Stephanie Eckman, Frauke Kreuter, Confirmation Bias in Housing Unit Listing, *Public Opinion Quarterly*, Volume 75, Issue 1, Spring 2011, Pages 139–150, https://doi.org/10.1093/poq/nfq066

Author Notes

STEPHANIE ECKMAN is a Senior Researcher with the Institute for Employment Research of the Federal Employment Office in Nürnberg, Germany. FRAUKE KREUTER is an Associate Professor in the Joint Program for Survey Methodology at the University of Maryland and head of the statistical methods research department at the Institute for Employment Research. Address correspondence to Stephanie Eckman, Institute for Employment Research, Regensburger Straße 104, 90478 Nürnberg, Germany; email: stephanie.eckman@iab.de.

Acknowledgments

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 James Lepkowski, 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. An earlier version of this paper won the 2009 student paper competition held by the Washington, D.C. chapter of the American Association for Public Opinion Research.

Abstract

Using an experimental repeated listing design, this paper demonstrates the presence of confirmation bias in dependent housing unit listing. We find evidence that when provided with an initial listing to update in the field, listers tend not to add missing units or delete inappropriate units. The listers are biased towards confirming the initial list as correct. 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.

Introduction

Several vendors now offer inexpensive access to address databases derived from the United States Postal Service's Delivery Sequence File (DSF). These address lists have attracted the interest of researchers who wish to reduce the costs of compiling address frames for mail or in-person surveys (Iannacchione et al., 2003; Battaglia et al., 2008; Link et al., 2008). However the lists do contain errors: they tend to undercover rural areas and new construction (O'Muircheartaigh et al., 2003; Staab and Iannacchione, 2003; Dohrmann et al., 2006; O'Muircheartaigh et al., 2006; Dohrmann et al., 2007; O'Muircheartaigh et al., 2007; Montaquila et al., 2010).

For this reason, survey organizations increasingly use dependent listing – also called update or enhanced listing – to overcome such errors before sample selection. In dependent listing, listers are provided with a list of addresses believed to lie inside the selected area. This initial listing is called the *input list* (our term) and may come from the DSF or from a previous listing of the area. Listers travel around the segment making corrections to the list to match what they find in the field. Dependent listing is used by the U.S. Census Bureau and the National Opinion Research Center and also in the current cycle of the National Survey of Family Growth (U.S. Census Bureau, 2006; Groves et al., 2009; Harter et al., 2010).

Although dependent listing is used to correct errors on the input list, the method may itself suffer from what we call *confirmation bias*: the tendency for listers to preserve the errors in the input list. Such confirmation bias is already known in other stages of the survey process. When a second coder reviews earlier work and can see the value assigned by the first, she is more likely to assign the same code, confirming the work of the first coder, than if she did an independent second coding (Biemer and Lyberg, 2003). In dependent interviewing, when the interviewer and/or respondent in a later wave knows the response recorded in the first wave, the two interviews tend to capture similar responses (O'Muircheartaigh, 2004; Lynn and Sala, 2006).[1]

[1] We follow Biemer and Lyberg in our use of the term confirmation bias. This term is not meant to imply

There is likely a similar phenomenon in dependent listing that carries errors on the input list into the final frame. We ask the following questions:

- Is a lister using the dependent method likely not to add units that are missing from the input list even though they exist in the field?
- When an inappropriate unit is on the input list, does a dependent lister tend not to delete it?

Our experiment shows, for the first time, evidence that confirmation bias exists in dependent listing. If these results hold up in larger studies, they will have important implications for household frame construction and coverage investigations.

Study Design

To study confirmation bias we implemented an experiment in the Survey Practicum course of the University of Michigan's Program in Survey Methodology. In this class, students work through all phases of a survey: questionnaire design, sample design, listing, interviewing, data preparation and analysis. In the spring of 2009, we introduced errors into the input list given to the dependent listers to examine whether those errors persist after field updating.

Students in this course were trained in both traditional[2] and dependent listing. They read the University of Michigan Survey Research Center (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.

---

that the behavior necessarily leads to bias in survey estimates. Use of this phrase in the survey field is similar to its use in social psychology where confirmation bias means "looking for the presence of what you expect" (Klayman, 1995, p. 386).

[2] In *traditional listing* – also called scratch listing – listers travel around each selected block in the segment, recording the address or description of every housing unit, whether occupied or not, without the aid of an existing list (Kish, 1965; Survey Research Center, 1969, 1976).

