Judgments under Uncertainty: Assessing the Quality of Interviewer-Generated Paradata

James M. Dahlhamer
jdahlhamer@cdc.gov

Renee Gindi
rgindi@cdc.gov

National Center for Health Statistics
3311 Toledo Road
Hyattsville, MD 20782

Nancy Bates
nancy.a.bates@census.gov

Rachael Walsh
rachael.walsh@census.gov

US Census Bureau
Center for Survey Measurement
4600 Silver Hill Rd.
Washington, DC 20233

26th International Workshop on Household Survey Nonresponse
Leuven, Belgium

DRAFT—WORK IN PROGRESS[1]

Do Not Cite Without Permission of the Authors

[1] The content of this paper reflects the opinions of the authors and not those of the National Center for Health Statistics, Centers for Disease Control and Prevention, or US Census Bureau. This paper is provided to inform interested persons of research in progress and does not reflect any official policy of these agencies.

INTRODUCTION

As response rates continue to decline in the U.S. National Health Interview Survey (NHIS), survey staff are exploring the use of interviewer observations of sample units and their environs to address possible nonresponse bias. When the observations correlate with both the propensity to respond and key survey estimates, they can be especially useful in nonresponse weighting adjustments, potentially reducing nonresponse bias and variance in the estimates (Little and Vartivarian, 2005; West, 2013). And since the observations are recorded early in data collection (first personal visit attempt on a household), they may hold promise for responsive or adaptive survey designs, which often use auxiliary data in daily response propensity models for case prioritization (Peytchev et al., 2010; Tabuchi et al., 2009). However, since interviewer observations are largely judgments and may be incorrect (Casas-Cordero, 2013; Sinibaldi et al., 2013; West and Kreuter, 2013; West et al., 2014), their effectiveness in bias reduction strategies can be compromised. Previous research, for example, has demonstrated limited ability of interviewer observations to substantially shift survey estimates when used in nonresponse adjustments (West, 2013; West et al., 2014), findings that have been attributed to large errors in the observations (West et al., 2014).

In this paper, three observations collected by U. S. Census Bureau interviewers using data from the first six months of the 2013 NHIS are evaluated: whether or not one or more adult householders are employed, whether or not all householders are over the age of 65, and whether or not there are children less than six years of age residing at the household. Data from these observations are compared to survey data collected from responding households. Agreement between observational and survey data is analyzed using multi-level, multinomial logistic regressions predicting false positives and false negatives (versus agreement). Model covariates include characteristics of contact attempts, sample units, interviewers, and neighborhoods.

PAST STUDIES OF THE ACCURACY OF INTERVIEWER OBSERVATIONS

Based on the few studies that have assessed the accuracy of interviewer observations of sample units/neighborhoods, predictors of accuracy can be classified into the following categories: characteristics of interviewers, circumstances of the observation, characteristics of sample units, and characteristics of neighborhoods/social environment.

Interviewer Characteristics

Both interviewer age and experience have been found to influence the accuracy of observational data. Sinibaldi et al. (2013) found that less interviewer experience, in terms of years worked, reduced agreement between observations of the presence of children in the housing unit and UK Census data. In addition, younger interviewers were more likely to make accurate observations with regard to type of housing unit and whether or not the household included a working adult (Sinibaldi et al., 2013). Older interviewers were less likely to perceive graffiti and damaged walls/peeling paint in the block face in the Los Angeles Family and Neighborhood Study (LAFANS) (Casas-Cordero et al., 2013). Interviewer characteristics such as race and marital status were also found to influence interviewer judgments in LAFANS (Casas-Cordero et al., 2013).

