Journal of Official Statistics, Vol. 28, No. 1, 2012, pp. 29–51

An Examination of Within-Person Variation in Response Propensity over the Data Collection Field Period

Kristen Olson[1] *and Robert M. Groves*[2]

Statistical examinations of deterministic and stochastic response propensity assert that a sample case's propensity is determined by fixed respondent characteristics. The perspective of this article, that of dynamic response propensities, differs, viewing sample cases' propensities as evolving over the course of the data collection. Each sample case begins the data collection period in a "base" response propensity. Each change in the data collection protocol which the survey organization subsequently makes might change that base propensity. This article examines four questions: (1) Is there any evidence that the average response propensities of sampled individuals vary over the data collection? (2) Is there any evidence that propensities are influenced in accordance with specific actions taken by the survey recruitment protocol? (3) Do these changes have fixed effects or do they also vary across sample units or across the data collection period? (4) Does the change in propensities coincide with changes in nonresponse bias of key survey estimates?

Key words: Nonresponse; responsive design; nonresponse bias; survival models; dynamic response propensities.

1. Introduction

As response rates decline in surveys of persons in the US and other countries (de Leeuw and de Heer 2002), survey researchers often choose to apply statistical weights to respondent data to reduce the effects of missing information on the nonrespondents. It can be shown that if respondents are assigned to groups that are relatively homogeneous both on their probability of being interviewed and the survey variable of interest, then an adjusted estimate of the respondent mean can have lower nonresponse bias than an unadjusted mean (Kalton 1981; Little and Rubin 2002). In practice, this often involves creating "weighting classes" based on a cross-tabulation of variables known on all sample cases (Bethlehem 2002). Alternatively, the researcher might build multivariate response propensity models and use them to estimate the

[1] University of Nebraska-Lincoln, Department of Sociology and Survey Research and Methodology Program, 703 Oldfather Hall, P.O. Box 880324, Lincoln, NE 68588-0324, U.S.A. Email: kolson5@unl.edu
[2] U.S. Census Bureau, Washington, DC, 20233 U.S.A. Email: Robert.M.Groves@Census.Gov
Acknowledgments: The views expressed in this article do not necessarily reflect those of the U.S. Census Bureau. We thank the U.S. National Science Foundation for support of both Olson (SES-0620228) and Groves (SES-0207435) and John VanHoewyk for analytic support. Groves is indebted to the National Center for Health Statistics for its access to the National Survey of Family Growth data. Vaughn Call graciously provided access to the Wisconsin Divorce Study data. Rene Bautista, Frauke Kreuter, John Loft, and Sonja Ziniel provided helpful comments on earlier drafts.

probability of being measured, weighting cases by a function of this probability (Little and Vartivarian 2005).

A careful reading of theoretical justification for traditional postsurvey adjustments notes that the estimated probabilities of response are conditional on the specific survey design and its recruitment protocol, where a survey recruitment protocol is the set of methods, rules and decisions implemented by a survey organization in an attempt to contact sample units and solicit their participation in a survey. That is, in conceptual replications of the survey, each respondent would be subjected to the same recruitment protocol. The given realization of the data set is viewed as one random selection from that set of possible realizations.

In fact, recruitment protocols themselves are not fixed in practice. Field personnel routinely apply different protocol features to different sample units in attempts to raise their propensities to respond. Sample persons receive very different sets of experiences with regard to contacting them and obtaining their cooperation (Olson 2007). That is, some features are applied to all sample units (e.g., the survey topic, sponsor, sending of an advance letter), while others are applied to only a subset of sample units (e.g., different interviewer introductory behavior, use of persuasion letters, increases in incentives, changes in interviewers). As a result, the ending state of propensities of a sample case may vary over conceptual realizations of a survey design, simply because all of the possible actions of the survey design may vary over replications. When each sample unit can receive different subsets of recruitment protocol features, the idea of conceptual realizations of a survey design becomes complicated.

The above observations have implications for weighting class construction. However, we first must establish that there is evidence of within-person variation in response propensities, that specific actions taken by a survey organization can influence response propensities, that these actions may not have a constant effect, and that actions taken can affect survey estimates. This article aims to answer some simple questions motivated by the above perspective:

1. Is there any evidence that response propensities of individuals vary over the course of the data collection?
2. Is there any evidence that propensities are influenced by specific actions taken in accordance with the survey recruitment protocol?
3. Is there any evidence that the influence of the recruitment protocol changes over the course of the data collection period or differs over sampled units?
4. Does the change in propensities coincide with changes in the nonresponse bias properties of key survey estimates?

We argue that the stochastic perspective on survey nonresponse (Lessler and Kalsbeek 1992) needs to be expanded to account for variation within sample persons in their likelihood to participate in a survey. The dynamic approach requires a different method of estimation of response propensities than is typically used (discrete time hazard models instead of logistic regression models). We use predicted probabilities from discrete time hazard models to illustrate Questions 1 through 3 above. We also illustrate how the conventional method for estimating response propensities fails to account for the influence of alternative protocol components, addressing the fourth research question.

2. Background

The importance of the recruitment protocol in influencing the likelihood of contacting a sample unit or a sample unit cooperating has been documented in countless experimental and observational analyses within the survey methodological literature. Examples of recruitment protocol components that influence response propensities are the number of call attempts (e.g., Curtin et al. 2000; McCarty et al. 2006), call timing (e.g., Brick et al. 1996; Hoagland et al. 1988; Weeks et al. 1987), mode and mode switches (de Leeuw 2005), and incentives (Singer 2002). In practice, different recruitment protocol components are used for different sample units, whether intentionally (incentives) or unintentionally (call timing) applied. For example, sample units may receive additional calls at different times of the day and on different days of the week, receive incentives as a refusal conversion tactic, or be approached with a shortened survey or with a different mode from the initial request, among other changes in design features. Further, and most interesting for the perspective of this article, persons vary in their reactions to different features of the recruitment protocol (e.g., those uninterested in the survey topic show greater incentive effects; see Baumgartner et al. 1998).

There is past research that studied respondent differences as a function of how much effort was required to obtain their participation. This approach, sometimes called a "continuum of resistance model," asserts that sampled persons who require additional effort to bring into the respondent pool (the late or difficult respondents) have lower fixed likelihoods of participating in a survey and are more similar to remaining nonrespondents than those measured with lower levels of effort (the early or easy respondents) (e.g., Fitzgerald and Fuller 1982; Lin and Schaeffer 1995). Statistical applications of the continuum of resistance model make similar assumptions, focus almost exclusively on call attempts, but also may incorporate additional information about fixed characteristics of the sampled person (Drew and Fuller 1980; Alho 1990; Wood et al. 2006). Interestingly, some statistical applications assume that response propensities decrease with additional levels of effort (e.g., Elliott et al. 2000) while others assume that response propensities increase (e.g., Tångdahl 2004), differing with regard to whether conditional (e.g., Elliott et al. 2000) or marginal (e.g., Tångdahl 2004) propensities are examined. The perspective taken in most of this research is that a person's propensity to respond is fixed, and those with lowest fixed propensities are measured only with extraordinary effort. As a result, levels of effort such as numbers of calls prior to interview or the need for refusal conversion have been used as proxy indicators for the fixed response propensity (Dunkelberg and Day 1973).

The increasing availability of paradata (Couper 1998) about survey administrative activities allows researchers to track the likelihood of a case being contacted and interviewed over the course of a data collection period. These paradata document the heterogeneity in recruitment protocol components experienced by different cases, not captured by the simple number of call attempts or refusal conversion. Some are never exposed to the survey request (the noncontacts), and may be completely unaware of their membership in the sample. Others have engaged in repeated conversations with interviewers over many contacts. Some of these sampled persons have engaged in considerable deliberation about the burdens and benefits of participation.

In short, often there is no single recruitment protocol consistently assigned to more than one sample person. Further, which components of the recruitment design are actually applied is determined by a complicated and sometimes unpredictable set of circumstances.

