University of Nebraska - Lincoln
DigitalCommons@University of Nebraska - Lincoln

Sociology Department, Faculty Publications | Sociology, Department of

1-1-2013

Paradata for Nonresponse Adjustment

Kristen Olson
University of Nebraska - Lincoln, kolson5@unl.edu

Follow this and additional works at: http://digitalcommons.unl.edu/sociologyfacpub

Olson, Kristen, "Paradata for Nonresponse Adjustment" (2013). *Sociology Department, Faculty Publications.* Paper 203.
http://digitalcommons.unl.edu/sociologyfacpub/203

This Article is brought to you for free and open access by the Sociology, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Sociology Department, Faculty Publications by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Published in *The Annals of the American Academy of Political and Social Science,* vol. 645 no. 1 (January 2013), pp. 142-170; doi: 10.1177/0002716212459475

Copyright © 2013 American Academy of Political and Social Science.
Used by permission.

Paradata for Nonresponse Adjustment

Kristen Olson

Kristen Olson is an associate professor of sociology and survey research and methodology at the University of Nebraska–Lincoln. Her research focuses on survey nonresponse, paradata, and interviewer effects in sample surveys.

Abstract
Survey researchers and practitioners use nonresponse adjustment weights to mitigate the effects of survey nonresponse on sample estimates. One challenge in creating these weights is finding useful auxiliary data that predict both the probability of participating in the survey and the survey variables of interest. This article reviews the use of paradata for nonresponse adjustment. Five different types of paradata are considered: neighborhood observations, observations of the sampled housing unit, observations of persons in the sampled housing unit, call records, and observations about the interviewer-householder interaction. Empirical evidence about the predictive value of these paradata for predicting both participation and survey variables is examined. Challenges of using these paradata are also identified, along with outstanding issues and opportunities related to the use of paradata for nonresponse adjustment.

Keywords: nonresponse, weighting, paradata, nonresponse adjustment, auxiliary variables

With unit nonresponse rates rising systematically over the past three decades (see Battaglia et al. 2008; de Leeuw and de Heer 2002; Steeh 1981; Steeh et al. 2001), survey practitioners have increasingly relied on two options to reduce the effects of nonresponse on survey estimates. The first has been to introduce new design features—financial incentives, additional interviewers, different modes of administration, and greater efforts to recruit previously uncontacted or uncooperative sample members. But these approaches have led to increased costs and lengthened field periods (Groves et al. 2002). The second such effort—using statistical adjust-

ment methods such as weighting—has led to increased reliance on auxiliary data (Kreuter et al. 2010; Little 1982, 1986). The goal of nonresponse adjustment weighting is to minimize nonresponse bias and potentially reduce the variance of sample estimates, or at least not to inflate the variance of estimates. The challenge to practitioners is how to identify auxiliary variables to meet the joint goals of bias and variance reduction.

Identifying variables that are available for respondents and nonrespondents requires looking outside the survey questionnaire for additional information. A good auxiliary nonresponse adjustment variable is one that strongly predicts the likelihood of participating in the survey and is also strongly associated with the survey variables of interest (Groves 2006; Kalton and Flores-Cervantes 2003; Little 1986; Little and Vartivarian 2003, 2005). Sample-based nonresponse adjustment methods use information available on both respondents and nonrespondents to derive case weights that can be applied to sample members to compensate for bias introduced by nonresponse (Kalton 1983). These weights are an estimate of the sampled unit's "response propensity," an unknown probability of participating in the survey that is assessed whether the unit actually participated in the survey. What Couper (1998) calls "paradata"—information on respondents and nonrespondents that is available from sources outside the survey questionnaire itself—is increasingly looked to as a source of auxiliary information for creating unit nonresponse weights.

No single formal definition exists for paradata (Couper and Lyberg 2005; Kreuter and Casas-Cordero 2010; Lyberg 2009), although paradata are distinct from metadata, which are global information about a survey such as the response rate or coverage rate (Lyberg 2009; but see Scheuren [2000] on use of the term "macro paradata"). Originally defined as process data that come "for free" as an offshoot of a computerized survey environment (Couper 1998, 2005), paradata have expanded to include any information recorded by interviewers that is external to data from the questionnaire itself (Groves and Heeringa 2006; Groves et al. 2009; Kreuter and Casas-Cordero 2010; Lepkowski, Axinn, et al. 2010). Paradata in this context include call records documenting the day, time, and outcome of each contact attempt; the mode of contact attempt; and any observations made by interviewers or field listers about the sampled unit or its neighborhood (Chearo and Van Haitsma 2010; Groves and Heeringa 2006; Kennickell 2000; Kreuter et al. 2010; Taylor 2008), and are available for both respondents and nonrespondents. Paradata collected during a survey interview, including keystroke files, timestamp data, and item-missing data indicators (Couper 1998), are available on respondents only and cannot be easily used for unit nonresponse adjustment.

This article examines the use of paradata collected about or during the survey recruitment process for unit nonresponse adjustment weights. Although paradata can be used when imputing for item nonresponse, this will not be the focus of this article. Paradata collected during the question-answering process also will not be considered here. The article begins by reviewing the characteristics of useful auxiliary variables for unit nonresponse adjustment and then examines the strengths and weaknesses of paradata for purposes of nonresponse adjustment before mov-

ing on to identify the particular challenges that arise from using paradata to create nonresponse adjustment weights. Finally, it explores outstanding issues and opportunities related to the use of paradata for nonresponse adjustment.

Characteristics of Useful Auxiliary Variables

A useful statistical conceptual scheme for understanding survey nonresponse is the "stochastic" model (Bethlehem 2002; Groves 2006; Lessler and Kalsbeek 1992). Under this model, all sampled units have a nonzero probability of participating in the survey, a "response propensity," whether or not they actually participate. This model contrasts with the "deterministic" model, in which all sampled units are assumed in binary fashion to be either respondents or nonrespondents. The nonresponse bias of an unadjusted respondent mean ($\bar{Y}_R$) under the stochastic model can be expressed as a function of the covariance between the response propensity (P) and the survey variables of interest (Y) and the average response propensity ($\bar{P}$) (see Bethlehem 2002; Kreuter et al. 2010):

$$\text{Bias}(\bar{Y}_R) \approx \frac{Cov(P,Y)}{\bar{P}} = Corr(P,Y) * SD(Y) * CV(P)$$

In this case, the $Cov(P,Y)$ covariance term arises when Y is a cause of P (sometimes called a nonignorable nonresponse or a unit not missing at random) or when a third variable Z is a common cause of both P and Y (sometimes called an ignorable nonresponse or a unit missing at random [see Groves 2006; Little and Rubin 2002]). The search for good auxiliary variables is thus a search for good Z variables (common causes).

Since a sampled unit's response propensity is not observed, it must be estimated from information obtained from both respondents and nonrespondents. A common method for estimating response propensities uses a logistic regression model to predict the dichotomous outcome of survey participation versus nonparticipation as a function of auxiliary variables (Little 1982, 1986; Raghunathan 2004). Predicted probabilities for each sampled unit estimated from this model constitute estimates of response propensities, which are then transformed into response propensity weights by taking their inverse (Little and Rubin 2002; Raghunathan 2004). A second method derives class rather than individual weights by creating mutually exclusive and exhaustive groups from auxiliary variables available on respondents and nonrespondents (Kalton 1983) and uses the inverse of the response rate in each class as the weight. Both methods require auxiliary variables that are available for both respondents and nonrespondents.

An ideal auxiliary variable for sample-based nonresponse adjustment has four characteristics (Lessler and Kalsbeek 1992; Little 1982, 1986; Little and Vartivarian 2005; Sarndal and Lundstrom 2005). First, nonmissing values of the variable must be available for both respondents and nonrespondents. Second, the variable should be measured completely and without error for all sampled units. Third, this variable should be strongly associated ($R^2 \geq .48$) with important survey variables of in-

terest (Little and Vartivarian 2005). Fourth, the auxiliary variable should also be a strong predictor of survey participation, thus reducing nonresponse bias in an adjusted mean (Little and Vartivarian 2005). Auxiliary variables that predict only survey participation and not the survey variables of interest inflate the variance of the mean by a factor proportionate to the coefficient of variation of the weights without reducing the nonresponse bias of the estimate (Kish 1992; Little and Vartivarian 2005).

It is surprisingly difficult to find auxiliary variables that meet all these criteria. The challenge for practitioners mostly lies in identifying useful auxiliary variables that jointly predict survey participation and the survey variables of interest. In multipurpose studies, this difficulty is amplified because there are many survey variables of interest. Traditionally, data from the sampling frame have been used as adjustment variables (Abraham, Maitland, and Bianchi 2006; Bethlehem 1988, 2002; Kalton and Flores-Cervantes 2003; Sarndal and Lundstrom 2008). Given the paucity of information available on most U.S. household frames, however, survey researchers are increasingly turning to other sources for these auxiliary variables. One source of information is paradata.

Categories of Paradata

Paradata collected during or about the survey recruitment process can be grouped into five categories. The first three categories refer to characteristics that are fixed over the course of the data collection period. These include observations made by interviewers about (1) the sampled unit's neighborhood, (2) the sampled unit's housing unit, and (3) persons in the sampled housing unit. The final two categories constitute data that change over the course of the data collection period. These include (4) call record information collected as part of the sample management system and (5) observations recorded by interviewers about their interaction with the sampled household at each contact. Only the fourth type of paradata, call record information, is consistent with a definition of paradata as process data. Nonetheless, data in the other four categories are frequently collected as part of call records or contact history forms in many surveys or disseminated through public use paradata files.

Observations about sampled individual's neighborhood

Interviewer observations about neighborhood characteristics for a sampled unit are collected in large-scale face-to-face surveys. Neighborhood observations, for example, are routinely collected in the European Social Survey (ESS), the Los Angeles Family and Neighborhood Study (LAFANS), the Survey of Consumer Finances (SCF), the National Survey of Family Growth (NSFG), the British Crime Survey (BCS), and the Health and Retirement Study (HRS) (see Callens and Croux 2003; ESS 2008; Groves and Heeringa 2006; HRS 2009; Kennickell 2000; Kreuter, Lemay,

and Casas-Cordero 2007; Lynn 2003; Peterson, Sastry, and Pebley 2007). Some surveys, such as the Project on Human Development in Chicago Neighborhoods (Interuniversity Consortium for Political and Social Research 2010) or LAFANS (Sastry and Pebley 2004), are specifically designed to collect neighborhood observations to evaluate the effects of neighborhood conditions on important social and economic outcomes. In these surveys, field interviewers observe and record characteristics such as the mix of residential and business units, the upkeep of the neighborhood, the wealth of the neighborhood, or the presence of trash and graffiti in the neighborhood surrounding the sampled housing units.