Circumstances of the Observation

An increasing number of contact attempts on a household has been shown to reduce the accuracy of observations of the presence of children and whether or not a working adult resides at the household (Sinibaldi et al., 2013; West and Kreuter, 2013). Furthermore, when similar observations were recorded during attempts resulting in a refusal, agreement with UK Census data decreased (Sinibaldi et al., 2013). And in LAFANS, observations recorded during evening hours produced more reports of cigar or cigarette butts in the area, while observations made on weekends resulted in lower overall ratings of the physical condition of residential buildings (Casas-Cordero et al., 2013).

Characteristics of Sample Units

Location in a multiple-unit building and the presence of physical access impediments both reduced the odds of a false positive judgment of the presence of children based on National Survey of Family Growth data (West and Kreuter, 2013). Interestingly, physical access impediments also reduced the odds of a false negative judgment. Sinibaldi et al. (2013) found that sample units in flats/apartments reduced the accuracy of observations regarding the type of housing unit, but increased the accuracy of whether the sample unit was part of a council or housing association housing estate. Other sample unit features found to impact accuracy include home ownership and the number of rooms (Sinibaldi et al., 2013).

Characteristics of Neighborhoods/Larger Social Environment

Location in an urban area increased the odds of a false negative judgment of the presence of children, while access problems in the larger neighborhood reduced the odds of a false negative judgment (West and Kreuter, 2013). In addition, working in areas of higher concentrations of black and Hispanic persons tended to increase the odds of both false positive and false negative judgments. In a

separate analysis, Sinibaldi (2010) found that working in areas of mixed racial composition increased the agreement between two sets of interviewer observations of the condition of residential structures. Finally, areas with concentrated disadvantage and/or higher concentrations of immigrants increased the odds of interviewers perceiving various indicators of physical disorder and residential decay in LAFANS, while areas with concentrated affluence were characterized by lower odds of perceiving these features (Casas-Cordero et al., 2013).

METHODS

Overview of the NHIS

The NHIS is an annual survey of the health of the civilian, noninstitutionalized household population of the United States, and is conducted by the National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention (CDC). Utilizing a multi-stage, clustered sample design, with oversampling of black, Hispanic, and Asian persons, the survey produces nationally representative data on health insurance coverage, health care access and utilization, health status, and health behaviors. Over 800 U. S. Census Bureau interviewers in 6 regional offices conduct in-person interviews (some telephone follow-up is allowed) using computer assisted personal interviewing. Currently, interviews are conducted in approximately 40,000 households yielding data on roughly 100,000 persons.

The Neighborhood Observation Instrument

In 2013, the NHIS began using the Neighborhood Observation Instrument (NOI) developed by the U.S. Census Bureau. NOI consists of 15 interviewer observation measures designed to correlate with both survey response and critical survey-specific estimates. Interviewers are instructed to record the observations (only once) on the first personal visit on which they can observe the sample unit or building within which the sample unit resides. Three of the 15 observations were amenable to explorations of measurement error with the NHIS: whether or not one or more adults in the household are employed, whether or not all residents of the household are over the age of 65, and whether or not there are children under the age of 6 residing in the household. Similar data are collected in the NHIS interview (see Table 1 for the observation and survey questions). For a full list and description of observations, see Walsh et al. 2013.

Data and Variables

The data used here are from the first six months of the 2013 NHIS. During this period, 862 interviewers recorded NOI observations on 19,486 households for which they also completed interviews. Hence, in addition to the observational data, these cases have age and employment data for household members. A three-category dependent variable was constructed for each of the three observations: 1) agreement (observation agrees with survey data), 2) false positive (FP) (e.g., a household without employed adults judged to have employed adults), and 3) false negative (FN) (e.g., a household with children under the age of 6 judged not to have children under the age of 6). Table 1 presents frequency listings for the dependent variables.

