does not take advantage of the experiment by comparing the manipulated to the unmanipulated units. $L_{unm}^{dep} - L_{del}^{dep}$ captures the difference in the listing likelihoods between the unmanipulated and manipulated cases, but does not account for any systematic difference between them.

The more appropriate estimator of the treatment effect is the difference-in-differences estimate, which adjusts for differences between the deleted and unmanipulated cases and takes advantage of our experiment:

$$D - in - D = (L_{unm}^{trad} - L_{del}^{trad}) - (L_{unm}^{dep} - L_{del}^{dep})$$

We use a similar difference-in-differences to estimate the extent of failure-to-delete confirmation bias.

While the difference-in-differences technique has intuitive appeal in this situation, it does not provide standard errors for the estimated effect sizes. To control for additional housing unit characteristics and to calculate standard errors, we also fit a cross-classified random

[5] In two of the segments it was not possible to add plausible units.

effects linear regression model for the failure-to-add analysis. The dataset used in these models is at the housing unit and listing level: each housing unit appears in the dataset twice, once for each of the two listings. The binary dependent variable in the model is whether a given lister listed a housing unit (1) or did not (0). The independent variables of interest are a dummy variable indicating method (traditional listing is the reference category), a dummy variable indicating whether the housing unit was deleted from the input listing, and the interaction of these two. By including the indicator of deletion at the housing unit level, we control for any unobserved attributes that make the suppressed units harder or easier to list. The cross-classified model allows us to account for the clustering of the observations into segments and listers (Raudenbush and Bryk, 2002, pp. 373–398). We fit the model in Stata 11 (StataCorp LP, 2009).

Because our dependent variable is dichotomous, a logistic model is an obvious choice for our analysis. However, it is the coefficient on the interaction term (method crossed with manipulation) that we are most interested in, and interpretation of interaction effects in nonlinear models is complex (Ai and Norton, 2003). For this reason, we use a linear probability model (as suggested by Wooldridge (2002, pp. 454–457) and Mood (2009)) to test the statistical significance of our interaction effect. In the Appendix, we report coefficient estimates from both the linear and logistic models.

Our analyses required matching the two listings of each segment. Due to the small number and size of the segments, we did the matching by inspection which allowed us to correct for spelling and typographical errors. The two listers used the same listing order which simplified the matching task. In a small number of cases we encountered difficult matching situations. The rule we applied in these cases called two listed housing units a match if they would lead to the same unit being approached for an interview. For example, if one lister recorded units A and B at an address, and another recorded units 1 and 2, we matched these two sets of cases. When one lister had two units at an address and another only one, we matched the single family home to the first unit and left the second unit unmatched because this is the instruction SRC interviewers receive when they find a selected single-family unit to be a multi-unit building. Incorrect matches and missed matches will bias our results below, but we believe them to be minimal.

4 Results

A total of 754 housing units were listed by the listers using traditional listing and 738 by those using dependent listing (see Table 1). The two frames included 787 unique housing units. Lister-pairs agreed about the inclusion of 704 units, for a disagreement rate of $(787 - 704)/787 = 11\%$. See Table 1. As expected, the disagreement rate is higher among multi-unit buildings (15%) than among single-family homes (8%).

Table 1: Counts of Listed Housing Units

	Traditional Listings	Dependent Listings	Listed by Both	Unique Listings	Disagreement Rate
Overall	754	738	704	787	11%
Ann Arbor	511	486	466	531	12%
Ypsilanti	243	252	238	256	7%
Single-Family	486	471	458	498	8%
Multi-Unit	268	267	246	289	15%

4.1 Failure-to-Add Error

If failure-to-add bias exists, units deleted from the input list should be listed less frequently by the dependent than the traditional listers. Table 2 gives listing rates for the unmanipulated and deleted cases and the two listing methods. The first row of this table refers to the addresses received from the vendor that were not deleted from the input listing. The traditional listers included 88.11% of these units on their frames and the dependent listers 89.48%. The second row refers to the 58 units deleted from the input listing. 81.03% of these units were added back by the dependent listers. A higher proportion, 93.10%, were listed by the traditional listers, who were not subject to the manipulation.[6]

Using the difference-in-differences technique, the effect of suppressing the housing units on the dependent listers is: $(88.11\% - 93.10\%) - (89.48\% - 81.03\%) = -13.4\%$. Deleting units from the input to the dependent listing led to a 13 percentage point decrease in the inclusion rate for the suppressed units. The failure-to-add errors are not confined to just a few segments or listers: they occurred in seven of the 14 segments and with seven of the 13 listers. There are few differences by structure type. The same difference-in-differences estimates are -13.6% for single-family units and -11.3% for multi-family units.