Neighborhood observations are theoretically available for all sampled persons, including noncontacted units, and are easily added to any in-person survey. However, ecological data for telephone, mail, or Web surveys may have to be obtained through other sources, such as merging census tract information to a sample frame (Campanelli, Sturgis, and Purdon 1997; Johnson et al. 2006; Kennickell 2005) or purchasing data from commercial sources (Schräpler, Schupp, and Wagner 2010). Field interviewers frequently visit a neighborhood before attempting initial contact with a sampled household (Morton-Williams 1993), and their observations are thus unobtrusive and require no interaction with the sampled unit. Not surprisingly, interviewers report little difficulty making these types of assessments (Lynn 2003).

Interviewers may fail to complete the observations altogether, of course (Matsuo, Billiet, and Loosveldt 2010); and different interviewers may perceive neighborhoods differently (Peterson, Sastry, and Pebley 2007; Raudenbush and Sampson 1999), which can lead to measurement error in the observational data. Item nonresponse rates for neighborhood observations vary greatly across surveys, ranging from under 5 percent (Casas-Cordero 2010; Lynn 2003) to more than 50 percent (Kreuter, Lemay, and Casas-Cordero 2007; Matsuo, Billiet, and Loosveldt 2010). One measure of the quality of neighborhood observations is the similarity in observations among different interviewers for the same neighborhood. Raudenbush and Sampson (1999, 7) found "modest" interobserver correlation coefficients ranging from 0.13 to 0.36 for a variety of scales created from neighborhood observations. Such measurement error naturally attenuates the association between the observations and survey outcomes of interest (Fuller 1987; Raudenbush and Sampson 1999). Attenuation can be reduced by increasing the number of neighborhood observers (Raudenbush and Sampson 1999), although using multiple raters quickly increases costs (Sastry and Pebley 2004).

Studies of "neighborhood effects" on social, economic, psychological, and health outcomes are becoming increasingly popular (Gee and Devon 2004; Leventhal and Brooks-Gunn 2003; Raudenbush and Sampson 1999), and recent reviews have shown that neighborhood circumstances affect a variety of outcomes, including crime-related behaviors, school achievement indicators, and, to a more limited degree, mental health measures (Leventhal and Brooks-Gunn 2000; Sampson, Morenoff, and Gannon-Rowley 2002). These findings suggest that neighborhood observations are particularly useful for making nonresponse

adjustments in crime surveys or studies of educational achievement among children, as neighborhood observations are most useful for nonresponse adjustment when the association between observations and survey outcomes is large ($R^2 \geq .48$). Such an effect size, however, is not often found (Diez Roux 2001), which may require the development of new neighborhood observations for purposes of nonresponse adjustment that focus on variables related to socioeconomic status (Leventhal and Brooks-Gunn 2000), alcohol or tobacco use, and the frequency of adults and children playing together (Sampson, Morenoff, and Gannon-Rowley 2002, 465–71).

Neighborhoods are hypothesized to affect survey cooperation through mechanisms such as urbanicity, population density, crime, and social disorganization (Groves and Couper 1998, 176–78). In the ESS, for example, observations about the presence of litter or vandalism in the neighborhood are significantly associated with survey participation (Kreuter, Lemay, and Casas-Cordero 2007) as well as respondent contactability and cooperation (Matsuo, Billiet, and Loosveldt 2010). The strength of these associations varies tremendously across countries, however. Interviewer observations about an area's safety are also significantly associated with both contactability and cooperation in six surveys in the United Kingdom (Durrant and Steele 2009) and with participation in the main interview in the NSFG (Lepkowski, Mosher, et al. 2010). Although aggregate information on tracts or zip code areas is often used for nonresponse adjustment in both face-to-face and telephone surveys (Campanelli, Sturgis, and Purdon 1997; Johnson et al. 2006), paradata about the neighborhood gathered by interviewers or listers are infrequently used for adjustment purposes (Kreuter, Lemay, and Casas-Cordero 2007; see also review in Casas-Cordero 2010), even in surveys that use interviewers to capture neighborhood characteristics (Sastry and Pebley 2003).

The form of models used to estimate response propensities from neighborhood data has received little empirical attention in the survey methodological literature. While studies of ecological effects on social, economic, and health outcomes use hierarchical or random effects models, propensity models that draw on area observations often ignore potential clustering effects. Those studies that do use hierarchical models tend to look at area characteristics as predictors of participation, contactability, or cooperation, but do not use estimates for adjustment purposes (Callens and Croux 2003; Durrant and Steele 2009; O'Muircheartaigh and Campanelli 1999; Steel, Tranmer, and Holt 2003). Although neighborhood observations have been used in adjustment models, little attention has been paid to the nested nature of the observations and how models should account for geographic clustering.

Neighborhood observations can be collected relatively easily in face-to-face surveys, but they are virtually impossible to obtain (without incurring huge costs) in other survey modes. Neighborhood observations may be particularly effective as adjustment variables in studies about crime or educational achievement, but item nonresponse in the neighborhood observations and variability across observers potentially limits their usefulness. When multiple sampled units are in the same

neighborhood, it may be necessary to account for clustering in propensity models, although this remains an empirical question.

Observations about the sampled housing unit

Observations about the sampled housing unit eliminate analytic issues related to clustering while still being fairly unobtrusive. Such observations typically include an assessment of the housing unit's quality relative to others in the neighborhood, whether the unit is in a multiunit structure, if it has access impediments, and whether it is in a safe condition. These sorts of observations have been made on surveys as diverse as the American National Election Studies (ANES), the British Survey of Social Attitudes (BSSA), the ESS, the NSFG, the SCF, the BCS, the Residential Energy Consumption Survey (RECS), and the National Survey of Drug Use and Health (NSDUH) (see Groves and Heeringa 2006; Kennickell 2000; Kreuter et al. 2010; Lepkowski, Mosher, et al. 2010; Lynn 2003; Peytchev and Olson 2007; Sinibaldi 2008, 2010; U.S. Energy Information Administration 2009; Wang et al. 2005). Some surveys also make housing unit observations specific to the content of the questionnaire, such as observations about the presence of political lawn signs in the ANES (National Election Studies 2004) or the presence of bars on the windows or burglar alarms in the BCS (Lynn 2003). As with neighborhood observations, housing unit observations can be made in face-to-face surveys but not in telephone, mail, or Web surveys; and they can easily be made for all sampled units, including noncontacts and refusals, for relatively little additional marginal cost (Lynn 2003). Unlike area characteristics, however, proxies cannot be directly obtained from census data.

Since interviewers in this case serve as reporters, correlated interviewer variance (Kish 1962) and item-missing data on the housing unit observations (Kennickell 1999; Lynn 2003) is possible. In addition, missing data on some housing units can be a problem when they are located in a gated community or locked building and cannot be directly observed. Surprisingly few surveys that collect housing unit observations report missing data rates on these observations. Nonetheless, in the case of the 2004 ANES, observation of whether the sampled person lived in a single-family home was complete for all respondents and nonrespondents (Peytchev and Olson 2007). Similarly, item nonresponse rates for housing unit observations were less than 2 percent in the BCS (Lynn 2003). In contrast, observations of the same characteristic were missing for all sampled units in Norway on round one of the ESS, although item-missing data rates were lower in other countries and varied by type of housing unit observation (Kreuter, Lemay, and Casas-Cordero 2007). Little empirical evidence exists about correlated interviewer variance and other forms of measurement error.

Characteristics of housing units often predict both contactability and cooperation in sample surveys. Ubiquitous predictors of contactability in face-to-face surveys include whether the sampled unit is in a multiunit structure versus a single-family home, or is in a locked building or has other access impediments (Blohm,

Hox, and Koch 2006; Campanelli, Sturgis, and Purdon 1997; Groves and Couper 1998; Kennickell 2003; Lynn 2003; Maitland, Casas-Cordero, and Kreuter 2009; Sinibaldi 2008; Stoop 2005). The condition of the housing unit relative to others in the area is also significantly associated with noncontact and refusal rates (Durrant and Steele 2009; Lynn 2003; Sinibaldi 2008). In the ESS and ANES, however, the correlation between being in a multiunit structure versus a single-family house and survey participation is small, explaining less than 4 percent of the variance. A two-stage approach to modeling survey participation, in which contact and cooperation are modeled separately, may yield stronger associations, especially in contact models (Groves and Couper 1995). Few other studies report the degree to which variation in response propensity is explained by housing unit characteristics alone.

Certain survey variables have been shown to be related to characteristics of the housing unit, such as socioeconomic status, energy use, technological sophistication, and exposure to environmental hazards. For example, single-family detached housing units are more likely to have occupants who own rather than rent the housing unit; those occupants also tend to have higher median incomes and higher housing costs than persons living in multiunit housing (U.S. Census Bureau 2008). Persons living in multiunit structures are more likely to be heated by electricity than natural gas, to use a personal computer, and to own more than one television (U.S. Energy Information Administration 2008). Housing unit characteristics also affect exposure to lead paint and environmental triggers for asthma (Jacobs et al. 2002; Saegert et al. 2003). In the ANES, the strongest associations for living in a single-family home and survey variables were observed for number of years living in the home ($r = .29$), age ($r = .22$), and putting off medical treatment ($r = -.20$), but political activity variables had much weaker associations (Peytchev and Olson 2007). In round one of the ESS, the correlation between housing unit observations and survey variables varied tremendously across eight survey variables and did not exceed an absolute value of $r = .3$ (Kreuter, Lemay, and Casas-Cordero 2007). Thus, housing unit characteristics predict many survey outcomes, although the strength of the association is often quite weak.

In sum, with sufficient monitoring, housing unit observations offer a potentially rich source of information for nonresponse adjustments. Unfortunately, observations about housing units can be made only for in-person surveys. Analytically, housing unit observations are available for each sampled unit and are free from concerns about neighborhood clustering, so that standard logistic regression methods and weighting class adjustments can be used. Although housing unit observations can be recorded for all sampled units, contacts and noncontacts, interviews and refusals, they often have modest predictive power for important survey variables. Although significantly associated with survey participation, housing unit observations tend to explain little variance in overall survey participation.

Observations about members of the housing unit

Potentially rich sources of paradata for nonresponse adjustment come from observations about the members inhabiting selected housing units. These observations fall into two categories: observations on demographic characteristics and observations that proxy for key survey variables. If measured without error, interviewer observations about characteristics of housing unit members will be identical to the respondent's report of those same characteristics. Information from interviewer observations of key survey variables could be used directly to assess nonresponse bias for statistics computed from those characteristics. Since a strong association with the survey outcome variables is key to being a useful auxiliary variable (Kreuter and Olson 2011; Little and Vartivarian 2005), proxy observations of important survey variables meet that goal, at least in theory. An interviewer-observed Y variable would also be the most effective variable for creating weighting classes that are homogeneous on that characteristic of interest (Kalton 1983; Kalton and Flores-Cervantes 2003).