The portion of the training devoted to dependent listing emphasized the importance of checking the input list for errors–adding, deleting and confirming units to ensure that the list matched what they saw in the field. The listers were instructed to add in all units that are inside the segment but not already on the list and to delete units from the list that do not exist, are outside the segment boundaries, or are not residential. The trainers informed the listers that the input list for this study came from postal addresses and that the lists could contain errors, but did not give the listers any information about the amount and nature of the error in the input lists or the manipulations we performed.

For the listing study, we selected seven segments in Ann Arbor and seven in Ypsilanti, Michigan. Each segment contained one or more Census blocks and at least 30 housing units, according to 2000 Census counts. We purposefully selected two segments, one in each city, which contained trailer parks, and three segments where a majority of the units are in multi-unit buildings. Both of these types of units are known to be associated with listing errors (Childers, 1992; Subcommittee on Survey Coverage, 1990; Chakrabarty and Torres, 1992; Barrett et al., 2002). Three segments contained only single family homes, and the remaining six segments contained the usual combination of single and multi-family housing that is common in these cities.

Overall we have 14 segments and 14 listers. Working separately, one lister in each segment used traditional listing, and the other lister in that segment used dependent listing. This design resulted in two listings of each segment, one using each of the two listing methods. Listers were randomly assigned to segments and to listing methods within those segments. We controlled the randomization so that each lister was assigned a segment in each of the two cities and used each of the two methods. Figure 1 illustrates the design for four segments and four listers.

[Figure 1 about here.]

Listers were provided with color maps of their assigned areas showing the boundaries of the blocks to list. All listings were completed within one week but we do not have data on which listing was completed first and which second in each segment or by each lister.

Twelve of the 14 listers worked alone. The other two listers accompanied each other in their respective assignments. Thus four segments were listed by this pair, two in Ypsilanti and two in Ann Arbor (and two via traditional listing and two via dependent listing). Fortunately, they did not do both the traditional and dependent listings of the same segment. We were unaware of this violation of our experimental design until after the listing work was completed. To ensure that this lister pair, or any individual lister, does not dominate the findings given below, all analyses were replicated with a jackknife-type procedure in which one lister (or lister pair) was dropped at a time.

The input to the dependent listing was based on the DSF provided through Marketing Systems Group (MSG). MSG licenses the DSF, geocodes all of its addresses, and creates frames from those addresses. These frames contain a Census block identifier for each address. We purchased all available residential addresses for the blocks in our segments from MSG. The procedures we used for ordering addresses and preparing them for the input list correspond to those used in the most recent cycle of the National Survey of Family Growth. We received 790 addresses from MSG for our selected blocks, 482 in Ann Arbor and 308 in Ypsilanti (see Table 1).

To test the confirmation bias hypothesis, we manipulated the input list in each segment. These manipulations were not random but covered several scenarios, while maintaining plausibility. We added and deleted single-family homes, units in multi-unit buildings, and entire street segments. We deleted at least one unit from each segment and added units to all but two of the segments.[3] The manipulations were not informed by a previous listing of

[3] In two of the segments it was not possible to add plausible units: one was in Ann Arbor and consisted entirely of a single development with sequential address numbers; the other was an Ypsilanti segment where many units were already deleted.

the area, but we did verify that none of the units we created and added truly existed.

In total, we deleted 58 addresses that we received from MSG and left 732 unmanipulated. We added in 24 addresses. The resulting 756 addresses (469 in Ann Arbor, 287 in Ypsilanti) became the input list for the dependent listing of our segments. Table 1 summarizes the manipulations.

[Table 1 about here.]

Results

A total of 754 housing units were listed by the listers using traditional listing and 738 by those using dependent listing (see the bottom section of Table 1). The frame created by the dependent listers includes two types of units: those from the input list which were confirmed by the dependent listers, and those which were added to the list by the listers.

If confirmation bias exists, units deleted from the input list should be listed less frequently by dependent listers than traditional listers (a *failure-to-add* error). Likewise, the units we added to the input list should be listed more frequently by dependent than traditional listers (a *failure-to-delete* error).

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 732 unmanipulated addresses. The traditional listers included 88.1% of these units on their frames and the dependent listers 89.5%. That is, as expected, the unmanipulated addresses are not error free: more than ten percent of these addresses were not listed by the traditional listers or were removed by the dependent listers. A larger investigation of the quality of the MSG address lists is outside the scope of this paper, but the error rate found here is in line with other studies of the coverage of address databases (O'Muircheartaigh et al., 2006, 2007; Montaquila et al., 2010).

