

Nonresponse and Measurement Error in Mixed-Mode Designs

by

Rachel M. Levenstein

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Survey Methodology)
in the University of Michigan
2010

Doctoral Committee:

Professor James M. Lepkowski, Chair
Research Professor Mick P. Couper
Professor Don A. Dillman, Washington State University
Assistant Professor Frauke Kreuter, University of Maryland

Acknowledgements

This dissertation would not have been completed without the help of many people. I would first like to thank my committee, particularly my chair, Jim Lepkowski. Jim has been a wonderful (and extremely patient) mentor. Many thanks to the other members of my committee—Mick Couper, Frauke Kreuter, and Don Dillman.

I am very thankful for the other faculty in the Program in Survey Methodology who did not end up on the committee but played a strong role my graduate career. Bob Groves had a great deal of input on the design of this study. Fred Conrad has been a terrific mentor and colleague and helped get me through the difficult first couple of years in the program. I would also like to thank Kirsten Alcsér and Beth-Ellen Pennell for their mentorship and support over the past few years.

This work was supported by the U.S. Census Dissertation Fellowship. Jennifer Barber, Bill Axinn, and Mick Couper provided access to the Relationship Dynamics and Social Life data; Heather Gatny helped me work with the data. Mark Trappmann provided access to the Panel Arbeitsmacht und Soziale Sicherung data. He and his staff provided support for the data and also helped me feel quite at home during my month in Nuremberg. Stefan Bender and Jörg Heining provided access to the Integrated Employment Biographies data.

The best part about graduate school has been my colleagues at SMP. Many graduate students provided a great deal of comments and support during and before the dissertation process. I'd like to thank them for many dissertation breakfasts and happy hours.

Many other people—too many to name—provided a lot of moral support over the past few years. I would like to thank Takashi Kushida, Akira Kushida, and all of my instructors and friends at AYANA for reminding me that even if things are difficult, showing up and working very hard can be progress.

Finally, I would like to thank my family for their support. I am especially thankful for my fiancé Mike. He was a constant source of support and had an almost psychic ability to know when he should come to my office bearing chocolate. Thanks for following me to the Arctic Circle.

Table of Contents

Acknowledgements.....	ii
List of Tables	vii
List of Figures	x
List of Appendices	xi
Abstract	xii
1. Introduction.....	1
2. Literature Review.....	4
2.1. Introduction.....	4
2.2. Nonresponse and nonresponse bias	5
2.2.1. An overview.....	5
2.2.2. Nonresponse in sequential mixed-mode designs	8
2.2.3. Other outcomes: Lateness and mode in a sequential design.....	10
2.2.4. Wave nonresponse in panel surveys	11
2.2.5. Sequential designs in panel surveys.....	13
2.3. Measurement error bias.....	16
2.3.1. Introduction.....	16
2.3.2. Measurement bias and mode.....	17
2.4. Nonresponse and measurement bias	19
2.5. Research questions in this dissertation	20
2.5.1. Response propensity, lateness, and mode propensity	20
2.5.2. Nonresponse bias in sequential designs.....	28
2.5.3. Nonresponse and measurement bias	30
2.6. Summary	35
3. Data and Methods	36
3.1. Relationship Dynamics and Social Life.....	36
3.1.1. Sample design	36
3.1.2. Data collection methods and journal period structure	37
3.1.3. Unit nonresponse	42
3.1.4. Outcomes and predictors.....	44
3.1.5. Data processing and model formulation	49
3.1.6. Missing data compensation: Sequential regression multiple imputation.....	49
3.1.7. Analytic models in Chapters 4 and 5	53
3.2. Panel Arbeitsmarkt und Soziale Sicherung.....	66
3.2.1. Sample design	67
3.2.2. Data collection methods.....	68
3.2.3. Outcomes and predictors.....	72

3.2.4. Linking procedures	79
3.2.5. Eligible panel respondents	82
3.2.6. Item missing data	82
3.2.7. Imputation procedures	83
3.2.8. Analytic models	83
4. Attrition Propensity and Mode Propensity in a Multimode Panel Survey.....	92
4.1. Introduction.....	92
4.2. Results.....	94
4.2.1. Propensity to respond.....	94
4.2.2. Lateness of response	98
4.2.3. Propensity to use a mode	102
4.3. Discussion	105
4.3.1. Response rates and response propensity	105
4.3.2. Lateness of response	108
4.3.3. Mode of response.....	109
4.3.4. Limitations	110
4.3.5. Conclusions.....	113
5. Wave nonresponse bias in sequential mixed-mode designs	114
5.1. Introduction.....	114
5.2. Results.....	116
5.2.1. Estimation of overall wave nonresponse bias.....	116
5.2.2. Differences in wave nonresponse bias between modes	117
5.2.3 Changes in δ_j across journal periods	122
5.3. Discussion.....	124
5.3.1. Overall wave nonresponse bias.....	124
5.3.2. Differences in wave nonresponse between modes.....	126
5.3.3. Changes in δ_j across journal periods	126
5.3.4. Limitations	126
5.3.5. Conclusions.....	127
6. Nonresponse bias and measurement error	129
6.1. Introduction.....	129
6.2 Results.....	130
6.2.1. Descriptive measures	130
6.2.2. Measurement bias	135
6.2.3. Nonresponse and measurement bias	136
6.2.4. Consent to link survey and record data.....	139
6.3. Discussion	139
6.3.1. Nonresponse bias	141
6.3.2. Measurement bias	143
6.3.3. Nonresponse bias and measurement bias.....	145
6.3.4. Summary and conclusions	146
7. Conclusions.....	147

7.1. Introduction.....	147
7.2. Research findings.....	147
7.2.1. Response propensity, latency, and mode	147
7.2.2. Nonresponse bias	148
7.2.3. Nonresponse bias and measurement error	149
7.2.4. Summary	151
7.3. Future research.....	151
7.3.1. Response behaviors in mixed-mode surveys	151
7.3.2. Wave nonresponse bias.....	152
7.3.2. Wave nonresponse bias and measurement error	152
References.....	193

List of Tables

Table 1. Contact protocol in RDSL.	39
Table 2 . Association between the lateness of response and mode of response in RDSL.	45
Table 3. Source, frequency, and item missing data for privacy concerns, topic salience, and education status variables from RDSL.	48
Table 4. Fraction of missing information (FMI) for target variables in RDSL.	52
Table 5. Association between the consent to link PASS with IEB data in Wave 1 and Wave 2.	72
Table 6. Wave 2 IEB variables used for nonresponse bias and measurement bias analyses.	78
Table 7. Wave 2 PASS survey data: descriptives.	79
Table 8. Response propensity models for participating in CATI, CAPI, and in either mode.	90
Table 9. Expected directions of the effects of predictors of response under three mechanisms.	93
Table 10. Odds ratios and standard errors for response propensity models.	99
Table 11. Odds ratios and standard errors of lateness models (3.4) and (3.5).	101
Table 12. Odds ratios and relative risk ratios for models predicting the likelihood of mode use: Effects of journal period and mean days in previous journal periods.	104

Table 13. Multinomial logistic regression predicting the likelihood of inbound or outbound CATI vs web response.....	105
Table 14. Findings from bivariate response propensity models in Chapter 4.	117
Table 15. Mean overall wave nonresponse bias estimates in RDSL: Baseline variables.	119
Table 16. Mean overall wave nonresponse bias estimates in RDSL: Variables in all journals.....	119
Table 17. Tests of overall wave nonresponse bias significance: Odds ratios, relative risk ratios, and standard errors in multinomial and logistic models. θ	120
Table 18. Change in overall nonresponse bias across journal periods: Odds ratios, relative risk ratios, and standard errors in multinomial and logistic regression model contrasts (θ_A)	120
Table 19. Mean change in difference in bias between Treatments 1 and 2: multinomial and logistic regression model contrasts (θ_D) and standard errors.	123
Table 20. Tests for changes in the effect of the sequential design on wave nonresponse bias across journal periods (change in δ_j): Contrasts and standard errors.	123
Table 21. Descriptive statistics for 13 IEB variables in the nonresponse bias analysis.	131
Table 22. Signed and relative nonresponse bias before and after mode switch, and the difference in bias between modes.	132
Table 23. Bivariate nonresponse bias models.....	134
Table 24. Regression models to determine if $\delta = \mathbf{0}$	135
Table 25. Measurement bias within and across modes.....	137
Table 26. Signed measurement bias models with predictors of participation in a mode.	138

Table 27. Relationship between signed measurement bias and propensity scores:

Estimates of β_1 140

Table 28. Consent to link IEB and PASS, Waves 1 and 2..... 141

List of Figures

Figure 1. Five scenarios: Patterns of nonresponse between treatment groups.	16
Figure 2. Journal periods and response for four hypothetical respondents.....	41
Figure 3. Maximum number of journal periods by replicate.....	43
Figure 4. Response rates over journal periods.....	44
Figure 5. Simulated experiment comparing a web single mode design to a web→CATI sequential design using one web→CATI mixed-mode survey.....	65
Figure 6. PASS flow chart with sample sizes.....	70
Figure 7. Data availability summary for IEB, PASS, and paradata.....	73
Figure 8. Simulated experimental design.....	85
Figure 9. Distribution of late responses.....	98
Figure 10. Proportion of completes in each mode.....	103
Figure 11. Distribution of mode responses between panel respondents.....	103
Figure 12. Nonresponse bias across journal periods: Coital contraception.....	121

List of Appendices

Appendix A. Supplemental tables and derivations.	154
Appendix B. Question wording in Relationship Dynamics and Social Life	175
Appendix C. Nonresponse bias in Panel Arbeitsmarkt und Soziale Sicherung.....	180

Abstract

Nonresponse and Measurement Error in Mixed-Mode Designs

by

Rachel M. Levenstein

Chair: James M. Lepkowski

Because of declining response rates in surveys, survey researchers are increasingly using data collection methods that can increase response rates while containing costs. One such method is the use of sequential mixed-mode designs, in which the mode of administration is switched to a second mode within the survey field period. This dissertation examines the use of sequential designs in two panel surveys: The Relationship Dynamics and Social Life survey (RDSL) and the Panel Arbeitsmarkt und Soziale Sicherung (PASS) in terms of nonresponse behaviors, nonresponse bias, and measurement bias.

We find that past behavior—the likelihood of response, the likelihood of responding late enough to warrant a mode switch, and the likelihood of responding in a particular mode—is strongly related to current behaviors in RDSL. The likelihood of

response is related to the frequency and timing of response in the past; the likelihood of responding late in past waves is positively related to the likelihood of responding late at a current wave; and the likelihood of using a mode is positively related to the number of times the individual responded via that mode in the past. Nonresponse in the panel survey is not a significant problem after responses are combined from both modes. Finally, the sequential design either does not affect or it reduces the nonresponse bias.

In Wave 2 of PASS, nonresponse bias, measurement bias, and the relationship between response propensity and measurement bias is examined. As above, the sequential design either does not affect or it reduces the bias. Although measurement bias is a significant problem, it is unaffected by the mode of response. No relationship between response propensity and measurement bias is found within or between modes.

Chapter 1

Introduction

Survey research is relied upon to give an accurate description of the population. Because the output generated by surveys can have consequences for our society, providing accurate estimates is crucial. Survey estimates are rarely, if ever, error-free, and efforts must be taken to reduce survey error and improve data quality where possible.

In the past, survey researchers have used a variety of tools to maintain costs while either improving data quality or minimizing the risk of poor data quality. In the 1970s and 1980s, a push was made to move to telephone interviewing because it was much less expensive than the more common face-to-face mode (Hochstim, 1967), and many elements of survey error and data quality were often equivalent to face-to-face surveys (Groves & Kahn, 1979). It was clear that changing the mode—the method of administration—could be a powerful tool for survey researchers.

Today, these methods are continually evolving. The use of technology has greatly impacted survey research. For example, web surveys have permitted survey researchers to interview an incredibly large number of people for very little financial cost (Couper, 2000). Computer assisted methods, such as computer-assisted telephone interviewing (CATI) and computer-assisted personal interviewing (CAPI) have generated paradata (Couper, 1998) that have allowed researchers to evaluate interviewer behaviors (Couper

et al., 1997), reasons for nonresponse (Groves & Couper, 1998), and how well efforts to reduce nonresponse are working (Groves & Heeringa, 2006).

Another invaluable tool to survey researchers is the use of multiple modes. Different kinds of mixed-mode designs can reduce coverage, nonresponse, and measurement error (de Leeuw, 2005; Dillman *et al.*, 2008; Dillman *et al.*, 2009a). This dissertation will examine one particular kind of mixed-mode design that is generally employed to increase response rates (de Leeuw, 2005): the sequential mixed-mode design.

This dissertation investigates nonresponse rates, nonresponse bias, and measurement bias in two surveys that implement sequential mixed-mode designs. One survey, the Relationship Dynamics and Social Life (RDSL) survey, is a panel survey with many waves of data collection. The other, the Panel Arbeitsmacht und Soziale Sicherung (PASS), is also a panel survey, but this dissertation only examines one wave of the survey.

Panel surveys are valuable to researchers because they afford us the ability to study change within an individual over time. RDSL was developed, in part, to capture changes in attitudes and behaviors that can lead to pregnancy in young women. PASS aims to estimate changes in attitudes and behaviors that can lead to program participation and exits from program participation.

This dissertation investigates the use of sequential mixed-mode designs in both of these surveys. Specifically, the following questions are addressed:

1. Do sequential designs affect the propensity to respond, the propensity to respond late, and/or the propensity to use a particular mode? Do any of these effects change across waves in a panel survey?
2. Do sequential designs affect nonresponse bias of various means? Does the effect of the mode switch on nonresponse bias of these estimates change over time in a panel survey?
3. Is there a relationship between nonresponse bias and measurement error in a single wave of a panel survey?

Chapter 2 presents a literature review on nonresponse bias and measurement error in mixed-mode designs and panel surveys. In Chapter 3, the methods for the rest of the dissertation are detailed. Chapters 4, 5, and 6 address research questions 1, 2, and 3, respectively. Chapter 7 summarizes the findings and outline directions for future research.

Chapter 2

Literature Review

2.1. Introduction

In cross-sectional surveys, nonresponse rates have generally been increasing over the past few decades (Atrostic *et al.*, 2001; de Leeuw & de Heer, 2002). These trends may be present in panel surveys, as well. The Panel Study of Income Dynamics (PSID) lost roughly a quarter of its original cohort to attrition by 1998 (Fitzgerald *et al.*, 1998), while the German Socio-Economic Panel lost over 30% of the original sample (Spiess & Pannenberg, 2003).

Given response to an initial or baseline interview and agreement to be contacted in future waves, the rate of attrition may be increasing—much like cross sectional surveys. Atrostic *et al.* (2001) found that the rate of attrition in later cohorts was greater than the rate of attrition in earlier cohorts in several rotating panel surveys.

This is of concern to survey researchers for a number of reasons. First, when response rates are low, the analytic sample size decreases. More importantly, when the response rate is below 100%, there is a potential for nonresponse bias. Although low response rates do not necessarily mean that nonresponse bias is occurring, survey researchers often attempt to maximize response rates (Groves, 2006; Groves & Peytcheva, 2008). As Groves (2006) notes, attempts to maximize response rates can actually

increase survey error. It is therefore important to understand the consequences of design changes that can increase response rates.

Survey researchers have many tools at their disposal to attempt to increase response rates, such as incentives (Singer, 2002), repeated call attempts (Groves & Couper, 1998), call scheduling techniques (Greenberg & Stokes, 1990; Kalsbeek *et al.*, 1994), and interviewer training (Groves & McGonagle, 2001). Another way that survey researchers can increase response rates is to switch the mode part-way through the field period. These designs—called sequential mixed-mode designs (de Leeuw, 2005; Hochstim, 1967)—capitalize on what we know about response rate and cost differences across modes.

This dissertation focuses on nonresponse bias and measurement bias as the sources of mode differences in survey means¹. This chapter will outline possible mechanisms for nonresponse bias and measurement bias in sequential mixed-mode designs in panel surveys.

2.2. Nonresponse and nonresponse bias

This section will introduce the concept of nonresponse bias. Why do we care about response rates? When would we expect specific kinds of estimates to be affected by nonresponse? And, how might design features implemented to reduce nonresponse rates influence nonresponse bias?

2.2.1. An overview

¹ Coverage bias can also drive mode effects (Groves *et al.*, 2009), although it is not investigated in this dissertation.

Nonresponse bias can be a serious concern in surveys. Probability sampling allows us to make inference from a sample to the population by assuming a 100% response rate and no coverage error (Groves & Couper, 1998), a clear impossibility in household surveys for logistical and ethical reasons.

An estimate of nonresponse bias of a sample mean \bar{y} is the difference between the mean of the respondents (\bar{y}_r) and the full sample mean (\bar{y}_n):

$$(2.1) \quad NRbias(\bar{y}) = \bar{y}_r - \bar{y}_n = \frac{m}{n}(\bar{y}_r - \bar{y}_m)$$

where $\frac{m}{n}$ = the nonresponse rate and \bar{y}_m = the mean of the nonrespondents (Groves & Couper, 1998). The bias is therefore not just a function of the number of respondents and nonrespondents; nonresponse bias exists when the respondents are different from the nonrespondents on that variable.

Model (2.1) implicitly assumes that there are respondents and nonrespondents, and we do not know which group a sample person belongs to until the survey request is made (Lessler & Kalsbeek, 1992). However, response propensity—the estimated probability of responding to a given survey request—is likely to be more of a stochastic phenomenon (Bethlehem, 2002; Groves et al., 2009). Bethlehem represents nonresponse bias as the relationship between the stochastic propensity to respond (p) and the survey variable (Bethlehem, 2002):

$$(2.2) \quad NRbias(\bar{y}) = \frac{Cov(y, p)}{\bar{p}}$$

The propensity to respond is unknown because we can never know nor measure all of the contributors to response (Lessler & Kalsbeek, 1992). We can, however,

estimate p using logistic regression models and auxiliary data that exist for respondents and nonrespondents. The denominator of (2.2) is the mean response propensity, which is often estimated as the response rate.

Nonresponse bias can occur when p is correlated with y directly or indirectly. For example, estimates of volunteerism tend to be positively biased because volunteers are much more likely to participate in surveys than individuals who do not volunteer (Abraham *et al.*, 2009). Groves (2006) calls this situation—in which the variable of interest causes the propensity to respond—the Survey Variable Cause Model of nonresponse.

Nonresponse bias can also occur when there is a third variable—let us call this Z —that is correlated with y and p (Groves, 2006). If the survey variable of interest is something related to volunteerism—say, level of education (McPherson & Rotolo, 1996; Okun & Eisenberg, 1992; Wilson, 2000)—then the nonresponse bias of level of education (y) is likely to be significant because volunteers are more likely to participate, and because education (Z) is related to volunteerism.

Nonresponse can be due to failures in contact or cooperation². In some cases, we will not know if the nonresponse is due to noncontact or refusal. This knowledge is often based on the mode of contact attempt. For example, a sample person may not complete a web survey because she has not received the survey request, or she refused to participate. In most cases, we will not be able to tell whether contact or cooperation was problematic. In a face-to-face survey, however, this should be clear; an interviewer knows if contact

² Other sources of nonresponse include not speaking the language or having some physical or mental disability that prevents response. Because these reasons for nonresponse are generally applicable to a small proportion of sample cases, they are not investigated here.

has been made.³ What we know about contact and cooperation in each contact attempt is often dependent on the mode of contact attempt.

2.2.2. Nonresponse in sequential mixed-mode designs

Sequential designs are used, in part, because response rates are higher in some modes than others. For example, face-to-face surveys tend to have higher response rates than telephone (Groves et al., 2009; van Campen *et al.*, 1998), and telephone surveys tend to have higher response rates than web (Lozar Manfreda *et al.*, 2008). However, the increase in response rates comes at a cost. Telephone surveys tend to be much more expensive than web surveys (e.g., Couper, 2000, 2005), and face-to-face surveys are often much more expensive than telephone surveys (Groves et al., 2009). Because these higher-performing modes are much more expensive than the lower-performing modes, sequential designs only utilize these modes for sample cases who have not responded to the lower-performing mode. In a sequential mixed-mode design survey that uses telephone as a second mode⁴, only web nonrespondents are attempted via the telephone.

Sequential mixed-mode designs have been useful tools to help increase response rates while containing costs. The American Community Survey (ACS) uses a mail→telephone→face-to-face⁵ design. A mail questionnaire is sent to all sample cases.

³ There are situations in which contact may have been made, but the sample person refused. For example, the interviewer may have seen people inside a house, but no one responded to the doorbell. This might be noncontact—if the sample person did not hear the doorbell ring. It could also have been a refusal, in which the sample person refused by not answering the door.

⁴ An arrow will be used to denote the moving from a primary mode to a secondary mode. For example, a survey that uses web first and telephone to attempt web nonrespondents will be referred to as a web→telephone sequential design.

⁵ The sampling frame of the ACS is a list of addresses. Telephone numbers can sometimes—but not always—be linked to the addresses using commercial lists. When a household does not have a telephone in such a list and they fail to respond to mail, a personal visit may be required. That is, for households with an available telephone number, the sequential design of the ACS is mail→telephone→face-to-face, but for households without a telephone number, the sequence is simply mail→face-to-face. (U.S. Census Bureau, 2009).

A subsample of nonrespondents for whom a telephone number can be located is selected and attempted via telephone. Finally, a subsample of nonrespondents to mail and telephone are attempted in person. Response rates jump from about 55% in mail to about 60% in telephone to about 96% in face-to-face (Griffin, 2009).

These increases in cumulative response rates across modes in a sequential design can be found in other types of surveys. In a market research survey, Dillman et al. (2008) found that using a telephone→mail sequential design increased the response rate from 43% after the telephone phase to 80% at the end of the field period. Other mode combinations were less successful; mail→telephone resulted in only a seven percentage-point increase in the response rate. Fowler et al. (2002) observed a 20 percentage-point increase in a mail→telephone survey of health plan members. And Beebe et al.(2007) found that both web→mail and mail→web sequential designs both resulted in a 15-point increase in the response rates from the first mode to the second in a survey of physicians.

The effect of sequential designs on nonresponse bias is not clear, even if response rates tend to be higher than designs with a single mode but multiple call attempts. Are these designs bringing in respondents who are substantively different than respondents to the initial mode as each sequence increases the response rates? And, if so, why is this occurring?

Let us first consider a simpler design, in which one randomly selected group of sample cases are attempted via telephone, and another is attempted via face-to-face. Let us also assume that, as is relatively more common in Europe than the United States (Steeh & Piekarski, 2008), the cell phone numbers are included in samples in telephone surveys. We can expect differences between modes in nonresponse bias when the mode

affects the relationship between y and p . This is similar to the Common Cause model of nonresponse with Z =the mode of contact attempt; here, however, the mode is not directly related to y .

Consider a telephone → face-to-face survey in which telephone can include cell phones. The relationship between being employed and the likelihood of contact should be stronger in the face-to-face follow-up mode than the telephone mode because unemployed people tend to be at home more often (Groves & Couper, 1998), and being at-home is likely to be more important to gain contact in face-to-face than it is in telephone. One cannot complete a face-to-face survey when the sample person is not at home, but one can complete a telephone survey via cell phone. When telephone becomes a more “mobile” mode—when sample persons can be contacted via cell phone—being at home is less important as a predictor of contact than in a landline telephone survey.

It is not clear that a sequential mixed-mode design would behave differently from a concurrent mixed-mode design. Those who receive the second mode will generally have lower response propensities than those who do not receive the second mode. In the above example, however, at-home patterns will still be more important in face-to-face than telephone in a telephone → face-to-face study.

2.2.3. Other outcomes: Lateness and mode in a sequential design

But response propensity is not the whole story in a sequential mixed mode design. At some defined point in the field period of a sequential design, the mode is switched. Individuals who respond after that point are considered in this dissertation to be “late”. Mode and lateness of response are not always identical; many surveys like RDSL or the ACS permit the use of earlier modes after the switch and later modes before the switch.

We cannot assume that they have the same causes. This dissertation will investigate lateness and mode of response separately in RDSL.

2.2.4. Wave nonresponse in panel surveys

Nonresponse in panel surveys differs from cross-sectional surveys on a number of important dimensions. This section summarizes the literature on wave nonresponse in panel surveys.

In panel surveys, the decision to respond does not occur once, as it does in cross-sectional surveys. Individuals can drop out temporarily or permanently. Temporary wave nonresponse means that an eligible sample person (hereafter referred to as a “panel respondent”) drops out of one wave but responds again in later waves. Permanent wave nonresponse, often referred to as attrition, means that the panel respondent drops out but never returns to the sample. Some studies permit temporary wave nonresponse, but others only attempt panel respondents who had participated in all prior waves. In these studies, the denominator of the response rate at each wave is sometimes dependent on response to a previous wave. For example, early waves of the PSID only attempt sample cases if they have participated in the previous wave (M. S. Hill, 1992). In this case, the propensity to participate in Wave t conditions on response in Wave $t-1$. Thus, we see that response rates—conditional on eligibility in each wave—tend to *increase* across waves of a panel survey (Kalton *et al.*, 1998; Zabel, 1998).

Wave response rates in panel surveys tend to be higher than response rates in cross-sectional surveys; however, panel surveys have problems in location as well as

contact and cooperation⁶ (Lepkowski & Couper, 2002). Location is necessary because sample persons must be contacted across a longer period of time; contact information must be kept up to date in order to facilitate contact. Lepkowski and Couper (2002) hypothesized that the time between waves can be inversely related to the likelihood of location. Similarly, nonresponse in a previous wave might affect the likelihood of location; having had a nonresponding journal at Wave $t-1$ would likely lead to a decrease in the likelihood of location at Wave t , as contact information becomes more and more out of date.

Provided that the contact information of the individual has been correctly updated (i.e., that location of the panel respondent is successful), the likelihood of contact tends to be quite high (Lepkowski & Couper, 2002). As discussed above, however, we only know if contact has been made if the survey is interviewer-administered. In self-administered modes like web, location, contact, and cooperation are often indistinguishable; for example, nonresponse in a web survey could be due to a problem in location (i.e., a bad email address) or a problem in contact (i.e., the message was caught by a spam filter) or a problem in cooperation (the individual refused to participate). We typically cannot distinguish these causes.

In interviewer-administered surveys, interviewers can often use information from previous waves to inform contact attempts. For example, interviewers can learn about at-home patterns and modify contact attempts based on those patterns. It seems logical that the more waves of information are available for the interviewer to utilize, the more likely contact is achieved (Laurie *et al.*, 1999).

⁶ When the contact information in a sampling frame is outdated, location may be a problem in cross-sectional surveys. However, location is more problematic in a panel survey because panel respondents must be tracked over a longer period of time.