Demographic characteristics are ubiquitously measured on sample surveys and are often used in nonresponse adjustments because they provide information on individuals and subgroups whose representation in the final sample pool is important (Bethlehem 2002; Cobben and Schouten 2008; Groves 2006). The ESS, the GSS, the SCF, the BCS, the National Health Interview Survey (NHIS), and the National Health and Nutrition Examination Survey (NHANES) incorporate observations of age (Matsuo, Billiet, and Loosveldt 2010; Sinibaldi 2010), gender (Matsuo, Billiet, and Loosveldt 2010), race (Burns et al. 2001; Lynn 2003; Saperstein 2006; Smith 1997, 2001), income (Burns et al. 2001; Kennickell 2000), and the presence of non-English speakers (Bates, Dahlhamer, and Singer 2008; National Center for Health Statistics 2009) for both respondents and nonrespondents. Although many surveys use demographic characteristics reported in the survey questionnaire for poststratification, these variables are collected in the survey questionnaire and will not be considered here.

Interviewer observations of demographic characteristics are not free from item nonresponse or measurement errors, of course. Item nonresponse on observations of demographic characteristics can vary from virtually nonexistent to fairly high. For example, item nonresponse rates for observations about gender and age in the 2004 ESS range from no missing data in Finland to almost 65 percent missing in Germany (Matsuo, Billiet, and Loosveldt 2010). Interviewer observations of age in the United Kingdom's National Survey of Sexual Attitudes and Lifestyles (NatSAL) were missing for 37 percent of households (Sinibaldi 2010). In the BCS, observations of race were missing on 76 percent of the cases, primarily on respondents (Lynn 2003).

Measurement error in observations of demographic characteristics has been investigated by comparing interviewer observations to reports from respondents themselves. These comparisons have focused largely on observations about household members' race, in which interviewer observations match about 95 percent of

white and black respondents' reports of race, but only around 50 percent for respondents who identify with other races (Hahn, Truman, and Barker 1996; Saperstein 2006; Smith 1997, 2001). The match rate is also affected by the interviewer's race (Hill 2002). Less is known about measurement error in observations of householder age and gender. In the NatSAL, about 76 percent of the observations of age matched the reported age within 10-year age groups (Sinibaldi 2010). Although interviewers are frequently instructed to observe respondent gender, surprisingly little research has been conducted on the accuracy of these assessments (Kenney McCulloch, Kreuter, and Calvano 2010).

Demographic characteristics have frequently been examined as predictors of survey participation. Lower response rates have been found among the young and the elderly, men, households of lower and higher socioeconomic status, racial-ethnic-language minorities, immigrants, and persons with lower levels of education (Brehm 1993; Callens and Croux 2003; Campanelli, Sturgis, and Purdon 1997; Collins et al. 2000; Goyder 1987; Goyder, Warriner, and Miller 2002; Green 1996; Groves and Couper 1998; Holbrook, Krosnick, and Pfent 2008; Kennickell 1999; Korinek, Mistiaen, and Ravallion 2005, 2007; Moonesinghe, Mitchell, and Pasquini 1995; Purdie et al. 2002; Smith 1984; Stoop 2005; Voogt 2004). Although frequently examined, these demographic characteristics explain little of the variation in response propensity (Kreuter and Olson 2011; Kreuter et al. 2010). Additionally, a recent meta-analysis revealed that a difference between respondents and nonrespondents on demographic variables may not necessarily indicate nonresponse bias on key survey estimates (Peytcheva and Groves 2009).

A growing set of observations about members of the housing unit come from interviewer-recorded proxy measures of characteristics that form key survey estimates. For example, on the NSFG, interviewers are asked to report whether the selected person is in an active sexual relationship and whether there are children in the household (Kreuter et al. 2010; West 2010). At the National Centre for Social Research in the United Kingdom, interviewers in the NatSAL dress rehearsal were asked to record whether there was evidence of smoking in the household, as well as the two characteristics observed in the NSFG (Sinibaldi 2010). Related to these characteristics is whether a contacted household refuses because of a topic-related issue, such as lack of interest in politics in the ANES (Peytchev and Olson 2007), or health-related reasons in the NHIS (see Dahlhamer and Simile 2009).

As the observation of proxies of important survey variables is quite new, very limited research has been conducted on the quality of these auxiliary variables. However, one measure of quality is item nonresponse. Item nonresponse on these observed characteristics can arise from multiple sources. First, housing units that have never been contacted will not have an observation about members of the sample unit or a selected respondent; these observations are available only for contacted households. Second, interviewers may choose not to complete their observation forms owing to perceptions that the observation is superfluous or irrelevant (Lynn 2003) or because of discomfort about making the obser-

vations (Sinibaldi 2010). Third, question wording and computerization of the observation instrument may encourage item nonresponse in the observations. The NSFG and NatSAL, for example, measure identical characteristics about sampled households (presence of children and sexual relationships), but the NSFG had no item nonresponse, whereas in the NatSAL item nonresponse rates ranged from about 10 percent to almost 45 percent, although some of these item-missing cases can be attributed to noncontacts (Sinibaldi 2010). On the NSFG survey, interviewers were required to provide a valid answer to these questions;[1] on the NatSAL, interviewers were permitted to record that a characteristic was not observed, that they never met the respondent, or that they "can't say" what the answer would be.

Another measure of quality is the association between an interviewer's observation and the respondent's report of the survey variable. Evaluations of observations that are proxies of important survey characteristics show significant associations with the survey characteristics they were intended to measure (Groves, Wagner, and Peytcheva 2007; Kreuter et al. 2010; Sinibaldi 2010; West 2010). The associations, however, are far from perfect (Groves, Wagner, and Peytcheva 2007; Kreuter et al. 2010; Sinibaldi 2010; West 2010). In the NSFG, for example, 79 percent of interviewer guesses about whether a selected female was in a sexually active relationship matched reports obtained through the interview, with "yes" guesses (the selected respondent was assessed as being in a relationship) being better than "no" guesses (the selected respondent was assessed as not being in a relationship).

Householder statements of not being interested in politics were significantly, but modestly, associated with being interested in following campaigns, voting intentions, and caring about who will win the elections (Peytchev and Olson 2007). These assessments are limited because information on these characteristics is not known for the nonrespondents; whether the interviewer guesses for nonrespondents are equivalent in quality cannot be assessed through this method. Additionally, interviewers can vary substantially in the quality of their observations, leading to an increased variance in these observations and potentially limiting the usefulness of these characteristics (West 2010). What makes certain interviewers better at recording these observations than others is unknown.

The novelty of these observations of proxy survey variables means that little is known about their performance as nonresponse adjustment variables. In the NSFG, observations of sexual activity and having children in the household are significantly associated with survey participation (Groves, Wagner, and Peytcheva 2007; West 2010). However, these extra variables explain only 0.6 percent additional variance when added to existing propensity models (West 2010). In the ANES, observations about not being interested in politics were made by only 5 percent of the contacted households and were only weakly correlated ($r = -.08$) with survey participation (Peytchev and Olson 2007). Thus, these variables may not be effective for reducing nonresponse bias in survey estimates (Little and Vartivarian 2005).

Multiple implementation challenges arise for the observation of either demographic characteristics or observations of proxy survey variables. To be useful for nonresponse adjustment, the measurement error properties of these characteristics should be the same for respondents and nonrespondents (Groves and Couper 1995). Thus, recording the observations must occur before an interview is conducted as the respondent's answers may influence the interviewer's observation. Practically, the number of observations of important survey variables that can be made is limited to minimize disrupting the flow of the interview. In a multipurpose survey with dozens of key survey variables, only a few variables can be observed. Additionally, to maintain a constant measurement environment, interviewers' ability to change the observation after conducting the interview must be restricted. This restriction is much easier to accomplish in a computer-assisted interview than in a paper-and-pencil interview.

Another implementation challenge arises due to the need to train interviewers on the task of completing these observations (Groves et al. 2009; Lepkowski, Axinn, et al. 2010). This training is auxiliary to the task of contacting a case, obtaining cooperation, and administering a survey interview. Proxy observations are directly counter to the task of standardized survey interviewing in which all questions must be asked as written. As such, many interviewers and field managers resist including proxy observations of survey variables, especially for sensitive characteristics (Sinibaldi 2010).

Analytically, observations of either demographic characteristics or proxy survey variables can be made on contacted units, but may be quite difficult to make for noncontacted units. This means that item-missing data on interviewer observations of the sampled person for noncontacts needs to be imputed or a two-stage adjustment method used in which the predictors for contacting a case are different from those for obtaining cooperation (Groves and Couper 1995). Additionally, evidence of interviewer variance and interviewer bias has been found in observations of members of sampled housing units. How these measurement errors affect adjustments and potential fixes require more research. Nevertheless, as a potentially strong predictor of important survey variables, thus fulfilling one of the key criteria for a useful auxiliary variable, interviewer observations on characteristics of housing unit members have great promise for paradata for nonresponse.

Information from call records

Call record data are the traditional form of "process data" related to survey participation, sometimes called "level of effort" data (Biemer 2009). Call records consist of information recorded in sample management systems about the attempts made to a sampled unit as part of the recruitment process (Couper 1998; Hansen 2008; Heerwegh 2002; Kreuter and Casas-Cordero 2010; Laflamme 2008a). These data typically include a record for each recruitment attempt made to a sampled case (sometimes labeled a call), the time of the attempt, the date of the attempt, iden-

tification codes for the interviewer (in interviewer-administered surveys), and attempt-level result codes. Although these data are collected for a survey conducted at one point in time, call record data typically are not rectangular and have a structure that resembles longitudinal data (Singer and Willett 2003). That is, sampled cases with different numbers of call attempts have different numbers of observations. A case with only two attempts, for example, will have two observations, whereas a case with twenty attempts will have twenty observations. Call record data may also indicate other call-level events, such as the mode of the attempt (telephone, in-person, Web, or mail), whether an answering machine message was left (in telephone surveys), whether the interviewer left a "sorry I missed you" card (in face-to-face surveys), or whether the case was offered incentives, among other information (Chearo and Van Haitsma 2010; Stussman, Dahlhamer, and Simile 2005).

Although call record data were traditionally recorded through paper-and-pencil forms (Bates 2003), the growth of computerized sample management systems and the usefulness of call record data for sample management purposes (Lepkowski, Axinn, et al. 2010) have made such data routinely available in most surveys (Couper 1998; Groves and Couper 1998; Groves and Heeringa 2006; Hansen 2008; Scheuren 2000; Stoop 2005; Taylor 2008). Information about recruitment attempts and their outcomes can be kept for any survey mode, whether interviewer-administered or self-administered. What data are recorded in call records varies tremendously across organizations (Couper 1998), however, leading some to appeal for call records to be standardized (Chearo and Van Haitsma 2010).

The measurement properties of call record data are largely unstudied. Call record data are collected for the purposes of sample management (Hansen 2008), not analysis, and, as such, may be filled with measurement errors (Couper 1998). Secondary analyses of call record data have encountered duplicate call attempts, missing call attempts, call attempts recorded at improbable times (e.g., 2:00 a.m.), and inconsistencies in outcome codes (e.g., a final outcome of "never contacted" when call attempts with outcomes of refusal were listed for the case). A recent survey of interviewers at a large research organization found that at least 15 percent report making some kind of error in call records, such as recording calls for the incorrect case or failing to record "drive-by" attempts (Wang and Biemer 2010). The same study also found that interviewers are more likely to underreport calls than overreport them (Wang and Biemer 2010). How often this happens and why it happens require future empirical investigation.