Based on prior studies, measures capturing interviewer characteristics, circumstances of the observation, characteristics of sample units, and characteristics of the neighborhood were included in the analysis. Interviewer-level measures included gender, years worked on the NHIS, education level, and whether or not the interviewer is an intermittent employee. An interviewer is considered intermittent if they sustain lag times of six months or more between assignments. Variables tapping the circumstances of the observation include the time of day and day of week observations were made and the outcome of the attempt during which the observation was recorded. Characteristics of the sample unit include presence of graffiti in the block face, physical condition of the sample unit, presence of an access barrier, whether or not the unit was in a multi-unit structure, whether or not householders ever expressed a soft refusal, whether or not householders ever expressed privacy concerns, and the number of households for which observations were recorded prior to the current household. Finally, neighborhood/larger social environment measures included urbanicity of the sample unit (MSA status) and four Census tract measures. The Census tract measures were included to explore the possibility that neighborhood-level features influence interviewer judgments, especially when the sample unit provides little in the way of visual clues. The tract-level measures include the proportion of the tract population under the age of 5 (Census 2010), the proportion of the tract population over the age of 64 (Census 2010), the vacancy rate for the tract (Census 2010), and the poverty rate for the tract (American Community Survey 2006-10). Coding of all independent variables is shown in Table 2.

Statistical Analysis

Overall agreement, false positive, and false negative rates for each observation are presented (see Table 1), followed by the results of multilevel multinomial logistic regressions of agreement that include random interviewer effects. These models allow for the inclusion of interviewer-level predictors and the estimation of appropriate standard errors. In the models, associations with false positive and false negative outcomes were estimated using correct judgments as the reference category. Unconditional models for each observation were fit, providing initial estimates of the variance across interviewers in the accuracy of the observations. Next,

models were fit that included fixed effects for the interviewer attributes (results not shown). Finally, full models including fixed effects for all model variables were fit.[2] All models were estimated using PROC GLIMMIX in SAS 9.2.

RESULTS

For ease of presentation, the following shorthand is used for the three outcomes: EMPLOYED for "one or more employed adults in the household;" OVER65 for "all residents of the household are over the age of 65;" and CHILDREN for "the presence of children under the age of 6 in the household."

Table 1 presents the false positive, false negative, and overall agreement rates for each of the three observations. Overall agreement was high, ranging from 89.2% for OVER65 to 80.7% for EMPLOYED. However, error remains and varies depending on the observation. The false negative rate is 50.5% and 41.9% for the CHILDREN and OVER65 observations respectively, but only 9.9% for EMPLOYED. Conversely, the false positive rate is 44.5% for EMPLOYED but less than 10% for the other two observations. The distributions of the dependent variables for the multilevel models are also shown in Table 1.

Moving to the multivariate results, estimating empty or unconditional models for the three observations under analysis produced significant variability in the intercepts for false positive and false negative outcomes (results not shown). This indicates that the distributions of the dependent variables changes substantially across interviewers, i.e., some interviewers are more likely to produce false positives while others are more likely to produce false negatives. For comparison to the full models (see Table 2), the interviewer intraclass correlation coefficients (ICCs) for the empty models are as follows: EMPLOYED (FP=0.05, FN=0.26); OVER65 (FP=0.24, FN=0.08); and CHILDREN (FP=0.14, FN=0.06).

Table 2 presents coefficients for the full set of model covariates for all three outcomes. A set of key findings are highlighted. Reviewing the results for all three models reveals that the interviewer attributes had little to no impact on agreement. Neither the gender of the interviewer nor the number of years worked on the NHIS were significantly associated with agreement for any of the observations. Education level and whether or not the interviewer is an intermittent employee had only minor impacts in the OVER65 and EMPLOYED models respectively.

Focusing on the characteristics of households, strong, but varied, impacts of interviewers' assessments of the physical condition of the sample unit are observed. Judging a property to be in less than excellent condition resulted in increased odds of a false negative judgment on the EMPLOYED observation. Lacking other visual clues, a more dilapidated structure may provide the impression that the occupants are not employed. At the same time, assessments of very poor/poor or fair condition reduced the odds of a false negative judgment on the OVER65 observation. And finally, assessments of less than excellent increased the odds of a false positive judgment with the CHILDREN observation.