The second row of Table 2 refers to the 58 units we deleted from the input list. 81.0% of these units were added back by the dependent listers. A higher proportion, 93.1%, was listed by the traditional listers, who were not subject to the manipulation.

To assess the presence of failure-to-add error, we therefore use a difference-in-differences estimator, a technique commonly used with panel data to derive treatment effects from non-randomized designs (Angrist and Pischke, 2009, pp. 221–247). Here, the technique helps to control for any ways in which the unmanipulated and deleted units differ that could affect their propensity to be listed.

In this analysis, the estimator is the difference between two differences: the difference in the listing rates of the unmanipulated and deleted cases as listed by the traditional listers ($TL_{\text{unm}} - TL_{\text{del}}$), and the difference in the listing rates of the unmanipulated and deleted cases as listed by the dependent listers ($DL_{\text{unm}} - DL_{\text{del}}$). The difference-in-differences estimate of the effect of deleting units is:

$$\begin{aligned} D - \text{in} - D &= (TL_{\text{unm}} - TL_{\text{del}}) - (DL_{\text{unm}} - DL_{\text{del}}) \\ &= (88.1\% - 93.1\%) - (89.5\% - 81.0\%) \\ &= -13.4\% \end{aligned}$$

That is, deleting units from the input to the dependent listing led to a 13 percentage point decrease in the inclusion rate for those cases.

[Table 2 about here.]

The difference between single-family units and those in multi-unit buildings is minimal, as shown in Table 2. We use the presence of an apartment identifier to indicate those units in multi-unit buildings, as listers in our study did not record this observation for each unit.[4] The difference-in-differences estimates are -13.6% for single-family units and -11.3% for multi-family units.

Table 3 compares the listing rates for the unmanipulated and the added units.[5] Of the 24 housing units we added to the input list, 16.7% were listed by the dependent listers. One of the added cases was listed by a traditional lister. Surprisingly, this unit was *not* confirmed by the lister using dependent listing.[6] The difference-in-differences estimate of the effect of adding units on the dependent listers is (86.7% − 4.2%) − (88.3% − 16.7%) = 10.9%. Adding units to the input list 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.

The differences by structure type are striking, though they are based on small sample sizes. All of the instances of failure-to-delete error occurred in multi-unit buildings. The estimates are 0% for single-family units and 20.7% for multi-family units.

[Table 3 about here.]

So far we have presented only point estimates of the confirmation bias phenomena. To capture the uncertainty in our estimates due to the small dataset and the inexperience of our listers, we conducted a sensitivity analysis. Dropping the listings of each lister in turn gives 13 difference-in-difference estimates of the effect of each manipulation.[7] The estimates of failure-to-add error range from -9.3 to -18.0 percentage points, and the estimates of failure-to-delete error range from 3.1 to 16.0 percentage points. These

[4] In this table and throughout the paper, all units in the trailer parks are considered single-family homes. We do not have a housing unit level indicator of which units are trailers.

[5] The two segments where no cases were added are not included in Table 3, leading to unmanipulated counts in this table that do not match Tables 1 and 2.

[6] The added unit listed by the traditional lister is a ninth unit we added to an eight unit building. We visited this building ourselves and spoke to a resident who confirmed that the building contained eight units, two on each of four floors. However, the building does have nine mailboxes, which likely confused the traditional lister in this segment.

[7] Here we treat the two students who worked together as one lister.

jackknife results indicate that our findings of confirmation bias are not due to just one lister.

To adjust for additional housing unit characteristics in the estimate of failure-to-add error and to obtain a standard error on the point estimate, we fit a multi-level linear probability model.[8] (Given the small number of added cases we do not provide regression results for the failure-to-delete scenario.) The dataset used to estimate the model is at the housing unit and listing level: each unit appears in the dataset twice, once for each of the two listings. The binary dependent variable in the model is whether a housing unit was listed (1) or not (0) by a given lister. The main independent variables of interest are method (traditional listing is the reference category), a deleted unit indicator (at the housing unit level), and the interaction of these two. Including the indicator of deletion at the housing unit level adjusts for unobserved attributes that make the deleted units harder or easier to list. Also included are indicators at the housing unit level for multi-unit buildings and segment level indicators for city (Ann Arbor is the reference category) and the two segments that contained trailer parks. Random effects account for the clustering of the observations by segment and listers in the calculation of standard errors on the coefficient estimates.