The main purpose of keeping call record data is to monitor response rates so as to understand how decisions made in the field affect contact and cooperation and to implement changes and decisions in the course of data collection. It has been found, for example, that the largest changes in response rates happen during the first few contact attempts, with diminishing returns over the course of the field period (Bates 2003; Jocelyn et al. 2009; Laflamme 2008a). Likewise, it has also been found that first and second calls made during weekday evenings and on weekends are more positively associated with contact propensity compared

with calls made during the day on weekdays, in both telephone and face-to-face studies (Bates 2003; Groves and Couper 1998; Hoagland, Warde, and Payton 1988; Laflamme 2008b; Odom and Kalsbeek 1999; Weeks et al. 1980; Weeks, Kulka, and Pierson 1987).

Electronic call records also facilitate the implementation of responsive survey designs (Groves and Heeringa 2006; Laflamme 2008a, 2008b; Laflamme, Maydan, and Miller 2008; Lepkowski, Axinn, et al. 2010; Sinibaldi 2008; Wagner 2010). A critical assumption in responsive designs is that response propensities are changeable or dynamic (Olson and Groves 2012). In responsive designs, propensities change when new design features are targeted toward sample units who are less likely to participate, such as members of underrepresented demographic subgroups. Call record data feature prominently in monitoring these potential changes in nonresponse rates and nonresponse bias in key survey estimates over the course of the data collection field period (Tangdahl 2004, 2005, 2006; Wagner 2010). They also figure importantly when making decisions about which cases to subsample (Elliott, Little, and Lewitzky 2000).

Use of call record data for purposes of nonresponse adjustment requires aggregating information across all calls made to a case or using different modeling approaches (Couper 1998; Heerwegh 2002; Kreuter and Casas-Cordero 2010; Kreuter and Kohler 2009; Olson 2006). At this time, there is no single agreed on method for how to incorporate call record data into nonresponse adjustments. Almost all of the adjustment methods that use call record data draw on one of two theoretical models—either a "continuum of resistance model" or a "classes model" (Lin and Schaeffer 1995).

The continuum of resistance model hypothesizes that the number of call attempts to obtain a completed interview is linearly related to a sampled person's unwillingness to participate in an interview and is also linearly related to important survey characteristics. When this model holds, the "last" or "most difficult" respondents can be used to represent information that would have been obtained from nonrespondents (Bates and Creighton 2000; Filion 1975; Fitzgerald and Fuller 1982; Lahaut et al. 2003; Lin and Schaeffer 1995).

The classes model, in contrast, uses both the number of call attempts and call outcomes to categorize respondents and nonrespondents into groups of easy and hard refusals and easy and hard noncontacts (Lin and Schaeffer 1995). Unlike the strength of the relationship between contact attempts and survey participation, very little variance in important survey variables is explained by the number of call attempts (Kreuter and Kohler 2009; Lin and Schaeffer 1995). Furthermore, a wide literature examining differences between "late" and "early" and "easy" versus "difficult" respondents shows that some estimates differ for these two groups while others do not (see reviews in Groves 2006; Groves and Peytcheva 2008).

One of the first approaches to using information from call attempts for nonresponse adjustment, the Politz-Simmons method, showed that reports of the number of nights that a respondent had been at home during the past week could be used

to account for the "not at homes" (Kalton 1983; Politz and Simmons 1949). Limitations in the Politz-Simmons method include potential misreporting by respondents and failure to meet the model's assumptions in practical situations (Ward, Russick, and Rudelius 1985); these have led to the development of alternative statistical models based on the number of calls made to a sampled person, sometimes incorporating characteristics of the sampled unit.

One approach that uses call record information in propensity models simply includes the number of call attempts made to a case as a predictor in a logistic regression propensity model (Little 1982, 1986). The number of call attempts is clearly related to survey participation; more resistant or difficult-to-contact respondents generally receive more call attempts (Blom 2009; Campanelli, Sturgis, and Purdon 1997; Groves and Couper 1998; Lin, Schaeffer, and Seltzer 1999; Olson 2006). In fact, when added to models predicting survey participation, information from the call records tends to explain more variability in participation than any other predictor (Beaumont 2005; Dahlhamer and Simile 2009).

An alternative approach identifies patterns of call outcomes made to different cases. An example of these patterns is the proportion of noncontacts out of all calls made, rather than simply number of call attempts (Kreuter and Kohler 2009). Concerns about endogeneity of the number of contact attempts with the outcome of survey participation and "censoring" for cases that are never contacted or interviewed have led to the use of discrete time hazard models that change the outcome to the conditional probability of an interview on a given call, given no contact or participation on prior calls (Durrant and Steele 2009; Groves and Heeringa 2006; Kennickell 1999; Olson and Groves 2009).

Other probability-based models have been used to estimate response probabilities at each call attempt as a function of respondent characteristics, sometimes permitting a "hard core" nonresponding group (Alho 1990; Anido and Valdés 2000; Colombo 1992; Drew and Fuller 1980; Potthoff, Manton, and Woodbury 1993; Wood, White, and Hotopf 2006). A recent expansion of the callback models uses latent class models, characteristics of respondents and nonrespondents, and reports from the survey to create weights based on the level of effort exerted to the case (Biemer 2009; Biemer and Wang 2007; Biemer and Link 2006). There is little consensus about how these models can be incorporated into adjustment, however.

Overall, although call record data are consistent with the definition of paradata, they are the most problematic for purposes of nonresponse adjustment. With multiple observations per sampled unit, concerns about causality, and the need for complicated data reduction or transformation decisions, information contained in call records is not necessarily easily incorporated into nonresponse adjustments. Yet call records are invaluable for monitoring progress during data collection and for understanding how design decisions affect contact and cooperation. Recent research into the use of call record data, however, has shown promising possibilities for nonresponse adjustments for univariate estimates and for potentially incorporating information about design decisions into adjustments.

Observations about the interviewer-respondent interaction

Survey researchers have long been interested in the reasons that sample units choose not to participate in surveys (Deming 1953; Morton-Williams 1993). Systematic collection of these reasons, sometimes called "doorstep statements," has become widespread in computerized sample management systems. Typical comments given include statements such as "I'm not interested" and "I'm too busy"; or questions asked about the survey itself, such as "What is this all about?" Although not as common as call records themselves, many surveys, including the NHIS, the Consumer Expenditure Interview Survey (CEIS), the NSFG, the British Household Panel Study (BHPS), the National Survey of Health and Stress (NSHS), and the ANES record householder questions or statements to the interviewer at the time of recruitment (Bates, Dahlhamer, and Singer 2008; Campanelli, Sturgis, and Purdon 1997; Couper 1997; Groves and Couper 1998; Groves et al. 2009; Peytchev and Olson 2007; Safir and Tan 2009; Taylor 2008). An alternative type of contact observation is the interviewer's subjective assessment of the householder's reluctance or willingness to participate on future contacts (Campanelli, Sturgis, and Purdon 1997; Stoop 2005). These contact observations may be recorded on the first contact with the sampled household or on every contact with the household. Contact observations can be recorded by interviewers in face-to-face or telephone surveys but are not possible in self-administered surveys.

Contact observations are usually recorded in closed ended or "check all that apply" questions answered by interviewers when the interaction with a selected householder is finished. In general, little is known about the measurement properties of these interviewer observations. As with all the observations, interviewers may choose not to complete the observations at all, leading to item-missing data. Evaluating the accuracy of the recorded information requires audio recordings, an expensive and rare procedure (Couper and Groves 2002; Morton-Williams 1993). Campanelli, Sturgis, and Purdon (1997) compared taped interviewer-householder interactions to record of the interaction and found low concordance between the two sources. Between 31 and 68 percent of the statements agreed between the two sources, depending on how narrowly the statements were categorized. Both the interviewer observation and the coded tape predicted participation, although the strength of the relationship differed across the two sources. The contact observations as recorded by interviewers, then, may be picking up general impressions of the householder from nonverbal cues as well as verbal statements.

Complications arise when using contact observations as nonresponse adjustment variables. First, similar to observations of demographic or other characteristics of householders, contact observations are available only for contacted households and missing for noncontacted households. Adjustment models thus either need to impute contact observations for noncontacted households or separately model contact and cooperation when creating adjustment weights (Groves and Couper 1995).

Second, although detailed statements collected in contact observations may be useful for understanding heterogeneity across householders in their recruitment concerns, each individual statement may be uttered infrequently, leading to the need to aggregate across multiple statements. For example, time delay statements such as "I'm too busy," "I don't have time right now," or "I need time to think about it" may be aggregated into an overall "Time delay statements" indicator (Campanelli, Sturgis, and Purdon 1997; Groves and Couper 1998; Maitland, Casas-Cordero, and Kreuter 2009).

Third, some households may have only one contact with a contact observation recorded, usually for the attempt on which an interview is obtained. This leads to endogeneity concerns when predicting survey participation. Last, other households may have multiple contacts with the interviewer, with a different observation made at each contact, requiring simplification such as taking the first or last contact observation or aggregation across multiple contacts (Henly and Bates 2006; Lepkowski, Mosher, et al. 2010; Lepkowski et al. 2006; Maitland, Casas-Cordero, and Kreuter 2009; Moon, Rose, and Steel 2005).

Contact observations are nonetheless strong predictors of survey participation. In general, householders who ask questions tend to be more likely to ultimately participate in the survey, whereas statements such as "I'm not interested" or "I'm too busy" and interviewer predictions of unwillingness tend to be negatively associated with participation (Campanelli, Sturgis, and Purdon 1997; Couper and Groves 2002; Dahlhamer and Simile 2009; Groves and Couper 1998; Maitland, Casas-Cordero, and Kreuter 2009; Morton-Williams 1993; Olson et al. 2006).

This pattern is not surprising, as contact observations are designed to be predictors of survey participation, not of survey variables themselves (Groves and Couper 1998). Although there are demographic differences in the characteristics of household members who make different types of statements (Campanelli, Sturgis, and Purdon 1997; Couper 1997; Groves and Couper 1998), contact observations of the interviewer-householder interaction tend to be weak predictors of key survey variables (Maitland, Casas-Cordero, and Kreuter 2009; Peytchev and Olson 2007). For example, across items in the National Health Interview Survey, Maitland and colleagues (2009) found that the average correlations between contact observations and the survey variables were r = .02 to .03, with maximum correlations no greater than r = .12. Similar magnitudes were found by Peytchev and Olson (2007) in the ANES. Thus, contact observations may be limited on a key criterion for being a good auxiliary variable.

In general, contact observations of statements made by the householder about why he or she may or may not want to participate in a sample survey request, not surprisingly, strongly predict participation. Thus, they meet one criterion for being a useful auxiliary variable for nonresponse adjustment, but limitations arise because they are not strong predictors of survey variables and are not measured for all sampled survey units. Additionally, as with call records, analytic decisions have to be made about simplifying, summarizing, or aggregating observations recorded at multiple contacts.