A second household characteristic significantly associated with all three outcomes was whether or not the residence was in a multi-unit structure. If so, a significant increase in the odds of a false negative judgment is observed for the EMPLOYED measure. Interviewers appear to make assumptions that adults residing in apartment complexes, senior housing, or other types of multi-unit structures are less likely to be employed. The multi-unit structure measure has error-reducing properties in the OVER65 model. While there is no association with false negatives relative to agreement, households in multi-unit structures were associated with reduced odds of a false positive judgment. Finally, multi-unit structures produce counterintuitive results for the CHILDREN outcome. While reducing false positives, households in multi-unit structures were associated with increased odds of false negative judgments.

Like other studies, significant associations between the neighborhood-level measures and error in interviewer judgments emerged, although the patterns are not always intuitive. As expected, increased odds of a false positive for the CHILDREN measure are observed when the household is located in a neighborhood with a higher proportion of children under the age of 5. Without the aid of other visual clues, interviewers working areas with more young children may erroneously judge a given household to have young children. However, increased odds of false negatives, relative to agreement, are also observed in neighborhoods with higher proportions of young children. This is less intuitive. This same pattern emerges for the OVER65 model. Increased odds of both false positive and false negative judgments are observed when the household is located in a neighborhood with a greater proportion of older adults. Finally, the results for the EMPLOYED model were also perplexing. Households in neighborhoods with higher poverty rates were associated with increased odds of a false positive, while no association was observed between this measure and false

[2] Although some interviewers worked a small number of cases (i.e., less than 10), all were retained in the analyses presented here. Consistent with simulation studies (see Bell et al., 2010), a sensitivity analysis revealed few to any substantive changes in the results when interviewers working less than 10 or 20 cases were removed.

negatives. The tract-level vacancy rate measure also produced mixed results consistent with the age-related, tract-level measures. Households in neighborhoods with the highest vacancy rates had both increased odds of a false positive and false negative judgment.

Finally, the strongest and most persistent associations with each of the three observations involved the measure capturing the outcome of the attempt on which NOI was completed. For all three observations, the completion of NOI during an attempt in which an interview was completed, compared to a noncontact, greatly reduced the odds of a false negative judgment, relative to agreement. And while the associations were not as strong, we observed the same pattern for attempts resulting in a contact but no interview. Outcomes resulting in contact also greatly reduced the odds of a false positive judgment for the EMPLOYED and CHILDREN observations. Unfortunately, these results are essentially an artifact of the measurement process. If an interviewer makes contact with a household and is able to complete the interview on that attempt, the NOI instrument will not launch until after the interview.[3] Interviewers are not instructed to open the NOI instrument prior to attempting contact with the household.[4]

DISCUSSION

Consistent with prior research, there was reasonably high agreement between the interviewer observations and data collected in the NHIS. But like those studies, significant interviewer error was present, appears to be systematic in nature, and varies by observation. Not surprisingly then, some model covariates had differential impacts on the error across the three outcomes of interest, increasing or decreasing one type of error for one observation while having the opposite or no effect for another. Certain visual clues may be beneficial for some observational questions while reducing clarity for others.

Related to the last point, interviewer characteristics included in the models had minor effects at best on the error present in the observations, reinforcing the idea that the error has much more to do with the complexity of the task. To further complicate matters, the strongest predictors of error were non-modifiable features such as the physical condition of the sample unit, whether or not the household is located in a multi-unit structure, or the outcome of the attempt during which the observation was recorded. The latter is particularly concerning since it violates a key objective of data collection: ensure uniformity of the measurement process. This could be addressed, but only with changes in field procedures and instrument programming.

Finally, neighborhood-level features are important for understanding the error properties of interviewer judgments of sample unit characteristics. When the sample address yields little in the way of useful information, interviewers fall back on their knowledge of the areas in which they are working, what West and Kreuter (2013) refer to as the representativeness heuristic. As shown, this produced mixed results, sometimes reducing one form of error while increasing another.