The linear probability model, with additional controls for housing unit characteristics (multi-unit vs. single family home) and segment level characteristics (city, trailer park), results in an estimated interaction effect for listing method and manipulation of -0.15 with a standard error of 0.04 ($z=-3.33$, $p<0.001$). The probability that a case will be listed by a dependent lister is 15 percentage points lower for cases that are not on the input listing, controlling for additional segment and housing unit characteristics. This estimated interaction effect is very close to the overall difference-in-differences estimate above and is statistically significant. See the Appendix for details.

[6] The difference in listing rates for the unmanipulated and deleted cases among the traditional listers indicates that the housing units we chose to delete were *easier* to list than those that we did not delete, which supports our use of the difference-in-differences estimator, as it adjusts for this fact.

Table 2: Failure-to-Add Error: Listing Rates

		Housing Units	Traditional Listings	Dependent Listings	D in D Estimate
Overall	Unmanipulated Units	732	88.11%	89.48%	-13.4%
	Deleted Units	58	93.10%	81.03%	
Single family	Unmanipulated Units	491	85.54%	85.54%	-13.6%
	Deleted Units	44	95.45%	81.82%	
Multi family	Unmanipulated Units	241	93.36%	97.51%	-11.3%
	Deleted Units	14	85.71%	78.57%	

4.2 Failure-to-Delete Error

Table 3 compares the listing rates for the unmanipulated and the added units. Of the 24 housing units we added to the input listing, 16.67% were confirmed by the dependent listers.[7] One of the added cases was listed by a traditional lister.[8] The difference-in-differences estimate of the effect of adding units on the dependent listers is $(86.70\% - 4.17\%) - (88.26\% - 16.67\%) = 10.9\%$. Adding units to the input listing led the dependent listers to include those units on their frames at a rate that is 11 percentage points higher than it would have been without the manipulation. (Given the small number of manipulated cases we do not provide regression results for this analysis.) The failure-to-delete errors occurred in three segments and with three different listers (two of these segments and listers also contained failure-to-add errors).

Table 3: Failure-to-Delete Error: Listing Rates

		Housing Units	Traditional Listings	Dependent Listings	D in D Estimate
Overall	Unmanipulated Units	639	86.70%	88.26%	10.9%
	Added Units	24	4.17%	16.67%	
Single family	Unmanipulated Units	407	82.80%	82.80%	0%
	Added Units	12	0%	0%	
Multi family	Unmanipulated Units	232	93.53%	97.84%	20.7%
	Added Units	12	8.33%	33.33%	

[7] Here we drop the two segments where no cases were added, which explains why the unmanipulated counts in table 3 do not match table 2.

[8] Surprisingly, this particular line was *not* confirmed by the lister using dependent listing. The added unit was the ninth unit in a multi-unit building. Our own examinations of this building suggested that the ninth unit is a basement apartment with its own entrance in the rear of the building.

The differences by structure type are striking, despite the small sample size. All of the instances of failure to delete confirmation bias occurred in multi-unit buildings. The same difference-in-difference estimate for the single-family units is 0% and for multi-family units it is 20.7%.

5 Discussion and Conclusion

This paper provides the first evidence of confirmation bias in housing unit listing. In these segments, inaccuracies in an input listing tended not to be corrected by listers using dependent listing. We find evidence from this experiment for failure-to-add confirmation bias. With respect to building structure, we find failure-to-delete confirmation bias to be stronger in multi-unit buildings. However, there is no evidence that failure-to-add confirmation bias is stronger in multi-unit buildings.

We acknowledge several shortcomings to our study. It involved segments in only Ann Arbor and Ypsilanti Michigan. Ann Arbor is a college town and contains a good deal of student housing made up of formerly single-family homes converted into several apartments, the sort of small multi-unit buildings that are hard to list accurately. Ypsilanti is also a college town though not quite as well-off as Ann Arbor, and it also contains many small multi-unit buildings. Due to the targeted segment selections and the confinement of our study to southeast Michigan, we cannot make claims about confirmation bias in the country as a whole, but this preliminary work does suggest that additional research is warranted.