Challenges in Using Paradata for Adjustment

The challenges of using paradata for nonresponse adjustment, as reviewed here, are many. They include problems related to item nonresponse, measurement errors in auxiliary variables, and myriad analytic difficulties. Although tools to address each of these issues have been developed for other purposes, they have received little attention in the area of nonresponse adjustment weights or for paradata in particular.

As with any type of data collected, paradata may be subject to item nonresponse. Little empirical attention has been given to how item nonresponse in auxiliary data may be addressed in weighting adjustments. When adjusting for item nonresponse in questionnaire data, survey practitioners increasingly turn to multiple imputation methods (Little and Rubin 2002; Rubin 1986). Multiple imputation for auxiliary weighting variables leads to additional questions, however. Should weights be created for each imputed auxiliary dataset? Should weights from each imputed dataset be provided to the user or should they be averaged across datasets? If multiple weights are provided, how do users obtain estimates from these weights? Alternatively, methods such as hot deck imputation, nearest neighbor imputation, or mean imputation may be used to fill in the missing values (Kalton 1983; Marker, Judkins, and Winglee 2002). Although any single imputation method necessarily fails to reflect the uncertainty involved in the adjustment, having only one imputation simplifies the creation of weights. Yet misclassification of sampled units through the imputation procedure may lower the utility of the auxiliary variable.

Since no single variable is a strong predictor of both participation and the survey variables, multiple auxiliary variables are typically used when creating weights. This fix, however, may yield additional problems. The bivariate strength of any relationship among the auxiliary variables, the survey variables, and participation does not, on its own, indicate how effective an auxiliary variable will be for reducing the root mean square error (RMSE) of a survey estimate in the presence of other auxiliary variables. In a simulation study, the direction of the relationship (for ordinal or continuous auxiliary variables) among the auxiliary variables, participation, and the survey variables was shown to be just as important to consider as the strength of the predictors for bias and variance reduction (Kreuter and Olson 2011). In fact, including fewer auxiliary variables may be better for reducing RMSE when the auxiliary variables are only weakly predictive of the important survey variables.

Measurement errors in paradata are poorly understood. Although some work has been done to evaluate the accuracy of interviewer estimates of respondent characteristics, these evaluations are conducted only on the respondent pool. No information is obtained about the accuracy of the guesses on the nonrespondents. Furthermore, variability across interviewers in their observations of characteristics of sample units is documented (O'Muircheartaigh and Campanelli 1998; Raudenbush and Sampson 1999; West 2010) but has received little empirical attention. Out-

standing questions include: Are certain types of paradata subject to higher or lower rates of interviewer variability? Can interviewers be trained to reduce variability on these observations? How does this variability affect the effectiveness of these observations as adjustment variables? Future research on these questions would greatly improve understanding of the quality of paradata.

Finally, the nonrectangular nature of many paradata files leads to the information being simplified and aggregated for use in traditional logistic regression models. When doing this simplification, paradata may differ in interpretation for respondents and nonrespondents. For example, the commonly used measure "total number of call attempts" (Beaumont 2005; Dahlhamer and Simile 2009; Kreuter and Kohler 2009; Olson 2006) differs in meaning for respondents and nonrespondents. For respondents, it represents the number of calls required to obtain an interview; for nonrespondents, it is the total number of calls made to the case before the field period ended or a maximum number of call attempts was achieved.

Alternatively, using information from paradata files "as is" yields a file with one observation per call attempt or per day of the field period. Although statistical models have been investigated that use some of the information from call record data for adjustment purposes, they tend to focus on the number of call attempts and exclude other information. To the extent that characteristics of the survey recruitment process, such as timing of contact attempts (Weeks et al. 1980), mode of the attempt (Olson and Groves 2009), or presence of an incentive (Groves, Presser, and Dipko 2004), predict survey participation, and persons who are receptive to these design features vary on survey variables of interest, incorporating recruitment characteristics in adjustment models should yield "better" adjusted estimates.

Issues in Using Paradata for Adjustment

Paradata are not a panacea for nonresponse adjustment variables. Many outstanding issues remain regarding the collection and use of paradata. First, although statistical theory suggests maximizing the relationship between the auxiliary variables and the survey variables for the best (lowest RMSE) results, most surveys are multipurpose. In a survey with one hundred key survey variables, survey designers may find it difficult to select only one or two as the "most important" for purposes of paradata development. Furthermore, these items may be variables that are difficult to observe on all sampled households, especially in attitudinal surveys. Identifying proxy variables for constructs such as "self-efficacy" or "trust in the media" may be more difficult than observing whether there are children or smokers in the household. Even if plausible proxy observations can be developed for the contacted units, observations cannot be made on noncontacted sampled units. Given the paucity of surveys collecting proxy observa-

tions of important survey variables, however, it is still unknown what types of observations are feasible and useful. This poses a great opportunity for innovation in paradata development.

Second, the routine collection of paradata may require changes to sample or field management systems. These changes may have substantial cost implications if infrastructure systems need to be developed. Organizations that do not currently maintain electronic call records will need to invest in an existing system or develop a new system to capture call-level information. Furthermore, existing sample management systems may not permit flexibility in the types of information collected at each call, such as statements made by householders or observations of characteristics of the householders. Getting "buy in" from other members or management in the survey organization about collecting this information may pose an additional hurdle. When building new sample management systems, however, an opportunity arises for designing and capturing information systematically that has been recorded largely in qualitative interviewer notes, such as interviewer strategies for contacting neighbors, leaving notes, and driving by the house in face-to-face studies, or information conveyed on answering machine messages in telephone surveys. Such new information may lead to better understanding about field decisions made by interviewers and potentially has use for nonresponse adjustment as new adjustment models using call-record data become more widespread.

Third, paradata development is easiest in face-to-face surveys where an interviewer can be asked to observe and record information about sampled neighborhoods or households. In telephone surveys, interviewers can be asked to make a guess about sampled persons based on voice characteristics or what is said at the time of recruitment but not about the neighborhood or housing unit. The United States Postal Service Delivery Sequence File, a list of all addresses that receive postal mail, is increasingly being adopted as a frame for household surveys, with a concomitant change in mode to self-administered surveys (Iannacchione, Staab, and Redden 2003; Link et al. 2008). Without interviewers to observe information about housing units, available paradata are limited to those about the recruitment process itself. These process data may be clearly related to the propensity to respond but less predictive of the survey variables of interest. As such, useful nonresponse adjustment variables for mail surveys may be sought from census data, purchased from commercial sources, or sought from a rich frame rather than paradata.

Process data also pose analytic difficulties. Although previous studies have aggregated these data, it is not clear what the best method for aggregation may be. When aggregated, concerns about endogeneity arise. That is, does the process data cause the outcome of survey participation, or does survey participation cause the process that is recorded in the paradata? For purposes of understanding why people participate in surveys, this issue clearly matters. For purposes of creating nonresponse adjustment weights, it may be less of a concern (Beaumont

2005; Kreuter and Kohler 2009). Similarly, although alternative modeling forms have been used for purposes of adjustment with process data, none of the alternative models is commonly used or accepted at this time. One reason for this could be that advanced statistical methods used in these alternative forms such as latent class models exceed the knowledge of many users of survey data. An additional reason could be that most of these models have been applied to estimates of means and proportions (Biemer 2009; Biemer and Wang 2007; Colombo 1992; Drew and Fuller 1980, 1981), but not to more complex estimands such as correlations or regression coefficients. Last, the alternative modeling method of hazard models has been used to understand the relationship between the process data and participation (Kennickell 1999; Olson and Groves 2009; Stoop 2005) but has not been directly extended to creating adjustment weights. This lack of consensus provides great opportunities for research to investigate bias and variance properties of different methods.

A final outstanding issue related to paradata is that they are infrequently released to the public. Understanding the measurement properties and usefulness of paradata for purposes of nonresponse adjustment requires having access to paradata. For secondary data analysts, having access to paradata is rare unless they happen to also work at a survey research organization that collects survey data and creates paradata files. The NHIS has started releasing paradata regularly for analysts on the NCHS website (Taylor 2008), but this survey is unusual among large-scale government surveys. Without public access to paradata files with information available for respondents and nonrespondents, it is unlikely that we will have much growth in the understanding about the quality or utility of paradata. Growing availability of paradata also may encourage survey data users to create their own nonresponse adjustment weights. User-driven weight creation may lead to insights that can improve survey estimates and lead to future adjustments for the "official" weights put on data files. At the same time, however, this may lead to new weights that increase the root mean square error of sample estimates relative to those that are released as part of the public use data file. Yet de-identified public use paradata files would open opportunities for survey organizations to conduct analyses that use data that may not be possible to collect given time and budget constraints at such organizations.

In sum, paradata have potential for being powerful auxiliary variables for nonresponse adjustment. Although traditionally developed to predict survey participation, new classes of paradata recently have been developed that predict key survey variables as well. Additional research is needed to understand the error properties of paradata and their effectiveness and utility. This research will require up-front investment by survey organizations and funding agencies to further understanding of when certain types of paradata are useful.

Note

1. Personal communication with Brady West, August 31, 2010.

References

Abraham, Katharine G., Aaron Maitland, and Suzanne M. Bianchi. 2006. Nonresponse in the American Time Use Survey: Who is missing from the data and how much does it matter? *Public Opinion Quarterly* 70 (5): 676–703.

Alho, Juha M. 1990. Adjusting for nonresponse bias using logistic regression. *Biometrika* 77 (3): 617–24.

Anido, Carmen, and Teófilo Valdés. 2000. An iterative estimating procedure for probit-type nonresponse models in surveys with call backs. *TEST* 9 (1): 233–53.

Bates, Nancy. 2003. Contact histories in personal visit surveys: The Survey of Income and Program Participation (SIPP) methods panel. In *JSM proceedings*, Survey Research Methods Section, 7–14. Alexandria, VA: American Statistical Association.

Bates, Nancy, and Kathleen Creighton. 2000. The last five percent: What can we learn from difficult/late interviews? In *JSM proceedings*, Survey Research Methods Section, 1–7. Alexandria, VA: American Statistical Association.

Bates, Nancy, James Dahlhamer, and Eleanor Singer. 2008. Privacy concerns, too busy, or just not interested: Using doorstep concerns to predict survey nonresponse. *Journal of Official Statistics* 24 (4): 591–612.

Battaglia, Michael P., Meena Khare, Martin R. Frankel, Mary Cay Murray, Paul Buckley, and Saralyn Peritz. 2008. Response rates: How have they changed and where are they headed? In *Advances in telephone survey methodology*, eds. James M. Lepkowski, Clyde Tucker, J. Michael Brick, Edith de Leeuw, Lilli Japec, Paul J. Lavrakas, Michael W. Link, and Roberta L. Sangster, 529–60. Hoboken, NJ: John Wiley.

Beaumont, Jean-Francois. 2005. On the use of data collection process information for the treatment of unit nonresponse through weight adjustment. *Survey Methodology* 31 (2): 227–31.

Bethlehem, Jelke. 1988. Reduction of nonresponse bias through regression estimation. *Journal of Official Statistics* 4 (3): 251–60.