We can deduce that if each sampled case receives a different collection of recruitment protocol components, then each sample case might also exhibit variation in response propensities during the data collection. We assert that sampled cases begin the data collection period with "base" response propensities, as yet unaffected by the future events that they will experience, but differing over sample members. The base propensity is determined by the fixed attributes of the case (e.g., residential setting, gender roles). This base propensity is inestimable in a practical sense because the participatory request has not been forwarded. We hypothesize that each action that the survey organization subsequently takes might change that base propensity – raising it or lowering it. When an advance letter from a prestigious institution is read, describing the laudable purposes of the research, the propensity might increase. When, in contrast, an interviewer visits and delivers a complicated message about a burdensome request, the propensity might decline. We also hypothesize that the actions taken by the survey organization are not consistent in their effects over time or over sampled persons. Response propensity might increase immediately after the aforementioned advance letter arrives, but its effectiveness may decay as the memory of the message fades. Thus, not only is the participation decision stochastic, but a sample unit's likelihood of participation changes as the protocol evolves.

That is, we hypothesize that sample units have more than one response propensity during the course of any single survey, all conditional on the experienced features of the recruitment protocol. As new design features are introduced, the probability of response, p_i, changes. We call this a *dynamic* view of response propensities. As the protocol evolves for a given sample person, so does his/her response propensity.

If response propensities do indeed vary within a given person (affected by the stimuli of different protocol features), then a new perspective on postsurvey adjustment might be necessary. Although survey adjusters acknowledge that the missing data mechanism (Little and Rubin 2002) must be specified correctly for an effective adjustment procedure, the specific features of the recruitment protocol that lead to the missingness are usually not explicitly incorporated into the adjustment scheme. The goal of most weighting adjustments is to create groups that are homogeneous on response propensities within the groups and are heterogeneous across groups. If sampled individuals themselves display internal variation, what does it mean when the adjustment goal is stated as identifying respondents with homogeneous response propensities? Additionally, how do applications of recruitment protocol components affect heterogeneity in response propensities across groups?

Additionally, an expression for nonresponse bias of a respondent mean is a function of the covariance of the final response propensities (that is, those traditionally estimated with a logistic regression model) with the survey variable of interest over the average final response propensity (Bethlehem 2002). The true value of a survey variable of interest for each respondent is constant over the data collection period. If changes in the probabilities of participating at each call also translate into changes in the distribution of the final response propensities over the field period, then the covariance between propensity and the survey variables also might change for survey variables related to the causes of survey participation.

3. Data Resources

We examine two surveys – the National Survey of Family Growth (NSFG), conducted by the University of Michigan for the National Center for Health Statistics, and the Wisconsin Divorce Study (WDS), conducted by the University of Wisconsin-Madison. To simplify the discussion of dynamic response propensities, we focus on two features of the recruitment protocol to illustrate within-person change in response propensities. First, to link this article to previous examinations of level of effort (e.g., Elliott et al. 2000), we examine how response propensities change over days in the field (NSFG) or over calls (WDS) for the entire sample and for various subsets of the sample. This analysis focuses on "Phase 1" of the recruitment protocol. Second, each study used an explicit change in recruitment protocol at the end of the survey field period. The recruitment protocol in each study was noticeably different from the one that had previously been used, it was designed to be differentially appealing to previous nonrespondents (consistent with the idea of complementary design features in a responsive design; Groves and Heeringa 2006) and was applied to persons that had not previously participated in the study. That is, this analysis focuses on "Phase 2" of the protocol.

The National Survey of Family Growth is a survey of U.S. household members 15 to 44 years of age. The sample consists of a stratified, multistage area probability design, with one selected respondent per eligible household and oversamples of 15- to 19-year-olds and racial/ethnic minorities. A screening interview of approximately three minutes is conducted with an adult household member to identify whether there are one or more 15- to 44-year-old members. The selected respondent is asked to complete a 60- to 80-minute interview about sexual and fertility experiences, partnering, and family formation events. The sample consists of a rotating design at the PSU-level, with approximately 33 areas being interviewed at any one time, and a rotation of 25 areas each year. The analysis of this article is based on 60 total primary areas and represents the first seven quarters of interviewing, from June 2006 to March 2008. Each sample selected for the NSFG is in the field for a quarter of the year. That is, every twelve weeks, a new sample is released for data collection; all field activities on that sample are completed in twelve weeks or 84 days, and a new, fresh sample is released.

This quarterly 84-day period is divided into Phase 1 (first 70 days) and Phase 2 (last 14 days). To focus field efforts on a limited subset of remaining nonrespondents, Phase 2 continues data collection effort on a probability subsample of cases that have not been interviewed in the first phase. In Phase 1 all sample persons are given \$40 in cash upon agreeing to be interviewed; at day 70, Phase 2 begins and all selected nonrespondents are mailed \$40 and then offered an additional \$40 if they complete the main interview. In this article, we will focus on the efforts to obtain a main interview; the likelihood of providing a successful screener will not be examined here. In the quarters examined here, there are approximately $n = 9{,}200$ male and female main interview respondents, with approximately 8,500 Phase 1 respondents and 700 Phase 2 respondents. The main interview response rate varies over the quarters of the survey, but averages about 75% (AAPOR RR1; AAPOR 2011), with less than 5% of the addresses never contacted.

The Wisconsin Divorce Study[3] is a survey of $n = 733$ divorced persons aged 18 and older in four counties in Wisconsin, conducted in 1995. The sample was selected from a list of extracted divorce records from 1989 and 1993 in these counties; one member of the divorced couple was randomly chosen to be the named respondent. The selected sample member was asked to participate in the "Life Events and Satisfaction Survey," containing questions on satisfaction with life and relationships, marital and cohabitation history, childbearing history, education and work history, satisfaction with current relationships, and demographics. The first recruitment requests were delivered by telephone (Phase 1). At the conclusion of the telephone field period, all telephone nonrespondents were followed up by means of a mailed questionnaire (Phase 2). Overall, the response rate after the two phases (AAPOR RR1) was 71 percent, with a contact rate of 80 percent and a cooperation rate of 88 percent.

4. Model Specification

We estimate discrete time hazard models, also known as event history models, to predict a successfully completed interview at a given time point for both surveys. We choose a discrete time hazard model over other forms of survival analysis because discrete time hazard models easily accommodate "ties" (e.g., an interview occurring at the same call for two sample units) (Singer and Willett 2003).

Discrete time hazard models estimate the conditional probability of an event occurring (an interview), $h(t_{ij}) = \Pr[T_i = j | T_i \geq j]$ as a function of covariates, where $T_i = j$ indicates that the interview occurred at time period (call attempts) j, given that it has not occurred during any call prior to call j (Singer and Willett 2003). Here, the outcome of interest is an interview and each time point is a call attempt.

Estimation of discrete time hazard models requires a data set that has a record for each time point that the sampled household is approached, regardless of the outcome of the call. For example, a sampled unit in the WDS who received five contact attempts before a successful interview on the fifth attempt appears in the data set five times; a sampled unit who received ten contact attempts, did not successfully complete an interview but was no longer called in Phase 1, appears in the data set ten times. Since some sampled units are successfully interviewed, the case base for model estimation shrinks across calls. For example, sample units who gave an interview at call one are not contacted again at call two. Additionally, field decisions are often made to stop calling sample units that have not successfully been interviewed and as such, those sample units are not part of the sample on which the model can be estimated (censored cases). An alternative approach to discrete time hazard models is estimating the probability of interview on each call separately using logistic regression methods, changing the sample on which the probability of an interview

[3] The Wisconsin Divorce Study was funded by a grant (HD-31035 & HD32180-03) from the National Institute of Child Health and Human Development, National Institutes of Health. The WDS was designed and carried out at the Center for Demography and Ecology at the University of Wisconsin-Madison and Brigham Young University under the direction of Vaughn Call and Larry Bumpass.

is estimated across calls (e.g., Groves and Couper 1998). One limitation with this approach is that as the number of calls becomes large, the number of cases to estimate a logistic regression on any given call becomes small. One strength of discrete time hazard models is their ability to "borrow strength" across time periods for model parameter estimates, minimizing the effects of small samples for later calls.