Limitations

There are limitations to the analyses presented here. First, the questions used in the survey instrument are not worded the same as the questions in the NOI. Therefore, lack of agreement may be an artifact of question wording, not the inability of interviewers to make accurate observations. Second, agreement could only be assessed using data from interviewed households. Little is known about the error present in observations of nonresponding households. And three, interviewers are instructed to attempt contact with the household before opening NOI. Observations recorded during attempts resulting in contact afford interviewers greater opportunities to observe and absorb visual cues that may enhance the accuracy of their assessments. Hence, the measurement process was not applied consistently across all households.

Discussion Questions

How can the accuracy of interviewer observations be improved? What types of training techniques should be applied? What types of analyses could be done to identify important visual cues that may improve the accuracy of their observations?

What are possible explanations for the counterintuitive associations between the neighborhood-level variables and agreement?

---

[3] Sensitivity analyses in which cases where contact was made on the attempt the observation was recorded were removed produced minor changes in the size and significance of coefficients. For the most part, the remaining key findings presented here held in these analyses.

[4] It is difficult to understand how complete agreement is not achieved during these attempts. What is more surprising is that completion of the observations during attempts resulting in an interview actually increased the odds of a false positive judgment for the OVER65 observation. This suggests error in the observational data, contact history data, or both. It is possible, for example, that interviewers making multiple attempts on a household on the same day may not record earlier attempts in the Contact History Instrument (CHI), especially if they end in a noncontact. Past analyses of CHI data have suggested that noncontact outcomes are under-recorded (Bates et al., 2010).

Given the error present in these observations, what are the implications for responsive design strategies (e.g., sample balancing, response propensity models) and post-survey nonresponse adjustments? How can these impacts be assessed?

Should field procedures be altered to ensure that the observations are always recorded prior to making a contact attempt on the household? How could this be done efficiently, especially given that listing operations for the NHIS will end in 2016?

REFERENCES

Bates, N., J. Dahlhamer, P. Phipps, A. Safir, and L. Tan. 2010. "Assessing Contact History Paradata Quality across Several Federal Surveys." Pp. 91-105 in *Proceedings of the 2010 Joint Statistical Meetings*.

Bell, B., G. Morgan, J. Kromrey, and J. Ferron. 2010. "The Impact of Small Cluster Size on Multilevel Models: A Monte Carlo Examination of Two-Level Models with Binary and Continuous Predictors." Pp. 4057-4067 of the *Proceedings of the Joint Statistical Meetings*.

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

Little, R. and S. Vartivarian. 2005. "Does Weighting for Nonresponse Increase the Variance of Survey Means?" *Survey Methodology*, 31:161-168.

Peytchev, A., S. Riley, J. Rosen, J. Murphy, and M. Lindblad. 2010. "Reduction of Nonresponse Bias in Surveys through Case Prioritization." *Survey Research Methods*, 4(1): 21-29.

Sinibaldi, J. 2010. "Measurement Error in Objective and Subjective Interviewer Observations." National Centre for Social Research, London, UK.

Sinibaldi, J., G. Durrant, and F. Kreuter. 2013. "Evaluating the Measurement Error of Interviewer Observed Paradata." *Public Opinion Quarterly*, 77: 173-193.

Tabuchi, T., F. Laflamme, O. Phillips, M. Karaganis, and A. Villeneuve. 2009. "Responsive Design for the Survey of Labour and Income Dynamics." In *Proceedings of Statistics Canada Symposium 2009, Longitudinal Surveys: From Design to Analysis*.

Walsh, R., J. Dahlhamer, and N. Bates. 2013. "Assessing Interviewer Observations in the NHIS." *Proceedings of the Joint Statistical Meetings.*

West, B. 2013. "The Effects of Errors in Paradata on Weighting Class Adjustments: A Simulation Study." Pp. 341-366 in F. Kreuter (ed.), *Improving Surveys with Paradata: Analytic Uses of Process Information*. New York: Wiley.