Experiences and outcomes in previous waves may be related to the likelihood of cooperation. Interviewer ratings of cooperation in previous waves are often positively related to cooperation (and contact) at a given wave (Lepkowski & Couper, 2002; N. Watson & Wooden, 2009). Item nonresponse in previous waves is also positively related to unit nonresponse in later waves (Laurie et al., 1999; Loosveldt *et al.*, 2002; Zabel, 1998). Panel respondents who were rated as interested in a Wave 1 interview were more likely to cooperate in Wave 2 (Lepkowski & Couper, 2002); similarly, panel respondents' rated enjoyment of the interview at Wave 1 is positively related to cooperation at Wave 2 (D. H. Hill & Willis, 2001; Lepkowski & Couper, 2002).

Because of the repeated measurements, the burden of each additional wave can decrease wave response rates (Fitzgerald et al., 1998; Zabel, 1998). This burden is not necessarily a linear effect, however; work on the PSID demonstrates that the response rates tend to drop early in the panel, then level off (Fitzgerald et al., 1998).

2.2.5. Sequential designs in panel surveys

When we have a sequential mixed-mode design that is employed in every wave of a panel survey, the picture becomes even more complex. First, as noted in the previous section, panel respondents tend to become more likely to participate in later waves than they were in earlier waves. Lateness of response and mode of response may be affected by behavior at earlier waves.

The nonresponse bias may be affected by the mode of response as well as time in sample. To date, I have found no literature has examined these moving parts of nonresponse bias in mixed-mode panel surveys simultaneously. This dissertation is a first

step in understanding the dynamics of response behavior and survey error in sequential mixed-mode designs in panel surveys.

Let us consider an example of this kind of design. Consider a panel survey that has many waves, and, at each wave, data collection begins in web and ends in telephone. In order to examine the effect of the sequential design on the nonresponse bias, we estimate the nonresponse bias of a variable of interest for the full web→telephone sample as well as the nonresponse bias if we had only permitted response in web. Let us call this difference in nonresponse biases δ_j .

Figure 1 displays the nonresponse bias estimates in this panel survey across waves under four scenarios, before the mode switch (i.e., only web responses contribute to the respondent mean) and after the mode switch (i.e., the full sample). δ_j is the difference between the lines at wave j and $\bar{\delta}$ = the average difference between the lines across all journal periods .

In Scenario 1, the nonresponse bias is greater if we stopped at web than it would be for the full sequential design—but the bias does not increase across waves. Here, $\bar{\delta} > 0$; on average, the nonresponse bias is reduced by including the telephone respondents in the respondent mean. However, the change in δ_j across waves is zero.

The change in δ_j across waves could be zero even if the nonresponse bias increases across waves. In Scenario 2, the change in δ_j across waves is zero, even though $\bar{\delta} \neq 0$ and the nonresponse bias increases across waves.

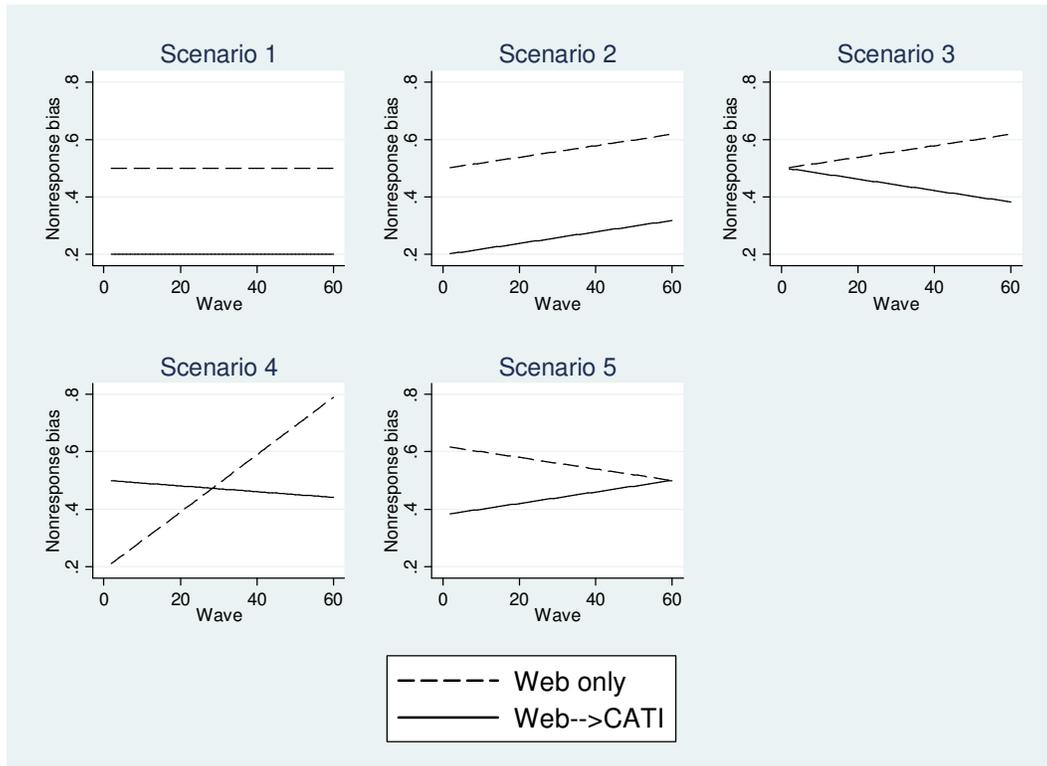
In the third scenario, the nonresponse bias at the first wave is the same regardless of whether we continued data collection with telephone ($\delta_1 = 0$). But as the number of

waves increase, the nonresponse bias for web only increases while the nonresponse bias for the full sequential design decreases. That is, δ_j increases in a linear fashion across waves. In Scenario 4, δ_j decreases until $\delta_{60} = -0.4$. This is still a linear trend. However, one could imagine some nonlinear trends in δ_j . This dissertation is a first look at δ_j and changes in δ_j . In such a first examination, it is useful to assume linearity, as indicated in Figure 1, as an initial approximation.

The fourth and fifth scenarios are situations in which δ_j decreases over waves. In Scenario 4, δ_j becomes less than zero after about wave 30. And Scenario 5 is a case in which δ_j tends towards zero. In both of these cases, the change in mean nonresponse bias may be equal to zero (as in Scenario 5) or it may increase linearly (Scenario 4). The nonresponse bias and the difference in nonresponse bias between the first mode and the full sequential design are independent estimates.

These scenarios are merely hypothetical situations. Chapter 5 examines the magnitude of the wave nonresponse bias, the change in wave nonresponse bias across waves of a panel survey, the effect of the sequential design on the wave nonresponse bias, and the change in the effect of the sequential design on the wave nonresponse bias. We can use the above scenarios as a framework for understanding these complex outcomes.

Figure 1. Five scenarios: Patterns of nonresponse between treatment groups.



2.3. Measurement error bias

2.3.1. Introduction

Like nonresponse bias, measurement bias can differ across modes in a sequential design as well as a concurrent mixed-mode design. Measurement bias is defined here as the difference between an estimated mean for respondents computed using reported values and the estimate computed using true values:

$$(2.3) \quad \epsilon_s = \frac{y_{riR} - y_{riT}}{r} = \bar{y}_{rR} - \bar{y}_{rT}$$

where y_{riR} = the reported value of y for respondent i , y_{riT} = the true value of y for respondent i , and r = the number of respondents.

This source of survey error can arise from an interviewer, the respondent, the questionnaire, or the mode (Groves, 1989). This section details why we might expect differences in measurement biases among modes.

2.3.2. Measurement bias and mode

We know from the literature on concurrent mixed-mode designs that differences in measurement bias between modes can be due to a number of reasons. First, reports of sensitive behaviors can differ between interviewer-administered modes, compared to self-administered modes. Tourangeau and Smith (1996) found that social desirability bias—the tendency to present oneself favorably—decreased in self-administered modes, compared to interviewer-administered modes. This effect of self-administration on the bias of sensitive questions has been replicated many times for many different kinds of estimates, such as drug use (Aquilino & LoSciuto, 1990; Beebe *et al.*, 2005; Tourangeau & Smith, 1996), sexual behaviors (Catania *et al.*, 1990; Gribble *et al.*, 1999; Tourangeau *et al.*, 1997; Tourangeau & Smith, 1996), and abortion (E. F. Jones & Forrest, 1992).

These effects can be found for all modes that are self-administered. Beebe *et al.* (2005) found this effect in mail surveys. Tourangeau and Smith (1996) compared estimates of sexual behaviors and drug use in computer-assisted self interviewing (CASI), audio computer-assisted self interviewing (A-CASI), and CAPI. They found that both ACASI and CASI resulted in more socially undesirable reports, compared to CAPI. Reports of sensitive behaviors tend to increase when respondents use web, compared to IVR or telephone (Kreuter *et al.*, 2009b).

Tourangeau *et al.* (2000) argue that sensitive questions have three elements that make them sensitive: they are subject to social desirability bias, the question itself is

intrusive, and the respondent would not want her responses disclosed to a third party. When an interviewer is present, the desire to present oneself in a positive light tends to be greater than when an interviewer is not present, and the salience of the risk of disclosure may be higher with an interviewer present.

But the presence of the interviewer is not the whole story with respect to mode differences in the measurement bias of responses to sensitive questions. Conducting the survey over the phone with an interviewer tends to result in larger measurement bias than conducting the survey with an interviewer in person (Holbrook *et al.*, 2003; Jäckle *et al.*, 2010). This could be due to a failure in the telephone mode of “sincerity of purpose” (de Leeuw, 1992, 2008). That is, an interviewer who conducts the survey in person may have more legitimacy than an interviewer over the phone. This gap in legitimacy between telephone and personal interviewing may also contribute to lower response rates in telephone, compared to face-to-face surveys (Groves & Couper, 1998).

Responses may also differ across modes because of the channels of communication. When the respondent can read the questions (as in web or mail modes), then visual characteristics of the instrument, such as ratings scales and directions, can influence responses (Dillman *et al.*, 2009b).

Primacy effects and recency effects are both types of order effects in closed-ended questions that may be related to the channels of communication. Primacy effects occur when respondents tend to select the first response option from a list of options, and recency effects occur when respondents tend to select the last response option. Recency effects can be more prevalent when the respondent hears the questions rather than reads it, and primacy effects can be more prevalent when the respondent reads the question rather

than hears it (Krosnick & Alwin, 1987; Schwarz *et al.*, 1992)—but this is not always the case (Dillman *et al.*, 1995).

Krosnick and Alwin (1987) see these order effects as one manifestation of “satisficing”, whereby respondents take cognitive shortcuts. Holbrook *et al.* (2003) found that other behaviors that may reduce cognitive burden, such as nondifferentiation (providing the same response in a series of questions—also known as “straightlining”), reporting having no opinion, and acquiescence, were more prevalent in telephone surveys than face-to-face.

2.4. Nonresponse and measurement bias

In a sequential mixed-mode design, nonresponse and measurement bias are, by design, linked. Early respondents use a different mode than later respondents, and, as discussed in section 2.3.2, the measurement bias of the second mode may be different than the measurement bias of the first mode. This presents an interesting problem; if estimates between the first and second mode differ, we do not know if lateness of response (e.g., lower propensity) is related to measurement bias or if the mode is related to measurement bias.

To date, there is little evidence available on this relationship. Using data from the Wisconsin Divorce Study, Olson (2006) found that measurement bias was lower for estimates including low-propensity respondents for some estimates (for length of marriage and time since divorce), but not all (number of marriages). Kreuter *et al.* (2009) found that as the number of contact attempts increased in PASS, the measurement bias of receiving unemployment and occupation did not change substantially; however, the measurement bias of income did.

2.5. Research questions in this dissertation

This dissertation will investigate how response behavior, nonresponse bias, and measurement bias are affected by the sequential design in two different panel surveys. In the first survey, changes across observations in the panel in response behavior and nonresponse bias are addressed. In the second, a single wave of a panel survey will be investigated with respect to nonresponse bias and measurement bias. This section will identify specific mechanisms for specific variables in each survey.

2.5.1. Response propensity, lateness, and mode propensity

Chapter 4 will examine the propensity to respond, the propensity to respond late, and the propensity to respond in a particular mode in the Relationship Dynamics and Social Life Survey (RDSL). RDSL is a panel survey implementing a sequential mixed-mode design at each observation (hereafter referred to as a journal period). See Chapter 3 for details on the design of RDSL.

Response propensity. The likelihood of response can be driven by a multitude of different mechanisms. This dissertation examines two mechanisms of nonresponse: privacy concerns and topic salience. A strong correlate of response propensity, educational status, is also examined.

Privacy concerns can reduce the likelihood of response. In a mail survey conducted by the U.S. Census Bureau, Dillman et al. (1993) found that asking for a social security number decreased response rates. In surveys that are not self-administered, the survey introduction can sometimes trigger privacy concerns, leading to a reduction in the likelihood of response (Bates *et al.*, 2008; Conrad *et al.*, 2006; Couper *et al.*, 2008).

Topic salience is also related to the likelihood of response. A great deal of research has found that individuals who are interested in the survey topic are more likely to respond when the topic is emphasized in the survey introduction (Brick *et al.*, 2006; Goyder, 1987; Groves & Couper, 1998; Groves *et al.*, 2004; Groves *et al.*, 2000).

Educational status may be predictive of response propensity for a number of reasons. First, education level is related to positive attitudes towards surveys (Price & Stroud, 2005); individuals with more education tend to have more positive attitudes towards surveys. Although some research indicates a positive relationship between positive attitudes towards surveys and the likelihood of response (Rogelberg *et al.*, 2001), Goyder (1986) found a *negative* relationship between attitudes towards surveys and response propensity. However, Price and Stroud (2005), Rogelberg *et al.* (2001), and Goyder (1986) used responses to a survey to assess attitudes towards surveys. Any evidence of the effect of attitudes towards surveys on response propensity is difficult to interpret in the presence of nonresponse bias in the estimates of these attitudes (Goyder, 1986, 1987).

Those who have more education may have higher levels of civic duty (Keeter *et al.*, 2002). Civic duty can be defined as a feeling of responsibility towards one's community. Actions related to this feeling include voting (Campbell, 1954; Campbell *et al.*, 1960; Keeter *et al.*, 2002; Putnam, 2000), volunteering (Keeter *et al.*, 2002; Putnam, 2000; Zimmerman & Rappaport, 1988), membership in an organization (Keeter *et al.*, 2002; Putnam, 2000), and some kind of political involvement, such as contacting a representative or joining a political organization (Putnam, 2000).

There is evidence that these components of civic duty are related to response. Volunteers are more likely to participate in surveys than non-volunteers (Abraham et al., 2009); voters are more likely to participate than non-voters (Knack, 1992; Voogt & Saris, 2005); individuals who are involved in an organization are more likely to participate than those who are not involved (Couper *et al.*, 1998; Groves et al., 2000) . Because educational status is related to things like voting, volunteering, and membership in social and religious organizations (Abraham et al., 2009; Egerton, 2003; Keeter et al., 2002), panel respondents who are in school should be more likely to participate at any given journal period than those who are not in school.

All of these theories regarding the link between education and response propensity point to a positive relationship: as level of education increases, the likelihood of response increases.

This research examines indicators of privacy concerns, topic salience, and education and their effects on the likelihood of response. In addition, the effects of these indicators on response propensity may change across waves of a panel survey. The variables described below are plausible indicators of these latent constructs that were collected in RDSL.

Concerns about privacy might stem from *living with a parent* or living with a partner. A respondent might not respond to a survey request if privacy is not guaranteed because a parent or partner might see or hear her responses. That is, living with a parent should be negatively related to the likelihood of response if privacy concerns drive the decision to respond. While living with a parent or partner does not directly cause a lack of privacy, these living situations might indicate the potential for lack of privacy.

In a panel survey, panel respondents learn what questions will be asked of them. If they have something personal to report at a given wave, then the likelihood of response should decrease. One example of this may be *having a sexual partner*. The RDSL questionnaire asks panel respondents if they have had a sexual partner since the previous journal period. Some research has shown that questions about having a sexual partner are often seen as sensitive (Bradburn & Sudman, 1979; Tourangeau et al., 1997), which may reduce the likelihood of response for individuals who have a sexual partner (Tourangeau et al., 2000). However, it is likely that the “sincerity of purpose” (de Leeuw, 1992) has been established in earlier waves, and the effect of having a sexual partner diminishes across journal periods.

Similarly, use of contraception may trigger privacy concerns in early journal periods, but this effect quickly tapers off as confidence in the legitimacy of the survey organization is established. If privacy concerns are the reason for nonresponse, the use of contraception should be negatively related to response propensity. In RDSL, use of contraception is estimated separately for *noncoital* methods (e.g., birth control pills) and for *coital* methods (e.g., condoms). Coital methods must be used at the time of intercourse, while noncoital methods are used consistently over time. Although each of these methods might trigger privacy concerns in earlier journal periods, coital contraception should have a larger effect on response propensity because it necessitates a report of having a sexual partner. Noncoital contraception may be used for reasons other than birth control. Reporting the use of such contraception is still expected to trigger privacy in early journal periods, but not to the same extent as coital contraception.

The second mechanism thought to influence response propensity in RDSL is topic salience. For example, women who have had a change in a relationship status between the previous observation and the current attempt should have an increased likelihood of participation under this hypothesis because the topic of romantic relationships is particularly salient to them.

However, this effect of salience could act in opposition to privacy concerns. For example, women who have a sexual partner may have interest in the topic, or they could have privacy concerns. Or, both of these effects could cancel each other out. This may also be the case for use of contraception, pregnancy intentions, and pregnancy avoidance. Women who are using contraception may be more interested in the topic, increasing the likelihood of response. Or, women who are using contraception may have more privacy concerns than those who are not using contraception. Women who want to become pregnant may be interested in the topic more than women who want to avoid pregnancy; or they may have more privacy concerns. The effect of topic salience on response propensity is not expected to change across journal periods.

In addition to privacy concerns and topic salience, educational status may be related to response propensity.. A great deal of literature indicates that individuals with higher levels of education are more likely to participate than those with lower levels of education (Goyder, 1987; Green, 1996; Groves *et al.*, 2000; Stoop, 2005), although there are exceptions (Groves & Couper, 1998). Fitzgerald et al. (1998) found that educational status at the first interview of the PSID was related to contact at later waves. Other research on the PSID (Lillard & Panis, 1998) indicate that educational status at wave $t-1$ is related to the likelihood of participation at wave t ; individuals who have less than a

high school education were less likely to participate than those who have more than a high school education. In the European Community Household Panel, this pattern emerges—but only for some European countries (Behr *et al.*, 2005).

The target population in RDSL consists solely of 18- and 19-year old women. Given the homogeneity in age of this group of individuals, the level of education completed is also likely to be homogeneous, relative to the general population. However, current *school enrollment* status and *type of school* (if enrolled) may be heterogeneous within this group and also may affect the likelihood of response. These variables serve as a proxy for level of education. Women who are enrolled in school full-time or part-time should be more likely to participate than women who are not enrolled in school. And women who are enrolled in a four-year college should be more likely to participate than women who are still in high school.

I expect the effect of full-time enrollment to have a stronger positive effect on response propensity than being in school part-time. Likewise, being in a four-year college should have a stronger positive effect on response propensity than being enrolled in a two-year junior or community college or other kind of school.

The effect of educational status on wave response propensity may or may not change across journal periods. If civic duty is the source of an effect of education on propensity, then we have no reason to believe that education affects response propensity differently in earlier than later waves. If social cohesion with the sponsor is affecting the likelihood of response via educational status, then the effect of education on propensity might diminish across journal periods. If positive attitudes towards surveys increase response propensities relative to negative attitudes towards surveys, then the effect of

educational status should diminish over time as the burden of the survey lowers these positive attitudes (Goyder, 1986). However, given the dearth of literature on the effect of educational status on changes in response propensity across waves in a panel survey, all of these hypotheses are really speculations.

In sum, indicators of privacy concerns, topic salience, and educational status will be examined with respect to response propensity. Overall effects on propensity as well as time-varying effects of these indicators will be examined.

Lateness of response. In a sequential design, the mode of contact attempt is switched after a panel respondent is late a specified number of days. The decision to switch the mode is made based on the length of the field period and the assumed behavior of sample persons. For example, if sample persons are thought to respond right away and the field period is quite short, then the mode switch may occur quite early in the field. Other considerations with respect to the timing of the switch include cost.

If the first call in the second (more expensive) mode occurs too early in the field period, then more calls will be made in this expensive mode than might be necessary. On the other hand, if the switch occurs too late in the field period, the efficacy of the sequential design may be lost. In a panel survey, this cost implication occurs at every wave. If we have an inefficient sequential design at each wave, then the costs are a problem at each wave. It is therefore important to make informed decisions on when the mode is switched and if the mode should be switched at all.

This brief discussion has assumed that panel respondents take roughly the same amount of time to respond at each wave in a panel. However, there are no existing data to back up this claim. In addition to estimating response propensity across journal

periods in RDSL, Chapter 4 will examine whether panel respondents are consistently late in their responses or if they learn to be better respondents via the sequential design. As panel respondents are exposed to the design, they may change their behavior to avoid the prompt of the second mode. Given the dearth of literature on the use of sequential designs in panel surveys, this study is exploratory.

Although we have no direct evidence on the impact of a repeated sequential design in a panel survey on the lateness of response, we do know that panel respondents use experiences in previous waves to inform their behavior at a given wave. Some work from the Survey of Income and Program Participation (SIPP) shows that incentives at early waves can increase response rates for a number of waves, even if the incentives are only provided at Wave 1 (James, 1997; Mack *et al.*, 1998). If incentives can increase participation in later waves, perhaps sequential designs can operate in the same manner.

In Chapter 4, the consistency of the lateness of response is evaluated; are panel respondents more likely to be late respondents if they have been late before, or does the switch in mode serve as negative reinforcement? If the latter is true, then survey organizations might find that it is more cost-effective with respect to response rates to eliminate the sequential design in later waves.

Mode of response. In some sequential designs, only the first mode can be used until the predetermined date of the mode switch; after that date, only the second mode can be used. In such a design, response propensity, lateness, and mode of response are linked. In practice, this is rarely the case. In the ACS, mail questionnaires will be accepted and recorded as valid responses even after the switch to telephone (U.S. Census Bureau, 2009). In RDSL, panel respondents can use web or telephone throughout each journal

period. A toll-free telephone number is provided in the email or text request; some panel respondents call into the survey lab (“inbound CATI⁷”) early in the field period. If an interviewer is unavailable when the panel respondent calls into the lab, an interviewer can call her back (“outbound CATI”). Likewise, the web study is available at any time.

Because the lateness of response and the mode of response are not strictly independent, I examine these outcomes separately. But the relationship between past behavior and current behavior is expected to be the same as lateness of response: panel respondents are expected to be more likely to respond via a particular mode if they have a history of participating in that mode.

2.5.2. Nonresponse bias in sequential designs

Although the behavior of the panel respondent (response, lateness of response, and mode of response) is important to understand, we know little about the dynamics of nonresponse bias within such a complex survey. This research examines the nonresponse bias of the same variables thought to be predictors of response propensity (see Section 2.3.1). Overall wave nonresponse bias—that is, the difference in \bar{y} between respondents and the full sample—should behave much like the bivariate response propensity models.

The change in wave nonresponse bias across waves is also expected to behave like the response propensity models outlined in Section 2.5.1. Effects of privacy concerns should decrease across journal periods, for example.

The nonresponse bias analyses—while important—are of less interest than the effect of the sequential design on the nonresponse bias. That is, when we include the

⁷ Groves et al. (2009) distinguishes between the mode of administration and the technology of that mode. The mode of response in a CATI survey is telephone, and the technology is CATI. The shorthand of CATI will be used when describing the second mode in RDSL as well as the first mode in PASS.

respondents for the second mode in the estimate for the respondent mean, are we improving our estimate? Let us call the difference in the nonresponse bias between a single-mode study and the full sequential design:

$$(2.4) \quad \delta = (\bar{y}_{r1} - \bar{y}_n) - (\bar{y}_r - \bar{y}_n)$$

where \bar{y}_{r1} = the mean of y of individuals who responded via the initial mode, \bar{y}_n = the full sample mean of y , and \bar{y}_r = the mean of y for all respondents. In the case of RDSL, we have estimates of δ at each wave (j). Here, $\delta_j = (\bar{y}_{rwebj} - \bar{y}_{nj}) - (\bar{y}_{rj} - \bar{y}_{nj})$ where $\bar{y}_{r,web}$ = the mean of y of individuals who responded via web.

We know little about this effect of the sequential design on nonresponse bias, and even less about possible changes in δ_j across waves. In RDSL, δ could decrease as time in sample increases. We see this effect in Scenarios 4 and 5 in Figure 1. In Scenario 4, the nonresponse bias of the full sample increases, but δ is decreasing over time. In Scenario 5, the nonresponse bias stays close to zero, and δ decreases across waves. In both scenarios, the error is not improving with the addition of the second mode, but costs are likely to be higher than if web was the only mode. or 5 in Figure 1).

δ_j may decrease across waves for indicators of privacy concerns. The mean of having a sexual partner for web respondents might be greater than the mean for the entire sample earlier in the panel than later in the panel. At the same time, the mean of having a sexual partner for all respondents may be *lower* earlier in the panel. Later in the panel, privacy concerns diminish, and we may find relatively more women with a sexual partner responding via CATI than in previous waves.

For topic salience, there is not much evidence or reason to believe either way that δ_j would change as time in sample increases. However, Voogt and Saris (2005) found that topic salience is related to early response; as RDSL is not a pure sequential design, we cannot necessarily conclude that those who are interested are more likely to participate via web.

Educational status is also not expected to differ with respect to δ_j or changes in δ_j across journal periods. We might expect individuals who are in school to use web because they are more likely to have internet access than those who are not in school (Fortson *et al.*, 2007; S. Jones, 2002; S. Jones *et al.*, 2009); however, 92% of respondents to the baseline interview reported being able to complete the web survey using their own internet access, and the other 8% were dropped from the sample⁸. We still may expect that individuals in school will use web more often than those who are not in school because students tend to use the web more often than the general population (Fortson *et al.*, 2007). Over the course of a panel survey, however this effect may diminish as those who are not in school might learn to incorporate internet use into their routine.

2.5.3. Nonresponse and measurement bias

So far, we have addressed nonresponse bias and measurement bias separately. Here, we examine the relationship between response propensity and measurement bias in a telephone→face-to-face sequential design in a single wave (Wave 2) of PASS. Nonresponse bias, measurement bias, and the relationship between the propensity to respond and the measurement bias are examined. We are not examining panel

⁸ Further sample restrictions are detailed in Chapter 3.

characteristics, such as changes in nonresponse bias over time; therefore, the analyses in Chapter 6 resemble that of a cross-sectional survey.

First, nonresponse bias will be estimated in the telephone mode and the full sample. As in the case of RDSL, the change in nonresponse bias between the first mode and the full sample (δ) is expected to be nonzero. The magnitude and direction of δ should vary among variables of interest.