Our goal is to estimate a conditional probability of an interview for each person in the survey, given both fixed characteristics of the sampled persons and time-varying characteristics of the recruitment protocol, for each time point of the survey – a day in the field for the NSFG and a call for the WDS. More formally, denote a set of call indicators as T, a set of time-varying protocol characteristics at call j as C_j and a set of time-invariant respondent and ecological characteristics as R. Time-invariant protocol characteristics can be included in a model only when they vary over sample units (e.g., random assignment of incentives or an advance letter). We estimate the probability of an interview at each call for each sampled person, given that they have not previously provided an interview:

$$p_{ij,c} = pr\,(\text{Interview at } T = j \,|\, \text{no interview at } T < j) = f(T, C_1, C_2, \ldots, C_{j-1}, C_j, R)$$

To obtain these estimated propensities at each call, we identify covariates available on both respondents and nonrespondents for each survey, including protocol characteristics and respondent and ecological characteristics. Some protocol characteristics are constant across all sampled persons (e.g., survey topic) and thus their individual contribution to propensity cannot be estimated. In the NSFG, we estimate the conditional probability that the next call will generate a main interview for each day of Phases 1 and 2 for all active sample persons. Similarly, for the WDS, we estimate the conditional probability that the next contact attempt (telephone call or mail out) will generate a completed interview at each contact attempt. We choose calls rather than days as the time points in the WDS because it is a telephone survey; the sample was released in replicates over multiple months rather than the release of all cases at the beginning of each quarter in the NSFG. In the NSFG, the time-invariant predictors include indicators at the sample segment level – urbanicity of the residence, presence of commercial units in the neighborhood, presence of non-English speakers, perceived safety issues for the neighborhood, presence of a multi-unit structure, and physical impediments to entrance. These time-invariant predictors also included observations at the sample person level – whether the chosen person was a teenager, a woman, an African-American, and whether the screener was administered in Spanish, and whether the interviewer judged the person to be in an active sexual relationship. Time-varying protocol characteristics include number of prior calls and contacts on the case, whether prior calls ever yielded a contact with the respondent, whether the sample person had posed questions or given various statements to the interviewer. All of these variables had been found predictive of the likelihood of a main interview being taken on the next call (see Groves et al. 2005). Model coefficients are presented in Appendix Table A.

In the WDS, time-varying protocol characteristics included in the model are the number of previous contacts, the number of days since the prior call, whether the call was made on a weekday or weekend, whether the contact attempt was made via telephone or mail, and an interaction effect between the number of prior contacts and whether the request was made by mail. The fixed respondent characteristics are attributes available from the frame, including age, education, the number of children for whom the respondent was awarded custody, their ex-spouse was awarded custody, or for whom they share custody, and whether they lived in Wisconsin at the time of the interview. Census data, merged on at the zip code level, include urbanicity, age composition, race composition, mobility, and commuting time composition of the zip code. Model coefficients are presented in Appendix Table B.

Finally, we examine whether the sample-based estimates change over the two phases. First, for the Wisconsin Divorce Study, we estimate a logistic regression model predicting the probability of an interview at the end of the telephone phase and the telephone and mail phases combined. That is, we estimate a response propensity model typically used for adjustment (Little 1986) at the end of each phase. Coefficients that differ between models estimated at the end of Phase 1 and Phases 1 and 2 combined indicate that the addition of Phase 2 brought in sampled persons to the respondent pool who previously had not cooperated with the Phase 1 request and that this recruitment was differential over subgroups. We then examine four key survey estimates for the WDS (length of marriage, number of marriages, months elapsed since the divorce, and the respondent's age) and four key estimates for the NSFG (number of abortions, number of life-time opposite sex partners for men and for women, and number of pregnancies) at the end of Phase 1 and the end of Phase 2. These analyses allow us to examine the question of whether the changes in response propensities also correspond with changes in survey estimates.

5. Results

We now turn to each of our four research questions. All of the analyses presented below in Sections 5.1 through 5.3 use estimated probabilities from the discrete time hazard models. Coefficients for these models are presented in Appendix Tables A and B.

5.1. Is there Variation in Propensities to Respond Within Individuals Over the Course of a Data Collection Period?

In most surveys, the application of the greatest number of new protocol components comes at the beginning and the end of data collection (e.g., the close out period). All of the prenotification and initial contact features occur at the beginning, and all of the refusal conversion attempts occur at the end. We speculate that the largest change in individual response propensities to participate at the next recruitment request occurs immediately when a recruitment protocol component is applied. We hypothesize that the application loses its effectiveness over time, either because the respondent no longer considers this new component in his or her decision or the persons for whom the component was attractive are culled out. For example, an advance letter is maximally effective if the interviewer approaches the sampled household soon after it is read; sample units that are approached later are less likely to remember the letter or to associate it with the interviewer making the recruitment request. We would hypothesize that the largest within-person variation in response propensities would be observed at the beginning and end of a field period in most surveys.

The first examination of within-person variation in response propensity concerns whether the estimated probability of participating in the survey at each call changes systematically as the data collection period progresses. We examine the average response propensity for all sampled cases and for a subset of cases who are not interviewed during the first phase in each survey. If sampled persons who do not participate in Phase 1 have a uniformly low response propensity that does not change over time, then we may see shifts in the average estimated probability of giving an interview for the overall sample but no change for the subset who are interviewed during the first phase. If, in contrast, the initial recruitment protocol declines in its effectiveness for all sampled persons over the field period, then we will see a decline in the average response propensity for all sampled persons, whether or not they provided an interview at the beginning of the field period.

First, in the National Survey of Family Growth, we examine the average estimated probability of a main interview on each day of the field period. Obtained from the aforementioned discrete time hazard model, we estimate a probability of a main interview for each day of the data collection period for each person eligible to be attempted on that day. We plot the mean across all sample persons of these estimated day-by-day probabilities to examine overall shifts in the likelihood of participating. Figure 1 is a plot of mean estimated probabilities for each sampled person that the next call will produce an interview, among sample persons who were successfully screened and eligible to give a main NSFG interview. The top plotted line (beginning at about .36 probability on day 1) is the mean person-level estimated propensity among all cases that were active on a given day, but provided a main interview by day 70 (the last day of Phase 1) of the survey. On day 70 of the quarterly data collection period, a probability subsample of remaining active cases is drawn. For these cases, a new recruitment protocol begins, with a notable

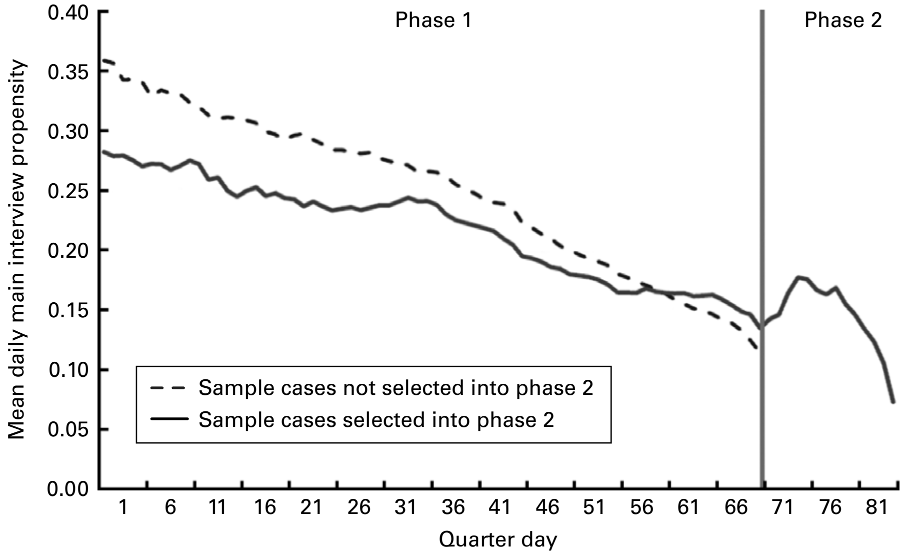

Fig. 1. Mean Estimated Response Propensity by Day of Quarterly Field Period, for sample cases not selected into Phase 2 and sample cases selected into Phase 2, National Survey of Family Growth

feature of increased incentives. The lower plotted line (beginning at about .28 probability on day 1) is the estimated mean conditional probability for these cases that did not grant an interview during Phase 1 and were selected into Phase 2.