West, B. and F. Kreuter. 2013. "Factors Affecting the Accuracy of Interviewer Observations: Evidence from the National Survey of Family Growth." *Public Opinion Quarterly*, 77(2): 522-548.

West, B., F. Kreuter, and M. Trappmann. 2014. "Is the Collection of Interviewer Observations Worthwhile in an Economic Panel? Survey: New Evidence from the German Labor Market and Social Security (PASS) Study." *Journal of Survey Statistics and Methodology*, 2: 159-181.

Table 1. Agreement between Three Observational Measures and Survey Data: NHIS, 2013, Months 1-6

EMPLOYED:		
Survey Question:	"Which of the following {was person} doing last week?" 1. Working for pay at a job or business; 2. With a job or business but not at work; 3. Looking for work; 4. Working, but not for pay, at a family-owned job or business; or 5. Not working at a job or business and not looking for work. Answers of 1, 2, or 4 were coded as employed.	
Observation Question:	Based on your observation, would you say at least one adult resident of the sample unit is employed? Yes or no.	
	Agreement Statistics:	**Dependent Variable:**
	Overall agreement: 80.7%	Agreement (ref): 80.7%
	False positive rate (FP / FP + TN): 44.5	False positive: 12.1
	False negative rate (FN / FN + TP): 9.9	False negative: 7.2
OVER65:		
Survey Question:	What is {person's} age?	
Observation Question:	How old would you estimate the residents of the sample unit to be? 1. All occupants under the age of 30; 2. All occupants over the age of 65; or 3. Other age composition. The responses were recoded to all occupants over the age of 65 versus other.	
	Agreement Statistics:	**Dependent Variable:**
	Overall agreement: 89.2%	Agreement (ref): 89.2%
	False positive rate (FP / FP + TN): 5.5	False positive: 4.7
	False negative rate (FN / FN + TP): 41.9	False negative: 6.2
CHILDREN:		
Survey Question:	What is {person's} age?	
Observation Question:	Based on your observation, does the SAMPLE UNIT have…indication that children under 6 (including babies) may live at the unit (visible toys, car seat, strollers, outdoor swing/play set for example)? 1. Yes; 2. No; or 3. Unable to observe the sample unit. Answers of 3 were removed from the analysis.	
	Agreement Statistics:	**Dependent Variable:**
	Overall agreement: 85.9%	Agreement (ref): 85.9%
	False positive rate (FP / FP + TN): 7.5	False positive: 6.4
	False negative rate (FN / FN + TP): 50.5	False negative: 7.7

FP = false positive, TP = true positive, FN = false negative, TN = true negative

Table 2. Parameter Estimates from Multilevel Multinomial Logistic Regression Models, with Random Interviewer Effects, Predicting Accuracy of Interviewer Judgments: NHIS, 2013, Months 1-6 (unweighted)

	EMPLOYED		OVER65		CHILDREN	
Independent Variable	**FP Coefficient**	**FN Coefficient**	**FP Coefficient**	**FN Coefficient**	**FP Coefficient**	**FN Coefficient**
Intercept	-1.87***	-3.40***	-3.79***	-2.31***	-3.01***	-2.69***
Interviewer Characteristics						
Education (ref=High school/G.E.D)						
Bachelor's degree or higher	-0.04	-0.03	-0.11	-0.04	0.03	-0.09
Some college	-0.08	0.09	0.36*	-0.03	0.07	-0.06
Male (vs. female)	0.07	0.16	-0.09	0.07	0.03	0.01
Years worked on the NHIS	-0.004	0.04	0.02	0.01	-0.01	0.01
Intermittent employee (vs. no)	-0.14*	-0.04	0.15	-0.12	0.04	0.08
Circumstances of the Observation						
Obs. recorded on weekend (vs. no)	0.08	0.05	0.06	0.03	-0.02	-0.05
Time of day obs. recorded (ref=afternoon)						
Morning	0.16*	-0.07	0.02	-0.01	0.05	-0.10
Evening	0.06	-0.06	-0.18	-0.04	-0.07	-0.03