Bethlehem, Jelke. 2002. Weighting nonresponse adjustments based on auxiliary information. In *Survey nonresponse*, eds. Robert M. Groves, Don A. Dillman, John L. Eltinge, and Roderick J. A. Little, 275–88. New York, NY: John Wiley.

Biemer, Paul. 2009. Incorporating level of effort paradata in nonresponse adjustments. JPSM Distinguished Lecture Series, University of Maryland, College Park. Online at http://www.jpsm.umd.edu/jpsm/events/specialevents/distinguished_lecture_2009_05_08/index.htm (accessed June 4, 2012).

Biemer, Paul P., and Michael W. Link. 2006. A latent call-back model for nonresponse. Paper presented at the 17th International Workshop on Household Survey Nonresponse, Omaha, NE.

Biemer, Paul, and Kevin Wang. 2007. Using callback models to adjust for nonignorable nonresponse in face-to-face surveys. In *JSM proceedings*, Survey Research Methods Section, 2889–96. Alexandria, VA: American Statistical Association.

Blohm, Michael, Joop Hox, and Achim Koch. 2006. The influence of interviewers' contact behavior on the contact and cooperation rate in face-to-face household surveys. *International Journal of Public Opinion Research* 19 (1): 97–111.

Blom, Annelies G. 2009. Nonresponse bias adjustments: What can process data contribute? ISER Working Paper Series, Institute for Social & Economic Research, Essex, UK.

Brehm, John. 1993. *The phantom respondents: Opinion surveys and political representation*. Ann Arbor: University of Michigan Press.

Burns, Nancy, Donald R. Kinder, Steven J. Rosenstone, Virginia Sapiro, and American National Election Studies. 2001. *ANES National Election Studies 2000: Pre/post-election study*. Ann Arbor: Center for Political Studies, University of Michigan. Online at http://www.electionstudies.org/ (accessed June 4, 2012).

Callens, Marc, and Christophe Croux. 2003. Nonresponse in the Belgian Fertility and Family Survey. EconPapers, Swedish Business School, Orebro University. Online at https://lirias.kuleuven.be/bitstream/123456789/118362/1/OR_0308.pdf (accessed June 4, 2012).

Campanelli, Pamela, Patrick Sturgis, and Susan Purdon. 1997. *Can you hear me knocking? An investigation into the impact of interviewers on survey response rates*. London: Survey Methods Centre, Social and Community Planning Research, University of London.

Casas-Cordero, Carolina. 2010. Disentangling neighborhood effects on survey participation. Paper presented at the annual meeting of the American Association of Public Opinion Research, 11–13 May, Chicago, IL.

Chearo, David, and Martha Van Haitsma. 2010. Standardizing paradata. Paper presented at the annual meeting of the American Association for Public Opinion Research, 11–13 May, Chicago, IL.

Cobben, Fannie, and Barry Schouten. 2008. *An empirical validation of r-indicators*. Voorburg/Heerlen: Statistics Netherlands.

Collins, Rebecca L., Phyllis L. Ellickson, Ron D. Hays, and Daniel F. McCaffrey. 2000. Effects of incentive size and timing on response rates to a follow-up wave of a longitudinal mailed survey. *Evaluation Review* 24 (4): 347–63.

Colombo, Richard A. 1992. Using call-backs to adjust for non-response bias. In *Survey and statistical computing*, eds. Andrew Westlake, Randy Banks, Clive Payne, and Terry Orchard, 269–77. Amsterdam: North Holland.

Couper, Mick P. 1997. Survey introductions and data quality. *Public Opinion Quarterly* 61 (2): 317–38.

Couper, Mick P. 1998. Measuring survey quality in a CASIC environment. In *JSM proceedings*, Survey Research Methods Section, 41–49. Alexandria, VA: American Statistical Association.

Couper, Mick P. 2005. Technology trends in survey data collection. *Social Science Computer Review* 23 (4): 486–501.

Couper, Mick P., and Robert M. Groves. 2002. Introductory interactions in telephone surveys and nonresponse. In *Standardization and tacit knowledge: Interaction and practice in the survey interview*, eds. Douglas W. Maynard, Hanneke Houtkoop-Steenstra, Nora Cate Schaeffer, and Johannes van der Zouwen, 161–78. New York, NY: John Wiley.

Couper, Mick P., and Lars Lyberg. 2005. The use of paradata in survey research. Paper presented at Proceedings of the 55th Session of the International Statistical Institute, Sydney, Australia.

Dahlhamer, James, and Catherine M. Simile. 2009. Subunit nonresponse in the National Health Interview Survey (NHIS): An exploration using paradata. In *JSM proceedings*, Government Statistics Section, 262–76. Alexandria, VA: American Statistical Association.

de Leeuw, Edith, and Wim de Heer. 2002. Trends in household survey nonresponse: A longitudinal and international perspective. In *Survey nonresponse*, eds. Robert M. Groves, Don A. Dillman, John L. Eltinge, and Roderick J. A. Little, 41–54. New York, NY: John Wiley.

Deming, W. Edwards. 1953. On a probability mechanism to attain an economic balance between the resultant error of response and the bias of nonresponse. *Journal of the American Statistical Association* 48 (264): 743–72.

Diez Roux, Ana V. 2001. Investigating neighborhood and area effects on health. *American Journal of Public Health* 91 (11): 1783–89.

Drew, J. H., and Wayne A. Fuller. 1980. Modeling nonresponse in surveys with callbacks. In *JSM proceedings*, Survey Research Methods Section, 639–42. Alexandria, VA: American Statistical Association.

Drew, J. H., and Wayne A. Fuller. 1981. Nonresponse in complex multiphase surveys. In *JSM proceedings*, Survey Research Methods Section, 623–28. Alexandria, VA: American Statistical Association.

Durrant, Gabriele B., and Fiona Steele. 2009. Multilevel modeling of refusal and non-contact in household surveys: Evidence from six UK government surveys. *Journal of the Royal Statistical Society Series A – Statistics in Society* 172 (2): 361–81.

Elliott, Michael R., Roderick J. A. Little, and Steve Lewitzky. 2000. Subsampling callbacks to improve survey efficiency. *Journal of the American Statistical Association* 95 (451): 730–38.

European Social Survey. 2008. European Social Survey contact forms. Online at http://www.europeansocialsurvey.org/index.php?option=com_content&task=view&id=121&Itemid=159 (accessed June 4, 2012).

Filion, F. L. 1975. Estimating bias due to nonresponse in mail surveys. *Public Opinion Quarterly* 39 (4): 482–92.

Fitzgerald, Robert, and Linda Fuller. 1982. I can hear you knocking but you can't come in: The effects of reluctant respondents and refusers on sample surveys. *Sociological Methods & Research* 11 (1): 3–32.

Fuller, Wayne. 1987. *Measurement error models*. New York, NY: John Wiley.

Gee, Gilbert C., and C. Payne-Sturges Devon. 2004. Environmental health disparities: A framework integrating psychosocial and environmental concepts. *Environmental Health Perspectives* 112 (17): 1645–53.

Goyder, John. 1987. *The silent minority: Nonrespondents on sample surveys*. Boulder, CO: Westview Press.

Goyder, John, Keith Warriner, and Susan Miller. 2002. Evaluating socio-economic status (SES) bias in survey nonresponse. *Journal of Official Statistics* 18 (1): 1–11.

Green, Kathy E. 1996. Sociodemographic factors and mail survey response. *Psychology & Marketing* 13 (2): 171–85.

Groves, Robert M. 2006. Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly* 70 (5): 646–75.

Groves, Robert M., and Mick P. Couper. 1995. Theoretical motivation for post-survey nonresponse adjustment in household surveys. *Journal of Official Statistics* 11 (1): 93–106.

Groves, Robert M., and Mick Couper. 1998. *Nonresponse in household interview surveys*. New York, NY: John Wiley.

Groves, Robert M., Don A. Dillman, John L. Eltinge, and Roderick J. A. Little, eds. 2002. *Survey nonresponse*. New York, NY: John Wiley.

Groves, Robert M., and Steven G. Heeringa. 2006. Responsive design for household surveys: Tools for actively controlling survey errors and costs. *Journal of the Royal Statistical Society A* 169 (3): 439–57.

Groves, Robert M., William D. Mosher, James M. Lepkowski, and Nicole G. Kirgis. 2009. Planning and development of the continuous National Survey of Family Growth. *Vital Health Statistics* 1 (48): 1–64.

Groves, Robert M., and Emilia Peytcheva. 2008. The impact of nonresponse rates on nonresponse bias: A meta-analysis. *Public Opinion Quarterly* 72 (2): 167–89.

Groves, Robert M., Stanley Presser, and Sarah Dipko. 2004. The role of topic interest in survey participation decisions. *Public Opinion Quarterly* 68 (1): 2–31.

Groves, Robert M., James Wagner, and Emilia Peytcheva. 2007. Use of interviewer judgments about attributes of selected respondents in post-survey adjustment for unit nonresponse: An illustration with the National Survey of Family Growth. In *JSM proceedings*, Survey Research Methods Section, 3428–31. Alexandria, VA: American Statistical Association.

Hahn, Robert A., Benedict I. Truman, and Nancy D. Barker. 1996. Identifying ancestry: The reliability of ancestral identification in the United States by self, proxy, interviewer and funeral director. *Epidemiology* 7 (1): 75–80.

Hansen, Sue Ellen. 2008. CATI sample management systems. In *Advances in telephone survey methodology*, eds. James M. Lepkowski, Clyde Tucker, J. Michael Brick, Edith D. de Leeuw, Lilli Japec, Paul J. Lavrakas, Michael W. Link, and Roberta L. Sangster, 340–58. Hoboken, NJ: John Wiley.

Health and Retirement Study. 2009. *HRS 2006 final release codebook*. Online at http://hrsonline.isr.umich.edu/modules/meta/2006/core/codebook/h06_00.html (accessed June 4, 2012).

Heerwegh, Dirk. 2002. Describing response behavior in Web surveys using client side paradata. Paper read at International Workshop on Web Surveys, 17–19 October, Mannheim, Germany.

Henly, Megan, and Nancy Bates. 2006. Using call records to understand response in panel surveys. In *JSM proceedings*, Survey Research Methods Section, 4128–35. Alexandria, VA: American Statistical Association.

Hill, Mark E. 2002. Race of the interviewer and perception of skin color: Evidence from the Multi-City Study of Urban Inequality. *American Sociological Review* 67 (1): 99–108.

Hoagland, Rebecca J., William D. Warde, and Mark E. Payton. 1988. Investigation of the optimum time to conduct telephone surveys. In *JSM proceedings*, Survey Research Methods Section, 755–60. Alexandria, VA: American Statistical Association.

Holbrook, Allyson L., Jon A. Krosnick, and Alison Pfent. 2008. The causes and consequences of response rates in surveys by the news media and government contractor survey research firms. In *Advances in telephone survey methodology*, eds. James M. Lepkowski, Clyde Tucker, J. Michael Brick, Edith D. de Leeuw, Lilli Japec, Paul J. Lavrakas, Michael W. Link, and Roberta L. Sangster, 499–528. Hoboken, NJ: John Wiley.