The graph illustrates that the average probability of obtaining an interview changes over the course of the data collection. The pattern of change is interpretable. The probability of granting a main interview in the NSFG is higher during the early days of the data collection than during the later days of the data collection. This is consistent with the above hypothesis that the maximal effect of a recruitment protocol element occurs immediately upon its application and declines in effectiveness over time. The cases remaining at the end of the Phase 1 data collection are those with low estimated probabilities. That is, cases with higher response propensities are interviewed, and thus are no longer part of the case base used to estimate the probability of an interview on later dates. The dashed line in Fig. 1 varies in case base, that is, is based on different cases, as we move from left to right on the *x*-axis. The lower solid line in Fig. 1 tries to repair that: it contains estimated probabilities for the subset of those sampled cases that did not give a main interview over the entire course of the data collection in Phase 1. If changes in estimated response propensities occurred only because people with high likelihoods of participating left the sample pool as the data collection progressed, then the solid line should be flat. However, sampled persons who were not interviewed during Phase 1 also show dramatic declines in their estimated conditional probabilities of interview during Phase 1.

Not surprisingly, those interviewed in the first phase begin the field period with higher estimated propensities (reflecting their base propensities being larger) than those not interviewed in the first phase. This indicates that, on average, respondent and household characteristics of those who are not interviewed during Phase 1 tend to be associated with lower main interview response rates in this phase concerning such characteristics as urbanicity, age, and single-person household status. As the field period progresses, the cases who were not interviewed during Phase 1 have repeated contacts that were unsuccessful, thereby reinforcing the lower predicted base probabilities of interview.

In the Wisconsin Divorce Study, we examine the predicted probability of an interview on each call of the telephone field period. A similar graph for the Divorce Study (Fig. 2) shows the probability of obtaining an interview for each call made during the field period on the *y*-axis and the call number on the *x*-axis. As with the NSFG, we separate the sample into two subsets – (1) sampled persons who provide an interview during Phase 1 (the telephone phase) and (2) sampled persons who do not provide an interview during Phase 2 (the mail phase). The dashed line represents sampled persons that were interviewed on the telephone. As with NSFG, as we move from left to right on the *x*-axis, there is a decline in the mean propensity that the next call will produce an interview from as high as an estimated 0.10 probability at the beginning of the field period to less than an estimated 0.01 probability at the end of the field period. The average estimated response propensity at each call for the WDS is substantially less than that for the NSFG because the response rate at the end of the phone phase is about 48 percent, consistent with many telephone surveys, whereas the main interview response rate, conditional on successful completion of a screener, at the end of Phase 1 for the NSFG is between 60 and 70 percent, depending on the quarter and the subgroup. The lower solid line represents sampled persons who

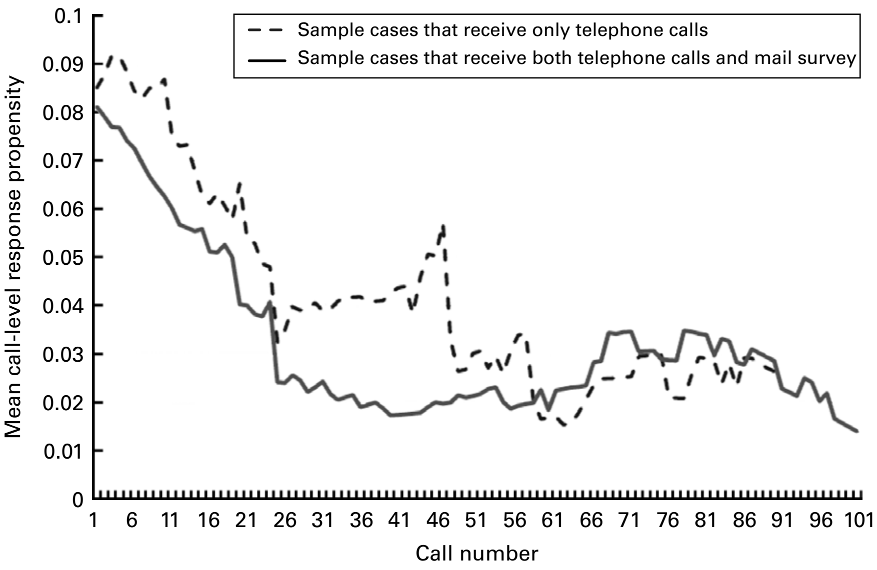

Fig. 2. Mean Estimated Response Propensity over Calls Made During the Telephone Field Period, Wisconsin Divorce Study

were attempted, but not interviewed, on the telephone and consequently were sent mail questionnaires. The pattern for this group of sampled persons is the same, decreasing over the course of the data collection period. As in the NSFG, this subset of cases starts out with lower base response propensities and stays lower through the calling period. From the model coefficients, we can deduce that these cases are more likely to not have kids, to live in different types of areas, and to have been called on a Friday or Saturday.

On average, the predicted probability of obtaining a completed mail questionnaire is 0.46 (not shown on graph), a dramatic increase. Although the mail survey was sent out on the same day for all remaining cases, at what point this mail-out occurred in the progression of call attempts varied. For example, among the 366 sampled persons who were sent a mail survey, roughly half of the nonresponding sampled cases were called three or fewer times (largely due to late release of sample lines in the telephone field period); about 10 percent of the cases received 20 or more calls before being sent the mail survey.

Figure 1 and Fig. 2 show impressive change in mean probabilities over time. Thus, there is a change in the average estimated probability of responding over the course of a data collection field period. However, a limitation of Figs. 1 and 2 is that the number of sampled persons for whom the propensity is calculated for each day or call varies over days or calls, depending on the active cases of the moment. Thus the case base varies over the *x*-axes in Figs. 1 and 2 above. To examine whether individual propensities are directly affected by changes in the recruitment protocol, changes in propensities for the same group of sampled persons should be estimated. That is, a better contrast than examining changes in propensities over time would be to examine the average propensities for the same group of cases before and after an intervention deliberately intended to change response propensities for the group to which it is applied.

5.2. Is there Any Evidence That Propensities Are Influenced by Specific Actions Taken the Survey Recruitment Protocol?

To answer the question of whether specific actions of the design can change propensities, we compare the person-level Phase 1 – Phase 2 propensities for each study. If the average Phase 2 propensities are statistically different from the average Phase 1 propensities for the same group of sampled persons, then we have evidence that the survey recruitment protocol affects response propensities. One way to examine this for the NSFG is to compare the estimated propensity on the last day an interviewer worked each case during Phase 2, whether or not an interview actually occurred, with the estimated propensity that the sampled person will complete an interview on the 70th day in Phase 1. The difference between these estimated mean propensities indicates whether the Phase 2 protocol yielded a net increase in propensities. To the extent that the Phase 2 protocol declines in effectiveness over the course of Phase 2, this will attenuate the difference between Phase 1 and Phase 2 propensities. We look only at the estimated propensities from the models presented in Section 5.1 above for the cases that were subsampled into the second phase. Among this group, the mean probability that the next call will generate a main interview on the last day of Phase 1 is 0.134; the mean estimated propensity for the last day on which the case was attempted during Phase 2 is 0.157, giving an increase in mean propensity of 0.022 (ste $= 0.0059$, $p < .001$). The Phase 2 protocol has a net effect of increasing estimated propensities, even though these cases exhibited very low propensities by the end of Phase 1.