Because telephone includes a substantial number of cell phones, being at home is more important for face-to-face than telephone⁹. And access impediments differ between face-to-face and telephone. Differences in the nonresponse bias between telephone and the full sample are expected when the variable of interest is related to an access impediment or at-home pattern, and these access impediments or at-home patterns are more or less important in telephone compared to face-to-face.

For example, *gross monthly income* may be related to the likelihood of contact more often in face-to-face than in telephone. Answering machines, voice mail, and caller ID are common across the income spectrum—particularly because cell phones are extremely common in Germany (Kuusela *et al.*, 2008). However, access impediments in face-to-face such as gated communities and locked apartment buildings are more common among higher-income households than lower-income households (Sanchez *et al.*, 2005; Uhrig, 2008). Because being at home is more important in face-to-face than in telephone, I would expect more nonresponse bias in the full sample for gross income than in telephone.

⁹ In PASS, contacts in CATI can also include contacts via cell phone.

Being *employed, unemployed, employed full-time, and employed part-time* should also have more nonresponse bias in the full sample than in telephone alone. Each of these variables is related to at-home patterns. Being employed can reduce the time spent at home both during weekdays as well as nights and weekends (Groves & Couper, 1998). Part-time employment often leads to unpredictable hours (Walsh, 2005), which may make contact more difficult in face-to-face than telephone. *Length of unemployment* may also have more nonresponse bias in the full sample than in telephone alone. As the length of unemployment increases, the time spent at home is likely to increase; individuals with shorter periods of unemployment may be at home less than those with longer periods of unemployment, who may have given up the job search (Wilke, 2005).

Disability status may also be related to the likelihood of contact; individuals with a disability may be at home more often because they are less likely to work than those without a disability (Brault, 2008). At the same time, disabled persons may have less predictable at-home patterns because of doctor's appointments, hospitalizations, and so on (Uhrig, 2008). Proportionally fewer individuals with a disability may respond; however, it is unclear if this would change across modes. We could expect the nonresponse bias to increase *or* decrease with the addition of face-to-face.

Living in the former East German states may be related to nonresponse. Since reunification, unemployment is quite high relative to the former West German states (e.g., Kronthaler, 2005; Uhlig, 2006). We could expect that nonresponse bias should follow the same patterns as the employment status indicators. In addition, Blank and Schmidt (2003) found that East Germans tend to feel more social isolation than West Germans—that they are less a part of German society. Given that PASS is sponsored by a federal

German agency, we could expect that nonresponse, in general, would be higher for residents of the former East German states than the former West German states.

Having *children in the household* may have a positive influence on the likelihood of response. Groves and Couper (1998) and Gfroerer et al. (1997) found that adults with children in the household are more likely to be at home than adults without young children in the household. The nonresponse bias may be improved by including face-to-face. Similarly, *marital status* may be related to the likelihood of response; married people tend to be at home more often than single people (Abraham *et al.*, 2006; Groves & Couper, 1998; Stoop, 2005).

Not only are women more likely to be at home than men, but being *female* is also related to increased cooperation. Because of the increased at-home patterns, this estimate should improve when we add face-to-face responses.

Being a *German national* may affect nonresponse for many of the same reasons as being a resident of the former East German states. Unemployment is relatively high among those who are not German nationals (Liebig, 2007). Further, social isolation with respect to the sponsor may reduce the likelihood of response.

Age is a common correlate of nonresponse. Households with older adults tend to be at home more often than households with younger adults (Groves & Couper, 1998; Stoop, 2005). Because of these different at-home patterns, I expect nonresponse bias to be lower in the full sequential design than in telephone.

Measurement biases of some of these estimates are expected to differ. As detailed above, income is a cognitively challenging and sensitive question. Gross income should be improved by the design, as the face-to-face mode can often (Aquilino & Lo Sciuto,

1990; de Leeuw & van der Zouwen, 1988)—but not always (Hochstim, 1967; Sykes & Collins, 1988)—result in more accurate estimates of sensitive behaviors. Interviewers in face-to-face modes, relative to telephone, can help reduce the likelihood of satisficing (Holbrook et al., 2003).

Other sensitive questions that are expected to differ between CATI and CAPI with respect to measurement bias are disability status, employment, employed part-time, employed full-time, being unemployed, and length of unemployment. The PASS target population consists of benefit recipients in Germany. I would assert that, for these individuals, being disabled, being unemployed, being employed part-time, and being unemployed for a long period of time are socially undesirable. I would expect that measurement bias would be greater in CATI than CAPI. Marital status and presence of children should not differ across modes with respect to measurement bias.

But, as briefly noted above, there may be a relationship between nonresponse and measurement bias. In this framework, earlier respondents have less measurement bias than later respondents. But in the sequential design in PASS, earlier respondents are using CATI, while later respondents use CAPI.

The link between response propensity and measurement bias is expected to be stronger in CATI compared to CAPI. Consider two different response propensities: a) the propensity to participate in CATI, compared to participating in CAPI or not at all; and b) the propensity to participate in CAPI conditional on nonresponse to CATI, compared to not responding at all. Within CATI, I would expect more dramatic changes in measurement bias across levels of propensity (a). Within CAPI, however, the

measurement bias—regardless of propensity (b)—should be reduced relative to CATI. Changes across propensity (b) levels may be too small to estimate.

2.6. Summary

This chapter addressed the dynamics of sequential mixed-mode designs in cross-sectional surveys as well as panel surveys. We know little about response behavior—aside from response rate increases—and we know less about the dynamics of nonresponse bias. Further, the relationship between nonresponse and measurement bias in sequential designs is inherently linked.

Three general research issues are presented in this research. First, the dynamics of panel respondent behavior are examined. Is response propensity affected by privacy concerns, topic salience, or education status? Is the likelihood of being a late respondent at any wave related to earlier behaviors? And is the mode of response at any wave related to the mode of response in earlier waves?

The second research issue concerns nonresponse bias in a panel survey implementing a sequential design at each wave. Is the sequential design decreasing the nonresponse bias? And is this effect changing over time?

Third, the link between nonresponse and measurement bias is examined. When we create a situation in which response propensity may be linked with measurement bias, are earlier respondents more accurate than later respondents? Does a relationship between response propensity and measurement bias change depending on the mode of administration?

Chapter 3

Data and Methods

Two datasets will be used in this dissertation: the Relationship Dynamics and Social Life Survey and the Panel Arbeitsmarkt und Soziale Sicherung. This chapter describes these datasets in detail, including the necessary coding of variables and manipulation of files, and detail the analytic methods used in the rest of the dissertation.

3.1. Relationship Dynamics and Social Life

The Relationship Dynamics and Social Life (RDSL) is a study of 18- and 19-year old women from a county in Michigan. Key estimates of interest to the investigators are the use of contraception, relationship history, and pregnancy history. Because of the inherent cognitive difficulties in longitudinal data collection of these variables (e.g., Brown & Sinclair, 1999), RDSL uses weekly journals to collect this information. The Survey Research Center (SRC) at the University of Michigan collected the data.

3.1.1. Sample design

A simple random sample of 1004 18- and 19-year old women was selected from Michigan public records. Sample cases were selected in four replicates of roughly 250 women each. Each replicate began roughly four months after the previous replicate began.

The first interview was conducted in person. At this baseline interview, respondents were asked about their romantic relationships, pregnancy status and history, sexual history, and demographics. Respondents were asked to consent to participation in future weekly journals that could be conducted via web or telephone, with an interviewer. They were also asked if they would be willing and/or able to participate in these weekly journals via web. Ninety-two percent (92%) of the respondents reported that they were willing and able to complete the diaries on their own computers. The remaining 8% of respondents were offered telephone. Only the 92% ($n=921$) able to complete the survey online on their own computers were retained in these analyses. By excluding the 8% who were in the telephone group, internet coverage is held constant across all panel respondents.

One woman had a severe mental illness, and an additional seven women only participated in the baseline interview; these eight women were dropped from the following analyses, leaving 913 18-19 year old women who had access to the web in the final sample.

Of these women, 41% were 18 years old; 50% were 19; and 9% were 20 years old. Sixty-six percent described themselves as White; 32% described themselves as Black or African American; and 3% scribed themselves as American Indian or Alaskan Native, Asian, or Native Hawaiian or other Pacific Islander. One panel respondent did not respond to this question.

3.1.2. Data collection methods and journal period structure

In each journal after the baseline interview, respondents were asked about changes in relationship status since the previous completed journal. Respondents were

asked about current intentions to become pregnant or avoid pregnancy in the next month. Respondents were also asked about their sexual behavior—including having any sexual partners and use of contraception—since the previous journal or in the previous week, whichever was shorter. Four times per year, respondents were asked whether they were still in school, and if so, what kind of school. See Appendix B for the question wording of these survey questions.¹⁰

For each journal, the method of invitation could be a text message, an email with a link to the web survey, or both. Respondents chose this invitation method during the baseline interview. Note that this method of invitation is distinct from the mode of completion, which was CATI or web.

Five days after the baseline interview was completed, the first journal became active (“Day 1”). If, by Day 3, the journal was not completed, the panel respondents received an invitation by the preferred method to participate in the journal. If the respondent still had not submitted the journal on Day 5, another invitation is sent (via email and/or text message). On Day 6, the panel respondent is considered “late”, and contact attempts began via telephone. On Days 8, 15, 20, and 24, telephone was attempted as well as additional emails and mailings to the panel respondent’s address if she had not responded by that day. See Table 1 for the modes of call attempt in each day of the journal period.¹¹

The design also permitted panel respondents to use either CATI or web at any point during the journal period to complete the journal. A panel respondent could choose

¹⁰ Although the remainder of the survey questions are not included here, they are available upon request.

¹¹ Halfway through the first replicate, this protocol was changed: The Day 7 CATI contact attempt was eliminated. Two months later, the email and letter was eliminated on Days 8, 15, and 20; the CATI contact attempt was also eliminated on Day 20.

to call into the SRC to complete a journal before or after Day 6. Let us call this “inbound CATI¹²”. But when an interviewer called the panel respondent and collected the journal by CATI, an “outbound CATI” interview was collected. The panel respondent could also complete the survey via web after Day 6.

Table 1. Contact protocol in RDSL.

Day	Type of reminder or call attempt
1	<i>Journal becomes active (start of journal period; 5 days after end of previous journal period)</i>
3	Email and/or text message on cell phone
4	Email and/or text message on cell phone
5	Email and/or text message on cell phone
6	Telephone call
7	Telephone call
8	Telephone call, email, mail
15	Telephone call, email, mail
20	Telephone call, email, mail
24	Telephone call, email, mail
30	<i>end of journal period</i>

A journal period is defined as the set of days that began with the first day a journal became active and the day the respondent completed the journal, or 30 days after the start of the journal. These journal periods began with the first day a journal became active, which occurs five days after the previous journal period ends; each journal period ends with either response or 30 days of nonresponse. These journal periods are not temporally distinct among panel respondents. In any given day, one panel respondent might be completing her fifth journal and another might be completing her eighth. Figure 2 displays a hypothetical set of journal periods for four respondents starting five days after baseline through 50 days after baseline. The beginning of each journal period is

¹² As noted in Chapter 2, some researchers differentiate between the mode of administration (telephone) and the technology of the mode (CATI). Even though CATI is not a mode, it is the way the mode is referred to by the staff in both studies. When discussing the use of a mode in RDSL and in PASS, the term “CATI” is used even if an interview was not completed.

denoted by “S”, and response is indicated by “R”. In the illustration, respondent *A* is relatively fast in the first journal period, completing the journal the day after it becomes active—even before an invitation is sent. Five days later, she completes the second journal the very same day it becomes active, and on Day 12 when the third journal becomes active she does not complete it until Day 3 of that third journal period, after receiving two email or text invitations. Respondent *A* has 8 journal periods for the fifty day period, and her average journal period length is 2.5 days.

Other panel respondents are not as fast. Respondent *B* only has four journal periods of an average 9.25 days; respondent *C* has two journal periods, lasting 22.5 days on average. Respondents *C* and *D* did not respond to their first journal periods. Those journal periods therefore lasted 30 days. Respondent *D* has not responded to her second journal period as of the 50th day.

In order to reduce wave nonresponse, RDSL offered incentives. Women who responded to the baseline interview were offered \$35 on a debit card. For each complete journal, women received \$1; after five consecutive on-time journal submissions, respondents were given \$5.

3.1.3. Unit nonresponse

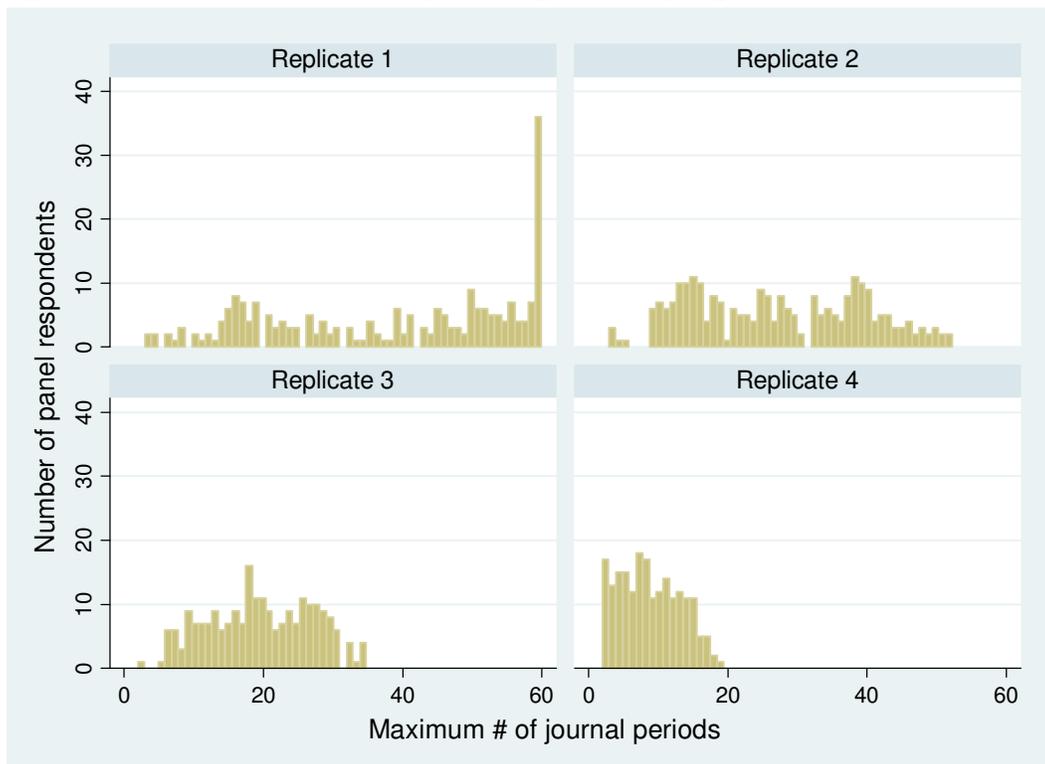
At each journal period, the response rate was computed. The denominator of each response rate is the total number of eligible panel respondents at that journal period. At journal period 2, the denominator of the response rate is the 913 women who had completed the baseline interview and had agreed to participate in the panel. After journal period 2, the denominator decreases in size (see bottom of Figure 4).

This drop in the number of eligible panel respondents occurs for three reasons. First, data collection was incomplete at the time of this analysis. The full field period in RDSL was 2 ½ years; however, only 16 months of data collection were available at the time of this analysis. The data were collected in four replicates, separated by about four months. Panel respondents in the last replicate were only in the field for a maximum of about four months, while panel respondents in the first replicate were in the sample for a maximum of about 16 months. These analyses could have included only the first four months of data collection within each replicate to maintain a constant time in the field; however, 53% of journal periods would have been excluded. Thus, all replicates were included to increase the sample size. Replicates were similar with respect to the predictor variables detailed below.

Figure 3 shows the total number of journal periods by replicate. Panel respondents from Replicate 1 were able to reach a maximum of 60 journal periods, while Replicate 4 reached a maximum of 19 journal periods. Within each replicate, the total number of journal periods has a wide range. Even in Replicate 1, a few panel respondents have as few as three journal periods; this is due to the panel respondent refusing to participate in all future journals or to the panel respondent not being located.

In Replicate 1, there is a spike at journal period 60, where 36 panel respondents had a total of 60 journal periods. Recall that only Replicate 1 had reached 16 months of data collection at the time of this analysis. We would expect a similar spike in the maximum number of journal periods for Replicates 2, 3, and 4 after those panel respondents completed 16 months of data collection.

Figure 3. Maximum number of journal periods by replicate.



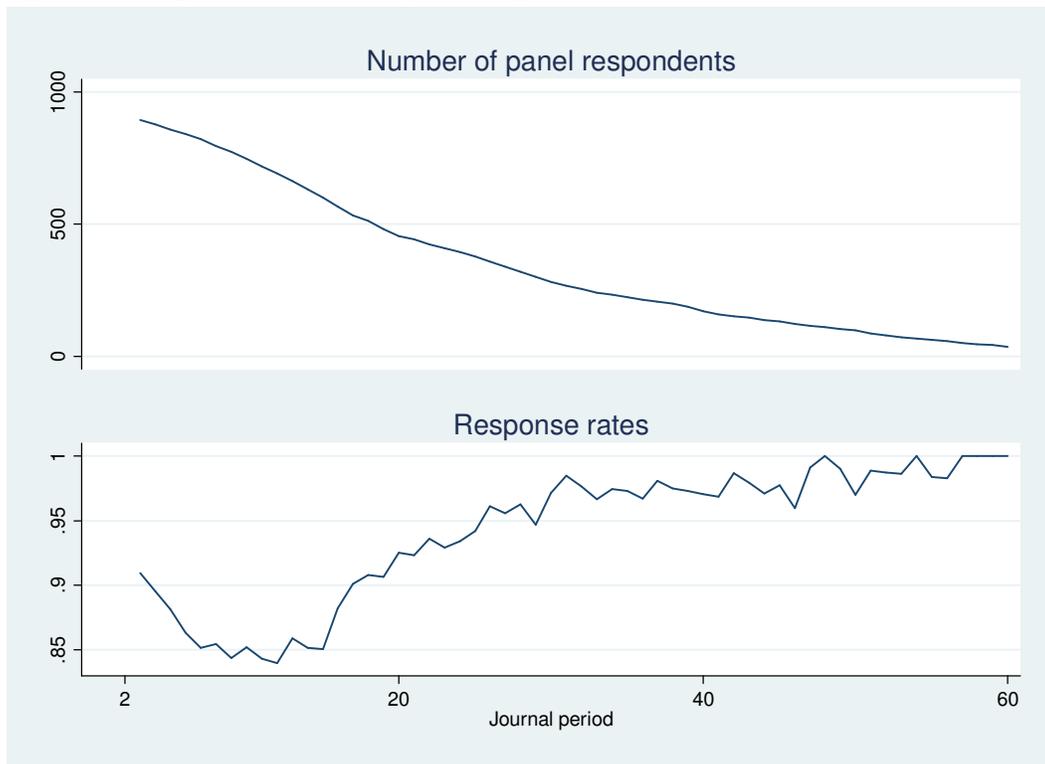
Second, some panel respondents are faster and more consistent respondents. Because the start, end, and length of journal period depends on responding behavior, the total number of journal periods will necessarily be larger for faster respondents than slower respondents, given the fixed 1 ½ year field period used in this analysis.

The third reason that the sample size declined across journal periods is that some panel respondents dropped out permanently. They either could no longer be found or they

explicitly refused to participate. This is not a major contribution to the decline in the response rate; only about 3.6% permanently left the panel.

Response rates were quite high across journal periods. The response rates also increased across journal periods; the first 15 journal periods see a decline in response rates, followed by an increase (see top of Figure 4). The lowest within-journal response rate was 84% in journal period 12, and the highest response rate was 100%, occurring at journal periods 48, 54, and 57 through 60.

Figure 4. Response rates over journal periods.



3.1.4. Outcomes and predictors

Chapter 4 examines response propensity, lateness of response, and mode of response in the RDSL. Response is a binary indicator: $r=0$ for nonresponse by Day 30 and $r=1$ for response. The lateness of response is a binary indicator of responding after at least one CATI prompt—that is, responding from Day 6 to Day 30 ($L=1$) or responding

by Day 5 ($L=0$). The mode of response is a three category indicator: web ($C=0$), inbound CATI ($C=1$), or outbound CATI ($C=2$), although some analyses simplify this outcome into a two category indicator (web vs CATI). Many panel respondents (about 50%) were consistent early respondents and consistent web respondents (about 33%).

Although the design of RDSL is a web→CATI sequential design, late respondents used web more often than inbound CATI or outbound CATI. As shown in Table 2, the majority of all responses were early. Before the switch, the proportion of outbound CATI journals is slightly higher than the proportion of inbound CATI journals. After Day 5, most responses were still in web; however, inbound CATI and outbound CATI had roughly the same proportion of completed journals.

Table 2 . Association between the lateness of response and mode of response in RDSL.

	Mode of response			Row %
	Web	Inbound CATI	Outbound CATI	
Early response	16,046	897	1083	90.0
Late response	1,757	159	138	10.2
Column %	88.7	5.3	6.1	100

In Chapter 4, indicators of privacy concerns, topic salience, and education status are used to model the likelihood of response and the likelihood of being late (see Table 3). These variables include questions about education, sexual behaviors, relationships, and attitudes towards pregnancy. See Appendix A, Tables 1 and 2, for means of these variables. Although most of the variables of interest were collected at baseline and at subsequent journals, some were only collected at baseline or subsequent journals. This chapter only briefly outlines these variables and their purpose in the model. To see the question wording, see Appendix B. See Section 3.1.7 for information on the handling of missing data.

Having any sexual partners was collected at baseline and at subsequent journals. At baseline, this question referred to the respondent's lifetime. At subsequent journals, this question asked about having sexual partners since the last journal was collected or one week, whichever was shorter.

Current use of *non-coital contraception*—such as hormonal methods or an intra-uterine device (IUD)—was asked at baseline and subsequent journals. The reference period for the use of coital contraception is either since the previous responding journal or two weeks, whichever is shorter. Note that this reference period is somewhat longer than *coital contraception* and *having any sexual partner*.

A series of questions were asked at journals after baseline about the use of various types of *coital contraception*, such as condoms, diaphragms, etc. These questions asked about contraceptive use in the last week or since the previous journal was completed, whichever was shorter. The *coital contraception* variable used in the analysis in this dissertation is a binary indicator of use of any kind of coital contraception or no use of contraception.

Living with a parent was only collected at the baseline interview. In a check-all-that-apply format, the respondent was asked to specify whether she lived with biological, adoptive, step- or a foster parent. These categories were collapsed into a binary indicator: living with any kind of parent or not.

Pregnancy intentions and *pregnancy avoidance*—how much the respondent intended to become pregnant or avoid becoming pregnant, respectively, during the next month—were collected at baseline and at each subsequent journal period for panel respondents who reported not being pregnant. These questions asked respondents to rate

their attitudes on a six-point scale. For pregnancy intentions, 0 means they do not at all want to get pregnant, and 5 means that they really want to get pregnant. For pregnancy avoidance, 0 means they don't at all want to avoid getting pregnant and 5 means they really want to avoid getting pregnant. For both questions, the responses were unevenly distributed for both questions. Across all journals—including baseline—92% of respondents chose 0 for the pregnancy intention question (they do not at all want to get pregnant); 91% chose 5 for the pregnancy avoidance question (they really want to avoid getting pregnant). Because of the small amount of variation, pregnancy intentions and pregnancy avoidance were both recoded as binary variables. For pregnancy intentions at baseline and across all subsequent journals, those who reported any desire to get pregnant were assigned 1 for pregnancy intentions, and all other respondents were assigned 0. Similarly, for pregnancy avoidance at baseline and all journals, those who expressed a strong desire to avoid getting pregnant were assigned 1 for pregnancy avoidance, and those who did not were assigned 0.

Note that pregnancy intentions and pregnancy avoidance are highly correlated but are not perfectly correlated. These attitudes, according to the investigators, are different dimensions of attitudes towards pregnancy (Barber *et al.*, 2008), and women may be more or less ambivalent about either dimension. For this reason, women who intend to become pregnant do not necessarily want to *not* avoid pregnancy, either.

Relationship status was collected at baseline and at each journal period; however, the *change in relationship status* between the previous responding journal period and the current journal period is thought to be related to nonresponse—not the relationship status at any given journal period. A change in relationship status is defined as any one of the

following situations: getting married; getting engaged; beginning a romantic, physical, or emotional relationship; separating or getting a divorce; ending an engagement; or ending a romantic, physical, or emotional relationship. This is only available for post-baseline journals. See Appendix B, Figure B.1. for the question series text.

Being *currently in school* and *type of school* was collected at the baseline interview and every 3 months. For the first three months of journals—until the first quarterly journal was completed—the value for school enrollment and school type were imputed from the baseline interview.

Table 3. Source, frequency, and item missing data for privacy concerns, topic salience, and education status variables from RDSL.

Variable	Collected at baseline	Collected at subsequent journals	Frequency at baseline	Item nonresponse [†]
<i>Privacy concerns and topic salience indicators</i>				
Having any sexual partners	Yes	Yes	71%	0.37%
Use of non-coital contraception	Yes	Yes	39%	0.02%
Use of coital contraception	No	Yes	n/a	0.03%
Living with a parent	Yes	No	45%	0%
Pregnancy intention	Yes	Yes	9%	0.37%
Pregnancy avoidance	Yes	Yes	90%	0.38%
Change in relationship status	No	Yes	n/a	0%
<i>Education status indicators</i>				
Currently enrolled in school	Yes	Every 3 months	24%	0%
Not enrolled			12%	
Part-time enrollment			64%	
Full-time enrollment				
Type of school (if enrolled)	Yes	Every 3 months	17%	0%
High school or less			34%	
2 year junior/community college			42%	
4 year college			6%	
Voc, tech, trade, or other school				

[†] Item nonresponse across all available journals and baseline. If a variable was collected only at baseline, item nonresponse listed is from baseline only. If a variable was collected only at subsequent journals, item nonresponse listed is from those journals only.

3.1.5 Data processing and model formulation

Data from all journal periods from all panel respondents were combined so that each observation in the dataset was the survey data from a particular journal period for a particular panel respondent, including response and mode indicators. In the analyses below, the respondent was treated as a cluster.

Item nonresponse—conditional on response to a particular journal—was quite low for all variables of interest, but especially baseline variables. Table 3 presents the percent item missing data for each variable of interest. Item nonresponse does not surpass 1% for any variable.

3.1.6 Missing data compensation: Sequential regression multiple imputation

Chapter 4 examines the effect of the variables of interest in Table 3 on response propensity, while Chapter 5 investigates longitudinal changes in nonresponse bias of these same variables. In order to complete these analyses, data for respondents and nonrespondents must be available. As in most surveys, we do not have these data for nonrespondents. Although we do not know the true values, we can estimate them using imputation to compensate for unit nonresponse and item nonresponse. Data were imputed for item nonresponse as well as unit nonresponse. If a contact attempt was made in a journal period but the panel respondent did not respond, data for that journal period were imputed. If a panel respondent was unable to be found, *and* no contact attempts were made, that journal period was neither imputed nor included in the dataset.