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): 202–10.

Inter-University Consortium for Political and Social Research. 2010. *Project for Human Development in Chicago Neighborhoods*. Online at http://www.icpsr.umich.edu/icpsrweb/PHDCN/ (accessed September 8, 2010).

Jacobs, David E., Robert P. Clickner, Joey Y. Zhou, Susan M. Viet, David A. Marker, John W. Rogers, Darryl C. Zeldin, Pamela Broene, and Warren Friedman. 2002. The prevalence of lead-based paint hazards in U.S. housing. *Environmental Health Perspectives* 110 (10): A599–A606.

Jocelyn, Wisner, Owen Phillips, Beatrice Baribeau, and Amelie Levesque. 2009. Using paradata to manage nonresponse in the Survey of Labour and Income Dynamics. In *Proceedings of Statistics Canada Symposium 2008*. Online at http://www5.statcan.gc.ca/bsolc/olc-cel/olc-cel?catno=11-522-X200800010996&lang=eng (accessed June 4, 2012).

Johnson, Timothy P., Young Ik Cho, Richard T. Campbell, and Allyson L. Holbrook. 2006. Using communitylevel correlates to evaluate nonresponse effects in a telephone survey. *Public Opinion Quarterly* 70 (5): 704–19.

Kalton, Graham. 1983. *Compensating for missing survey data*. Ann Arbor: Survey Research Center, Institute for Social Research, University of Michigan.

Kalton, Graham, and Ismael Flores-Cervantes. 2003. Weighting methods. *Journal of Official Statistics* 19 (2): 81–97.

Kenney McCulloch, Susan, Frauke Kreuter, and Stephanie Calvano. 2010. Interviewer observed vs. reported respondent gender: Implications for measurement error. Paper presented at the annual meeting of the American Association for Public Opinion Research, 11–13 May, Chicago, IL.

Kennickell, Arthur B. 1999. Analysis of nonresponse effects in the 1995 Survey of Consumer Finances. *Journal of Official Statistics* 15 (2): 283–303.

Kennickell, Arthur B. 2000. Asymmetric information, interviewer behavior, and unit nonresponse. In *JSM proceedings*, Survey Research Methods Section, 238–43. Alexandria, VA: American Statistical Association.

Kennickell, Arthur B. 2003. Reordering the darkness: Application of effort and unit nonresponse in the Survey of Consumer Finances. In *JSM proceedings*, Survey Research Methods Section, 2219–26. Alexandria, VA: American Statistical Association.

Kennickell, Arthur B. 2005. Darkness made visible: Field management and nonresponse in the 2004 SCF. In *JSM proceedings*, Survey Research Methods Section, 3187–94. Alexandria, VA: American Statistical Association.

Kish, Leslie. 1962. Studies of interviewer variance for attitudinal variables. *Journal of the American Statistical Association* 57 (297): 91–115.

Kish, Leslie. 1992. Weighting for unequal Pi. *Journal of Official Statistics* 8 (2): 183–200.

Korinek, Anton, Johan A. Mistiaen, and Martin Ravallion. 2005. Survey nonresponse and the distribution of income. Policy Research Working Paper WPS3543. Washington, DC: World Bank. Online at http://documents.worldbank.org/curated/en/2005/03/5681006/survey-nonresponse-distributionincome (accessed June 4, 2012).

Korinek, Anton, Johan A. Mistiaen, and Martin Ravallion. 2007. An econometric method of correcting for unit nonresponse bias in surveys. *Journal of Econometrics* 136 (1): 213–35.

Kreuter, Frauke, and Carolina Casas-Cordero. 2010. Paradata. German Council for Social and Economic Data Working Paper Series, Working Paper 136. Online at http://www.ratswd.de/download/RatSWD_WP_2010/RatSWD_WP_136.pdf (accessed June 4, 2012).

Kreuter, Frauke, and Ulrich Kohler. 2009. Analyzing contact sequences in call record data: Potential and limitations of sequence indicators for nonresponse adjustments in the European Social Survey. *Journal of Official Statistics* 25 (2): 203–26.

Kreuter, Frauke, Michael Lemay, and Carolina Casas-Cordero. 2007. Using proxy measures of survey outcomes in post-survey adjustments: Examples from the European Social Survey (ESS). In *JSM proceedings*, Survey Research Methods Section, 3142–49. Alexandria, VA: American Statistical Association.

Kreuter, Frauke, and Kristen Olson. 2011. Multiple auxiliary variables in nonresponse adjustment. *Sociological Methods & Research* 40 (2): 311–32.

Kreuter, Frauke, Kristen Olson, James Wagner, Ting Yan, Trena M. Ezzati-Rice, Carolina Casas-Cordero, Michael Lemay, Andy Peytchev, Robert M. Groves, and Trivellore E. Raghunathan. 2010. Using proxy measures and other correlates of survey outcomes to adjust for nonresponse: Examples

from multiple surveys. *Journal of the Royal Statistical Society Series A – Statistics in Society* 173 (2): 389–407.

Laflamme, Francois. 2008a. Data collection research using paradata at Statistics Canada. In *Proceedings of Statistics Canada Symposium 2008*. Online at http://www.statcan.gc.ca/pub/11-522-x/2008000/article/10997-eng.pdf (accessed June 4, 2012).

Laflamme, Francois. 2008b. Understanding survey data collection through the analysis of paradata at Statistics Canada. In *AAPOR proceedings*, Survey Research Methods Section, 4217–24. Alexandria, VA: American Statistical Association.

Laflamme, Francois, Mike Maydan, and Andrew Miller. 2008. Using paradata to actively manage data collection survey process. In *JSM proceedings*, Survey Research Methods Section, 630–37. Alexandria, VA: American Statistical Association.

Lahaut, Vivienne M. H. C. J., Harrie A. M. Jansen, Dike van de Mheen, Henk F. L. Garretsen, Jacqueline E. E. Verdurmen, and Ad van Dijk. 2003. Estimating non-response bias in a survey on alcohol consumption: Comparison of response waves. *Alcohol & Alcoholism* 38 (2): 128–34.

Lepkowski, James M., William G. Axinn, Nicole Kirgis, Brady T. West, Shonda Kruger Ndiaye, William Mosher, and Robert M. Groves. 2010. Use of paradata in a responsive design framework to manage a field data collection. NSFG Survey Methodology Working Papers, No. 10-012. Atlanta, GA: Centers for Disease Control and Prevention.

Lepkowski, James M., William D. Mosher, Karen E. Davis, Robert M. Groves, and John Van Hoewyk. 2010. The 2006–2010 National Survey of Family Growth: Sample design and analysis of a continuous survey. *Vital Health Statistics* 2 (150). Atlanta, GA: Centers for Disease Control and Prevention.

Lepkowski, James M., William D. Mosher, Karen E. Davis, Robert M. Groves, John Van Hoewyk, and Jennifer Willem. 2006. National Survey of Family Growth, Cycle 6: Sample design, weighting, imputation, and variance estimation. *Vital Health Statistics* 2 (142). Atlanta, GA: Centers for Disease Control and Prevention.

Lessler, Judith T., and William D. Kalsbeek. 1992. *Nonsampling error in surveys*. New York, NY: John Wiley.

Leventhal, Tama, and Jeanne Brooks-Gunn. 2000. The neighborhoods they live in: The effects of neighborhood residence on child and adolescent outcomes. *Psychological Bulletin* 126 (2): 309–37.

Leventhal, Tama, and Jeanne Brooks-Gunn. 2003. Children and youth in neighborhood contexts. *Current Directions in Psychological Science* 12 (1): 27–31.

Lin, I-Fen, and Nora Cate Schaeffer. 1995. Using survey participants to estimate the impact of nonparticipation. *Public Opinion Quarterly* 59 (2): 236–58.

Lin, I-Fen, Nora Cate Schaeffer, and Judith A. Seltzer. 1999. Causes and effects of nonparticipation in a child support survey. *Journal of Official Statistics* 15 (2): 143–66.

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.

Little, Roderick J. A. 1982. Models for nonresponse in sample surveys. *Journal of the American Statistical Association* 77 (378): 237–50.

Little, Roderick J. A. 1986. Survey nonresponse adjustments for estimates of means. *International Statistical Review* 54 (2): 139–57.

Little, Roderick J. A., and Donald B. Rubin. 2002. *Statistical analysis with missing data*. New York, NY: John Wiley.

Little, Roderick J. A., and Sonya Vartivarian. 2003. On weighting the rates in non-response rates. *Statistics in Medicine* 22 (9): 1589–99.

Little, Roderick J. A., and Sonya Vartivarian. 2005. Does weighting for nonresponse increase the variance of survey means? *Survey Methodology* 31 (2): 161–68.

Lyberg, Lars. 2009. The paradata concept in survey research. Paper presented at the NCRM Paradata Network Meeting, Why Use Paradata in Social Survey Research? London: National Centre for Social Research. Online at http://www.natcen.ac.uk/ncrm-paradata-network/docs/Lyberg-paradata-concept.ppt

Lynn, Peter. 2003. PEDAKSI: Methodology for collecting data about survey non-respondents. *Quality & Quantity* 37 (3): 239–61.

Maitland, Aaron, Carolina Casas-Cordero, and Frauke Kreuter. 2009. An evaluation of nonresponse bias using paradata from a health survey. In *JSM proceedings*, Government Statistics Section, 370–78. Alexandria, VA: American Statistical Association.

Marker, David A., David R. Judkins, and Marianne Winglee. 2002. Large-scale imputation for complex surveys. In *Survey nonresponse*, eds. Robert M. Groves, Don A. Dillman, John L. Eltinge, and Roderick J. A. Little, 329–42. New York, NY: John Wiley.

Matsuo, Hideko, Jaak Billiet, and Geert Loosveldt. 2010. *Response-based quality assessment of ESS round 4: Results for 24 countries based on contact files*. Leuven, Belgium: European Social Survey, University of Leuven.

Moon, Nick, Nickie Rose, and Nikki Steel. 2005. How could they ever, ever persuade you? Are some refusals easier to convert than others? In *JSM proceedings*, Survey Research Methods Section, 3937–42. Alexandria, VA: American Statistical Association.

Moonesinghe, Ramal, Susan Mitchell, and Daniel Pasquini. 1995. An identification study of nonrespondents to the 1993 Survey of Doctorate Recipients. In *JSM proceedings*, Survey Research Methods Section, 453–58. Alexandria, VA: American Statistical Association.

Morton-Williams, Jean. 1993. *Interviewer approaches*. Cambridge: Cambridge University Press.

National Center for Health Statistics. 2009. *National Health and Nutrition Examination Survey: Interviewer procedures manual*. Online at http://www.cdc.gov/nchs/nhanes/nhanes2005-2006/current_nhanes_05_06.htm (accessed June 4, 2012).

National Election Studies. 2004. *National Election Study 2004 codebook*. Ann Arbor: University of Michigan, Center for Political Studies.