A similar comparison in the divorce study involves the estimated propensity on the last day the case was active in Phase 1 (the telephone field period) and the estimated propensity in Phase 2 (the mail request). Here, the mean estimated probability that an interview will occur on the last active day of the telephone field period is 0.08 (se $= 0.0039$). The mean estimated probability that the mail survey for the same group of sampled persons will yield a completed interview is 0.46 (se $= 0.0062$), giving an increase in mean propensity of 0.38 (se $= 0.095$, $p < .0001$). This also corresponds to a highly significant coefficient in the WDS discrete time hazard model (beta $= 2.47$, se(beta) $= 0.194$, $p < .0001$). The mail mode switch had an overall effect of increasing the mean propensity of the cases that had not yet completed an interview in the phone survey.

Thus, in both surveys we have clear evidence that a change in recruitment protocol affects an individual's response propensity. The increase in incentive in Phase 2 of NSFG and the mode switch in the WDS both yielded significant increases in the average propensity of the sample pool who received these protocol components. This is notable because this group generally had low propensity to participate at the end of Phase 1 in both studies.

An alternative method for examining influences of the recruitment protocol is to look at the coefficients on time-varying protocol characteristics in the estimated discrete time hazard model. A time-varying protocol characteristic is one that is applied across sampled respondents and that varies in how it is applied over time within sampled respondents. A time-varying protocol characteristic common to multiple surveys is the day on which contact attempts are made with sampled respondents. In the WDS, for example, we see that calls made on a Friday or Saturday decrease the probability of an interview relative to calls made on a Sunday (beta = -0.369, se(beta) = 0.183, $p < .05$). Thus, not only do large protocol changes such as an incentive increase or a mode switch yield dramatic differences in the likelihood to respond at a given point in the field period, but small changes such as the day of the week on which a call is made also yield differences in the likelihood of obtaining an interview.

5.3. Is there Any Evidence That the Effectiveness of These Specific Survey Recruitment Protocol Changes Varies Over Time or Over Sample Units?

We now examine whether these changes in estimated propensity are consistent over time and sample units. Our perspective outlined above asserts that the effect of the intervention declines over time. It also asserts that the effectiveness of an intervention may differ depending on characteristics of the sample units, in terms either of fixed characteristics or of previously experienced protocol components. We now turn to these two questions, examining the first in the NSFG and the second in the WDS.

Given the design of the NSFG we can ask whether the effects of the interventions themselves change over the course of the post-intervention period. In the NSFG, we elaborate the difference between the estimated propensities during Phase 2 and the estimated propensities at the end of Phase 1 by examining this difference for each day of the Phase 2 fourteen-day field period. Figure 3 presents this Phase 2 – Phase 1 difference in estimated propensities. The cases in the figure include only those that were sampled into Phase 2 as active main cases, that is, those who were not interviewed during the first phase. The x-axis of the figure is the day of the 14-day data collection period of Phase 2, the y-axis is the estimated propensity for the last day the case was approached during Phase 2 minus the propensity for the case on the last day of Phase 1. For each day, the case base for the estimated increase is cases that were finalized (that is, gave an interview or were designated a final refusal) on that day of Phase 2. The lighter lines about the estimated increase reflect two standard errors away from the sampled case-based estimate.

The shape of the estimated increase shows growth in the Phase 2 – Phase 1 propensities over the first five days of Phase 2. We believe that this reflects the time required to receive

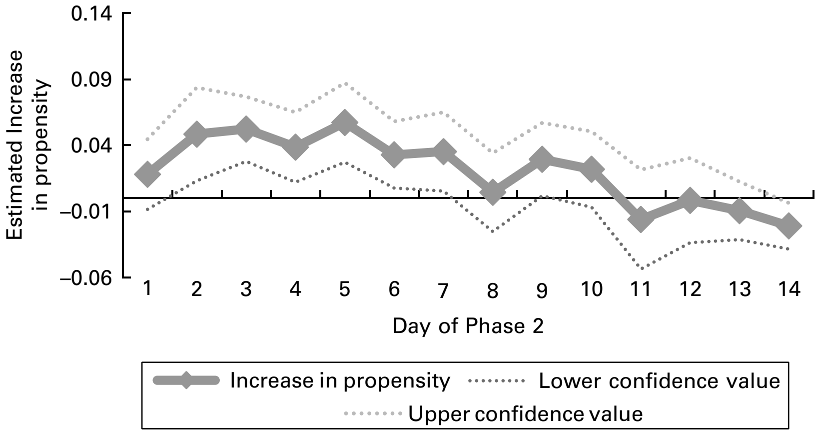

Fig. 3. Estimated Increase in Mean Propensity, National Survey of Family Growth, from Phase 1 to Phase 2

the priority mailing containing the new incentive offer. After this point, the increase in propensities for the Phase 2 protocol over Phase 1 propensities declines, as the remaining cases are less receptive to benefits of the higher incentive and refusal conversion efforts proceed. As expected, Phase 2 ends with a case base with mean propensities that are even lower than the cases they exhibited at the end of Phase 1. They have been exposed to the Phase 2 new incentive structure, rejected it, and there are no new features of the Phase 2 recruitment protocol that could raise their propensities.

Our perspective also permits the introduction of the new protocol component to be modified by characteristics of a sampled unit. One characteristic relevant to multiple surveys is the package of recruitment protocol components experienced by the respondent prior to the introduction of the new protocol feature. One examination of this is to ask whether the amount of effort exerted on a case in Phase 1 affects the effectiveness of the mode switch in Phase 2 of the divorce study, among those who did not answer in Phase 1. We expand the divorce study to examine the mean increase in propensities by the number of the last call attempt in the telephone mode. The telephone survey was in the field for approximately three and a half months. The last telephone call for some cases was during the first month; other cases were called through the last month. The last call number was not randomly assigned.

When we look at the average difference in response propensities between the last call made before the mode switch and the mode switch, we see a decreasing trend (Fig. 4). That is, although the net effect of the mode switch was positive for all sample units as expected, the effectiveness of the protocol switch was greater among cases who did not receive extra effort in Phase 1. We also take as evidence a significant interaction effect between the protocol switch and the number of previous contact attempts in the response hazard model. The coefficient on the interaction term is significant and negative, indicating decreased effectiveness of the mode switch depending on the number of previous contacts (beta $= -0.2231$, se $= 0.07$, $p = .0011$). That is, an additional recruitment request made in a different mode is most effective for those whose Phase 1 recruitment protocol was the least developed.

Thus, although the change in the recruitment protocol increased response propensities, the change was not uniform over time or sampled persons. In the NSFG, the effectiveness of the new recruitment protocol used in Phase 2 declined over the course of Phase 2,

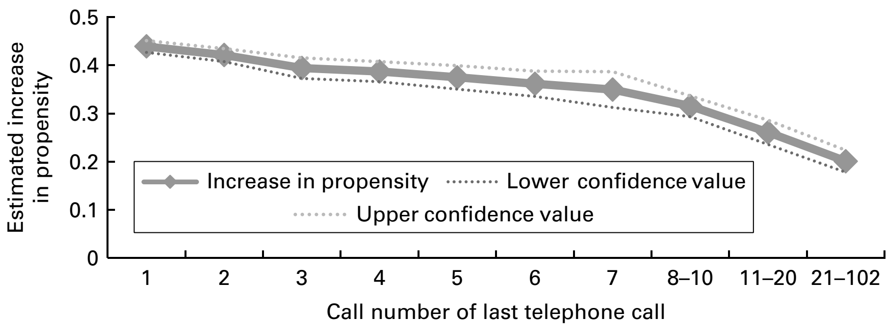

Fig. 4. Estimated Increase in Mean Propensity, Wisconsin Divorce Study, from Telephone Mode to Mail Mode

mirroring the decline in Phase 1 propensities over time. In the WDS, the effectiveness of the mode switch differed depending on the level of effort previously exerted on the case. This is consistent with previous findings showing that an additional contact attempt raises contact and cooperation rates earlier in the field period (after only one or two calls) rather than later (e.g., Groves and Couper 1998, p. 83). We now examine whether either of these protocol changes also changed the characteristics of the respondent pool on survey statistics of interest.