Imputation is a procedure by which sets of plausible values are generated to compensate for item nonresponse and, sometimes, unit nonresponse. Imputation results in a complete-case dataset that can reduce the nonresponse bias of certain estimates. Single

imputation imputes a value for each missing datum, based on models utilizing auxiliary data. However, the standard errors resulting from a singly imputed dataset may be artificially low (Raghunathan *et al.*, 2001). To avoid this issue, data were imputed five times¹³, and multiple imputation procedures were used.

Sequential regression multiple imputation (SRMI) was used in IVEWARE (Raghunathan *et al.*, 2001)—a SAS program that utilizes SRMI to compensate for missing data—was used for the imputation. SRMI is a method of imputing multiple datasets by using a regression model to predict missing values of a variable conditional on all available nonmissing information. After one variable is imputed, it is used for the imputation model for another variable. This procedure is repeated for all variables, beginning with the variable with the fewest missing values and ending with the variable with the most missing values.

The dataset was converted to a “wide” format, in which each observation represented a particular panel respondent and each column was a particular variable at a particular journal period. The sequential regression models could thus utilize information from other journal periods as predictors, taking advantage of higher between-journal correlations to obtain more accurate and reliable predicted values.

Including all variables in the dataset as a source for predictors in each sequential regression resulted in a lack of convergence in some of the regression models due to a high degree of collinearity. Therefore, specific covariates were selected that could create relatively good-fitting regression models. Spearman correlations were run to determine how correlated the variables of interest were with those variables at adjacent journal

¹³ Although five is an arbitrary number of datasets, (Raghunathan, 2004) found that five imputations are generally sufficient.

periods. Cross-wave correlations were quite high, as might be expected (tables available upon request). Each variable of interest collected at journal period 2 (y_{i2}) was regressed on the same variable of interest at journal period 3 (y_{i3}). The residuals from these regression models were correlated with each of the other variables in the dataset at journal period 2 in order to estimate how much extra variance would be explained by adding each of these other variables, beyond simply adding y_{i3} to the regression model. For each variable of interest, the variables with the five highest correlations with the residual were selected. This procedure was replicated at journal periods 12 and 13¹⁴. After the five highest covariates were selected for each run, some were removed to reduce potential collinearity problems. A total of 37 variables were selected, 35 of which were collected only at the baseline interview.

IVEWARE has the capability to incorporate restrictions and bounds in the imputation procedure. Some variables are only asked of subsets of the sample; for example, coital contraceptive use is only asked of individuals who report having at least one sexual partner. These restrictions were added where appropriate. Numerical boundaries were specified for continuous or count variables to eliminate implausible values. By using restrictions and bounds, implausible values are less likely to be imputed.

The robustness of multiple imputation depends on the data being “missing at random” (Rubin, 1987); conditional on observed characteristics, respondents are no different from nonrespondents on a statistic of interest. That is, the distribution of y values for respondents and nonrespondents is identical, conditional on a set of x variables.

¹⁴ Journal periods 2, 3, 12, and 13 were selected somewhat arbitrarily for this test. Journal periods 2 and 3 were selected because they are at the beginning of the panel. Journal periods 12 and 13 were selected because they occur partway through data collection, and most of the original panel respondents had reached these journal periods (76% for journal period 12 and 73% for journal period 13).

In RDSL—as in most surveys—the missing at random assumption is untestable because we do not know the distribution of y for nonrespondents. However, some diagnostics are available to estimate possible violations of this assumption. One such diagnostic is the fraction of missing information (Rubin, 1987), which is the ratio of the between-imputation variance to the total variance of the imputed data. The fraction of missing information is a measure of how certain we are of the precision of the imputation procedures; it is a measure of the reliability of imputation procedures (Wagner, 2010). A low fraction means greater reliability. Because of the relatively low incidence of missing data in RDSL and the extraordinarily large amount of auxiliary information on panel nonrespondents, the fractions of missing information are quite low (see Table 4).

Table 4. Fraction of missing information (FMI) for target variables in RDSL.

Variable	FMI at baseline	FMI at journals
<i>Privacy concerns and topic salience indicators</i>		
Having any sexual partners	0.0003	0.0004
Use of non-coital contraception	0.0003	0.0000
Use of coital contraception	(n/a)	0.0001
Living with a parent	0.0000	(n/a)
Pregnancy intention	0.0010	0.0001
Pregnancy avoidance	0.0003	0.0005
Change in relationship status	(n/a)	0.0001
<i>Education status indicators</i>		
Currently enrolled in school		
Not enrolled	0.0003	0.0001
Part-time enrollment	0.0004	0.0005
Full-time enrollment	0.0003	0.0003
Type of school (if enrolled)		
High school or less	0.0004	0.0003
2 year junior/community college	0.0007	0.0001
4 year college	0.0002	0.0001
Voc, tech, trade, or other school	0.0003	0.0007

The imputation procedure created five datasets with fully-completed data. Analyzing the multiply-imputed datasets involves running the same analysis on each dataset, using complete-case statistical methods, and generating point estimates and standard errors. The multiply imputed point estimate reported is the average of the point estimates across imputed datasets, and the standard errors are a function of the within- and between-imputation standard errors (Li *et al.*, 1991a; Li *et al.*, 1991b; Meng & Rubin, 1992; Rubin, 1987). STATA's *mim* program was used to combine each of these datasets and generate estimates from the multiple datasets.

3.1.7. Analytic models in Chapters 4 and 5

Analytic models in Chapter 4. In Chapter 4, several types of models are used to examine response propensity, the lateness of response, and the mode of response. For response propensity, the predictors include those described above, as well as the change in the substantive predictors over journal periods.

For all models used in Chapters 4 and 5, the data were stacked so that each observation in the dataset was a particular journal period for a particular respondent. The within-panel respondent correlation was accounted for using STATA's *svy* command.

Three sets of models are estimated related to *response propensity*. First, a set of bivariate models estimate the net effect of the indicators of privacy concerns, topic salience, and education status on the likelihood of response. Second, the effect of time in sample is examined using process data. Third, the change in the effect of the indicators of privacy concerns, topic salience, and education status on the likelihood of response is estimated.

To estimate the effect of the predictors described above on the likelihood of response, the following logistic regression model¹⁵ was estimated:

$$(3.1) \quad \ln\left(\frac{p_{rij}}{1-p_{rij}}\right) = \beta_0 + \beta_1 X_{ij} + \varepsilon_{ij}$$

where p_{rij} = the probability of response at journal period j for panel respondent i and X_{ij} = the value of a predictor from Table 3 for panel respondent i at journal period j . If $\beta_1 > 0$, then the odds of responding are greater when $X_{ij} = 1$ than when $X_{ij} = 0$. For example, for having a sexual partner at baseline, if $\beta_1 > 0$, then the odds of participating are greater for women who have a sexual partner at journal period j than those who do not have a sexual partner.

Second, the change in the propensity to participate across journal periods was examined. A simple logistic regression model with only a main effect for journal period (JP_{ij}) could have been fitted. However, the effect of journal period on the likelihood of response is confounded with the design of RDSL. Panel respondents who are generally slower to respond must have fewer journal periods than those who are generally faster to respond. Any model predicting change across journal periods must avoid confounding the journal period indicator with the speed of response. One way to do this is to include the average length of journal periods from journal period 1 to $j-1$. Let us denote $MNLEN_{ij-1}$

¹⁵ Hazard models are often used to estimate the time to a single event, such as a death. Some models, such as the Andersen-Gill model, can be used to estimate the time to multiple events (Andersen & Gill, 1982). But this dissertation is concerned only with the presence or absence of the event of response, late response, or response in a particular mode—not the time to one of these events. A hazard model treating time as a continuous variable would therefore not be appropriate here. Discrete time hazard models could also be used, treating time as a journal period indicator; however, these models are computationally quite similar to logistic regression models, which have a clearer interpretation. Logistic regression models will therefore be used in Chapter 4 to estimate the propensity to respond, the propensity to respond late, and the propensity to respond in a particular mode.

as the mean journal period length from journal period 1 to journal period $j-1$. The response propensity model then becomes:

$$(3.2) \quad \ln\left(\frac{P_{rij}}{1-P_{rij}}\right) = \beta_0 + \beta_1 JP_{ij} + \beta_2 MNLEN_{ij-1} + \varepsilon_{ij}$$

A positive coefficient for journal period indicates that, as journal periods increase, the likelihood of response decreases. β_2 is of less substantive interest; however, it is expected to be less than zero. That is, as the time it has taken a panel respondent in the past to respond increases, the likelihood of response at journal period j decreases.

Next, the change in the effect of X_{ij} across journal periods was estimated. Again, $MNLEN_{ij-1}$ was added to this model to control for the fact that higher journal periods, by design, included only panel respondents who had responded relatively quickly and consistently:

$$(3.3) \quad \ln\left(\frac{P_{rij}}{1-P_{rij}}\right) = \beta_0 + \beta_1 X_{ij} + \beta_2 JP_{ij} + \beta_3 MNLEN_{ij-1} + \beta_4 X_{ij} JP_{ij} + \varepsilon_{ij}$$

If $\beta_4 > 0$, then as journal periods increase, the difference in propensity between women who, for example, have a sexual partner and do not have a sexual partner increases. If $\beta_4 < 0$, then as journal periods increase, the difference in response propensity between women who have a sexual partner and those who do not decreases.

The second set of models in Chapter 4 will examine the *lateness of response*—that is, the effect of having past late responses on current late response. For these models, nonrespondents were dropped from the analysis. First, the effect of time in sample and previous behavior on the likelihood of being a late respondent is estimated. Second, the change in the effect of previous late responding on the likelihood of current late

responding is estimated. Third, the impact of indicators of privacy concerns, topic salience, and education status on the lateness of response is examined.

The variable $LATEJ_{ij-1}$ was estimated as the total number of completed journals that were submitted after Day 5, up to journal period $j-1$.

A main effects model will estimate the effect of a history of lateness and time in sample on the likelihood of being late at the current journal:

$$(3.4) \quad \ln\left(\frac{p_{Lij}}{1-p_{Lij}}\right) = \beta_0 + \beta_1 LATEJ_{ij-1} + \beta_2 JP_{ij} + \beta_3 MNLEN_{ij-1} + \varepsilon_{ij}$$

where p_{Lij} = the probability of responding late for panel respondent i at journal period j .

β_2 is the change in the propensity to be late across journal periods; $MNLEN_{ij-1}$ is again added to the lateness model because higher-numbered journal periods will only be available for panel respondents who are rarely late. If $\beta_2 > 0$, then the propensity to respond after Day 5 (compared to responding before Day 5) increases across journal periods—that is, panel respondents are more likely to be late at later journal periods than earlier ones. If $\beta_2 < 0$, then panel respondents are becoming faster respondents as time in sample (journal period) increases.

A separate model estimates the interaction between time in sample and a history of lateness on lateness at the current journal:

$$(3.5) \quad \ln\left(\frac{p_{Lij}}{1-p_{Lij}}\right) = \beta_0 + \beta_1 LATEJ_{ij-1} + \beta_2 JP_{ij} + \beta_3 MNLEN_{ij-1} + \beta_4 LATEJ_{ij-1} JP_{ij} + \varepsilon_{ij}$$

If $\beta_4 > 0$, then the difference among levels of $LATEJ_{ij-1}$ in the change in the odds of being a late respondent compared to an early respondent decreases across journal periods. In that case, a history of being late has less of an effect on the propensity to be late as

journal periods increase. If $\beta_4 < 0$, then a history of lateness has a stronger effect on current late response in later journal periods than in earlier journal periods.

But some indicators of privacy concerns and topic interest may be related to lateness of response. First, a bivariate model is estimated:

$$(3.6) \quad \ln\left(\frac{p_{Lij}}{1-p_{Lij}}\right) = \beta_0 + \beta_1 X_{ij} + \varepsilon_{ij}$$

Women who have privacy concerns are expected to be more likely to be late than women who do not have such concerns; at the same time, women who are interested in the topic may respond more quickly than those who have low levels of topic salience. But do these effects alter the relationship between a history of lateness and current lateness?

The predictors X_{ij} are added to model (3.5):

$$(3.7) \quad \ln\left(\frac{p_{Lij}}{1-p_{Lij}}\right) = \beta_0 + \beta_1 X_{ij} + \beta_2 LATEJ_{ij-1} + \beta_3 JP_{ij} + \beta_4 MNLEN_{ij-1} + \beta_5 LATEJ_{ij-1} JP_{ij} + \varepsilon_{ij}$$

A third set of analyses in Chapter 4 will examine the panel respondents' *propensity to use a mode*. As an initial step, mode is treated as a binary variable (CATI vs web). The effect of time in sample on the likelihood of using CATI at journal period j is estimated. Second, the effect of using CATI in the past on current CATI use is estimated. Then, these models are repeated, distinguishing inbound CATI from outbound CATI.

The effect of time in sample on use of mode is estimated using a logistic regression model:

$$(3.8) \quad \ln\left(\frac{1-p_{wij}}{p_{wij}}\right) = \beta_0 + \beta_1 JP_{ij} + \beta_2 MNLEN_{ij-1} + \varepsilon_{ij}$$

where p_{wij} =the probability of participating in web, compared to participating in CATI. As in the lateness models, the analyses investigating mode of response include only respondents. $MNLEN_{ij-1}$ is again added to the model to control for the selection mechanism described above.

This dissertation investigates whether past uses of a mode influences current uses of a mode. Still treating mode as a binary variable, the following model is fit:

$$(3.9) \quad \ln\left(\frac{1-p_{wij}}{p_{wij}}\right) = \beta_0 + \beta_1 CATISUM_{ij-1} + \beta_2 MNLEN_{ij-1} + \varepsilon_{ij}$$

where $CATISUM_{ij-1}$ =the total number of CATI journals completed from baseline through journal period $j-1$. This predictor, like the dependent variable, treats mode as a binary variable. As the number of journals completed via CATI increase, it is expected that the likelihood of using web decreases. That is, β_1 should be less than zero.

But mode in RDSL is not binary in nature. Panel respondents can call into the survey lab (inbound CATI) or an interviewer can call them from the survey lab (outbound CATI). Models (3.8) and (3.9) are repeated, treating inbound CATI and outbound CATI as separate modes:

$$(3.10) \quad \ln\left(\frac{p_{Cij}}{p_{wij}}\right) = \beta_0 + \beta_1 JP_{ij} + \beta_2 MNLEN_{ij-1} + \varepsilon_{ij}$$

where p_{Cij} =the probability of using one of the CATI modes (inbound: $C=1$; outbound: $C=2$) compared to using web at journal period j .

$$(3.11) \quad \ln\left(\frac{p_{Cij}}{p_{wij}}\right) = \beta_0 + \beta_1 ICSUM_{ij-1} + \beta_2 OCSUM_{ij-1} + \beta_3 MNLEN_{ij-1} + \varepsilon_{ij}$$

where $ICSUM_{ij-1}$ = the total number of inbound CATI journals completed from baseline through journal period $j-1$ and $OCSUM_{ij-1}$ = the total number of outbound CATI journals completed from baseline through journal period $j-1$.

Consider the two logits in model (3.11). If $\beta_1 > 0$ when the outcome variable is inbound CATI response compared to web, then having a history of inbound CATI journals is related to current use of inbound CATI. If $\beta_2 > 0$ when the outcome is outbound CATI compared to web, then having a history of outbound CATI journals is related to current use of outbound CATI.

As in the response propensity models and the lateness models, the use of a mode should be related to some indicators of topic interest and privacy concerns. This is examined in both a binary logistic regression model as well as a multinomial logistic regression model:

$$(3.12) \quad \ln\left(\frac{1 - P_{wij}}{P_{wij}}\right) = \beta_0 + \beta_1 X_{ij} + \beta_2 CATISUM_{ij-1} + \beta_3 MNLEN_{ij-1}$$

and

$$(3.13) \quad \ln\left(\frac{P_{cij}}{P_{wij}}\right) = \beta_0 + \beta_1 X_{ij} + \beta_2 ICSUM_{ij-1} + \beta_3 OCSUM_{ij-1} + MNLEN_{ij-1}.$$

If the addition of X into model (3.11), as shown in model (3.13), has the effect of changing the coefficients for $ICSUM_{ij-1}$ or $OCSUM_{ij-1}$, then the privacy concerns or topic salience indicators explain previous behavior, in part.

Analytic models in Chapter 5. Chapter 5 examines the nonresponse bias of the variables listed in Table 3. Nonresponse bias within each journal period j is estimated as:

$$(3.14) \quad \bar{y}_{r,j} - \bar{y}_{nj}$$

where $\bar{y}_{r,j}$ = the mean of the estimate y for all respondents at journal period j and \bar{y}_{nj} = the mean of all panel respondents at journal period j . Across journal periods, then, we can estimate the mean bias by taking the average of the nonresponse bias across journal periods.

In order to test whether the wave nonresponse bias estimates are significantly different from zero, a series of regression models are estimated, each treating the panel respondent as a cluster using STATA's *svy* command. In these models, we ignore any effect of journal period or mode on nonresponse bias; we estimate only whether the net difference in y (or categories of y) are significantly different from zero. In these models, the predictors used in Chapter 4 (listed in Table 3) become the dependent variable (y).

Most of these models are logistic regression models with binary outcomes (e.g., having a sexual partner versus not having a sexual partner):

$$(3.15) \quad \ln \left[\frac{p(y_{ij})}{1-p(y_{ij})} \right] = \beta_0 + \beta_1 r_{ij} + \varepsilon_{ij}$$

where r_{ij} = an indicator for response for panel respondent i at journal period j and

$p(y_{ij})$ = the probability of $y_{ij} = 1$ (for example, having a sexual partner). In this model,

β_1 = the difference in the log odds of $y_{ij} = 1$ (e.g., having a sexual partner) between

respondents and nonrespondents—which is equivalent to $\bar{y}_r - \bar{y}_m$. We can test the

following contrast to estimate if $\frac{\sum_{j=1}^{J-1} (\bar{y}_{r,j} - \bar{y}_{nj})}{J-1} = 0$:

$$(3.16) \quad \theta = \frac{\beta_1}{J-1} \left(\sum_{j=1}^{J-1} \frac{m_j}{n_j} \right)$$

If θ is not significantly different from zero, then the mean nonresponse bias across all journal periods is not significantly different from zero.

For being enrolled in school and type of school, a binary logistic regression model is inadequate because each has more than two categories. For these variables, a multinomial logistic regression is fit:

$$(3.17) \quad \ln \left[\frac{p(y_{kij})}{p(y_{0ij})} \right] = \beta_0 + \beta_1 r_{ij} + \varepsilon_{ij}$$

where y_{kij} = category k of y at journal period j and y_{0ij} = the base outcome of y at journal period j . Because (3.17) is a multinomial model, there are separate estimates of β_1 for each outcome of y , excluding the reference category. Each estimate of β_1 equals the difference between respondents and nonrespondents in the relative risk of having the value for category k compared to the reference category. For example, for being enrolled in school, the categories are being enrolled part-time and being enrolled full-time; the reference category is not being enrolled in school. When k = being enrolled part-time, β_1 is the difference in the relative risk of being enrolled in school between respondents and nonrespondents.

Similar to the logistic models, the estimated nonresponse bias for category k is:

$$(3.18) \quad \theta_k = \frac{\beta_1}{J-1} \left(\sum_{j=1}^{J-1} \frac{m_j}{n_j} \right)$$

If $\theta_k = 0$ (or, for the multinomial models, $\theta_k = 0$) then the overall wave nonresponse bias of $y = 0$. If $\theta_k > 0$, then the overall wave nonresponse bias is greater than zero; for example, the proportion of respondents who have a sexual partner is greater than

the proportion of all panel respondents who have a sexual partner. If $\theta. < 0$, then the nonresponse bias is negative.

But these estimates do not account for time in sample, nor do they tell us about changes in wave nonresponse bias across journal periods. Models (3.15) and (3.17) can be extended as follows for the binary variables:

$$(3.19) \quad \ln \left[\frac{p(y_{ij})}{1-p(y_{ij})} \right] = \beta_0 + \beta_1 r_{ij} + \beta_2 JP_{ij} + \beta_3 r_{ij} JP_{ij} + \varepsilon_{ij}$$

For the education variables, this takes the multinomial form:

$$(3.20) \quad \ln \left[\frac{p(y_{kij})}{p(y_{0ij})} \right] = \beta_0 + \beta_1 r_{ij} + \beta_2 JP_{ij} + \beta_3 r_{ij} JP_{ij} + \varepsilon_{ij}$$

In both models, β_3 = the mean change in the overall wave nonresponse bias $(\bar{y}_{r.j} - \bar{y}_{m.j})$ across journal periods. As in (3.17) and (3.18), we can test a combination of coefficients to test whether the mean change in nonresponse bias is significantly different from zero:

$$(3.21) \quad \theta_{\Delta} = \frac{\beta_3}{J-1} \sum_{j=1}^{J-1} \frac{m_j}{n_j}$$

And, as above, this can be extended for the multinomial models by running this contrast on each logit.

This study compares the nonresponse bias of web respondents (i.e., treating CATI respondents as nonrespondents) to the nonresponse bias of all web and CATI respondents (i.e., at the end of the field period).

$$(3.22) \quad \delta_j = (\bar{y}_{rwebj} - \bar{y}_{nj}) - (\bar{y}_{r.j} - \bar{y}_{nj})$$

where δ_j = the change in nonresponse bias before and after the switch at journal period j ;
 \bar{y}_{rwebj} = the mean of a particular estimate y for web respondents at journal period j ; \bar{y}_{nj} = the overall sample mean of y at journal period j ; and $\bar{y}_{r,j}$ = the mean of all respondents at journal period j . This calculation is straightforward, but we do not have a good method to determine if δ_j is significantly different from zero. We can extend the methods above; a combination of coefficients in a regression model can be used to test this.

As an illustration, let us consider a different kind of mixed mode design, in which an individual is randomly assigned to web or CATI; within panel respondent, the mode is never switched in this “concurrent” mixed mode design. Assume that this is a cross-sectional survey. In order to estimate significant differences in nonresponse bias between CATI and web in such a design, we could estimate the following linear regression model:

$$(3.23) \quad y_i = \beta_0 + \beta_1 r_i + \beta_2 M_i + \beta_3 r_i M_i + \varepsilon_i$$

where r_i = an indicator of response (versus nonresponse), and M_i = an indicator for individuals *assigned* to web (versus CATI). To test whether the nonresponse bias differs across modes—that is, whether $(\bar{y}_{r,web} - \bar{y}_{n,web}) - (\bar{y}_{r,CATI} - \bar{y}_{n,CATI})$ —we could test the

$$\text{contrast } \frac{m_{web}}{n_{web}} (\beta_1 + \beta_3) - \frac{m_{CATI}}{n_{CATI}} (\beta_1), \text{ where } m_{web} = \text{the number of nonrespondents assigned}$$

to the web mode; n_{web} = the number of total sample persons assigned to web; m_{CATI} = the number of nonrespondents assigned to CATI; and n_{CATI} = the total number of sample persons assigned to CATI. See Appendix A, Equation 1 for the derivation.

However, this method is inadequate for estimating the change in nonresponse bias between waves in a sequential mixed-mode design. Model (3.23) does not work within a

sequential design because those who could participate in web will necessarily be respondents. The model is not estimable under a sequential design.

An alternative is to design an experiment. Sample cases would be assigned randomly to one of two treatment groups: in one group (Treatment 1), sample cases would be permitted to respond only via web. In the other group (Treatment 2), individuals could participate in the sequential web→CATI mixed mode design like RDSL. Treatment 1 would have exactly the same number of prompts and calls as Treatment 2, but only web responses would be counted as responses. In such a design, we could compare the change in the nonresponse bias of the web study to the nonresponse bias of the full web→CATI study. In Chapter 5, this experiment is simulated using two independent multiple imputation procedures.¹⁶

In Section 3.1.7, methods were detailed on an imputation procedure used to estimate missing data for unit nonresponse and item missing data. We will call this Treatment 2; unit nonresponse was imputed only for nonrespondents.

The data were imputed a second time to simulate a study in which only web responses were counted as responses. Let us call this Treatment 1 (see Figure 5; imputed data are shaded in gray). The imputation procedures in Treatment 1 were identical to the procedures in Treatment 2; however, in Treatment 1, CATI respondent data were removed and treated as missing data. The data were imputed again using the methods detailed in Section 3.1.7. These imputations were based only on the web responses, responses at previous or later journal periods, and baseline data.

¹⁶ An inherent assumption of this method is that the nonresponse bias in a web-only survey would be the same as the nonresponse bias of web respondents in a sequential design. We have no reason to suspect this assumption is violated.

Figure 5. Simulated experiment comparing a web single mode design to a web→CATI sequential design using one web→CATI mixed-mode survey.

Treatment 1 [$T_i = 1$]	Treatment 2 [$T_i = 0$]
Web respondents [$G_{1i} = 1$] [$G_{2i} = 0$]	Web respondents and CATI respondents [$G_{1i} = 0$] [$G_{2i} = 1$]
CATI respondents and nonrespondents [$G_{1i} = 0$] [$G_{2i} = 0$]	Nonrespondents [$G_{1i} = 0$] [$G_{2i} = 0$]

As described in Section 3.1.7, five imputed datasets were generated. The data were stacked so that each observation is a particular imputation for a treatment for a panel respondent at a particular journal period. Therefore, the data contain $5 \times 2 \times \sum_{j=1}^{60} n_j$ observations, where n_j = the number of eligible panel respondents at journal period j .

The following logistic regression model is used to estimate whether the mean change in nonresponse bias between Treatment 1 and Treatment 2 was significantly different from zero ($H_0 : \bar{\delta}_{T_1, T_2} = 0$). For binary variables, this model took a logistic form:

$$(3.24) \quad \ln \left[\frac{p(y_{ij})}{1 - p(y_{ij})} \right] = \beta_0 + \beta_1 T_i + \beta_2 G_{1i} + \beta_3 G_{2i} + \beta_4 JP_{ij} + \beta_5 T_i JP_{ij} + \beta_6 G_{1i} JP_{ij} + \beta_7 G_{2i} JP_{ij}$$

where G_{1i} = an indicator for response in Treatment 1, and G_{2i} = an indicator for response in Treatment 2 (see Figure 5). For the multinomial variables, a multinomial model was fitted with the same covariates.

We can use the coefficients of these models to address whether the nonresponse bias between treatment groups was, overall, significantly different from zero. For example, the respondent mean of Treatment 1 at journal period j is equal to $\beta_0 + \beta_1 + \beta_2 + \beta_4 j$. The following linear combination of coefficients can be tested to determine if $H_0 : \bar{\delta} = 0$:

$$(3.25) \quad \theta_D = \frac{1}{59} \left[\beta_2 \sum_{j=2}^{60} \left(\frac{m_{j1}}{n_{j1}} \right) + \beta_6 \sum_{j=2}^{60} \left(\frac{m_{j1}j}{n_{j1}} \right) - \beta_3 \sum_{j=2}^{60} \left(\frac{m_{j2}}{n_{j2}} \right) - \beta_7 \sum_{j=2}^{60} \left(\frac{m_{j2}j}{n_{j2}} \right) \right]$$

See Appendix A, Equation 2 for this derivation. As in the previous set of models, separate contrasts are estimated for the multinomial models.