The interaction term in the third row of Table 4 captures the failure-to-add error of interest. The estimated coefficient is -0.145 with a standard error of 0.04 ($z = -3.33$, $p < 0.001$). Deleting cases from the input list decreases the probability that those housing units will be listed by listers using the dependent method by 14.5 percentage points. This estimated interaction effect is very close to the overall difference-in-differences estimate above and is statistically significant.

The indicator for the two segments which contained trailer parks (one in each city) is also

[8] Since interpretation of interaction effects in nonlinear models is complex (Ai and Norton, 2003), we interpret the results of a linear probability model (as suggested by Mood, 2010). Table 4 presents results from both linear and logistic regressions.

significant. Housing units in these two segments were significantly less likely to be listed than units in other segments. However, not all units in these two segments were in trailer parks, and we cannot be sure about what is driving this result.

[Table 4 about here.]

Discussion and Conclusion

This paper provides the first evidence of confirmation bias in housing unit listing. Inaccuracies in the input list tend not to be corrected by listers using dependent listing. This study finds support for *failure-to-delete* error and *failure-to-add* error, though the latter results are based on a small number of cases.

We acknowledge several shortcomings of our study. First, it involves only segments in two college towns in Michigan that contain many small multi-unit buildings. These buildings are known to be difficult to list because listers have trouble determining how many units the buildings contain (Subcommittee on Survey Coverage, 1990; Chakrabarty and Torres, 1992). We suspect this is because small buildings often have a less organized presentation (multiple entrances, doorbells, etc.) than large buildings but cannot test this hypothesis here.

Second, analyses required manual matching of the two listings, and this is a possible source of error in our study. In the few difficult situations we encountered, we matched two cases 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 pairs. When one lister included two units at an address and another only one, we matched the single family home to the first unit, because this is the instruction SRC interviewers receive when they find a selected single-family unit to be a multi-unit building.

Third, the listers were not experienced professionals, though they did receive three hours of training in the classroom and in the field and had previous instruction in sampling and survey methods. These three hours are about half the training time that new listers on the National Survey of Family Growth (NSFG) receive. However, the NSFG listing training covers material the students did not need, such as use of the listing software. Furthermore, the NSFG listers are also trained as interviewers in the same week, which might reduce retention of the listing instructions. We also recognize that the student listers may be less motivated to do good work than professional listers, although the opposite is also possible: a lister far away from the central office can assume that no one is going to check her work, while a student knows that a review of his local listing is possible and may believe that his grade depends on how well he lists.

In light of these shortcomings, we cannot generalize from this study to the nation as a whole, but the findings here indicate that confirmation bias warrants additional investigation. On the one hand, the lack of experience and the short training given to the listers could mean that the rates of confirmation bias we find in this study are an upper-bound on the confirmation bias phenomenon. On the other hand, the fact that our listers worked in small, familiar cities with low crime rates and many single-family homes could indicate that the results are a lower bound. Future work should study the extent of the phenomena in a national study with experienced listers.

If our findings hold up in larger studies, they will have several implications for surveys. Most directly, studies using dependent listing run the risk of transmitting errors on the input list to the final frame. Additionally, if the kinds of units undercovered and overcovered by the input list are different than those which are properly covered, confirmation bias can contribute to coverage bias in survey estimates (Wright and Tsao, 1983; Groves, 1989; Lessler and Kalsbeek, 1992). The confirmation bias findings also have implications for the half-open interval procedure, which is used to rectify undercoverage on a frame. During data collection, interviewers are asked to look near their selected cases for any housing units missing from the frame. This procedure is essentially dependent listing and interviewers may tend not to find missing units.

The confirmation bias phenomenon also has implications for evaluations of the coverage of housing unit frames. Many previous studies, including those that assess the coverage of frames derived from the DSF, use dependent listing to create a gold standard frame (Hansen and Steinberg, 1956; O'Muircheartaigh et al., 2003; Pearson, 2003; Thompson and Turmelle, 2004; Turmelle et al., 2005; O'Muircheartaigh et al., 2006, 2007). Failure-to-add and failure-to-delete error would lead these coverage studies to overestimate quality.[9] Future research should explore alternative approaches to gold standard frame construction. We suggest multiple listings by multiple methods, such as was used in this study. Another approach to studying coverage would use multiple dependent listings with varying input lists, such as DSF databases and E-911 lists (Marker and Fraser, 2010). If the errors on the input frames are independent then perhaps repeated dependent listing can approach a gold standard frame.