5.4. Does the Change in Propensities Coincide With Changes in the Nature of Nonresponse Bias on Key Survey Estimates?

One method for investigating changes in the risk of nonresponse bias at the end of the survey for different parts of the survey recruitment process is to estimate a logistic regression model predicting cooperation at each stage of the recruitment process and examine the change in estimated coefficients. For example, one could estimate a logistic regression predicting cooperation at each call to examine the effects of repeated call attempts on correlates of survey participation. Here, our interest is in the effects of the mode switch in the divorce study on correlates of survey participation. We estimate a logistic regression model predicting cooperation among all cases, where cases that completed a phone interview take the value of "1" and cases that did not complete a phone interview take the value of "0." We compare this to a logistic regression predicting cooperation among all cases, but here cases that completed either a phone interview or returned a mail questionnaire take a value of "1" and all remaining cases take a value of "0."

Table 1. Estimated Coefficients and Standard Errors for Logistic Regression Predicting Interview at the End of the Phone Field Period and at the end of the Phone and Mail Combined, Wisconsin Divorce Study

	Only phone completes			All completes		
	Beta	SE	Sig.	Beta	SE	Sig.
Intercept	−2.14	0.51	****	−0.45	0.53	
Age	0.021	0.009	*	0.012	0.010	
Female	0.30	0.16	+	0.26	0.18	
Education = missing	0.90	0.44	*	0.36	0.47	
Education = College	0.58	0.33	+	0.45	0.34	
Education = Some college	0.83	0.32	*	0.53	0.34	
Education = HS grad	0.45	0.29		0.08	0.29	
# Children, sole custody for R	0.28	0.12	*	0.23	0.13	+
# Children, joint custody	0.30	0.10	**	0.24	0.12	*
# Children, sole custody for spouse	0.05	0.13		0.16	0.14	
Married in Wisconsin	−0.09	0.18		0.04	0.20	
County of divorce	−0.04	0.17		0.01	0.18	
Currently live in Wisconsin	0.52	0.21	*	0.35	0.22	
Likelihood ratio test (12 d.f.)	39.20		****	19.33		+
Pseudo-R2	3.9%			2.2%		

$+p < .10$, $*p < .05$, $**p < .01$, $***p < .001$, $****p < .0001$.

At the end of the phone survey, older persons, women, persons with higher levels of education, persons with children, and those who lived in Wisconsin at the time of the field period were significantly more likely to be respondents than other persons. With the addition of the mail survey, no difference was observed by age, sex, education, or residence location; some differences remain across persons with children. That is, response rates for men and women, older and younger persons, and persons with more or less education were equalized with the mail survey.

We now turn to three substantive variables from the Wisconsin Divorce Study, the length of marriage, the number of marriages, and the time elapsed since the last divorce. We also examine the sampled persons' age as an example of an item included in the propensity model. These variables are calculated from divorce record data available for both respondents and nonrespondents. The mean of each variable for respondents and nonrespondents for the phone survey is compared to that for the phone and mail survey combined. Each is compared to the overall mean. The correlations between the predicted propensities from both of the above models with each survey variable are also estimated on the full sample (respondents and nonrespondents; not shown in Table 2).

Table 2 shows the means and standard errors for the survey variables for five groups – overall, respondents and nonrespondents before the mode switch, and respondents and nonrespondents after the mode switch. For example, the mean length of marriage for the full sample is 130.34 months. At the conclusion of the telephone survey, the mean length of marriage for respondents is 139.82 (se = 5.27) and for nonrespondents is 121.43 (se = 4.86), a statistically significant difference of 18.39 months between respondents and nonrespondents ($p < .01$), and an overestimate of the full sample mean of 9.48 months. The correlation between the estimated propensity and the mean length of marriage before the mode switch is 0.37 ($p < .0001$).

After the mode switch, the mean length of marriage is 134.17 months for respondents and 120.80 months for nonrespondents, a difference of 13.37 months ($p < .10$). Furthermore, the overestimate of the full sample mean has been reduced to 3.83 months and the correlation between estimated propensity and length of marriage is 0.32 ($p < .0001$), smaller than that before the mode switch.

Table 2. Means and Standard Errors, Total Sample, Respondents and Nonrespondents after the Phone Survey, and Respondents and Nonrespondents, Phone and Mail Surveys Combined, Wisconsin Divorce Study

	Length of marriage		Number of marriages		Months since divorce		Respondent's age	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Full sample	130.34	3.59	1.22	0.016	49.69	0.90	39.83	0.32
Before mode switch								
Respondents – phone only	139.82	5.27	1.19	0.021	51.10	1.28	40.57	0.46
Nonrespondents	121.43	4.86	1.24	0.024	48.37	1.26	39.13	0.45
After mode switch								
Respondents – mail + phone	134.17	4.29	1.20	0.017	50.44	1.06	40.06	0.38
Nonrespondents	120.80	6.53	1.27	0.036	47.84	1.71	39.24	0.61

The difference between respondents and nonrespondents is reduced after the mode switch for the mean number of months elapsed since the divorce and for the respondent age, but increases slightly for the mean number of marriages for the respondent. For the two items in which the difference between respondents and nonrespondents decreases after the mode switch, the correlation between estimated propensity and the survey variables also decreases. For the item in which the difference increases, the correlation also increases (in absolute magnitude). That is, the reduction in nonresponse bias has also been reflected in a smaller correlation between propensity and the survey variables.

Since there are no records for both respondents and nonrespondents in NSFG, we instead examine whether estimates based on the Phase 1 respondents are meaningfully different from estimates based on the Phase 2 respondents. Table 3 shows the means (in standard deviation units) of three key survey variables from the National Survey of Family Growth – the number of abortions, the number of lifetime opposite sex partners, and the number of pregnancies. For each variable, to permit comparison across estimates the difference between the full sample mean and the mean of each variable in Phase 1 and Phase 2 is expressed in standard deviation units of the full sample mean [(Phase i mean – Full sample mean)/Standard deviation of full sample], where $i = (1,2)$.

For Phase 1, among female respondents who reported completed pregnancies, the mean number of abortions was -0.047 standard deviation units below the full sample mean. Phase 2 yielded respondents who, on average, had more abortions than their Phase 1 counterparts, with a mean 0.13 standard deviations above the full sample mean. For the mean number of lifetime opposite sex partners, Phase 1 female respondents on average reported 0.05 standard deviation units more partners than the full sample, whereas Phase 1 male respondents reported -0.045 standard deviation units below the full sample mean. Phase 2 brought in females who had had fewer opposite sex partners and males who had more opposite sex partners than in Phase 1. Finally, Phase 1 yielded female respondents who, on average, had slightly fewer pregnancies than Phase 2 female respondents (-0.016 standard deviation units below the full sample mean compared to 0.044 standard deviation units above the full sample mean).

Table 3. Mean number of abortions, number of lifetime opposite sex partners, and number of pregnancies, in full sample standard deviation units, Phase 1 and Phase 2, National Survey of Family Growth

	Number of abortions[1]		Number of life time opposite sex partners[2]				Number of pregnancies	
	Females		Females		Males		Females	
	Mean	n	Mean	n	Mean	n	Mean	n
Phase 1	−0.0468	2870	0.0530	4717	−0.0453	3821	−0.0162	4750
Phase 2	0.1304	217	−0.1444	371	0.0913	331	0.0436	374

Notes: All results reported in standard deviation units of the full sample mean. The full sample mean is coded as "0" for all estimates. All results reflect probability of selection weights and double sample weights.