As noted in Chapter 2, the difference in the nonresponse bias between treatment groups may change across journal periods. Under model (3.24), we can also examine the change in nonresponse bias between web and the full sample ($H_0 : \delta_j = 0$) by testing the following linear combination:

$$(3.26) \quad \beta_6 - \beta_7 = 0$$

If this combination is not significantly different from zero, then the effect of the sequential design on the nonresponse bias does not change across journal periods; we have a situation similar to Scenario 1 or Scenario 2. The nonresponse bias may or may not be increasing for both treatment groups, but the rate of increase is the same for both groups.

3.2. Panel Arbeitsmarkt und Soziale Sicherung

The Panel Arbeitsmarkt und Soziale Sicherung (PASS) study is a panel survey of program participation in Germany conducted by the Institute of Employment Research

(Institut für Arbeitsmarkt- und Berufsforschung: IAB), a department of the German Federal Employment Agency (Bundesagentur für Arbeit: BA). The goal of the survey is to examine the social effects of receiving government benefits such as welfare and unemployment benefits, coping behaviors of people receiving benefits, and behaviors leading to employment. The target population of PASS includes benefit recipients as well as non-recipients; however, this dissertation will include only recipients in all analyses.

This dissertation uses PASS survey data linked to the Integrated Employment Biographies (IEB), an administrative record dataset consisting of data on employment, unemployment, and benefit reciprocity. Because these data sources can be linked, nonresponse and measurement bias can be estimated. As detailed in Chapter 2, some of the questions asked in PASS may be related to mode differences in nonresponse bias and measurement bias.

3.2.1. Sample design

PASS is a dual frame cluster sample (see Trappmann *et al.*, 2009 for details on sampling methods). The first frame was selected from the IEB. The second frame was selected from a commercial database from the private Microm vendor. These two frames were used to target individuals who were receiving benefits in July 2006 and individuals who were not. Microm cases were not selected from the IEB, although some individuals within that database could be linked using names, addresses, and so on. However, linking the individuals in the Microm sample to IEB would be difficult, and, in some cases, impossible. For this reason, only the first sample, selected from IEB, is included in the analyses in this dissertation.

300 postcodes were selected. 23,812 “benefit communities” were selected within these postcodes from the IEB. A benefit community consists of a person or group of persons who receive benefits. This is similar to a household or a family, although the BA has a distinct definition of a benefit community, while households and families can change over time (Duncan & Hill, 1985; McMillen & Herriot, 1985).

Per BA regulations, one adult within each benefit community is designated as the person in charge of communicating with the BA regarding their benefits, handling BA requirements, and so on. These individuals who act as gatekeepers between the benefit community and the BA were targeted for an interview. If the gatekeeper was unavailable for an extended period of time, another adult was substituted. All of these substitutions were treated as unit nonresponse—all of these survey responses were set to missing.

3.2.2. Data collection methods

PASS utilizes a sequential mixed mode design using CATI and CAPI at each wave. The survey population consists of households with and without working telephone numbers. If a working telephone number was available on the frame or through commercial directories or lists, CATI was attempted. If a sample household or individual within that household initially assigned to CATI was not contacted after a minimum 12 attempts, the case was switched to CAPI data collection¹⁷. The analyses in this dissertation include only the households who were assigned to CATI first, as the other households were only attempted in a single mode. Households who refused the CATI

¹⁷ The timing of the switch to CAPI was largely decided by production staff. Up to 100 contacts via CATI were made in CATI, but the minimum number of contacts in CATI before switching to CAPI was 12. If a telephone number was clearly wrong, sample households were immediately switched to CAPI.

attempt were not switched to CAPI; no further contacts were attempted. These cases are considered nonrespondents and are included in the analyses in this dissertation as such.

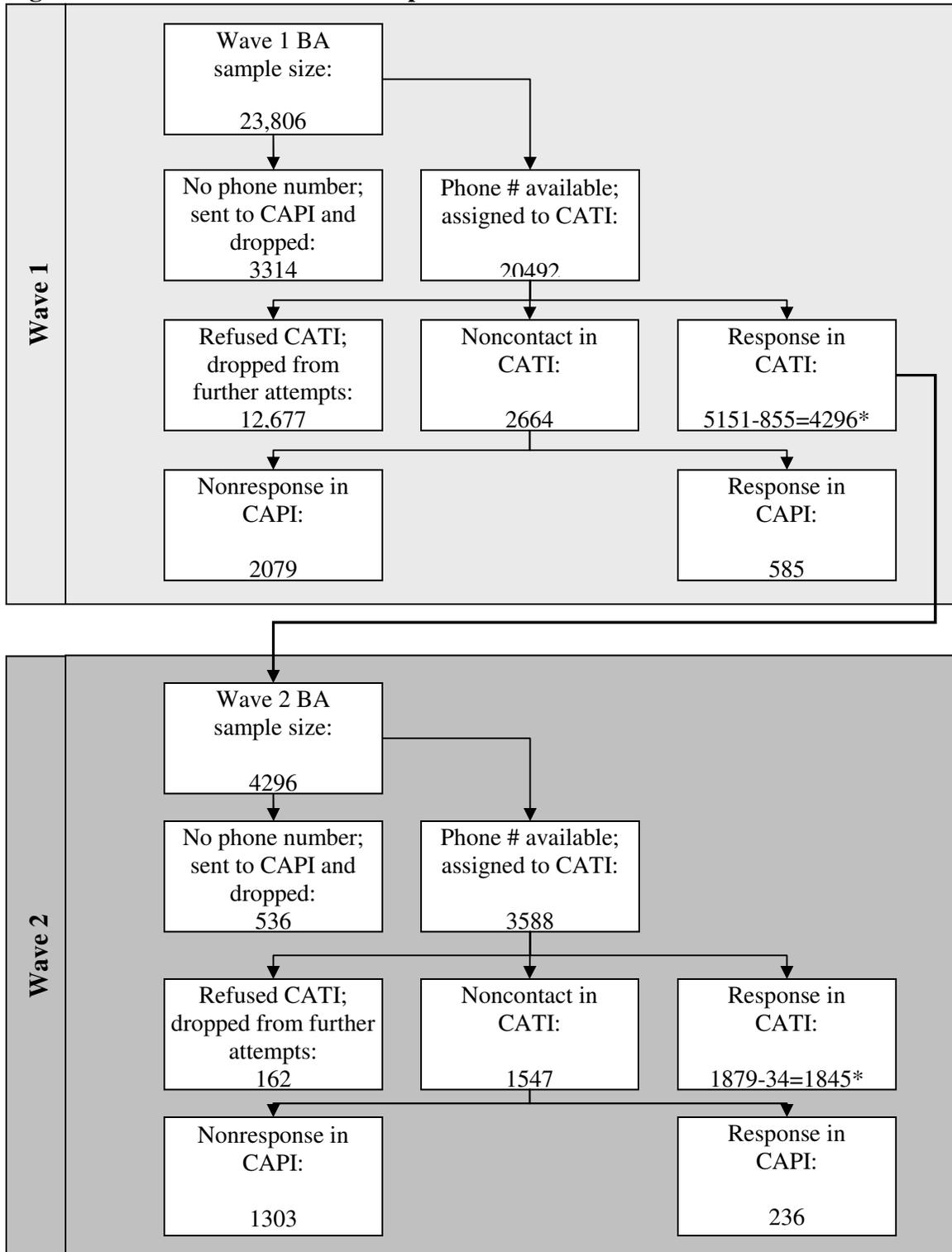
Only Wave 1 households in which at least one person responded were contacted for Wave 2 participation¹⁸; Wave 1 nonresponding households were dropped from further contact attempts after Wave 1. Furthermore, the mode of the first contact attempt was dependent on the mode of response at Wave 1: Wave 1 CAPI respondents were only attempted via CAPI in Wave 2¹⁹ and are also excluded from the analyses in this dissertation. See Figure 6 for the number of households in both waves at each point in the data collection process. This research will only examine outcomes and predictors from Wave 2.

Because this research focuses on the sequential mixed mode design, only benefit communities who were eligible for the sequential design will be retained. Conditional on having a telephone number and being selected from the BA frame, the Wave 1 response rate is 28%. Before the switch to CAPI, the response rate was about 25%. In Wave 2, the response rate is conditional on the gatekeeper's participation in Wave 1 in CATI *and* having a telephone number and a household at Wave 2. The CATI response rate in Wave 2 was 51%; the overall Wave 2 response rate, including both modes, was 58%.

¹⁸ At Wave 2, the sample was refreshed with additional benefit communities. This dissertation drops those cases from all analyses.

¹⁹ The reason for this exclusion was an assumption that CAPI respondents preferred CAPI or were simply unreachable via CATI.

Figure 6. PASS flow chart with sample sizes.



*Cases dropped because interviewed respondents were not the gatekeeper.

Two different questionnaires were administered to the gatekeeper: the household questionnaire and the person questionnaire²⁰. The household interview included the household listing and other characteristics of the household. The person interview asked about respondents' participation in the labor market, family structure, attitudes toward receiving benefits, and characteristics of their household and environment. Although this person questionnaire was administered to all individuals over the age of 14, only the gatekeepers' data will be used in these analyses because it is directly linkable to the IEB. No proxies were used. All person interviews of the gatekeepers were conducted in the same mode as the household interview (CATI or CAPI). For 98% of all responding gatekeepers, the person interview was conducted on the same day as the household interview, with the same interviewer.

In the person questionnaire, respondents were asked if they would consent to linking the IEB data and PASS survey data. In Wave 1, 3819 respondents of the original 4881 CATI and CAPI respondents consented to this link. Once a respondent consented to the link at Wave 1, they implicitly consented to the link at every subsequent wave. Only respondents who had not consented to the link at Wave 1 were asked for consent to link at Wave 2; those who had already consented were not asked this question. In the case of item missing data for this consent question (133 in Wave 1, 68 in Wave 2), respondents were treated as if they did not consent to the link. Table 5 presents the number of respondents who consented in both waves within the analytic sample (i.e., the 3588 individuals who had participated in CATI in Wave 1 and still had a telephone number

²⁰ See http://fdz.iab.de/de/FDZ_Individual_Data/PASS/Working_Tools.aspx for the German questionnaires with English translations.

and household at Wave 2). In Wave 2, consent rates did not vary significantly among modes (Wald F=2.91, p>.05).

Table 5. Association between the consent to link PASS with IEB data in Wave 1 and Wave 2.

				Wave 2		
		No consent	Consent	Item nonresponse	Unit nonresponse	Total
Wave 1	No consent	116	231	18	346	711
	Consent	n/a	1706 [†]	n/a	1038	2744
	Item nonresponse	5	29	18	81	133
	Total	121	1966	36	1465	3588

Note. [†]This includes the 120 respondents who already provided consent in wave 1.

Chapter 6 presents a brief investigation of correlates of consent and its effect on the estimation of measurement bias.

3.2.3. Outcomes and predictors

In order to estimate nonresponse bias, variables from the IEB dataset will be used. The nonresponse bias of certain employment characteristics, household characteristics, and personal characteristics will be investigated using Wave 2 data (see Table 6).

These data come from several administrative sources. First, every employer in Germany is legally obligated to report to the BA all entries and departures of employees from the establishment and the employees' pay. Self-employed persons and civil servants are, however, not included in the IEB. Each spell of unemployment, employment, receipt of benefits, and searches for unemployment using the local labor office is reported.

In order to estimate measurement bias, variables from PASS will be compared with variables from the IEB. See Table 6 and

Table 7 for the list of variables that are available in both datasets.

Three sources of data are used: data from PASS, data from IEB, and data from paradata (e.g., indicators for response, interviewer characteristics, etc). As noted above, consent was needed to link PASS with IEB data. But all individuals—responding or not responding—have IEB data. displays all of these data sources and the number of cases with data in each source.

Figure 7. Data availability summary for IEB, PASS, and paradata.

IEB	Paradata	PASS Wave 1 survey data	PASS Wave 2 survey data	PASS Wave 2 linked with IEB	
					participated in Wave 2 and consented to link (n=1966)
					participated in Wave 2 and did not consent to link (n=157)*
					participated in Wave 1 but did not respond in Wave 2 (n=3588)

*Includes 36 Wave 2 respondents who did not answer the consent question.

See Chapter 2 for a thorough discussion of why these variables were chosen for this analysis. The following variables will be used for an analysis of nonresponse bias and measurement bias:

Gross monthly income was collected in PASS as a continuous variable. Because gross income tends to have large amounts of item missing data (Moore *et al.*, 2000), unfolding brackets were used to reduce item nonresponse (Heeringa *et al.*, 1993). The

corresponding value in IEB is collected as a daily wage, multiplied by 30. This is only available for spells in which the person was employed.

In PASS, respondents were asked their *disability status*: disabled, not disabled, or applying for disability. Those applying for disability are legally considered able-bodied, but they may not be able-bodied during the application process. For the purposes of this study, these individuals are considered disabled²¹. Disability status was thus recoded as a binary variable where 1=disabled or applying for disability and 0=not disabled.

Employment status—being *registered as employed, unemployed, employed full-time, and employed part-time* was asked in PASS. In IEB, being registered as unemployed is collected from the local labor office. All other employment status variables are collected from employers as part of a mandatory reporting process. The PASS estimate of *length of unemployment* was asked of those who are currently unemployed. The corresponding IEB variable is collected from the local labor office.

Data from the local labor office was used to determine if the gatekeeper *lives in former East German states*. This was not collected via PASS; therefore measurement bias cannot be estimated. However, this may be a meaningful predictor of nonresponse.

The IEB also has information on whether the gatekeeper *has children in the benefit community*. In PASS, the gatekeeper was merely asked if she *has children*. Note that measurement bias may occur because IEB refers to the benefit community, while PASS refers to having children generally.

Marital status was collected at each wave in PASS. Although there were multiple categories (single, married, widowed, divorced, civil union), this variable was

²¹ Although this was an arbitrary decision, very few individuals were applying for disability; this decision is unlikely to affect any of the nonresponse bias or measurement bias analyses.

dichotomized into 1=*married* and 0=*not married*. Civil unions were not counted as equivalent to a marriage.

Although gender was collected at Wave 1 in PASS, estimating the measurement bias of such a salient, relatively fixed characteristic is unlikely to be meaningful. Being *female*—a binary variable—is used for nonresponse bias analyses only.

Being a *German national* may also be related to the likelihood of response, as well. This IEB variable was collected as the nationality of the individual. This was recoded into a binary variable indicating being a German national or not.

The PASS variable for *age* was collected at Wave 1 only and generated in subsequent waves. For this reason, measurement bias is not estimated for this variable. For the corresponding IEB variable, age is reported by employers or the local labor office.

The IEB data is used as a gold standard to estimate measurement bias and nonresponse bias. While these administrative records may have their own sources of measurement bias and missing data (Groves, 1989; Miller & Groves, 1985), having any information on possible true values for respondents and nonrespondents is invaluable to answering these questions.

While most IEB data is collected concurrently with either the date of the household interview or the first contact attempt, some data were slightly more outdated. Therefore, the time difference between PASS data collection and IEB collection may mean that differences between PASS and IEB are due to the time difference rather than measurement bias of the survey data. Some of these data, such as the employment data, are collected via a notification process—which is itself subject to some unknown measurement bias.

Some process variables—interviewer age, gender, education, and number of calls made—can be linked to both survey data as well as IEB data, regardless of whether a respondent consented to linking her survey data with IEB data. These paradata can help inform models related to nonresponse and measurement bias.

Outcomes for some measurement bias models are also estimated. Measurement bias is estimated in several different ways. First, the signed difference in y between PASS and IEB variables is estimated:

$$(3.27) \quad \varepsilon_{Si} = y_{i,PASS} - y_{i,IEB}$$

where $y_{i,PASS}$ =the PASS value of y for respondent i and $y_{i,IEB}$ =the IEB value of y for respondent i). The average bias across both modes is also estimated:

$$(3.28) \quad \bar{\varepsilon}_S = \frac{\sum_{i=1}^r \varepsilon_{Si}}{r}$$

where r =the number of respondents in Wave 2=1845 CATI respondents + 236 CAPI respondents=2081.

The average bias within each mode is estimated:

$$(3.29) \quad \bar{\varepsilon}_{S,CATI} = \frac{\sum_{i=1}^{1845} \varepsilon_{Si,CATI}}{r_{CATI}}$$

where r_{CATI} =the number of Wave 2 CATI respondents=1845.

$$(3.30) \quad \bar{\varepsilon}_{S,CAPI} = \frac{\sum_{i=1}^{236} \varepsilon_{Si,CAPI}}{r_{CAPI}}$$

where r_{CAPI} =the number of Wave 2 CAPI respondents=236.

For dichotomous categorical variables, the signed measurement bias is estimated in a similar fashion, although ε_s becomes a categorical variable with values 0 for agreement between PASS and IEB, 1 for an underestimate when $y_{i,IEB} = 1$ and $y_{i,PASS} = 0$, and -1 for an overestimate when $y_{i,IEB} = 0$ and $y_{i,PASS} = 1$. These three-category variables were transformed into two binary variables: an indicator for overestimate (ε_{Vi}), where 0=agreement or underestimate, and an indicator for underestimate (ε_{Ui}), where 0=agreement or overestimate. The average overestimate and underestimate measurement biases are calculated as in (3.28). The within-mode mean overestimate and underestimate measurement biases were estimated as in (3.29) and (3.30).

Absolute measurement bias (ε_{Ai}) is also estimated. For continuous variables:

$$(3.31) \quad \varepsilon_{Ai} = |y_{i,PASS} - y_{i,IEB}| = |\varepsilon_{Si}|$$

For categorical variables, ε_{Ai} = an indicator for *disagreement*. For example, if the IEB data shows that a particular panel respondent is unemployed, but the respondent reported that she is employed, then $\varepsilon_{Ai} = 1$. But if unemployment status on PASS agrees with unemployment status in IEB, then $\varepsilon_{Ai} = 0$. Table 6 contains means, standard errors, and item missing data rates for IEB data in Wave 2;

Table 7 includes means, standard errors, and missing data rates for PASS survey data. All standard errors were adjusted for the complex sample design using STATA's *svy* command and for multiple imputation using STATA's *mim* command.

Table 6. Wave 2 IEB variables used for nonresponse bias and measurement bias analyses.

Variable	Analysis [†]	Mean (std. err.)	Item missing data (%)
<i>Employment</i>			
Gross monthly income	NR, MB	1021.148 (52.331)	0
Currently registered as disabled	NR, MB	0.161 (0.044)	21
Registered as employed	NR, MB	0.349 (0.009)	1
Employed full-time	NR, MB	0.232 (0.008)	1
Employed part-time	NR, MB	0.129 (0.006)	1
Registered as unemployed	NR, MB	0.837 (0.007)	1
Length of unemployment	NR, MB	331.168 (8.318)	3
<i>Household characteristics</i>			
Lives in former East German states	NR	0.292 (0.009)	1
Has children in the benefit community	NR, MB	0.239 (0.010)	4
<i>Person characteristics</i>			
Marital status	NR, MB	0.158 (0.007)	3
Female	NR	0.469 (0.010)	1
German national	NR	0.928 (0.005)	1
Age (in years)	NR	43.341 (0.315)	5

[†]NR=Nonresponse bias analysis; MB=measurement bias analysis.

Table 7. Wave 2 PASS survey data: descriptives.

Variable	Mean (std. err.)	Item missing data
<i>Employment</i>		
	1581.132	27.65%*
Gross monthly income	(54.644)	
	0.148	3.20%
Currently registered as disabled	(0.009)	
	0.538	0.38%
Registered as employed	(0.01)	
	0.729	0.38%
Employed full-time	(0.022)	
	0.271	0.38%
Employed part-time	(0.022)	
	0.647	7.07%
Registered as unemployed	(0.012)	
	27.684	8.23%
Length of unemployment	(1.046)	
<i>Household characteristics</i>		
	0.558	3.39%
Has children in the benefit community	(0.010)	
<i>Person characteristics</i>		
	0.165	3.58%
Marital status	(0.008)	

*An additional 416 cases responded to the categorical income question. However, these cases were treated as item missing data although the categorical data were available.

3.2.4. Linking procedures

Because of the sensitive nature of the IEB data, all data analysis was conducted at the Research Data Center (Forschungsdatenzentrum: FDZ) at the BA in Nuremberg.

First, the PASS sample cases were selected from questionnaire data and process data. Only the gatekeepers' questionnaire data were retained. Households that had participated in CATI at Wave 1, that were part of the BA sample, and that still had a phone number and household at Wave 2 were kept in the dataset. All other cases—households from the Microm sample (13,340), BA households with no telephone number at Wave 1 (3314), those who did not respond at Wave 1 (14,756), and those who no longer had a telephone or a household at the beginning of the Wave 2 field period

(536)—were dropped. A total of 3588 households were retained and 31,946 were dropped.

Data from the household interviews were merged with data from the person interviews for each wave, creating a “wide” file with 3588 cases and variables for each wave. That is, each panel respondent had one observation, Wave 1 variables, and Wave 2 variables.

Because some respondents did not consent to the link between PASS survey data and IEB data, all PASS survey data of nonconsenting respondents were set to missing. These data were later imputed (see below). This was a better alternative to dropping those cases from analysis for a number of reasons: First, the cell sizes are already quite small, in some cases; only 236 gatekeepers responded via CAPI. Second, we have a great deal of information on these nonconsenters—data from the IEB for waves 1 and 2 as well as data from PASS for Wave 1. A weighting scheme could have been used, such as propensity weighting; however, we have a great deal of nonmissing auxiliary data available on the nonconsenting respondents, and weighting tends to underutilize these data (Raghunathan, 2004).

Next, the survey data of the restricted sample were linked to the IEB. The IEB data contains information for each gatekeeper and individuals in her benefit community. But, as briefly mentioned above, households do not directly correspond to benefit communities. For example, a relative may live in the household but is not receiving benefits at all. Or, individuals in the same benefit community may be living separately. Because of this discrepancy, we may have records in the IEB data that do not correspond with survey data. And we may have PASS survey data that does not have a corresponding

record in IEB. However, we can identify each gatekeeper in the survey data and the IEB. Because of this feature of the PASS design, all analyses are restricted to the gatekeeper—all other householders are dropped from the analysis.

Each gatekeeper could have many spells of unemployment, employment, or benefit reciprocity, each of which exists in the IEB. One spell had to be selected for each gatekeeper at each wave for the purpose of this analysis. For respondents, the spell that includes the date of interview was selected. But for nonrespondents, the selection of a particular spell is not as straightforward. As a rule, the spell that overlapped the most with the field period—the time between the first PASS contact attempt and the last—was selected. If there was no temporal overlap between an IEB spell and the field period, then the spell closest to the beginning of the field period will be selected.

In some cases, multiple spells could have been selected—for example, if a household was unemployed *and* receiving some kind of family benefit at the same time. Two different spells would have the same amount of overlap with the field period. In such cases, data from multiple spells of IEB data were combined. These spells were selected for both waves, and then restructured into a wide format so that each gatekeeper selected has one observation and variables for each wave.

The final dataset consists of one observation per panel respondent. Variables included PASS survey data for both waves, paradata for both waves, and IEB variables for both waves that coincided per rules discussed above. This contrasts with the RDSL dataset, in which each case was a time point for a particular panel respondent. In RDSL, the dataset has a “long” format, but in PASS, the dataset has a “wide” format.

3.2.5. Eligible panel respondents

As discussed above, households that are included in the analyses in this dissertation must meet several conditions: First, they must have been selected from the BA list. Second, they must have completed Wave 1 in CATI. And third, they must still have a household and a working telephone number at Wave 2.

A total of 3588 gatekeepers met these criteria. Of those, about 47% of the gatekeepers are female. The average age is 42.3. 93% are German nationals. 50% of Wave 2 CATI respondents are female, while only 45% of Wave 2 CAPI respondents are female. In Wave 2, the mean age of CATI respondents is 44; in CAPI, the mean age is 39.

3.2.6. Item missing data

Although we have no missing paradata, PASS survey data and IEB data may have some item missing data. See Appendix A, Tables A.3 and A.5 for item missing data rates in PASS and IEB. Data may be missing from PASS because the respondent refused to respond to a particular question, the respondent did not know the answer to the question, or the question was not administered by the interviewer by mistake.

For the IEB data, most of the data are nonmissing, although some data are only available for specific types of spells. For example, the number of children is only available for individuals who have received unemployment or a maintenance allowance during that spell (Bergbauer *et al.*, 2010). Note that the missing data rate for disability status is quite high (21%); this is because disability status is only collected in spells of looking for work or taking part in labor market programs (Bergbauer *et al.*, 2010).

3.2.7. Imputation procedures

For both waves, three kinds of data were imputed: Item nonresponse in PASS, all PASS data for respondents who did not consent to the link between IEB and PASS, and item missing data in IEB. Unlike RDSL, missing data due to unit nonresponse was *not* imputed. Although the analysis focuses on Wave 2 panel respondents, Wave 1 data were retained for the imputation models because of the likely strong correlations between wave data.

Sequential regression multiple imputation was used to eliminate item missing data. As discussed above, SRMI uses data available on all cases to impute for missing cases; the imputed values from those missing cases are then used for other imputation models for other variables. Because data from both waves for PASS, IEB, and paradata were included, these variables may have been used for the sequential regression imputation models. IVEWARE was used for the imputation procedures.

As in the case of RDSL, the fraction of missing information was estimated for each variable imputed to determine if the ratio of between-imputation variance to total variance was reasonably low (see Appendix A, Table A.6. All seem to be quite low. This is likely because information from Wave 1 was used in the imputation models.

3.2.8. Analytic models

In Chapter 6, nonresponse bias is estimated before the switch to CAPI and after. Regression models are used to estimate whether the overall nonresponse bias is significantly different from zero. A separate set of regression models will be used to determine whether the change in nonresponse bias, from CATI to CAPI, is significant.

We could estimate the nonresponse bias in several different ways: (1) the signed nonresponse bias $(\bar{y}_r - \bar{y}_n)$; (2) the absolute nonresponse bias $|\bar{y}_r - \bar{y}_n|$; (3) the relative nonresponse bias $\frac{\bar{y}_r - \bar{y}_n}{\bar{y}_n}$; and (4) the absolute relative nonresponse bias $\frac{|\bar{y}_r - \bar{y}_n|}{\bar{y}_n}$, where \bar{y}_r = the mean of a particular estimate of the variable y for respondents and \bar{y}_n = the mean of y for all sample cases. The absolute estimates of bias ignore the direction of the bias, and the relative bias measures account for the scale of y .