In closing, we emphasize that our findings do not imply purposefully erroneous actions by listers. In fact, the coding literature suggests appeal to authority as the likely explanation for confirmation bias: the first code appears as an authority which the second coder is reluctant to contradict (Biemer and Lyberg, 2003). In dependent listing, the input list may play such a role. It is also possible that the list may alter the perceptions of the listers, causing them to see, or not see, units in their segments. The list may make listers less conscientious, leading them to confirm because it makes the task easier. While we cannot disentangle these influences here, we encourage future studies to explicitly test these mechanisms.

[9] Some of the cited studies use a senior lister for the dependent check. If senior listers are less susceptible to confirmation bias, a hypothesis that is untested, then those studies may be less vulnerable to this critique.

References

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

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

Barrett, Diane F., Michael Beaghen, Damon Smith, and Joseph Burcham (2002). Census 2000 Housing Unit Coverage Study. In *Proceedings of the Section on Survey Research Methods, American Statistical Association*.

Battaglia, Michael P., Michael W. Link, Martin R. Frankel, Larry Osborn, and Ali H. Mokdad (2008). An Evaluation of Respondent Selection Methods for Household Mail Surveys. *Public Opinion Quarterly 72*(3), 459–469.

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

Chakrabarty, Rameswar P. and Georgina 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, Danny R. (1992). The 1990 Housing Unit Coverage Study. In *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 506–511.

Dohrmann, Sylvia, Daifeng Han, and Leyla Mohadjer (2006). Residential Address Lists vs. Traditional Listing: Enumerating Households and Group Quarters. In *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 2959–2964.

Dohrmann, Sylvia, Daifeng Han, and Leyla Mohadjer (2007). Improving Coverage of Residential Address Lists in Multistage Area Samples. In *Proceedings of the Section on Survey Research Methods, American Statistical Association*.

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

Groves, Robert M., William D. Mosher, James Lepkowski, and Nicole G. Krigis (2009). Planning and Development of the Continuous National Survey of Family Growth. *Vital Health Statistics Sep*(1), 1–64.

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

Harter, Rachel, Stephanie Eckman, Ned English, and Colm A. O'Muircheartaigh (2010). Applied Sampling for Large-Scale Multi-Stage Area Probability Designs. In Peter D. Marsden and James V. Wright (Eds.), *Handbook of Survey Research, Second Edition*. Emerald.

Iannacchione, Vincent G., Jennifer M. Staab, and David T. Redden (2003). Evaluating the Use of Residential Mailing Addresses in a Metropolitan Household Survey. *Public Opinion Quarterly 67*(2), 202–210.

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

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

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

Link, Michael W., Michael P. Battaglia, Martin R. Frankel, Larry Osborn, and Ali H. Mokdad (2008). A Comparison of Address-Based Sampling (ABS) Versus Random-Digit Dialing (RDD) for General Population Surveys. *Public Opinion Quarterly 72*(1), 6–27.

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

Marker, David A. and Whitney J. Fraser (2010). Are 911 Listings Good Enough to Use as a Sampling Frame for Area Surveys? In *Presented at the Joint Statistical Meetings, Vancouver, Canada*.

Montaquila, Jill M., Valerie Hsu, and J. Michael Brick (2010). Using a Match Rate Model to Predict Areas Where USPS-Based Address Lists May Be Used in Place of Traditional Listing. Under Review.

Mood, Carina (2010). Logistic Regression: Why We Cannot Do What We Think We Can Do, and What We Can Do About It. *European Sociological Review 26*(1), 67–82.

O'Muircheartaigh, Colm. A. (2004). Simple Response Variance: Estimation and Determinants. In Paul Biemer, Robert M. Groves, Lars Lyberg, Nancy Mathiowetz, and Seymour Sudman (Eds.), *Measurement Errors in Surveys*. Wiley-Interscience.

O'Muircheartaigh, Colm A., Stephanie Eckman, and Charlene 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, Colm A., Ned English, and Stephanie 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, Colm A., Ned English, Stephanie Eckman, Heidi Upchurch, Erika Garcia, and James 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, Jeffrey M. (2003). Quality and Coverage of Listings for Area Sampling. In *Proceedings of the Section on Government Statistics, American Statistical Association*.

Staab, Jennifer M. and Vincent G. Iannacchione (2003). Evaluating the Use of Residential Mailing Addresses in a National Household Survey. In *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 4028–4033.

Subcommittee 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, Gavin and Claude 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.