Odom, Dawn M., and William D. Kalsbeek. 1999. Further analysis of telephone call history data from the behavioral risk factor surveillance system. In *JSM proceedings*, Survey Research Methods Section, 398–403. Alexandria, VA: American Statistical Association.

Olson, Kristen. 2006. Survey participation, nonresponse bias, measurement error bias, and total bias. *Public Opinion Quarterly* 70 (5): 737–58.

Olson, Kristen, and Robert M. Groves. 2012. An examination of within-person variation in response propensity over the data collection field period. *Journal of Official Statistics* 28:29–51.

Olson, Kristen, Jennifer Sinibaldi, James M. Lepkowski, and David Garabrant. 2006. Analysis of a new form of contact observations. Paper presented at the annual meeting of the American Association for Public Opinion Research, Montreal, Canada.

O'Muircheartaigh, Colm, and Pamela Campanelli. 1998. The relative impact of interviewer effects and sample design effects on survey precision. *Journal of the Royal Statistical Society, A* 161 (1): 63–77.

O'Muircheartaigh, Colm, and Pamela Campanelli. 1999. A multilevel exploration of the role of interviewers in survey non-response. *Journal of the Royal Statistical Society, A* 162 (3): 437–46.

Peterson, Christine E., Narayan Sastry, and Anne R. Pebley. 2007. *The Los Angeles Family and Neighborhood Survey: Neighborhood observations codebook*. Santa Monica, CA: RAND Corporation. Online at http://www.rand.org/pubs/working_papers/WR240z13 (accessed August 23, 2010).

Peytchev, Andy, and Kristen Olson. 2007. Using interviewer observations to improve nonresponse adjustments: NES 2004. In *JSM proceedings*, Survey Research Methods Section, 3364–71. Alexandria, VA: American Statistical Association.

Peytcheva, Emilia, and Robert M. Groves. 2009. Using variation in response rates of demographic subgroups as evidence of nonresponse bias in survey estimates. *Journal of Official Statistics* 25 (2): 193–201.

Politz, Alfred, and Willard Simmons. 1949. An attempt to get the "not at homes" into the sample without callbacks. *Journal of the American Statistical Association* 44 (245): 9–16.

Potthoff, Richard F., Kenneth G. Manton, and Max A. Woodbury. 1993. Correcting for nonavailability bias in surveys by weighting based on number of callbacks. *Journal of the American Statistical Association* 88 (424): 1197–1207.

Purdie, D. M., M. P. Dunne, F. M. Boyle, M. D. Cook, and J. M. Najman. 2002. Health and demographic characteristics of respondents in an Australian national sexuality survey: Comparison with population norms. *Journal of Epidemiology and Community Health* 56 (10): 748–53.

Raghunathan, Trivellore E. 2004. What do we do with missing data? Some options for analysis of incomplete data. *Annual Review of Public Health* 25:99–117.

Raudenbush, Stephen W., and Robert J. Sampson. 1999. Ecometrics: Toward a science of assessing ecological settings, with application to the systematic social observation of neighborhoods. *Sociological Methodology* 29:1–41.

Rubin, Donald B. 1986. Basic ideas of multiple imputation for nonresponse. *Survey Methodology* 12 (1): 37–47.

Saegert, Susan C., Susan Klitzman, Nicholas Freudenberg, Jana Cooperman-Mroczek, and Salwa Nassar. 2003. Healthy housing: A structured review of published evaluations of U.S. interventions to improve health by modifying housing in the United States, 1990–2001. *American Journal of Public Health* 93 (9): 1471–77.

Safir, Adam, and Lucilla Tan. 2009. Using contact attempt history data to determine the optimal number of contact attempts. In *AAPOR proceedings*, Survey Research Methods Section, 5970–80. Alexandria, VA: American Statistical Association.

Sampson, Robert J., Jeffrey D. Morenoff, and Thomas Gannon-Rowley. 2002. Assessing "neighborhood effects": Social processes and new directions in research. *Annual Review of Sociology* 28:443–78.

Saperstein, Aliya. 2006. Double-checking the race box: Examining inconsistency between survey measures of observed and self-reported race. *Social Forces* 85 (1): 57–74.

Sarndal, Carl-Erik, and Sixten Lundstrom. 2005. *Estimation in surveys with nonresponse*. Chichester, UK: John Wiley.

Sarndal, Carl-Erik, and Sixten Lundstrom. 2008. Assessing auxiliary vectors for control of nonresponse bias in the calibration estimator. *Journal of Official Statistics* 24 (2): 167–91.

Sastry, Narayan, and Anne R. Pebley. 2003. Non-response in the Los Angeles Family and Neighborhood Survey. Labor and Population Program Working Paper Series 03-01. Santa Monica, CA: RAND Corporation.

Sastry, Narayan, and Anne R. Pebley. 2004. *The Los Angeles Family and Neighborhood Survey: Neighborhood observation forms and interviewer manual 2004*. Santa Monica, CA: RAND Corporation. Online at http://www.rand.org/pubs/drafts/DRU2400.6-1/DRU2400.6-1.pdf (accessed August 25, 2010).

Scheuren, Fritz. 2000. Macro and micro paradata for survey assessment. Paper presented at Working Session on Statistical Metadata, 28–30 November, Washington, DC. Online at http://www.unece.org/fileadmin/DAM/stats/documents/2000/11/metis/crp.10.e.pdf (accessed June 4, 2012).

Schräpler, Jörg-Peter, Juergen Schupp, and Gert G. Wagner. 2010. Individual and neighborhood determinants of survey nonresponse: An analysis based on a new subsample of the German Socio-Economic Panel, microgeographic characteristics and survey-based interviewer characteristics. SOEP Paper. Online at http://dx.doi.org/10.2139/ssrn.1588730 (accessed June 4, 2012).

Singer, Judith D., and John B. Willett. 2003. *Applied longitudinal data analysis*. New York, NY: Oxford University Press.

Sinibaldi, Jennifer. 2008. Exploratory analysis of currently available NatCen paradata for use in responsive design. Unpublished internal report. London: NatCen.

Sinibaldi, Jennifer. 2010. Analysis of 2010 NatSAL dress rehearsal interviewer observation data. Unpublished internal report. London: NatCen.

Smith, Tom W. 1984. Estimating nonresponse bias with temporary refusals. *Sociological Perspectives* 27 (4): 473–89.

Smith, Tom W. 1997. Measuring race by observation and self-identification. GSS Methodological Report No. 89. Chicago, IL: National Opinion Research Center.

Smith, Tom W. 2001. Aspects of measuring race: Race by observation vs. self-reporting and multiple mentions of race and ethnicity. GSS Methodological Report No. 93. Chicago, IL: National Opinion Research Center.

Steeh, Charlotte G. 1981. Trends in nonresponse rates, 1952–1979. *Public Opinion Quarterly* 45 (1): 40–57.

Steeh, Charlotte G., Nicole Kirgis, Brian Cannon, and Jeff DeWitt. 2001. Are they really as bad as they seem? Nonresponse rates at the end of the twentieth century. *Journal of Official Statistics* 17 (2): 227–47.

Steel, D. G., M. Tranmer, and D. Holt. 2003. Analysis combining survey and geographically aggregated data. *Analysis of survey data*, eds. R. L. Chambers and C. J. Skinner, 323–44. Chichester, UK: John Wiley.

Stoop, Ineke A. L. 2005. *The hunt for the last respondent*. The Hague: Social and Cultural Planning Office of the Netherlands.

Stussman, Barbara J., James M. Dahlhamer, and Catherine M. Simile. 2005. How interviewer strategies relate to contact and cooperation in the National Health Interview Survey, 2004. Paper presented at the Federal Committee on Statistical Methodology Research Conference, Washington, DC.

Tangdahl, Sara. 2004. Nonresponse bias for some common estimators and its change over time in the data collection process. Working Paper Series Paper No. 13, Orebro University, Sweden. Online at http://www.oru.se/PageFiles/15368/WP%2013%2004.pdf (accessed June 4, 2012).

Tangdahl, Sara. 2005. The variance of some common estimators and its components under nonresponse. Working Paper Series Paper No. 9, Orebro University, Sweden. Online at http://www.oru.se/PageFiles/15370/WP%209%2005.pdf (accessed June 4, 2012).

Tangdahl, Sara. 2006. On the evaluation of the cost efficiency of nonresponse rate reduction efforts—Some general considerations. Working Paper Series Paper No. 5, Orebro University, Sweden. Online at http://www.oru.se/PageFiles/15372/WP%205%2006.pdf (accessed June 4, 2012).

Taylor, Beth. 2008. The 2006 National Health Interview Survey (NHIS) paradata file: Overview and applications. In *JSM proceedings*, Survey Research Methods Section, 1909–13. Alexandria, VA: American Statistical Association.

U.S. Census Bureau. 2008. *American Housing Survey for the United States: 2007*. Current Housing Reports, Series H150/07. Washington, DC: Government Printing Office.

U.S. Department of Energy, Energy Information Administration. 2008. 2005 Residential Energy Consumption Survey—Detailed Tables, Table HC 2.4 Type of Housing. Washington, DC: Department of Energy.

U.S. Energy Information Administration. 2009. *Residential Energy Consumption Survey: Household questionnaire*. Washington, DC: U.S. Department of Energy.

Voogt, Robert. 2004. I'm not interested: Nonresponse bias, response bias and stimulus effects in election research. PhD diss., Department of Communication Studies, University of Amsterdam, the Netherlands.

Wagner, James. 2010. The fraction of missing information as a tool for monitoring the quality of survey data. *Public Opinion Quarterly* 74 (2): 223–43.

Wang, Kevin, and Paul Biemer. 2010. The accuracy of interview paradata: Results from a field investigation. Paper presented at the annual conference of the American Association for Public Opinion Research, 13–16 May, Chicago, IL.

Wang, Kevin, Joe Murphy, Rodney Baxter, and Jeremy Aldworth. 2005. Are two feet in the door better than one? Using process data to examine interviewer effort and nonresponse bias. Paper presented at the Federal Committee on Statistical Methodology Conference, Washington, DC.

Ward, James C., Bertram Russick, and William Rudelius. 1985. A test of reducing callbacks and not-athome bias in personal interviews by weighting at-home respondents. *Journal of Marketing Research* 22 (1): 66–73.

Weeks, M. F., B. L. Jones, R. E. Folsom Jr., and C. H. Benrud. 1980. Optimal times to contact households. *Public Opinion Quarterly* 44 (1): 101–14.

Weeks, Michael F., Richard A. Kulka, and Stephanie A. Pierson. 1987. Optimal call scheduling for a telephone survey. *Public Opinion Quarterly* 51 (4): 540–49.

West, Brady T. 2010. An examination of the quality and utility of interviewer estimates of household characteristics in the National Survey of Family Growth. Paper read at the annual conference of the American Association for Public Opinion Research, 13–16 May, Chicago, IL.

Wood, Angela M., Ian R. White, and Matthew Hotopf. 2006. Using number of failed contact attempts to adjust for non-ignorable non-response. *Journal of the Royal Statistical Society, A* 169 (3): 525–42.