A regression model is estimated to test whether the nonresponse bias of y is significantly different from zero:

$$(3.32) \quad y_i = \beta_0 + \beta_1 r_i + \varepsilon_i$$

where r_i = an indicator for nonresponse. Nonresponse bias models for gross monthly income, length of unemployment, and age were run as linear regression models. All other models were logistic regression models because the remaining y variables are binary; in

those cases, the dependent variable is $\ln\left(\frac{p(y_i)}{1-p(y_i)}\right)$.

In model (3.32) is identical to model (3.17) in the RDSL analysis. β_1 is the mean difference in y between respondents and nonrespondents $(\bar{y}_r - \bar{y}_m)$. Because the

nonresponse bias = $\frac{m}{n}(\bar{y}_r - \bar{y}_m)$, if $\frac{m}{n}\beta_1 = 0$, then the nonresponse bias of $y=0$. As in

RDSL, we can test the contrast:

$$(3.33) \quad \theta = \frac{m}{n}\beta_1$$

If $\theta=0$, then the nonresponse bias of y is 0.

The analysis in Chapter 6 will also test whether the nonresponse bias changes across modes. In RDSL—a web→ CATI sequential mixed-mode design—the data were imputed twice. One set of imputations treated all CATI respondents as nonrespondents, imputing item nonresponse and all variables for CATI respondents and nonrespondents (Treatment 1), and a second set imputed item nonresponse and all variables for nonrespondents (Treatment 2). Treatment 1 simulates a study that uses just the first mode, while Treatment 2 represents the sequential design. This study will take a similar approach.

For RDSL, two sets of imputation were completed because the estimates of the true values were generated via the imputation methods. In this chapter, however, we have record data on all sample cases, which serve as proxies for the true values. Therefore, the sequential regression multiple imputation procedure only needs to take place one time to eliminate item nonresponse issues in the IEB record data. Instead of running two separate sets of imputations, the data were duplicated, creating a dataset that includes all sample cases twice, differentiating the duplicates by Treatment. See Figure 8.

Figure 8. Simulated experimental design.

Treatment 1 $[T_i = 1]$	Treatment 2 $[T_i = 0]$
CATI respondents $\begin{bmatrix} G_{1i} = 1 \\ G_{2i} = 0 \end{bmatrix}$	CATI respondents and CAPI respondents $\begin{bmatrix} G_{1i} = 0 \\ G_{2i} = 1 \end{bmatrix}$
CAPI respondents and nonrespondents $\begin{bmatrix} G_{1i} = 0 \\ G_{2i} = 0 \end{bmatrix}$	Nonrespondents $\begin{bmatrix} G_{1i} = 0 \\ G_{2i} = 0 \end{bmatrix}$

As with RDSL, the change in nonresponse bias between CATI and CAPI can be estimated as:

$$(3.34) \quad \delta = (\bar{y}_{r1} - \bar{y}_{n1}) - (\bar{y}_{r2} - \bar{y}_{n2})$$

where \bar{y}_{r1} = the mean of y for respondents in Treatment 1 (i.e., CATI respondents only), \bar{y}_{n1} = the mean of y for all cases in Treatment 1, \bar{y}_{r2} = the mean of y for respondents in Treatment 2 (i.e., all CATI and CAPI respondents), and \bar{y}_{n2} = the mean of y for all cases in Treatment 2. In RDSL, the different imputations used for the two treatment groups might result in different \bar{y}_n values; here, $\bar{y}_{n1} = \bar{y}_{n2}$. We therefore only need to test whether $\bar{y}_{r1} - \bar{y}_{r2} = 0$.

To test this hypothesis, the following regression model is run for each IEB variable:

$$(3.35) \quad y_i = \beta_0 + \beta_1 T_i + \beta_2 G_{1i} + \beta_3 G_{2i}$$

where T_i = an indicator for Treatment 1; G_{1i} = an indicator for being a respondent under Treatment 1; and G_{2i} = an indicator for being a respondent under Treatment 2. As above, models for gross monthly income, length of unemployment, and age are run as linear regression models, while all other models are logistic regression models. The contrast to test if $\delta = 0$ is:

$$(3.36) \quad \theta_D = \beta_1 + \beta_2 - \beta_3$$

As above, if $\theta_D = 0$, then $\delta = \bar{y}_{r1} - \bar{y}_{r2} = 0$. In this case, the sequential design has no significant effect on nonresponse bias.

If the mixed-mode design was concurrent—that is, one group of respondents was randomly assigned to CATI and another group was randomly assigned to CAPI—then a

simple regression model could assess whether the measurement bias changed across modes:

$$(3.37) \quad ME(y_i) = \beta_0 + \beta_1 CAPI_i + \varepsilon_i$$

However, mode was not randomly assigned. We cannot assume that all differences between CATI and CAPI with respect to measurement bias are due to the mode; it is possible that this measurement bias differs because of the selection mechanism of the sequential design. This selection mechanism is controlled for using propensity modeling.

A logistic regression model is used to predict the likelihood of response in CAPI versus response in CATI. This model accounts for the complex design of PASS as well as the imputation using STATA's *svy* and *mim* commands. Predictors include variables from the IEB, PASS, and paradata thought to be related to mode propensity. Variables that are significantly related to response in CATI (compared to response in CAPI) were added to (3.37):

$$(3.38) \quad ME(y_i) = \beta_0 + \beta_1 CAPI_i + \beta_2 DISAB_i + \beta_3 IAGE_i + \beta_4 NUMFF_i + \varepsilon_i$$

where $DISAB_i$ = disability status from, the IEB data; $IAGE_i$ = age of the household interviewer; and $NUMFF_i$ = the number of family and friends reported by the respondent in Wave 1.

Finally, a relationship may exist between nonresponse and measurement bias. Perhaps those who are more likely to respond are generally better respondents—i.e., they have lower levels of measurement bias. And perhaps this relationship differs across modes in a sequential design.

First, let us include all cases and neglect the issue of mode. A logistic regression model is used to estimate the likelihood of response, compared to nonresponse. The covariates in this model include variables from IEB, Wave 1 PASS, and paradata thought to be related to response—that is, all available variables from all three data sources that were not variables of interest. It is more important to have a good fit in the propensity model than a parsimonious model; therefore, covariates were added one at a time and kept if the goodness of fit—determined by the *svylogitgof* command—increased. After adding and removing covariates, the following model was fit:

$$(3.39) \quad \ln\left(\frac{p_{ri}}{1-p_{ri}}\right) = \beta_0 + \beta_1 FEMALE_i + \beta_2 AGE_i + \beta_3 LENFIELD_{1i}$$

where p_{ri} = the probability of response; $FEMALE_i$ = an indicator for being female, as indicated by the IEB data; AGE_i = the age of the individual, as collected in PASS, Wave 1; and $LENFIELD_{1i}$ = the length of the field period, in days, for Wave 1.

A second propensity model is estimated, predicting the likelihood of participation in CATI, compared to not participating in CATI (i.e., participating in CAPI or not participating at all). The *svylogitgof* command is used, again, to determine which covariates created a model that was the most predictive of response in CATI (Archer & Lemeshow, 2006). The covariates in this model are:

(3.40)

$$\ln\left(\frac{p_{CATIi}}{1-p_{CATIi}}\right) = \beta_0 + \beta_1 FEMALE_i + \beta_2 AGE_i + \beta_3 RELIGACTIV_{1i} + \beta_4 SOCCLASS_{1i} + \beta_5 SOCISOL_{1i}$$

where p_{CATIi} = the probability of response in CATI; $RELIGACTIV_{1i}$ = an indicator for being active in a religious organization in Wave 1, collected in PASS; $SOCCLASS_{1i}$ = a rating of

what social class the respondent felt she was in at Wave 1, in PASS; and $SOCISOL_{1i}$ = a rating of how socially isolated the respondent feels, collected in Wave 1 PASS.

A third response propensity model is generated, limited to cases with at least one CAPI contact—that is, CATI nonrespondents. The covariates are selected in a similar fashion. The final model predicting the likelihood of responding in CAPI, compared to not responding at all—conditional on nonresponse to CATI—was:

(3.41)

$$\ln\left(\frac{p_{CAPIi}}{1-p_{CAPIi}}\right) = \beta_0 + \beta_1 FEMALE_i + \beta_2 EMPLOYED_{1i} + \beta_3 LENFIELD_{1i} + \beta_4 ACTIV_{1i} + \beta_5 RELIGACTIV_{1i} + \beta_6 SOCISOL_{1i}$$

where p_{CAPIi} = the probability of participation in CAPI, conditional on nonresponse to CATI; $EMPLOYED_{1i}$ = an indicator for being employed at Wave 1, provided in IEB; and $ACTIV_{1i}$ = an indicator for being active in any non-religious organization at Wave 1, collected in PASS.

See Table 8 for the coefficients and standard errors of each model. For the CAPI model (3.41), none of the covariates are significant predictors of response in CAPI; however, the fit is the best out of all reasonable combinations of covariates. It is likely that these predictors are not significant because the number of CAPI respondents is quite small.

Table 8. Response propensity models for participating in CATI, CAPI, and in either mode.

CATI response		CAPI response		Overall response	
Predictor	β (Std. Err.)	Predictor	β (Std. Err.)	Predictor	β (Std. Err.)
<i>FEMALE</i>	1.336(0.106)***	<i>FEMALE</i>	1.077(0.177)	<i>FEMALE</i>	1.330(0.107)***
<i>AGE</i>	1.012(0.005)*	<i>EMPLOYED</i>	0.840(0.169)	<i>AGE</i>	1.005(0.005)
<i>RELIGACTIV</i>	0.951(0.208)	<i>LENFIELD</i>	0.998(0.002)	<i>LENFIELD</i>	0.996(0.001)***
<i>SOCCLASS</i>	0.993(0.071)	<i>ACTIV</i>	0.961(0.278)		
<i>SOCISOL</i>	1.011(0.029)	<i>RELIGACTIV</i>	1.063(0.290)		
		<i>SOCISOL</i>	0.974(0.039)		

Regression models are used to estimate if the likelihood of response was related to measurement bias. First, predicted propensities of models (3.39), (3.40), and (3.41) were estimated. These propensities became dependent variables in a series of three models per y variable:

$$(3.42) \quad \ln[E(\hat{p}_{Ti})] = \beta_0 + \beta_1 ME_Z(y_i) + \varepsilon_i$$

where \hat{p}_{Ti} = the predicted propensity for model T —that is, the propensity model for CATI (3.40), CAPI (3.41), or across modes (3.39). The measurement bias term could be signed or absolute.

Chapter 4

Attrition Propensity and Mode Propensity in a Multimode Panel Survey

4.1. Introduction

A sequential mixed mode design in a cross-sectional survey often increases response rates. But in a panel survey, a panel respondent may experience the sequential design multiple times. This study will address the impact of time in sample on response propensity, lateness of response, and mode of response using data from RDSL.

As detailed in Chapter 2, indicators of privacy concerns, topic salience, and education are expected to predict response propensity. Section 2.4.1. described which RDSL variables should be positively related to response propensity under each assumed mechanism and which should be negatively related. These hypotheses are summarized in Table 9.

To date, little research has investigated if the effect of predictors of response propensity changes across observations in a panel survey. As described in Section 2.4.1, we could simultaneously expect that these predictors have a constant effect on response propensity across journal periods or an increasing effect on response propensity across journal periods. This study is exploratory in nature and tests these possibilities.

Table 9. Expected directions of the effects of predictors of response under three mechanisms.

Predictor	Education/school enrollment	Privacy concerns	Topic interest
<i>Privacy concerns and topic salience indicators</i>			
Having any sexual partners	n/a	negative	Positive
Use of non-coital contraception	n/a	negative	Positive
Use of coital contraception	n/a	negative	Positive
Living with a parent	n/a	negative	n/a
Pregnancy intention	n/a	negative	Positive
Pregnancy avoidance	n/a	positive	Negative
Change in relationship status	n/a	n/a	Positive
<i>Education status indicators</i>			
Currently enrolled in school			
Not enrolled	(reference)	n/a	n/a
Part-time enrollment	positive		
Full-time enrollment	positive		
Type of school (if enrolled)			
High school or less	(reference)		
2 year junior/community college	positive	n/a	n/a
4 year college	positive		
Voc, tech, trade, or other school	positive		

Second, the lateness of response will be examined. If a panel respondent is late enough to receive a CATI prompt at one journal period, does the likelihood of being late at subsequent journals increase? Late respondents may be perpetually late.

Lateness of response may also be affected by indicators of topic interest. Voogt and Saris (2005) found that voters were more likely to participate and participate early than nonvoters in an election study. They believed that topic salience drove both response propensity and the propensity to be late.

But in a panel survey, previous late responses and contemporaneous indicators of topic salience may not be independent effects. As described in Chapter 2, the indicators of topic salience may have a mediating effect on the likelihood of lateness of response.

As in the case of lateness of response, panel respondents may repeat their behavior and tend to participate in one particular mode. If panel respondents tend towards one particular mode across journal periods, then maybe it will be less cost-effective to include a sequential design in later journal periods.

Indicators of topic interest and privacy concerns may be related to the likelihood of participating in a mode. Individuals who have more privacy concerns should be more likely to participate via web compared to CATI, and this may have a mediating effect on the relationship between current mode of response and history of mode of response

Four general hypotheses will be tested in this study. 1) The characteristics of panel respondents are expected to influence the likelihood of response at any given journal period as specified in Table 9. 2) The effects of these characteristics are expected to change across journal periods. 3) The likelihood of being late at any given journal period is positively related to the number of previous late journal periods and is mediated by topic salience. 4) The likelihood of using a CATI mode (inbound or outbound) is positively related to the number of previous journals completed using that CATI mode and is mediated by privacy concerns as well as topic salience.

4.2. Results

4.2.1. Propensity to respond

As detailed in Chapter 3, a series of logistic regression models estimated the effect of the predictors on the likelihood of response. See Table 10 for the odds ratios and standard errors of each bivariate model (3.6) .

Note that Table 10 includes a single model for each group of predictors (e.g., Currently enrolled in school, Having any sexual partners, etc.). The “Baseline” models predict response at any journal period; the predictor (e.g., having a sex partner) was collected at baseline and is treated as a time-invariant covariate; the “All journals” models predict response at any journal period; the predictor here is collected at every journal period (i.e., a time-variant covariate)²². The “Baseline” models and the “All journals” models differ only in the predictor.

The overall propensity to respond, estimated by the bivariate models, is influenced by a number of variables. Being enrolled full-time in school at baseline, compared to not being enrolled, is positively related to the likelihood of participation. But being enrolled part-time at any post-baseline journal is *negatively* related to participation, compared to not being enrolled.

For panel respondents who are students, the type of school is sometimes related the likelihood of response. Students who are enrolled at a four-year college at baseline are more likely to participate than students who are enrolled in high school or less at baseline. Being enrolled in either a 2 year junior/community college or a 4-year college results in a greater likelihood of participation, compared to being enrolled in high school or less.

Being sexually active at baseline (having a sex partner) is negatively related to participation, although having a sex partner at journal period j is unrelated to the

²² Tables X and Y in Appendix A are structured in this way as well.

likelihood of response at journal period j . Living with a parent at baseline is positively related to participation. And a change in relationship status is negatively related to participation.

Pregnancy intentions and avoidance at post-baseline journals are both related to the likelihood of response. Women who have intentions to become pregnant are much less likely to participate than women who do not have pregnancy intentions. And women who wish to avoid pregnancy are much more likely to participate than women who do not wish to avoid pregnancy. It should be noted that pregnancy avoidance has very little variance: about 98% wanted to avoid pregnancy. This small amount of variance likely contributes to the large standard error in the bivariate models.

Response propensity is also driven, in part, by time in sample. Journal period is positively related to the likelihood of response (OR=1.012; std. err.=0.005, $p<.001$). the length of previous journal periods is negatively related to the likelihood of response (OR=0.713, std. err.=0.011, $p<.001$). That is, as the number of exposures to the survey design increases, the likelihood of response increases by about 1%, controlling for the time it had taken the panel respondent to respond in previous journal periods. And the panel respondents' history of being early or late also affects the likelihood of response; as the mean number of days within journal period (from journal period 2 to journal period $j-1$) increases, the likelihood of participating in journal period j decreases by about 29%. However, the estimate of interest is the coefficient for journal period. While the negative coefficient for the length of previous journal periods is significant, it is not central to the research questions in this paper. The research questions here address the effects of

indicators of privacy concerns, topic salience, and education status on response propensity.

Table 10 also includes the odds ratios and standard errors of each of the multivariate main effects models, which estimates the effect of each of the predictors, controlling for time in sample and the mean length of previous journal periods. As shown, some of the predictors are related to the likelihood of response. Having a sexual partner, having pregnancy intentions, having a change in relationship status, being in school part-time, and being enrolled in a vocational, technical, or other kind of school are all negatively related to the likelihood of response, controlling for time in sample and the length of previous journal periods. And wanting to avoid pregnancy, being enrolled in school full-time at the baseline interview, and being enrolled in a four-year college are all positively related to the likelihood of response, controlling for time in sample and the length of previous journal periods.

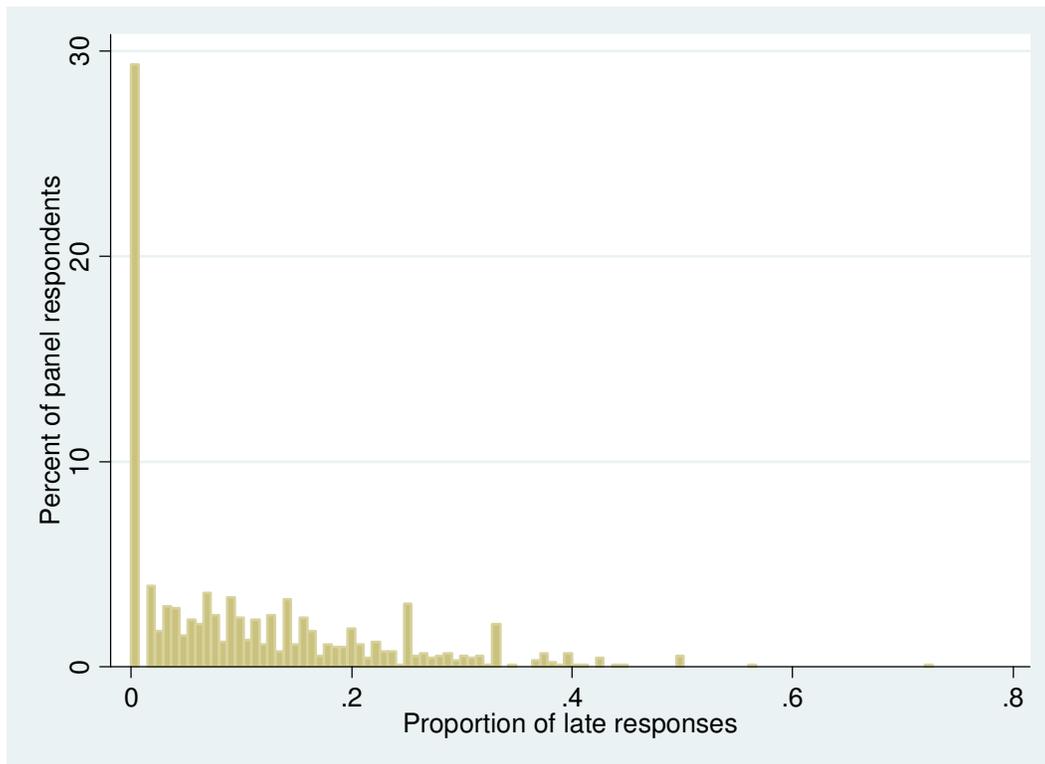
Although response propensity increases across journal periods, and some of the indicators of privacy concerns, topic salience, and education status predict response regardless of time in sample, does the effect of the predictors on the likelihood of response change over time? In general, this is not the case. None of the interactions between *JP* and the predictor (*X*) are significantly different from zero. However, use of coital contraception is marginally related to the change in propensity across journal periods. As journal periods increase, panel respondents who are *not* using coital contraception become marginally more likely to respond than panel respondents who are using contraception. The difference in response propensity widens across journal periods

between those who are not using coital contraception and those who are. See Appendix A, Table A. 9.

4.2.2. Lateness of response

The panel respondents in RDSL were not only unusually cooperative, but they also were unusually prompt. As shown in Figure 9, about 30% of panel respondents were never late for a journal, and about 20% of panel respondents were late for less than 10% of their responding journal periods²³.

Figure 9. Distribution of late responses.



²³ The pattern of lateness was slightly different for replicate 4 than the first three replicates; about 60% of panel respondents in replicate 4 were never late. This is likely happening because data collection was not complete for this replicate.

Table 10. Odds ratios and standard errors for response propensity models.

		Baseline		All journals		
		Bivariate model	Multivariate main effects model	Bivariate model	Multivariate main effects model	
<i>Privacy concerns and topic salience indicators:</i>						
66	Having any sexual partners	Sex partner <i>MNLEN</i> _{ij-1} JP	0.375(0.101)**	0.612(0.138)* 0.726(0.017)*** 1.022(0.007)**	0.493(0.216)	0.574(0.246) 0.714(0.011)*** 1.019(0.005)*
	Use of noncoital contraception	Use of noncoital contraception <i>MNLEN</i> _{ij-1} JP	1.330(0.274)	1.286(0.235) 0.723(0.017)*** 1.024(0.007)**	1.092(0.275)	0.965(0.269) 0.720(0.014)*** 1.026(0.028)
	Use of coital contraception	Use of coital contraception <i>MNLEN</i> _{ij-1} JP			0.623(0.155)	0.621(0.149) 0.723(0.021)*** 1.008(0.011)
	Living with a parent	Living with a parent <i>MNLEN</i> _{ij-1} JP	1.324(0.186)*	1.160(0.127) 0.713(0.011)*** 1.019(0.005)***		
	Pregnancy intention	Pregnancy intention <i>MNLEN</i> _{ij-1} JP	0.899(0.105)	0.905(0.095) 0.713(0.011)*** 1.020(0.005)***	0.039(0.010)***	0.040(0.013)*** 0.716(0.013)*** 1.020(0.006)*
	Pregnancy avoidance	Pregnancy avoidance <i>MNLEN</i> _{ij-1} JP	1.207(0.817)	1.639(0.780) 0.713(0.011)*** 1.020(0.005)***	4.577(2.459)*	4.436(2.348)* 0.713(0.011)*** 1.019(0.005)**

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

Table 10. Odds ratios and standard errors for response propensity models. (contd.)

		Baseline		All journals	
		Bivariate models	Multivariate main effects models	Bivariate models	Multivariate main effects models
Change in relationship status	Relationship change			0.101(0.020)***	0.127(0.0206)***
	$MNLEN_{ij-1}$				0.730(0.0150)***
	JP				1.021(0.0122)
<i>Education status:</i>					
Currently enrolled in school	Part-time enrollment	1.515(0.419)	1.352(0.289)	0.333(0.085)*	0.361(0.0914)*
	Full-time enrollment	1.837(0.361)**	1.544(0.253)*	3.963(2.103)	3.682(1.893)
	$MNLEN_{ij-1}$		0.724(0.017)***		0.718(0.010)***
	JP		1.021(0.007)**		1.017(0.005)*
Type of school	2 year junior/community college	1.446(0.437)	1.380(0.364)	3.125(1.374)	2.580(1.436)
	4 year college	2.007(0.617)*	1.853(0.572)	7.668(3.189)**	5.529(2.646)*
	Voc, tech, trade, or other school	1.409(0.652)	0.939(0.304)	0.315(0.170)	0.264(0.120)*
	$MNLEN_{ij-1}$		0.721(0.020)***		0.709(0.015)***
	JP		1.019(0.007)*		1.013(0.010)

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

As shown in Table 11, the likelihood of responding late at journal period j increases as the number of previous late journals increases, controlling for journal period and the length of previous journal periods. Being late in the past means that the panel respondent is likely to be late in the present journal period.

The interaction term in Model (3.5) is significantly less than one; the change in the effect of having a history of lateness on the likelihood of being a late respondent at the current journal across journal periods increases. That is, as journal periods increase, the effect of having a history of being late on current lateness increases.

Table 11. Odds ratios and standard errors of lateness models (3.4) and (3.5).

	Model (3.4)	Model (3.5)
Covariate	Odds Ratio (Std. Err.)	Odds Ratio (Std. Err.)
$LATEJ_{ij-1}$	1.333(0.021)***	1.583(0.044)***
JP_{ij}	0.983(0.002)***	0.991(0.003)**
$MNLEN_{ij-1}$	1.020(0.006)***	1.012(0.006)*
$LATEJ_{ij-1}JP_{ij}$		0.994(0.001)***

Note. Reference category is responding on time.

†p<.1; ***p<.000

Next, the predictors we used in the response propensity models were added to model (3.5); if the panel respondents' past behavior is the primary driver of current late response, then the covariates that were in model (3.5) should be the only significant predictors of current late response. And if X is a mediator between past lateness and current lateness, then the effect of $LATEJ_{ij-1}$ should disappear when X is included in the model.

As shown in Appendix A, table Table A. 1, past behavior explains most of the variation in current lateness of response. However, being enrolled in school full-time at baseline, being enrolled in school full-time, having a sexual partner at baseline, and

relationship change, are significant predictors of lateness in the bivariate models.

Coefficients for $LATEJ_{ij-1}$, JP_{ij} , and $LATEJ_{ij-1}JP_{ij}$ are roughly the same as in Table 11.

4.2.3. Propensity to use a mode

The majority of responses (mean=89%) in all journal periods are by web, while a small proportion are in CATI (6% in inbound CATI, 8% in outbound CATI). About 33% of panel respondents used web exclusively, most (73%) never used inbound CATI, and most (63%) never used outbound CATI (see Figure 11). A total of 20% of panel respondents used all three modes during the study²⁴.

In Figure 10, we see that panel respondents seem to increasingly move to web across journal periods. But perhaps panel respondents who have many journal periods tend to be faster respondents and tend to use web more often. On the other hand, panel respondents may simply become more efficient across journal periods; perhaps a telephone call from an interviewer helps encourage early responses in later journal periods. This section will address whether prior use of a particular mode predicts the current use of a mode in a given journal period.

As shown in Table 12, when treating mode as a binary variable, the use of CATI decreases significantly across journal periods. When we treat inbound CATI and outbound CATI as two separate modes, we see the same pattern for both inbound and outbound CATI: as journal periods increase, the relative risk of participating in inbound CATI decreases by 2% and the relative risk of participating in outbound CATI decreases by about 5%.

²⁴ As in the lateness of response, this pattern is slightly different among replicates; in replicate 4, 54% of panel respondents used web exclusively, while only 24% did so in replicate 1. Again, it is likely that this occurred because data collection in the later replicates did not finish at the time of this analysis.

Figure 10. Proportion of completes in each mode.