Turmelle, Claude, Jean-Francois Rodrigue, and Gavin 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.

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

Figure 1: Illustration of Experimental Design. The full design included 14 segments and 14 listers. T = traditional listing; D = dependent listing.

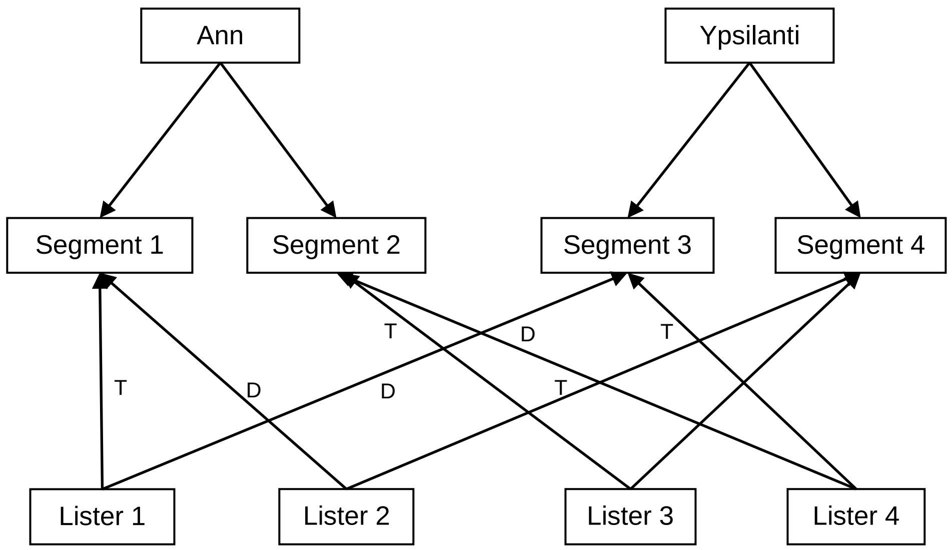

Table 1: Housing Units Counts, Before Manipulation, After Manipulation and After Listing

	Total	Ann Arbor	Ypsilanti
Before Manipulation	790	482	308
Deleted	58	26	32
Unmanipulated	732	456	276
Added	24	13	11
After Manipulation	756	469	287
Listing Counts			
Traditional Listings	754	511	243
Dependent Listings	738	486	252

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

		Housing Units	Traditional Listings	Dependent Listings	D in D Estimate
Overall	Unmanipulated	732	88.1%	89.5%	-13.4%
	Deleted	58	93.1%	81.0%	
Single family	Unmanipulated	491	85.5%	85.5%	-13.6%
	Deleted	44	95.5%	81.8%	
Multi-unit	Unmanipulated	241	93.4%	97.5%	-11.3%
	Deleted	14	85.7%	78.6%	

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

		Housing Units	Traditional Listings	Dependent Listings	D in D Estimate
Overall	Unmanipulated	639	86.7%	88.3%	10.9%
	Added	24	4.2%	16.7%	
Single family	Unmanipulated	407	82.8%	82.8%	0%
	Added	12	0%	0%	
Multi family	Unmanipulated	232	93.5%	97.8%	20.7%
	Added	12	8.3%	33.3%	

Table 4: Beta-Coefficients from Regression Models Predicting the Propensity of Units to be Listed (standard errors in parentheses)

	Linear (n=1580)	Logistic (n=1580)
Main Effects of Interest		
Method (Traditional=0, Dependent=1)	0.015	0.432
	(0.013)	(0.294)
Deleted (HU level)	-0.006	-0.105
	(0.031)	(0.617)
Interaction: Method $*$ Deleted	-0.145**	-1.994**
	(0.044)	(0.766)
Control: Housing Unit Characteristic		
Multi-unit	-0.020	-0.450
	(0.014)	(0.303)
Controls: Segment Level Characteristics		
City (Ann Arbor=0, Ypsilanti=1)	-0.100	-0.525
	(0.091)	(-.0726)
Trailer Park	-0.369**	-2.058*
	(0.130)	(1.015)
Intercept	0.999**	3.518**
	(0.068)	(0.581)
Random Effects		
Standard Deviation (lister)	0.028**	0.534
	(0.242)	(0.011)
Standard Deviation (segment)	0.169**	1.23
	(0.292)	(0.037)
Summary Statistic		
Wald χ^2 (6)	34.55	25.93

* $p < 0.05$, ** $p < 0.01$, two-tailed tests