Our use of students as listers in this study is vulnerable to critiques concerning their training and motivation when we try to generalize to other listing efforts. The listers were students who received three hours of classroom and field instruction in both listing methods, which is three fewer than new interviewers on the National Survey of Family Growth (NSFG), a large SRC study does regular listing. However, the NSFG interviewers are also trained on use of SRC's listing software and given a general introduction to the importance of coverage and the design of area-probability surveys that the students received elsewhere. In addition, the NSFG interviewers receive several days of interviewer training the same week they learn listing. For these reasons, we feel that the listers used in our study were well-prepared for the listing task and that the level of training they received was comparable to that given to professional listers. Our student listers may be less motivated than professional listers. A professional may believe that her continued employment depends on high-quality work, while a student might slack off when given a listing task. But the opposite is also possible: a lister in southern California can be quite sure that no one is going to check up on her listing work, while a student knows that a check on his work is entirely possible and may believe that his grade depends on how well he lists. Thus we do not believe that there are clear-cut motivational differences between our listers and professional listers.

While these findings still need more research, they do have several implications for surveys which use dependent listing, which include the Current Population Survey and most of the Census Bureau's other household surveys (U.S. Census Bureau, 2006). Surveys which rely on dependent listing to create housing unit frames are much more reliant on the quality of their input listing than has been thought: errors (of inclusion or exclusion) on the input listing are likely to be transmitted to the final frame due to confirmation bias. If the kinds of units undercovered and overcovered by the input listing are different than those which are properly covered, then confirmation error can introduce coverage bias into survey data (see Wright and Tsao, 1983; Groves, 1989; Lessler and Kalsbeek, 1992, on bias due to undercoverage and overcoverage). For this reason, we believe survey organizations using dependent listing should have a good understanding of the determinants of the quality of their input lists.

Our findings also suggest that those data collection organizations which use dependent listing to check the quality of a previous listing underestimate the error rate. Confirmation bias means that errors are likely not to be detected in the dependent check and the results favor the initial frame. In a similar way, the phenomenon of confirmation bias calls into question many of the estimates of undercoverage and overcoverage in previous listing studies which used dependent listing as a gold-standard frame (studies which use this technique include Hansen and Steinberg, 1956; Joncas, 1985; O'Muircheartaigh et al., 2003; Pearson, 2003; Thompson and Turmelle, 2004; Turmelle et al., 2005; O'Muircheartaigh et al., 2006, 2007).[9] These results may understate the net coverage rate in listed housing unit frames.

We have several ideas for future research into confirmation bias in dependent listing. Although we randomized the assignment of lister and method to the selected segments, the segments themselves were not randomly selected. Future studies should aim to make inferences about the expected size of the effect in nationally-representative studies. Future research should also expand upon our findings of building structure effects: what kinds of housing units, listers and blocks are most susceptible to confirmation bias? The introduced error rates in these segments were quite low. We suspect that there is some threshold error rate after which listers become more suspicious of the input list and less likely to show confirmation bias, but with the data at hand we cannot test for such an effect.

There several possible underlying mechanisms which could explain the confirmation bias phenomenon. The list may alter the perceptions of the listers, causing them to see, or not see, units that they otherwise would not. The list may make listers less conscientious, leading them to confirm because it makes the task easier. Or the list may serve as an authority that listers appeal to when confronted with complex or unclear listing situations. In this study we cannot disentangle the effects of these three influences, but future studies on listing quality should do so.

[9] Some, but not all, of the cited studies use a senior lister for the dependent check. If senior listers are less susceptible to confirmation bias (a hypothesis that we cannot speak to in this paper) then those studies may be less vulnerable to this critique.

Appendix: Regression Models

Table 4 presents estimates from the linear regression model mentioned in the text, alongside coefficients from a logistic model on the same dataset and variables. The dataset used in both models contains two observations for each housing unit, one for each of the two listings. The dependent variable in each model is a binary variable indicating whether a given lister included that unit on her frame. The independent variables are: a dummy indicating dependent listing, a dummy indicating that a unit was deleted from the input listing,[10] the interaction of these two, an indicator at the housing unit level for units in multi-unit buildings, and two variables at the segment level, indicators for Ypsilanti (versus Ann Arbor) and for the two segments that contain trailer parks. Both models include crossed random effects to account for how the cases are nested inside listers and segments. The linear model was fit with the `xtmixed` command and the logistic model with the `xtmelogit` command in Stata 11 (StataCorp LP, 2009).

The third row in this table refers to the interaction of dependent listing and deleting a housing unit from the input listing. It is this interaction effect that captures the failure-to-add (FTA) error that we are interested in. As discussed in text, the linear model simplifies interpretation of the coefficient on the interaction effect. The estimated coefficient is strongly significant and negative, as hypothesized. Deleting cases from the input listing decreases the propensity that those housing units will be listed by listers using the dependent method by 15 percentage points. In the logistic model, the estimated beta coefficient is also significant and negative, but interpretation of this coefficient (or its associated odds ratio, 0.136) is not straightforward. We include both models to demonstrate that the sign and significance patterns of the coefficients are similar. Together these two models support our findings in the body of the paper that failure-to-add error exists in dependent listing.