Outcome of attempt on which obs. made (ref=noncontact)						
Contact, non-interview	-0.62***	-0.37***	0.38***	-1.35***	0.13	-0.49***
Contact, interview	-1.21***	-1.04***	0.27**	-1.69***	-0.61***	-1.23***
Household Characteristics						
Graffiti observed in black face (vs. no)	0.12	0.37**	0.12	0.23	0.37**	0.09
Observed condition of the sample unit (ref=excellent)						
Very poor/poor	-0.0002	1.60***	-0.08	-0.52***	0.91***	0.20
Fair	0.07	0.69***	-0.05	-0.46***	0.55***	0.10
Good	0.04	0.22*	0.003	-0.05	0.21*	0.08
Access barrier (vs. no)	0.14	0.11	0.22	-0.15	-0.10	-0.22*
In multi-unit structure (vs. no)	-0.002	0.33***	-0.31**	-0.30***	-0.31***	0.17*
Refusal statements (vs. no)	0.21**	0.18	0.04	0.35***	0.07	-0.13
Privacy concerns (vs. no)	0.02	-0.002	0.001	0.06	0.12	-0.23**
# of observations prior to current one	-0.003***	0.001	-0.01***	0.0001	-0.002	-0.001
Area Characteristics/Social Environment						
MSA status (ref=non-MSA)						
MSA, central city	-0.02	-0.06	-0.07	-0.11	-0.37**	0.18
MSA, non-central city	-0.05	-0.19	0.12	-0.21*	-0.24*	0.22*
Under 5 years of age rate (Census tract) (ref = < 5.2631578947%)						
7.8313921242% +	-0.27**	-0.27*	0.14	-0.27*	0.46***	1.00***
6.4044022948% - < 7.8313921242%	-0.04	-0.16	0.16	-0.11	0.34**	0.67***
5.2631578947% - < 6.4044022948%	-0.04	0.04	0.08	-0.02	0.18	0.33***
Over 64 years of age rate (Census tract) (ref = < 8.7068521381%)						
16.194209891% +	0.22*	-0.12	0.64***	0.75***	-0.17	-0.12
12.352941176% - < 16.194209891%	0.15	-0.18	0.40**	0.61***	0.09	-0.15
8.7068521381% - < 12.352941176%	0.14	-0.08	0.16	0.46***	0.09	-0.15
Vacancy rate (Census tract) (ref = < 5.2210526316%)						
11.877037727% +	0.20*	0.30*	0.14	0.20	0.01	-0.09
7.7228327228% - < 11.877037727%	0.11	0.20	0.03	0.10	-0.02	-0.11
5.2210526316% - < 7.7228327228%	0.06	-0.01	-0.30*	0.09	0.05	-0.18*
Poverty rate (Census tract) (ref = < 6.412501684%)						
20.936480522% +	0.50***	0.07	0.03	-0.18	0.08	-0.01
12.16054904% - < 20.936480522%	0.19*	0.11	-0.08	-0.10	0.06	-0.01
6.412501684% - < 12.16054904%	0.14	-0.02	-0.09	-0.12	-0.05	-0.01
Interviewer Variance Component (S.E.)	0.13 (0.03)	1.08 (0.11)	0.92 (0.11)	0.15 (0.04)	0.43 (0.07)	0.13 (0.04)
Interviewer ICC	0.04	0.25	0.22	0.04	0.12	0.04
AIC	21955.02		14974.48		18204.64	
N	19,206		19,305		18,911	

*p < .05; **p < .01; ***p < .001

FP = false positive, FN = false negative, S.E. = standard error