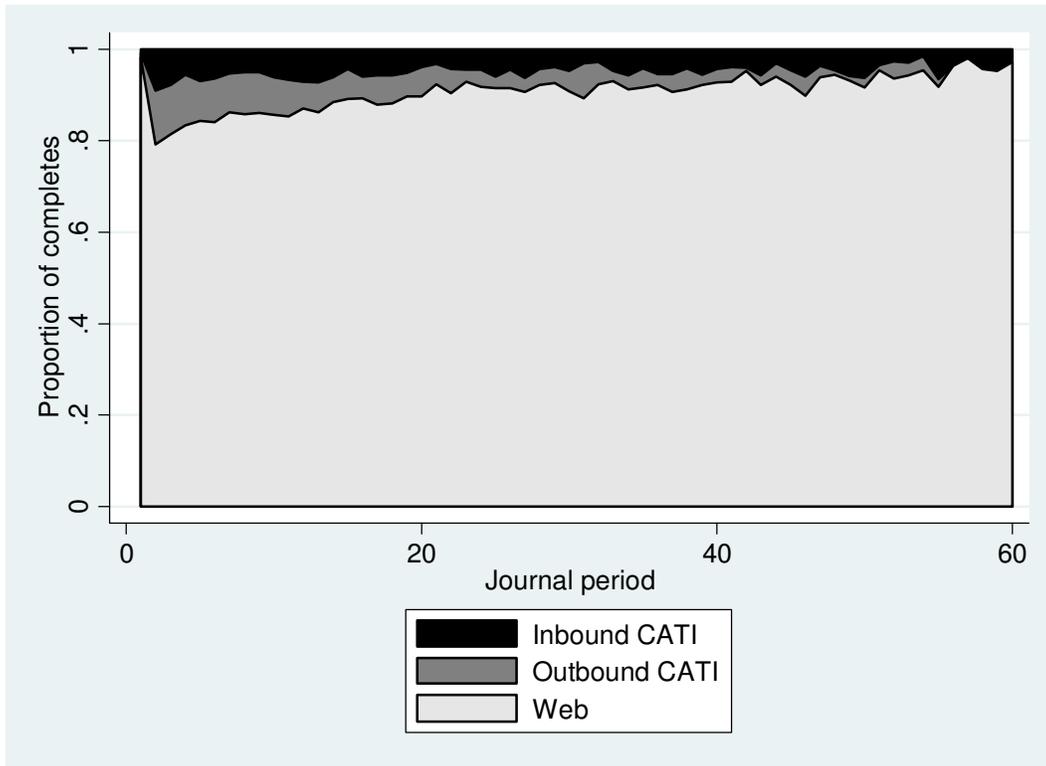


Figure 11. Distribution of mode responses between panel respondents.

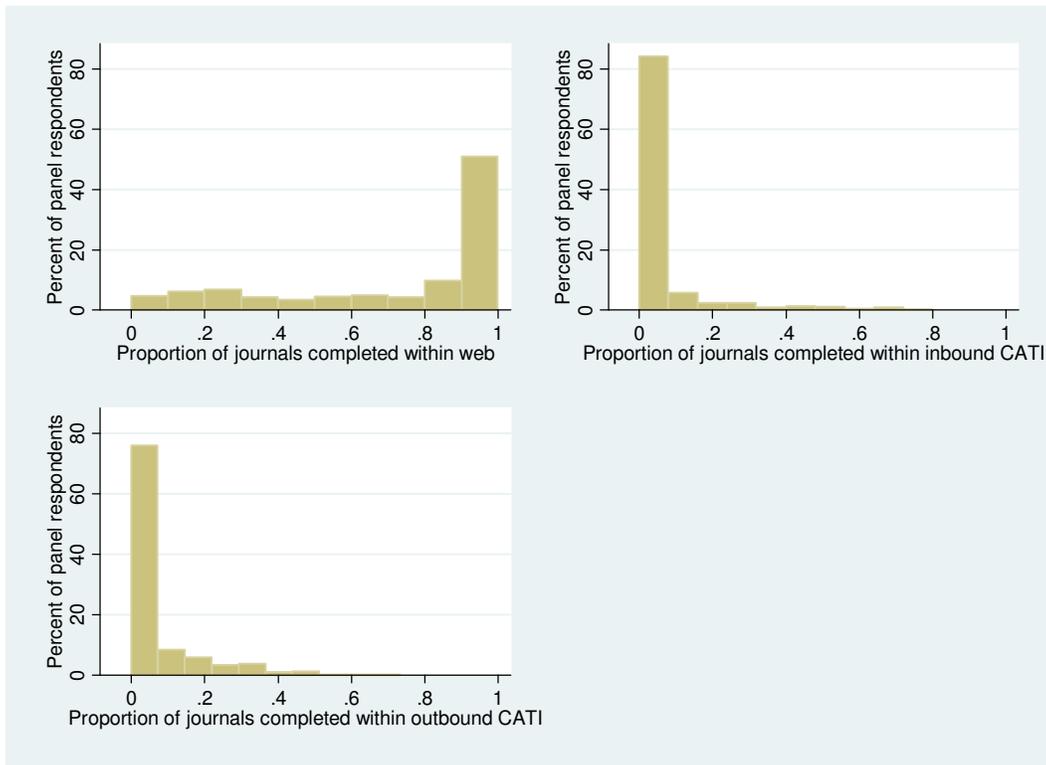


Table 12. Odds ratios and relative risk ratios for models predicting the likelihood of mode use: Effects of journal period and mean days in previous journal periods.

Covariate	Model 1: CATI vs web as outcome CATI	Model 2: Inbound, outbound CATI vs web as outcome	
	Odds ratio(std. err.)	Inbound CATI Relative risk ratio (std. err)	Outbound CATI Relative risk ratio (std. err)
Journal period	0.972(.007)***	0.985(.008)*	0.958(.006)***
$MNLEN_{ij-1}$	1.089(.017)***	1.070h(.021)***	1.102(.016)***

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

A logistic regression model (3.9) estimated the effect of the number of previous journals completed via CATI on the likelihood of participation in web in the current journal, controlling for the length of previous journal periods. Note that nonrespondents are excluded from this analysis. As the number of previous CATI responses increases, panel respondents become less likely to respond in web than CATI in the current journal (OR=0.803, std. err. =0.022, p<.001), controlling for the length of previous journal periods (OR=0.934, std. err.=0.015, p<.001).

As shown in Table 13, the number of previous inbound CATI responses increases, the odds of participating in inbound CATI increases, controlling for $MNLEN_{ij-1}$. But as the number of previous inbound CATI responses increases, the odds of participating in *outbound* CATI also increases. The covariate for outbound CATI behaves in a similar manner. Panel respondents who have used either CATI mode in the past are more likely to use either CATI mode in the present.

Indicators of topic interest, privacy concerns, and educational status does not mediate the relationship between a history of using these CATI modes and current use of CATI. When the predictors in Table 9 are added to the model, the coefficients presented in Table 13 are unaffected. These tables are available in Appendix A,

Table A. 11 and Table A. 12.

Table 13. Multinomial logistic regression predicting the likelihood of inbound or outbound CATI vs web response.

	Relative risk ratio (std. err.)	
	Inbound CATI	Outbound CATI
# Prev. inbound CATI journals	1.249(0.081)**	1.154(0.071)*
# Prev. outbound CATI journals	1.275(0.044)***	1.315(0.042)***
$MNLEN_{ij-1}$	1.052(0.022)*	1.078(0.016)***

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

4.3. Discussion

This study examined the dynamics of three related but distinct outcomes in a panel survey implementing a sequential mixed-mode design at each observation: response, lateness of response, and mode of response. As noted in Section 2.4.1, individuals who are late are attempted via the second mode. However, in RDSL and in many other surveys implementing a sequential design, late respondents can use the first mode, and early respondents can use the second mode. This study analyzes each outcome separately because they are three separate processes.

4.3.1. Response rates and response propensity

Response rates declined early in the field period and rose steadily after the 15th journal period. The declining rate of attrition in later journal periods is consistent with the literature on panel attrition. For example, Zabel (1998) found that nonresponse rates in three panel surveys declined across waves; in each of these surveys, the denominator of the nonresponse rate was estimated as the number of individuals who had responded to the previous wave. In RDSL, the denominator of the nonresponse rate is the total number

of individuals who had agreed to participate in the panel *and* were eligible for the particular journal period. As journal periods increased, the denominator decreased.

We see the same phenomenon in RDSL as in Zabel's (1998) study: nonresponse rates tend to decline across journal periods. Furthermore, the propensity to respond increases across journal periods even when we control for the fact that some panel respondents tend to be faster or slower to respond. In other words, panel respondents complete journals more consistently as time in sample increases.

One possible explanation for the increase in response propensity as time in sample increases is that panel respondents develop a relationship with the survey organization. Some research shows that liking the interviewer or the survey organization can increase the likelihood of response (Groves *et al.*, 1992). However, much of the literature on liking and response focuses on interviewer behaviors (Groves & Couper, 1998) or similarities between interviewers and potential respondents in demographic characteristics (Durrant *et al.*, 2010). It is unclear how the long-term relationship with the survey sponsor would affect the likelihood of response. One could test this by telling some panel respondents that the sponsor had changed part-way through a panel. If response propensity drops as a result, then we can conclude that switching the sponsorship partway through a panel decreases the response propensity.

Some of the hypothesized predictors of response were related to the likelihood of response in the bivariate models. But no clear patterns emerged that could tell us why this was the case. For example, being enrolled full-time at baseline is positively related to the likelihood of response at subsequent journals, but being enrolled full-time at post-

baseline journals is not related to the likelihood of response. Part-time enrollment at subsequent journals is negatively related to response.

Under the privacy concerns hypothesis, panel respondents who live with a parent at baseline should be *less* likely to respond, but we see the opposite effect. And having a sexual partner at any given journal period is not related to the likelihood of response. However, pregnancy intentions and pregnancy avoidance have the predicted effect on response propensity: having pregnancy intentions is negatively related to the likelihood of response and wanting to avoid pregnancy is positively related to the likelihood of response. Privacy concerns are likely not the primary driver of response propensity.

Topic interest, likewise, may not explain the bivariate findings in the response propensity models. If the salience of the topic increased the likelihood of response, then having any sexual partners, use of contraception, pregnancy intentions, and a change in relationship status would all be positively related to response. None of these effects were found.

An alternative explanation is that the instability in some panel respondents' lives is interfering with contact, cooperation, or both. Groves and Couper (1998) found that accessible at-home patterns influenced the likelihood of contact. Their model assumed that contact and cooperation occurs at roughly the same time—at least the same day. However, contact and cooperation in RDSL can be two separate events. In RDSL, contact occurs when an email, text message, or telephone call from an interviewer is made. At contact, the panel respondent might click on a link to the web survey or respond to a CATI interviewer's request. Or, the panel respondent might note the text message and wait until she is at her computer to complete the survey. She must have

access to a phone or the web when she decides to complete the journal. Some events in the panel respondent's life may interfere with the cooperation process by reducing her accessible patterns.

Some events related to instability within the panel respondent's life might reduce the likelihood of response. For example, a change in relationship status can change a panel respondent's routine, which could decrease the likelihood of response. She may be spending time with a new partner, or she may be spending time away from home with friends after a breakup. Because her daily routines are disrupted, her accessibility to the web or telephone may change; this can negatively impact the likelihood of response.

A change in routine may explain the various effects of the education variables. If a woman changes her educational status by dropping out or changing to a four-year college, her routine may change. The instability in her life could lead to a reduced likelihood of participation.

Although some of the predictors were significantly related to the likelihood of response in the bivariate models, we have no support for the speculation that the effect of these predictors might change across journal periods. Most of the substantive covariates and interactions in both the main effects models and the interaction models were not significant; this is likely because the covariate the length of previous journal periods accounted for most of the variance in the model. This covariate was included because panel respondents with larger numbers of journal periods had taken a shorter period of time in previous journals to respond than those with fewer numbers of journal periods.

4.3.2. Lateness of response

The odds of being a late respondent are greater when the panel respondent has been late in the past, controlling for journal period and the mean length of previous journal periods (i.e., $MNLEN_{ij-1}$). This effect decreases across journal periods.

If we assume a continuum of resistance model we would expect that the significant predictors of response propensity in the bivariate models would also be significant predictors of late response in the bivariate lateness models. Because of the coding of the dependent variables (response: 1=response, 0=nonresponse; lateness: 1=late response, 0=early response), we would expect the sign of the coefficient to be reversed. This is the case for being enrolled in school full-time at baseline, type of school (4-year college), having a sexual partner at baseline, pregnancy intentions, and a change in relationship status. But the bivariate response propensity and lateness models disagree for being enrolled in school full-time and part-time, type of school at baseline (4-year college), type of school (2 year junior or community college), noncoital contraception, and pregnancy intentions at baseline. Because we do not see a consistent agreement between the response propensity models and the lateness models, late respondents are dissimilar to nonrespondents with respect to the predictors.

4.3.3. Mode of response

The more a panel respondent responded via CATI in the past, the greater the likelihood of using CATI in the current journal. It does not seem to matter whether the calls in previous journals came from the panel respondent or from the interviewer; each is more likely to result in more current CATI responses—again, regardless of the direction of the call—than web. However, use of either of the CATI modes was relatively rare relative to web.

Some of these inbound calls may result from a message from an interviewer after Day 5. Or, some outbound calls might result from a message from a panel respondent before or after Day 5. This analysis did not differentiate these cases. Although we have data on the order of calls and outcome of calls, we have no way of knowing if the panel respondent heard the message from the interviewer, or just called in spontaneously.

This study makes no claims that panel respondents have a preference for web over CATI. Mode preference is a construct that is difficult to measure (Groves & Kahn, 1979) and is rarely correlated with the likelihood of response in that mode (Dillman *et al.*, 1994; Selfa & Sederstrom, 2006). In this study, we did not attempt to measure mode preference; we only know what mode the respondent used.

Panel respondents can, however, make three choices in each wave of a survey implementing a sequential mixed-mode design: 1) to respond at all; 2) to respond early or late (conditional on response); and 3) to respond in web or CATI (conditional on response). Some individuals may have chosen web because they believed it was the only mode available. Others may have chosen to participate in web because it was convenient at the time of the survey request. But from these data, we cannot conclude that panel respondents overwhelmingly “prefer” web to CATI.

4.3.4. Limitations

RDSL had a very small amount of nonresponse compared to cross-sectional surveys, but were roughly comparable to some panel surveys. The nonresponse rates were slightly higher than the Dutch Socioeconomic Panel (Winkels & Withers, 2000), the European Community Household Panel (D. Watson, 2003), and the PSID (Zabel, 1998), conditional on response to the first wave or journal period. Perhaps the

substantive variables did not predict response because there were so few cases of nonresponse, particularly in later journal periods.

Similarly, panel respondents overwhelmingly responded on time and in web. On average, only 10% of panel respondents were late respondents. And an average of only 15% of respondents to a particular journal period used CATI. Although RDSL contained a large number of observations compared to other panel surveys, the cell sizes may have been too small to detect many of the hypothesized effects outlined above.

The predictors available in RDSL may be weak proxies for privacy concerns and topic salience. We have no direct evidence that panel respondents would have privacy concerns if they knew they would be asked questions about sexual partners, for example. Nor do we have direct evidence that a change in relationship status makes the topic more salient to a panel respondent. However, it seems reasonable that these are indicators of salience and privacy concerns.

Furthermore, any effects of privacy concerns and topic salience may be more important in predicting response to the baseline interview, rather than subsequent journals. Mechanisms of attrition in panel surveys differ from cross-sectional surveys because the panel respondent has already cooperated once.

As noted in Section 4.4.1, changes in the indicators of privacy concerns, topic salience, and education status may better predict the likelihood of response. This study focused on contemporaneous variables as an exploratory analysis, and the effects of changes in these indicators (with the exception of change in relationship status) on the likelihood of response was out of scope of this paper.

Similarly, changes in the indicators may have a stronger mediating effect on the relationship between past late behavior and current late behavior than a simple snapshot of an indicator. For example, current lateness of response is negatively related to currently being enrolled in a four-year college. But those 4-year college students who are late may be more likely to leave a four-year college in the future than those who are not late. This could occur if a drop in the panel respondents' GPA—which is related to the likelihood of dropout (Stratton *et al.*, 2008)—is related to lateness at the current journal. If the GPA is related to lateness *and* to the likelihood of dropping out in the future, then the effect of having a history of lateness on current lateness would be mediated by leaving a 4-year college at later journal periods. This analysis is beyond the scope of this dissertation, but warrants investigation.

It is unclear if changes in the indicators of privacy concerns, topic salience, and education status would influence mode in the same way. Perhaps a change in an indicator is related to a change in the mode of response because of the change in routine related to that indicator. Future research should address how changes in predictors of the use of a particular mode affect the relationship between past behavior and current behavior.

Finally, an ideal survey would have information on respondents and nonrespondents at each observation in the panel for each predictor variable. However, RDSL does not have this information for nonrespondents, and most surveys on similar topics would not have this information available. As detailed in Chapter 3, RDSL used the considerable auxiliary information on respondents and nonrespondents to impute

values. Given the large amount of data available and the small fractions of missing information, the imputed values are likely a reasonable approximation of the true values.

4.3.5. Conclusions

In general, this study presented some good news for survey practitioners. First, as the time in sample increases, wave nonresponse either gets better or stays the same relative to the previous journal period.

Second, lateness of response was relatively low. This is good news because late responses in a sequential design can increase costs as survey organizations switch the mode of contact attempt. Among those who were late, most had been late in the past. Perhaps survey researchers can target these individuals at early waves to decrease the likelihood of this pattern.

Third, the mode of response was overwhelmingly web, which also helped keep costs for RDSL low. Panel respondents were able and willing to participate in a long-term panel survey with sensitive questions online. As noted above, we cannot determine from these data whether this is predominance of web is due to mode preference, however; but it is clear that panel respondents—at least in this special population—may be willing to use web for such a survey.

Although this study dealt with the behavior of panel respondents, we still know little about the dynamics of wave nonresponse bias in sequential mixed-mode designs. Chapter 5 will address this issue.

Chapter 5

Wave nonresponse bias in sequential mixed-mode designs

5.1. Introduction

Chapter 4 addressed issues surrounding the likelihood of response, the likelihood of being a late respondent, and the likelihood of participating in a particular mode in the RDSL sequential mixed-mode design. Some variables, such as having pregnancy intentions, were predictive of response propensity; however, the effect of such variables on response was constant across journal periods.

Instead of examining respondent behavior, the impact of the behavior on the statistic of interest will be examined here. Response, used as a dependent variable in Chapter 4, becomes the independent variable in Chapter 5. This chapter focuses on the impact of the sequential design on the wave nonresponse bias of privacy concerns, topic salience, and education status.

First, the overall wave nonresponse bias—that is, the wave nonresponse bias at the end of the field period at each journal period—is estimated. As discussed in Section 2.4.2 of Chapter 2, the overall wave nonresponse bias in the variables of interest in RDSL may increase or decrease across journal periods.

As noted in Chapter 2, the stochastic view of nonresponse estimates nonresponse bias as the covariance between a variable y and the response propensity p , divided by the mean response propensity. Although Chapter 4 estimated the numerator of the bias in the

bivariate propensity models, it did not address the nonresponse bias ($\text{cov}(y, p) / \bar{p}$).

However, given the high and somewhat stable response rates, the overall nonresponse bias of each of the variables used as predictors is expected to be similar to the coefficients in the bivariate models. For example, having any sexual partners at baseline was negatively related to the likelihood of response. The nonresponse bias of having a sexual partner at baseline should, similarly, be significantly less than zero. Refer to Table 14 for a summary of findings from Chapter 4.

This chapter not only estimates the nonresponse bias overall, but also estimates whether the nonresponse bias changes linearly across journal periods. Just as we would expect the coefficients of the bivariate models to be related to the overall nonresponse bias, we would also expect the coefficients of the interaction between indicators and journal period to be related to the change in nonresponse bias. Because the multivariate response propensity models in Chapter 4 also included the length of previous journal periods (i.e., $MNLEN_{ij-1}$) as a control variable, the relationship between the coefficient for the interaction in the multivariate propensity models and the change in the nonresponse bias across journal periods is not clear.

Second, the sequential design at each wave in RDSL certainly increases response rates; but is the wave nonresponse bias decreasing after targeting nonrespondents with CATI? For example, women who have pregnancy intentions may have some privacy concerns and thus may be less likely to participate in CATI compared to web than women who do not fall in this category. This could be because having pregnancy intentions is socially undesirable, and the panel respondent knows the question will be asked. She therefore may be more likely to participate in web or not respond, compared

to women without pregnancy intentions. If so, then the wave nonresponse bias of pregnancy intentions at that journal period should be greater in the full sample than if the study only included web respondents.

This difference in nonresponse bias between web and the full sample may change across journal periods. In the above example, these women who have pregnancy intentions may be less concerned about self-presentation in later journals than in earlier journals; perhaps their trust for the survey organization increases across journal periods as they develop a relationship with the organization.

In sum, this study aims to estimate the nonresponse bias at each journal period; the change in the nonresponse bias across journal periods; the difference in nonresponse bias between the web portion of the sequential design and the full sequential design; and the change in this difference across journal periods.

5.2. Results

5.2.1. Estimation of overall wave nonresponse bias

Significant nonresponse bias was found for some of these variables. When the linear combination is significantly less than zero, the nonresponse bias is negative. For example, women who had reported having had a sexual partner at the baseline interview were, in general, underrepresented in RDSL because of differential nonresponse.

Variables for which this occurs include being enrolled part-time, having a sexual partner at baseline, pregnancy intentions, and a change in relationship status.

When this linear combination is significantly greater than zero, the nonresponse bias is positive. This is only the case for being enrolled full-time at baseline. If we only

used data from respondents, we would overestimate the proportion of full-time students (compared to not being enrolled) by about .035.

Table 15 and Table 16 displays the means of each variable, the mean wave nonresponse bias including web responses as respondents, the mean wave nonresponse bias including web and CATI responses as respondents, and the mean difference in bias ($\bar{\delta}$). These estimates do not take into account time in sample—this is merely the average within-journal period nonresponse bias of each of these variables. But are any of these estimates of wave nonresponse bias significantly different from zero?

As described in Chapter 3, the linear combination θ . (or, in the case of the multinomial models, θ_k .) was estimated. Table 17 displays each of these estimates. This combination—not the coefficient—is of interest here. See Appendix A, Table A. 13 for the coefficients of the models.

Figure 12 displays the nonresponse bias at each journal period. As shown, the bias decreases until about journal period 20. After journal period 20, the bias is still below zero for most journal periods; however, it does not seem to decline further. This change in nonresponse bias may not be linear.

5.2.2. Differences in wave nonresponse bias between modes

Thus far, we have examined nonresponse bias irrespective of mode. However, the main goal of this chapter is to examine how the sequential mixed-mode design employed by RDSL affects the wave nonresponse bias, and how that effect may change across journal periods.

Table 14. Findings from bivariate response propensity models in Chapter 4.

Predictor	Findings from Chapter 2 (bivariate models)	
	Baseline	All journals
Currently enrolled in school		
Not enrolled	(reference)	(reference)
Part-time enrollment	n.s.	Negative
Full-time enrollment	Positive	n.s.
Type of school (if enrolled)		
High school or less	(reference)	(reference)
2 year junior/community college	n.s.	Positive
4 year college	Positive	Positive
Voc, tech, trade, or other school	n.s.	n.s.
Having any sexual partners	Negative	n.s.
Use of non-coital contraception	n.s.	n.s.
Use of coital contraception	(n/a)	n.s.
Living with a parent	Positive	(n/a)
Pregnancy intention	n.s.	Negative
Pregnancy avoidance	n.s.	Positive
Change in relationship status	(n/a)	Negative

n.s.=not significant at the $p<.05$ level; Positive=coefficient for predictor is significantly greater than zero at the $p<.05$ level; Negative=coefficient for predictor is significantly less than zero at the $p<.05$ level.

Table 15. Mean overall wave nonresponse bias estimates in RDSL: Baseline variables.

Variable	Full sample mean (Std. Err.)	Non-response bias: Web only	Non-response bias: Web+ CATI	Change in bias ($\bar{\delta}$)
Currently enrolled in school				
Not enrolled	0.241(0.016)	-0.028	-0.010	-0.018
Part-time enrollment	0.122(0.013)	-0.004	-0.002	-0.002
Full-time enrollment	0.638(0.019)	0.032	0.012	0.020
Type of school (if enrolled)				
High school or less	0.172(0.017)	-0.017	-0.006	-0.011
2 year junior/community college	0.343(0.022)	-0.007	-0.001	-0.005
4 year college	0.420(0.023)	0.032	0.008	0.024
Voc, tech, trade, or other school	0.064(0.011)	-0.008	-0.001	-0.007
Having any sexual partners	0.713(0.019)	-0.028	-0.014	-0.014
Use of non-coital contraception	0.403(0.025)	0.019	0.006	0.013
Living with a parent	0.447(0.020)	0.010	0.006	0.004
Pregnancy intention	0.547(0.020)	-0.013	-0.005	-0.009
Pregnancy avoidance	0.989(0.004)	0.000	0.000	0.000

Table 16. Mean overall wave nonresponse bias estimates in RDSL: Variables in all journals.

Variable	Full sample mean (Std. Err.)	Non-response bias: Web only	Non-response bias: Web+ CATI	Change in bias ($\bar{\delta}$)
Currently enrolled in school				
Not enrolled	0.246(.015)	-0.019	0.000	-0.019
Part-time enrollment	0.122(.010)	-0.004	0.000	-0.004
Full-time enrollment	0.632(.017)	0.023	0.000	0.023
Type of school (if enrolled)				
High school or less	0.118(.012)	-0.010	0.000	-0.010
2 year junior/community college	0.334(.021)	-0.010	0.000	-0.010
4 year college	0.479(.022)	0.024	0.000	0.024
Voc, tech, trade, or other school	0.069(.009)	-0.003	0.000	-0.003
Having any sexual partners	0.377(.015)	0.006	0.000	0.006
Use of non-coital contraception	0.342(.017)	0.014	0.001	0.013
Use of coital contraception	0.438(.021)	-0.029	-0.015	-0.015
Pregnancy intention	0.187(.011)	-0.063	-0.059	-0.004
Pregnancy avoidance	0.976(.004)	0.006	0.006	0.000
Change in relationship status	0.072(.004)	-0.001	0.000	-0.001

Table 17. Tests of overall wave nonresponse bias significance: Odds ratios, relative risk ratios, and standard errors in multinomial and logistic models. θ .

	Baseline	All journals
Currently enrolled in school		
Part-time enrollment	0.026(0.025)	-0.086(0.019)**
Full-time enrollment	0.035(0.016)*	0.104(0.068)
Type of school		
2 year junior/community college	0.017(0.024)	0.015(0.040)
4 year college	0.040(0.026)	0.086(0.056)
Voc, tech, trade, or other school	0.064(0.047)	-0.102(0.057)
Any sexual partner	-0.058(0.024)*	-0.045(0.037)
Use of non-coital contraception	0.010(0.017)	0.009(0.033)
Use of coital contraception		0.008(0.018)
Living with a parent	0.016(0.012)	
Pregnancy intention	-0.002(0.009)	-0.131(0.040)*
Pregnancy avoidance	0.019(0.056)	0.044(0.053)
Change in relationship status		-0.109(0.031)*

†p<.1; *p<.05; **p<.01; ***p<.001.