The only other significant variable in these models is the indicator for the two segments which contained trailer parks (one in Ann Arbor and one in Ypsilanti). Housing units in these two segments were significantly less likely to be listed and the effect is strong in magnitude. We recommend additional study to uncover what it is about trailer parks that makes them particularly hard to list.

[10] This indicator is at the housing unit level, as in the tables above.

Table 4: Failure to Add Error: Crossed-Effects Regression Models

	Coefficient (Std. Error)	
	Linear	Logistic
Dependent (vs Traditional)	0.01	0.43
	(0.01)	(0.29)
Deleted (vs Unmanipulated)	-0.01	-0.11
	(0.03)	(0.62)
Interaction: FTA	-0.15**	-1.99**
	(0.04)	(0.77)
Multi-unit (vs Single Family)	-0.02	-0.45
	(0.01)	(0.30)
Ypsilanti (vs Ann Arbor)	-0.10	-0.52
	(0.09)	(0.73)
Trailer Park	-0.37**	-2.06*
	(0.13)	(1.02)
N	1508	1508

Coefficient on constant term not shown
* p<0.05, ** p<0.01

References

Ai, C. and E. C. Norton (2003). Interaction Terms in Logit and Probit Models. *Economics Letters 80*(1), 123 – 129.

Angrist, J. and J. Pischke (2009). *Mostly Harmless Econometrics: an Empiricist's Companion*. Princeton University Press.

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*.

Biemer, P. P. and L. E. Lyberg (2003). *Introduction to Survey Quality*. Wiley-Interscience.

Chakrabarty, R. P. and G. Torres (1992). American Housing Survey: A Quality Profile. Technical Report H121/95-1, U.S. Department of Housing and Urban Development, Office of Policy Development and Research.

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

Groves, R. M. (1989). *Survey Errors and Survey Costs*. John Wiley and Sons.

Hansen, M. H. and J. Steinberg (1956). Control of Errors in Surveys. *Biometrics 12*(4), 462–474.

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

Jacobs, C. (1986). Interim Evaluation of Listing Process Audit. Unpublished memorandum to Housing Working Group, U.S. Bureau of Labor Statistics. [Cited in Subcomittee on Survey Coverage (1990)].

Joncas, M. (1985). Cluster Listing Check Program for the Redesigned LFS Sample. Unpublished report, Ottawa: Statistics Canada. [Cited in Subcomittee on Survey Coverage (1990)].

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.

Klayman, J. (1995). Varieties of Confirmation Bias. *Psychology of Learning and Motivation 32*, 385–418.

Lessler, J. T. and W. D. Kalsbeek (1992). *Nonsampling Error in Surveys*. John Wiley and Sons.

Lynn, P. and E. Sala (2006). Measuring Change in Employment Characteristics: The Effects of Dependent Interviewing. *International Journal of Public Opinion Research 18*(4), 500–509.

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

Mood, C. (2009). Logistic Regression: Why We Cannot Do What We Think We Can Do, and What We Can Do About It. *European Sociological Review*, jcp006. Advance Access published on March 9, 2009.

O'Muircheartaigh, C. A. (2004). Simple Response Variance: Estimation and Determinants. In P. Biemer, R. M. Groves, L. Lyberg, N. Mathiowetz, and S. Sudman (Eds.), *Measurement Errors in Surveys*. Wiley-Interscience.

O'Muircheartaigh, C. A., S. A. Eckman, and C. Weiss (2003). 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.

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

Raudenbush, S. and A. Bryk (2002). *Hierarchical Linear Models* (Second ed.). Sage.

StataCorp LP (2009). *Stata Statistical Software: Release 11*. College Station, TX: StataCorp.

Subcomittee on Survey Coverage (1990). Survey Coverage. Technical report, Federal Committee on Statistical Methodology.

Survey Research Center (1969). *Interviewer's Manual*. Institute for Social Research, The University of Michigan.

Survey Research Center (1976). *Interviewer's Manual* (Revised ed.). Institute for Social Research, The University of Michigan.

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*.

Turmelle, C., J.-F. Rodrigue, and G. Thompson (2005). Using the Canadian Address Register in the Labour Force Survey Implementation, Results and Lessons Learned. In *Proceedings of the Conference of the Federal Committee on Statistical Methodology*.

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

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. The MIT Press.

Wright, T. and H. J. Tsao (1983). A Frame on Frames: An Annotated Bibliography. In T. Wright (Ed.), *Statistical Methods and the Improvment of Data Quality*. Academic Press.