[1] Abortions among Rs reporting completed pregnancies.

[2] Life Time Opposite Sex Partners topcoded at 7 partners.

Not all subgroups respond to a change in the recruitment protocol uniformly. For example, in the WDS, there were significant differences in response rates across education groups before the mode switch; after the mode switch these differences were minimized. However, persons with children remained more likely to participate in the WDS both before and after the mode switch. Concurrently, a change in the recruitment protocol does not affect all sample estimates uniformly. This is consistent with previous research showing that there is great within-survey variation in nonresponse bias of unadjusted respondent means and proportions (Groves and Peytcheva 2008). In both the WDS and in NSFG, respondents brought in during Phase 2 were different from those recruited during Phase 1 on some characteristics (e.g., length of marriage, number of abortions) but less so on other characteristics (e.g., number of previous marriages, number of pregnancies). Thus, a change in recruitment protocol changes nonresponse bias of some but not all estimates.

6. Summary and Discussion

This article has addressed changes during a data collection period in individual propensities to be interviewed. Using discrete time hazard models, we were able to detect the expected overall decline in average probability of being interviewed on the next call, given a model that contains both fixed attributes and time-varying attributes in two surveys. Further, we showed that direct interventions of incentive changes and mode switches act to increase these conditional probabilities. We also examined whether these design features had a constant level of effectiveness over time and over sampled persons. We found that the effectiveness of new features in a recruitment protocol (e.g., the incentive increase in the NSFG) declined over time and also may be affected by the prior history of effort on the case (e.g., the mode switch in the WDS). Finally, we showed that the introduction of a new recruitment protocol component can affect the overall relationship between fixed characteristics of sampled persons and households and the probability of interview estimated using logistic regression as typically used for weighting adjustments. We also showed that the introduction of a new recruitment protocol component does not uniformly affect nonresponse bias properties of key survey variables.

That is, where the recruitment protocol ends affects which respondents are in the final data set. The individual components of the recruitment protocol, and the order in which they are administered, also affect who is in the final respondent data set. Thus, how these components are administered during data collection and the persons to whom they are administered affects the final adjustment procedures.

The practical implication of this is that tracking the propensities and measuring the impact of field actions on those propensities is merited from the postsurvey adjustment perspective. The design's interventions in the field can affect the distribution of the expected propensities to respond and thus the household and respondent characteristics that will play a large role in a postsurvey adjustment model. Postsurvey adjustments condition on the existing distribution of respondents and nonrespondents, and also on the recruitment protocol. Yet merely conditioning the postsurvey adjustments on how the field efforts happened to be spread over a given sample ignores the fact that the spread of those efforts can be under the control of the design. That is, the recruitment protocol affects who ends up as a final respondent or nonrespondent, information that is traditionally ignored in postsurvey adjustments.

7. Implications for Practice and Estimation

The importance of this perspective of dynamic response propensities is that how the sample cases' propensities evolve over the course of the data collection is not fixed by the survey design or fully determined by fixed respondent characteristics. Who ends up as a final respondent or nonrespondent thus depends on the collection of protocol features that they receive. Thus, the nonresponse bias properties of respondent-based estimates also depend on the recruitment techniques used, which techniques are applied to which sampled persons, and the characteristics of these sample persons. Similarly, the effectiveness of adjustment techniques also depends on how the recruitment protocol was applied. This work is distinct from a continuum of resistance model (e.g., Lin and Schaeffer 1995) in that we make no assumptions about those persons requiring more effort being more similar to remaining nonrespondents. It is also distinct from the stochastic propensity perspective in that we believe – and indeed empirically demonstrate – that response propensities are malleable for a given individual.

Based on these results, we believe that the specification of postsurvey adjustments should reflect variation in application and reaction to recruitment protocol components applied to different cases, especially when those components are related to the key survey measures. We encourage future research incorporating dynamic (i.e., conditional) participation probabilities for purposes of adjustment and estimation. Previous work in this area incorporating number of call attempts (e.g., Biemer 2009) or patterns of outcomes from call attempts (e.g., Kreuter and Kohler 2009) is a start. We would encourage examination of nonresponse models that more explicitly condition on different distinct phases of the data collection period, incorporate protocol components beyond call attempts such as changes in interviewers or modes, use of incentives, and so on, and account for the decline in attractiveness of different selectively applied protocol features over time.

We also recommend building sets of informative auxiliary variables that are useful predictors of dynamic propensities and survey variables and tracking them during the data collection period. This includes information about the timing and implementation of new protocol features, in addition to auxiliary variables that are known to be closely related to key survey variables. With these vectors of auxiliary variables, "phase capacity" (Groves and Heeringa 2006) can be monitored for decisions on when to implement planned new protocol features as in a responsive design, as can whether the observed sample matches known population benchmarks (i.e., a "balanced" sample, Särndal 2011) on comparable variables. Additionally, base and dynamic propensities can be more robustly estimated with extensive auxiliary data.

Finally, we suggest examining the covariance of estimated propensities and the survey variables as a proxy indicator of the nature of the effect of field interventions on postsurvey adjustments and nonresponse bias (Groves et al. 2008). Evaluating the effect of these interventions will be accomplished most readily when the field interventions are randomly assigned (as in a responsive design) or have clearly defined decision rules (e.g., implement after day 70). As survey goals balance increasing response rates with minimizing nonresponse bias, tools are needed that monitor changes in nonresponse bias, either for adjusted or unadjusted estimates. Since the covariance of response propensity and the survey variables is the numerator of the nonresponse bias expression, monitoring changes in this expression will also permit evaluation of changes in nonresponse bias. Calculating the covariance between an estimated response propensity and the survey variables for respondents alone must be viewed with some caution, given that there may be nonresponse bias in the covariance estimate itself. In general, the use of the covariance between the survey variables and an estimated propensity as a diagnostic tool with regard to nonresponse bias requires more empirical and theoretical development, such as recent work by Andridge and Little (2011) using a proxy-pattern mixture model.

Appendix Tables

Table A. Coefficients and Standard Errors for Discrete Time Hazard Model Predicting Main Interview in the National Survey of Family Growth, August 18, 2007

	Beta		Se
Intercept	−1.249	****	0.061
Number of calls	−0.180	****	0.006
Time-varying protocol characteristics			
Number of prior contacts	0.253	****	0.031
Ever had a prior contact	1.262	****	0.056
Ever had prior resistance	−1.028	***	0.275
Ever made a statement	0.360	***	0.101
Made a statement on last call	−0.863	****	0.101
Ever asked a question	0.005		0.075
Asked a question on last call	−0.116		0.080
Previously showed maximum resistance	−0.559	+	0.304
Nontime-varying respondent characteristics			
Teen	0.325	****	0.045
Sex	−0.016		0.031
Race	−0.006		0.034
Language	−0.206	**	0.069
Single person household	0.063		0.052
Presence of children	0.064	*	0.032
Interviewer believes the selected person is sexually active	−0.092	*	0.042
Nontime-varying area level characteristics			
Urban	−0.115	**	0.041
Residential	−0.016		0.032
Non-English	0.092		0.110
Spanish	−0.045		0.112
Safe condition	0.123	**	0.037
Multi-unit structure	0.028		0.037
Physical access impediments	−0.056		0.053

Note: NSFG models were reestimated every day of the data collection, fully interactive with the day of the data collection. This model was estimated on August 18, 2007. ****$p < .0001$, ***$p < .001$, **$p < .01$, *$p < .05$, $+\, p < .10$.

Table B. Coefficients and Standard Errors for Discrete Time Hazard Model Predicting Interview in the Wisconsin Divorce Study

	Beta		Se
Intercept	−3.559	****	0.568
Number of calls	−0.097	****	0.021
Number of calls^2	0.002	**	0.001
Number of calls^3	−0.00001	*	0.00001
Time-varying protocol characteristics			
Number of prior contacts	0.297	****	0.032
Mail survey	2.470	****	0.194
Number of prior contacts*mail survey	−0.223	**	0.068
Monday–Thursday call	−0.012		0.151
Friday–Saturday call	−0.369	*	0.183
Number of days between calls	0.001		0.003
Nontime-varying respondent characteristics			
Age	0.005		0.006
Female	0.102		0.109
# kids for which R has sole custody	0.107		0.074
# kids for which R and spouse share custody	0.158	*	0.068
# kids for which spouse has sole custody	0.052		0.084
Education = missing	0.495	+	0.300
Education = college	0.012		0.221
Education = some college	0.297		0.214
Education = high school graduate	−0.034		0.197
Live in Wisconsin	−0.192		0.157
Nontime-varying zip-code level characteristics			
% Urban	0.185		0.153
% 55 and older	−2.065	+	1.057
% non-White	−0.575		0.792
% lived in same house 5 years ago	1.692	**	0.653
% commute less than 15 minutes	0.967	+	0.531
% work at home	3.809	*	1.621

****$p < .0001$, ***$p < .001$, **$p < .01$, *$p < .05$, + $p < .10$.

8. References

AAPOR (2011). Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 7th edition: The American Association for Public Opinion Research.

Alho, J.M. (1990). Adjusting for Nonresponse Bias Using Logistic Regression. Biometrika, 77, 617–624.

Andridge, R.R. and Little, R.J.A. (2011). Proxy Pattern-Mixture Analysis for Survey Nonresponse. Journal of Official Statistics, 27, 153–180.

Baumgartner, R., Rathbun, P., Boyle, K., Welsh, M., and Laughland, D. (1998). The Effect of Prepaid Monetary Incentives on Mail Survey Response Rates and Response Quality. Paper presented at the Annual Conference of the American Association of Public Opinion Research, St. Louis, MO.

Bethlehem, J. (2002). Weighting Nonresponse Adjustments Based on Auxiliary Information. Survey Nonresponse, R.M. Groves, D.A. Dillman, J.L. Eltinge, and R.J.A. Little (eds), Chapter 18. New York: John Wiley and Sons, Inc.

Biemer, P. (2009). Incorporating Level of Effort Paradata in Nonresponse Adjustments. Paper presented at the JPSM Distinguished Lecture Series.

Brick, J.M., Allen, B., Cunningham, P., and Maklan, D. (1996). Outcomes of a Calling Protocol in a Telephone Survey. Proceedings of the American Statistical Association, Survey Research Methods Section, 142–149.

Couper, M.P. (1998). Measuring Survey Quality in a CASIC Environment. Proceedings of the American Statistical Association, Survey Research Methods Section, 41–49.

Curtin, R., Presser, S., and Singer, E. (2000). The Effects of Response Rate Changes on the Index of Consumer Sentiment. Public Opinion Quarterly, 64, 413–428.

de Leeuw, E. (2005). To Mix or Not to Mix Data Collection Modes in Surveys. Journal of Official Statistics, 21, 233–255.

de Leeuw, E. and de Heer, W. (2002). Trends in Household Survey Nonresponse: A Longitudinal and International Perspective. Survey Nonresponse, R.M. Groves, D.A. Dillman, J.L. Eltinge, and R.J.A. Little (eds). New York: John Wiley and Sons, 41–54.

Drew, J.H. and Fuller, W.A. (1980). Modeling Nonresponse in Surveys with Callbacks. Proceedings of the American Statistical Association, Survey Research Methods Section, 639–642.

Dunkelberg, W.C. and Day, G.S. (1973). Nonresponse Bias and Callbacks in Sample Surveys. Journal of Marketing Research, 10, 160–168.

Elliott, M.R., Little, R.J.A., and Lewitzky, S. (2000). Subsampling Callbacks to Improve Survey Efficiency. Journal of the American Statistical Association, 95, 730–738.

Fitzgerald, R. and Fuller, L. (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, 3–32.

Groves, R.M. and Couper, M.P. (1998). Nonresponse in Household Interview Surveys. New York: John Wiley & Sons, Inc.

Groves, R., Benson, G., and Mosher, W. (2005). Plan and Operation of Cycle 6 of the National Survey of Family Growth, (1) 42, Hyattsville, MD: National Center for Health Statistics.

Groves, R.M., Brick, J.M., Couper, M., Kalsbeek, W., Harris-Kojetin, B., Kreuter, F., Pennell, B.-E., Raghunathan, T.E., Schouten, B., Smith, T.W., Tourangeau, R., Bowers, A., Jans, M., Kennedy, C., Levenstein, R., Olson, K., Peytcheva, E., Ziniel, S., and Wagner, J. (2008). Issues Facing the Field: Alternative Practical Measures of Representativeness of Survey Respondent Pools. Survey Practice. Available at http://surveypractice.files.wordpress.com/2008/11/survey-practice-october-2008.pdf

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

Groves, R.M. and Peytcheva, E. (2008). The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis. Public Opinion Quarterly, 72, 167–189.

Hoagland, R.J., Warde, W.D., and Payton, M.E. (1988). Investigation of the Optimum Time to Conduct Telephone Surveys. Proceedings of the American Statistical Association, Survey Research Methods Section, 755–760.

Kalton, G. (1981). Compensating for Missing Survey Data. Ann Arbor (MI): Survey Research Center, The University of Michigan.

Kreuter, F. and Kohler, U. (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, 203–226.

Lessler, J.T. and Kalsbeek, W.D. (1992). Nonsampling Error in Surveys. New York: John Wiley & Sons, Inc.

Lin, I.-F. and Schaeffer, N.C. (1995). Using Survey Participants to Estimate the Impact of Nonparticipation. Public Opinion Quarterly, 59, 236–258.

Little, R.J.A. (1986). Survey Nonresponse Adjustments for Estimates of Means. International Statistical Review, 54, 139–157.

Little, R.J.A. and Rubin, D.B. (2002). Statistical Analysis with Missing Data. New York: John Wiley & Sons, Inc.

Little, R.J.A. and Vartivarian, S. (2005). Does Weighting for Nonresponse Increase the Variance of Survey Means? Survey Methodology, 31, 161–168.

McCarty, C., House, M., Harman, J., and Richards, S. (2006). Effort in Phone Survey Response Rates: The Effects of Vendor and Client-Controlled Factors. Field Methods, 18, 172–188.

Olson, K. (2007). An Investigation of the Nonresponse/Measurement Error Nexus, Ph.D.Thesis, University of Michigan, Survey Methodology.

Särndal, C.-E. (2011). The 2010 Morris Hansen Lecture Dealing with Survey Nonresponse in Data Collection, in Estimation. Journal of Official Statistics, 27, 1–21.

Singer, E. (2002). The Use of Incentives to Reduce Nonresponse in Household Surveys. Survey Nonresponse, R.M. Groves, D.A. Dillman, J.L. Eltinge, and R.J.A. Little (eds). New York: John Wiley & Sons, Inc, 163–177.

Singer, J.D. and Willett, J.B. (2003). Applied Longitudinal Data Analysis. New York: Oxford University Press.

Tångdahl, S. (2004). Nonresponse Bias for Some Common Estimators and Its Change over Time in the Data Collection Process. Technical Report, Örebro University.

Weeks, M.F., Kulka, R.A., and Pierson, S.A. (1987). Optimal Call Scheduling for a Telephone Survey. Public Opinion Quarterly, 51, 540–549.

Wood, A.M., White, I.R., and Hotopf, M. (2006). Using Number of Failed Contact Attempts to Adjust for Non-Ignorable Non-Response. Journal of the Royal Statistical Society, Series A, 169, 525–542.

Received June 2009
Revised August 2011