Note: reference category for Currently enrolled in school is “Not enrolled”. The reference category for Type of school is “High school or less”.

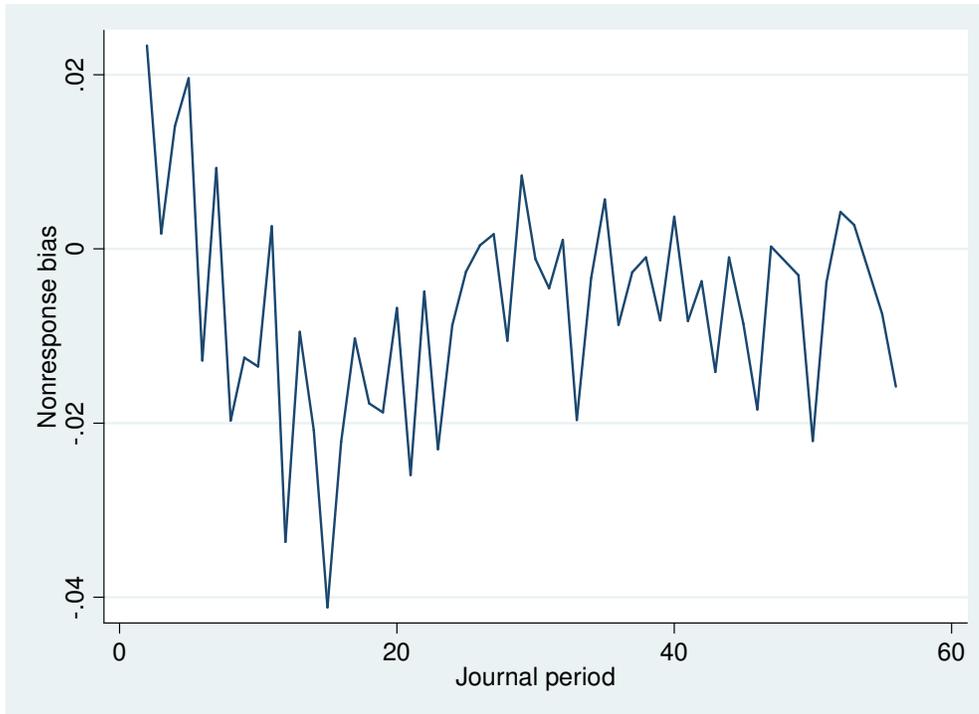
Table 18. Change in overall nonresponse bias across journal periods: Odds ratios, relative risk ratios, and standard errors in multinomial and logistic regression model contrasts (θ_A)

	Baseline	All journals
Currently enrolled in school		
Part-time enrollment	0.000(0.001)	0.002(0.001)
Full-time enrollment	0.000(0.001)	-0.002(0.003)
Type of school		
2 year junior/community college	0.000(0.001)	0.003(0.002)†
4 year college	0.000(0.001)	0.002(0.002)
Voc, tech, trade, or other school	-0.003(0.002)	0.002(0.002)
Any sexual partner	0.000(0.001)	0.001(0.001)
Use of non-coital contraception	0.000(0.001)	0.000(0.002)
Use of coital contraception		-0.002(0.001)*
Living with a parent	0.000(0.001)	
Pregnancy intention	0.000(0.001)	-0.005(0.003)
Pregnancy avoidance	0.000(0.002)	0.003(0.002)
Change in relationship status		-0.002(0.002)

†p<.1; *p<.05.

Note: reference category for Currently enrolled in school is “Not enrolled”. The reference category for Type of school is “High school or less”.

Figure 12. Nonresponse bias across journal periods: Coital contraception.



A first descriptive step is to test if the average difference in nonresponse bias between web and the full sequential design is significantly different from zero. In this analysis and all subsequent analyses in this chapter, the simulated treatment groups are used to estimate changes in bias between a hypothetical web study (Treatment 1) and a full web→CATI sequential design (Treatment 2).

Table 19 displays the linear contrasts θ_D with their standard errors. If they are significantly greater than zero, then the sequential design is *improving* nonresponse bias, relative to the simulated web study. If the contrasts are less than zero, then the nonresponse bias of the sequential design is greater than the nonresponse bias of the simulated web study. However, neither of these situations occurred. The average

difference in nonresponse bias between treatment groups was roughly zero for all variables.

5.2.3 Changes in δ_j across journal periods

As described above and in Chapter 3, the difference in nonresponse bias between the simulated web study and the full sequential design was estimated within each journal period (δ_j).

The figures in Appendix C contain plots of δ_j for each variable with the nonresponse bias in each treatment group. Although few linear trends in δ_j seem visually apparent, we see some variables with high variance in both nonresponse bias and delta in the early journal periods. For example, the figure for having any sexual partner after baseline shows that the nonresponse bias jumps from about -.1 to .1 for the first 20 or so journal periods. After journal period 20, the estimates of nonresponse bias tend toward zero. Similarly, δ_j has a great deal of variance until journal period 20, then tends toward zero in later journal periods. It is likely that this is an artifact of the selection mechanism in RDSL; women who have higher-numbered journal periods are more likely to participate than those who have not reached those journal periods.

As shown in Table 20, none of these contrasts were significantly different from zero. That is, for all variables, we cannot reject the null hypothesis that δ_j is constant over journal periods. Although this may be due to the selection mechanism, the same analysis was repeated for the first 20 journal periods only; the same null results were found. Therefore, δ_j does not change linearly across journal periods.

Table 19. Mean change in difference in bias between Treatments 1 and 2: multinomial and logistic regression model contrasts (θ_D) and standard errors.

	Baseline	All journals
Currently enrolled in school		
Part-time enrollment	-0.002(0.025)	0.003(0.023)
Full-time enrollment	0.007(0.016)	-0.006(0.027)
Type of school		
2 year junior/community college	-0.006(0.023)	-0.009(0.020)
4 year college	0.024(0.024)	-0.022(0.034)
Voc, tech, trade, or other school	-0.031(0.037)	0.010(0.019)
Any sexual partner	0.014(0.018)	0.008(0.033)
Use of non-coital contraception	0.008(0.017)	0.002(0.023)
Use of coital contraception		0.021(0.013)
Living with a parent	-0.003(0.010)	
Pregnancy intention	-0.008(0.010)	0.043(0.018)†
Pregnancy avoidance	0.001(0.032)	0.009(0.032)
Change in relationship status		0.019(0.015)

Table 20. Tests for changes in the effect of the sequential design on wave nonresponse bias across journal periods (change in δ_j): Contrasts and standard errors.

	Baseline	All journals
Currently enrolled in school		
Part-time enrollment	0.009(0.028)	-0.012(0.046)
Full-time enrollment	0.004(0.019)	0.014(0.047)
Type of school		
2 year junior/community college	-0.001(0.030)	0.008(0.026)
4 year college	0.009(0.029)	0.051(0.063)
Voc, tech, trade, or other school	0.042(0.044)	0.007(0.024)
Any sexual partner	0.008(0.024)	-0.026(0.025)
Use of non-coital contraception	0.003(0.020)	-0.019(0.026)
Use of coital contraception		0.023(0.018)
Living with a parent	0.007(0.013)	
Pregnancy intentions	-0.001(0.012)	0.024(0.047)
Pregnancy avoidance	-0.039(0.046)	-0.020(0.031)
Change in relationship status		0.011(0.051)

5.3. Discussion

Sequential mixed-mode designs certainly increase response rates, relative to utilizing just one mode (de Leeuw, 2005). This chapter examined how sequential mixed-mode designs used in many observations in a panel survey affect wave nonresponse bias.

5.3.1. Overall wave nonresponse bias

Wave nonresponse bias was first estimated as the overall difference in y between respondents and the full sample, neglecting any effects of mode or journal period. The overall nonresponse bias was significantly different from zero for a number of variables: being enrolled in school full-time at baseline, being enrolled part-time at any journal period, having a sexual partner at baseline, pregnancy intentions, and a change in relationship status. For each of these variables, the direction of the bias was the same as the coefficient on its respective propensity model in Chapter 4. For example, having a sexual partner at baseline is both negatively related to participation *and* has a nonresponse bias that is significantly less than zero. As noted in Section 5.1, we would expect this similarity, given the generally high response rates.

If privacy concerns drove the decision to respond, the nonresponse bias of having a sexual partner, using non-coital contraception, using coital contraception, living with a parent, and wanting to become pregnant would be expected to have nonresponse biases less than zero; wanting to avoid becoming pregnant was expected to have a nonresponse bias greater than zero. For example, women who had a sexual partner to report should be less likely to respond than women without a sexual partner because of concerns about her privacy.

In most cases, indicators of privacy concerns do not have nonresponse bias associated with them; for example, the nonresponse bias of using coital contraception or non-coital contraception is effectively zero. And the nonresponse bias of those living with a parent at baseline is *greater* than zero. But for having a sexual partner at baseline, and pregnancy intentions, the direction of the nonresponse bias is consistent with the privacy concerns hypothesis. As discussed in Chapter 4, privacy concerns are unlikely to be a mechanism of nonresponse in RDSL.

The second hypothesized mechanism of nonresponse is topic interest. Under this hypothesis, women who have a sexual partner, use contraception, intend to become pregnant, want to avoid pregnancy, or have a change in their relationship status are more likely to participate because the survey questions are relevant to them. This explanation for response does not hold for any of the variables studied here. If topic interest played a role in nonresponse, then the bias of having a change in relationship status should be greater than zero. That is, women who had a change in status should be more likely to participate than those who did not have a change in status because the topic is more salient for them. However, this study found no significant overall wave nonresponse bias for this variable or the other variables that should be affected by this salience.

Although some of these nonresponse bias estimates were significantly different from zero, the nonresponse bias of most variables did not change across journal periods. Only use of coital contraception had a nonresponse bias that changed across journal periods; as journal periods increased, the nonresponse bias decreased. It is unclear why this is the case for coital contraception but not for other similar variables, such as non-coital contraception or having a sexual partner.

5.3.2. Differences in wave nonresponse between modes

The difference in nonresponse bias between the simulated web study and the full sequential design was not significantly different from zero for all variables except for noncoital contraception. For this variable, the sequential design decreased the nonresponse bias relative to the simulated web study.

5.3.3. Changes in δ_j across journal periods

The difference in wave nonresponse bias between Treatment 1 and Treatment 2 was constant across journal periods for all variables. For all variables tested, this difference was zero. This is a special case of Scenario 1 described in Chapter 2: the nonresponse bias of the simulated web study was not significantly different from zero, nor was the nonresponse bias of the full sequential design. Both lines would essentially overlay each other on the x axis.

However, a great deal of variance in the initial journal periods may have masked any significant changes in δ_j .

5.3.4. Limitations

As noted in Chapter 4, we had no true values for these y variables. Multiple imputation served as a likely approximation for these values, but, as in most studies, data for nonrespondents are rare. This is a limitation for two reasons: first, the data on the nonrespondents is only as valid as the imputation models. While many auxiliary variables were used for the imputation models, this is inherently an imperfect method.

Second, the use of respondent data and imputed values does not take into account measurement error. Many of these variables often have measurement error associated

with them. Telescoping may result in over- or under-estimation of variables that change rapidly over time, such as use of coital contraception (Neter & Waksberg, 1964; Tourangeau *et al.*, 2000). Some of the survey questions are also sensitive; underestimates and overestimates may be an issue. And if these overestimates and underestimates are related to the mode—which they may be—then the nonresponse bias estimates presented here may be attenuated. Because we do not have true values, we can only speculate on the mechanism and direction of measurement error in RDSL.

The analyses in this chapter assumed that nonresponse bias or δ_j would change linearly across journal periods. However, nonlinear changes may have occurred. Future research should investigate if the nonresponse bias increases in early waves then decreases in later waves, for example.

5.3.5. Conclusions

In general, these findings are good news for survey researchers. The wave nonresponse bias was generally unaffected or improved by the sequential design, and few changes in δ_j were observed.

As discussed in Chapters 3 and 4, RDSL is different from many other panel surveys. The survey population is homogeneous with respect to age, gender, and level of education. The panel lasts for quite a long time, relative to many other panel surveys, and has many observations. This kind of design fosters a strong relationship between panel respondents and the survey organization, which may have resulted in the extraordinarily high response rates.

Panel respondents in RDSL also have varying numbers of journal periods, as described in Chapter 3. This design feature poses a challenge in the estimation of

longitudinal nonresponse bias, as it did in the propensity models in Chapter 4. Although a simple control variable was added to the models in Chapter 4, no such control was used here because the linear combinations of coefficients would not have been meaningful.

To test whether these analyses might be biased because of this selection mechanism, the regression models above were re-run using only the first 10 journal periods (not shown)²⁵. None of the findings changed. Although it is unlikely that the nonresponse bias estimates and δ_j were biased because of the selection mechanism, future work should attempt to hold the number of observations constant across panel respondents.

These sensitive questions are also thought to be related to the likelihood of response because of their sensitivity. Thus, there may be a relationship between nonresponse and measurement error, in which panel respondents who are more likely to respond have less measurement error than those who are less likely to respond. For example, women who have no pregnancy intentions were generally likely to respond than those who had pregnancy intentions. If women who have pregnancy intentions are more likely to misreport these intentions, then this relationship exists. Although this was not examined in the context of RDSL, Chapter 6 will investigate the relationship between nonresponse and measurement error.

²⁵ About 80% of panel respondents had at least 10 journal periods; by restricting these tests to the first 10 journal periods, we can see any effect of journal period on the bias or change in bias while limiting the reach of the selection mechanism.

Chapter 6

Nonresponse bias and measurement error

6.1. Introduction

As detailed in earlier chapters, sequential mixed-mode designs can dramatically increase response rates, relative to stopping call attempts in the first mode. This increase may be due to the additional calls made or to the mode switch itself. Regardless of the source of the response rate increase, high response rates do not necessarily correspond to reduced nonresponse bias (Groves & Peytcheva, 2008).

Nonresponse bias may be affected by the mode switch in a sequential design. Some individuals may be more likely to respond via the first or the second mode, and those individuals may vary on some statistic of interest. In this chapter, nonresponse bias of estimates of employment, household characteristics, and person characteristics are expected to vary across modes in a sequential mixed-mode design. See Chapter 2 for a detailed discussion on these variables and hypothesized impacts on nonresponse bias.

Similarly, measurement bias may be affected by the sequential design. As described in Chapter 2, some of these estimates may be subject to measurement error bias. This study will examine changes in measurement error bias across modes.

Finally, the likelihood of response may be related to changes in measurement bias within each mode. Individuals who respond earlier in the field period may have lower

levels of measurement bias because of a motivation to participate and participate well (Cannell & Fowler, 1963; Olson, 2006). And, this relationship may be stronger or weaker within CATI than within CAPI or the full responding sample.

The goal of this chapter is not to compare a sequential design to a single mode design, but rather to see the change across modes with respect to nonresponse and measurement bias and the relationship between the two error sources in the sequential context. The following specific hypotheses will be addressed: 1) In the case of PASS, nonresponse bias of all variables related to at-home patterns decreases across modes in a sequential design; 2) Measurement bias is improved by the CATI→CAPI sequential design in PASS; 3) Nonresponse and measurement bias have a stronger relationship with CATI respondents than CAPI respondents.

6.2 Results

6.2.1. Descriptive measures

Table 21 presents the means or proportions, standard deviations, and other measures for the 13 IEB variables used in the nonresponse bias estimation below. These estimates were generated using the imputed values. Standard deviations in the imputed datasets were combined and adjusted per Rubin (1987).

6.2.2. Nonresponse bias

Table 22 presents two estimates of signed nonresponse bias for each IEB variable . The absolute value of the nonresponse bias of CATI is greater than the absolute nonresponse bias of CATI+CAPI for all variables except for age. That is, the sequential mode has reduced nonresponse bias. However, the reduction is not guaranteed, although

the three instances where it increases (registered as employed, employed full-time, and married) all have small relative bias values. The increase in nonresponse bias for these three variables may not be significant.

Table 21. Descriptive statistics for 13 IEB variables in the nonresponse bias analysis.

IEB Variable	Mean	Std. Dev.	Mini-mum	Maxi-mum	Median	n
Gross monthly income	1021.148	83.469	0	5203.5	931.5	977
Currently registered as disabled	0.161	0.046	0	1	0	3534
Registered as employed	0.349	0.021	0	1	0	3534
Employed full-time	0.232	0.018	0	1	0	3534
Employed part-time	0.129	0.014	0	1	0	3534
Registered as unemployed	0.837	0.016	0	1	1	3534
Length of unemployment	331.168	15.294	0	1067.436	242	2557
Lives in former East German states	0.292	0.020	0	1	0	3534
Has children in the benefit community	0.239	0.017	0	1	0	3534
Married (vs unmarried)	0.158	0.014	0	1	0	3534
Female	0.469	0.022	0	1	1	3534
German national	0.928	0.011	0	1	1	3534
Age (in years)	42.29	0.704	16.74	66.49	40.20	3534

Although the relative bias did range in value for the CATI nonresponse bias, the range of relative nonresponse bias for the full sequential design was much smaller. In the three variables where the relative bias increased, the increase was quite large for two (registered as employed and employed full-time). There is no particular reason to expect such an increase for these variables, and with only three such variables, patterns across variables are not helpful in explaining the finding.

Table 22. Signed and relative nonresponse bias before and after mode switch, and the difference in bias between modes.

Variable	Signed bias			Relative bias		
	CATI	CATI+CAPI	Change	CATI	CATI+CAPI	Change
	$(\bar{y}_{r,CATI} - \bar{y}_n)$	$(\bar{y}_r - \bar{y}_n)$	δ	$\frac{\bar{y}_{r,CATI} - \bar{y}_n}{\bar{y}_n}$	$\frac{\bar{y}_r - \bar{y}_n}{\bar{y}_n}$	δ
<i>Employment</i>						
Gross monthly income	35.209	-6.799	42.008	0.035	-0.007	0.042
Currently registered as disabled	0.008	0.000	0.008	0.050	0.002	0.048
Registered as employed	0.009	-0.015	0.024	0.024	-0.039	0.063
Employed full-time	-0.002	-0.016	0.014	-0.008	-0.066	0.058
Employed part-time	0.007	-0.004	0.011	0.050	-0.028	0.078
Registered as unemployed	0.037	0.016	0.021	0.044	0.019	0.025
Length of unemployment	19.464	5.359	14.105	0.061	0.017	0.044
<i>Household characteristics</i>						
Lives in former East German states	0.019	0.009	0.01	0.067	0.031	0.036
Has children in the benefit community	0.007	0.001	0.006	0.019	0.004	0.015
<i>Person characteristics</i>						
Married (vs unmarried)	-0.003	-0.016	0.013	-0.013	-0.067	0.054
Female	0.056	0.039	0.017	0.106	0.072	0.034
German national	0.046	0.021	0.025	0.049	0.023	0.026
Age	-3.070	-0.446	-2.624	-0.082	-0.012	-0.07

Next, the simple regression model was run to test whether the nonresponse bias was significantly different from zero. Table 23 presents the nonresponse bias model results. The rows in the table are the results of separate regressions of the IEB variable in the first column on the response indicator r_i . Since there is a single predictor in each model, a t-test of the null hypothesis that the nonresponse bias is equal to zero is given for each IEB variable. Each regression model may include different numbers of individuals, listed on the far right column. Test statistics were computed using the weighting scheme described in detail in Chapter 3.

Table 23 shows that the difference between respondents and nonrespondents ($\bar{y}_r - \bar{y}_n$)—represented by β_1 —is significantly different from zero for some variables (being employed full-time, married, female, German national, and age). For example, women are 30% more likely to be respondents than nonrespondents. However, the nonresponse bias ($\bar{y}_r - \bar{y}_n$)—estimated as the contrast θ —is not significantly different from zero for any variables. That is, the proportion of respondents who have a white collar job is statistically the same as the proportion of all sample cases who have a white collar job.

Table 23 also displays the difference in nonresponse bias estimates between modes (δ). While the difference is quite large for gross monthly income, the size of these differences does not indicate whether they are statistically important. The change in the relative bias is only .042, which is smaller than a number of other variables.

Table 23. Bivariate nonresponse bias models.

Dependent variable	Coefficient (Std. Err.)	θ (Std. Err.)	<i>N</i>
Gross monthly income	-3.733(71.164)	-1.000 (1.000)	977
Currently registered as disabled	-0.047(0.284)	-0.019 (0.107)	3534
Registered as employed	-0.163(0.083) [†]	-0.067 (0.003)	3534
Employed full-time	-0.229(0.093) [*]	-0.094 (0.006)	3534
Employed part-time	-0.109(0.116)	-0.045 (0.006)	3534
Registered as unemployed	0.082(0.108)	0.034 (0.005)	3534
Length of unemployment	-1.869(13.351)	-0.768 (0.599)	2557
Lives in former East German states	0.010(0.088)	0.004 (0.002)	3534
Has children in the benefit community	-0.093(0.096)	-0.038 (0.022)	3534
Married (vs unmarried)	-0.283(0.122) [*]	-0.116 (0.033)	3534
Female	0.262(0.080) ^{**}	0.108 (0.004)	3534
German national	0.388(0.153) [*]	0.160 (0.015)	3534
Age	1.972(0.522) ^{***}	0.811 (0.004)	3534

[†]p<.10; *p<.05; **p<.01; ***p<.001.

The coefficients in this model are not substantively meaningful here; see Appendix A, Table A. 17 for these coefficients. Instead, the statistics of interest are the contrasts θ_D , shown in Table 24. If the contrast in each of these models is significantly different from zero, then the nonresponse bias in the hypothetical Treatment 1—a single-mode CATI design—is significantly different from the nonresponse bias in the full CATI→CAPI design in Treatment 2. If the contrast is greater than zero, then nonresponse bias is improved because the nonresponse bias of CATI (Treatment 1) is greater than the nonresponse bias of the overall CATI→CAPI sample (Treatment 2).

For almost all of these variables, this contrast is not significant. For example, the nonresponse bias of gross income is statistically identical between Treatments; that is, the nonresponse bias of gross income did not increase or decrease when households are attempted via CAPI.

Table 24. Regression models to determine if $\delta = 0$.

Variable	θ_D
<i>Employment</i>	
Gross monthly income	0.818
Current profession: blue collar	0.424
Current profession: white collar	3.252
Currently registered as disabled	1.001
Registered as employed	0.743
Employed full-time	0.643
Employed part-time	0.834
Registered as unemployed	1.148
Length of unemployment	-0.179
<i>Household characteristics</i>	
Lives in former East German states	1.032
Has children in the benefit community	0.810
<i>Person characteristics</i>	
Married (vs unmarried)	0.583
Female	1.723***
German national	2.068***
Age	4.447***

**p<.001

The exceptions to this rule are being female, being a German national, and age. The overrepresentation of women in CATI demonstrated in Table 22 and Table 23 is reduced when we add CAPI; that is, relatively more men participate in CAPI than in CATI. Similarly, we know that German nationals are more likely to participate than those who are not German nationals. But when CAPI attempts are made, more individuals who are not German nationals participate. And, as shown in Table 22 and Table 23, older individuals are more likely to respond in either mode than younger individuals. But when CAPI attempts are made, this age gap narrows. Proportionally more young people participate in CAPI than in CATI.

6.2.2. Measurement bias

See Table 25 for estimates of signed ($\bar{\epsilon}_S, \bar{\epsilon}_U, \bar{\epsilon}_V$) and absolute ($\bar{\epsilon}_A$) measurement bias (see section 3.2.3). Significance was determined from a series of t-tests. Almost all measurement bias estimates were significantly different from zero. However, this study is less concerned with the overall measurement bias and more concerned about *changes* in measurement bias across modes in this sequential design.

In the models shown in Table 26, we see that the coefficient for β_1 did not vary significantly between modes. That is, the average signed bias was the same between CATI and CAPI, controlling for the propensity to use CAPI.

6.2.3. Nonresponse and measurement bias

This chapter has examined nonresponse bias and measurement bias as though the two error sources are independent. However, they may be related; those who are more difficult to reach may be less motivated respondents (Cannell & Fowler, 1963; Kreuter et al., 2009a; Olson, 2006).

Although no significant differences in measurement bias between the modes were found, the relationship between nonresponse and measurement bias may be “washing out” any differences. This section will address the relationship between nonresponse and measurement bias within modes and across modes.

As shown in Table 27, measurement bias for most of these variables is not significantly related to propensity to respond CATI. Those who are more likely to respond have the same levels of measurement bias as those less likely to respond.

Table 25. Measurement bias within and across modes.

Variable	Measurement bias type	Measurement bias in CATI (Std. err.)	Measurement bias in CAPI (Std. err.)	Measurement bias across modes (Std. err.)
Gross monthly income	Signed bias	-30.887(147.263)	-448.724(775.108)	-60.043(156.329)
	Absolute bias	471.867(133.391) **	812.706(695.333)	495.997(116.661) ***
Currently registered as disabled	Underestimate	0.105(0.041) †	0.102(0.059)	0.105(0.042) †
	Overestimate	0.093(0.017) **	0.080(0.026) **	0.092(0.017) **
	Absolute bias	0.198(0.047) *	0.183(0.065) *	0.197(0.047) **
Registered as employed	Underestimate	0.169(0.010) ***	0.129(0.025) ***	0.245(0.008) ***
	Overestimate	0.040(0.005) ***	0.029(0.015)	0.0237(0.003) ***
	Absolute bias	0.209(0.011) ***	0.158(0.028) ***	0.269(0.009) ***
Registered as unemployed	Underestimate	0.239(0.029) ***	0.248(0.036) ***	0.240(0.028) ***
	Overestimate	0.017(0.005) **	0.006(0.007)	0.016(0.005) **
	Absolute bias	0.256(0.028) ***	0.253(0.036) ***	0.256(0.026) ***
Length of unemployment	Signed bias	304.374(9.950) ***	271.579(21.348) ***	301.204 (9.381) ***
	Absolute bias	307.047(9.867) ***	273.519(21.236) ***	303.805(9.302) ***
Has children in the benefit community/household	Underestimate	0.026(0.009) *	0.032(0.013) *	0.113(0.007) ***
	Overestimate	0.045(0.006) ***	0.074(0.019) ***	0.029(0.004) ***
	Absolute bias	0.071(0.008) ***	0.106(0.021) ***	0.142(0.006) ***

Table 26. Signed measurement bias models with predictors of participation in a mode.

Variable	Coef. (std. err.)				
	CAPI	Disability status	HH Iwer age	# family and friends	cons
<i>Employment</i>					
Gross monthly income	-846.298(997.691)	234.588(284.929)	7.583(11.059)	4.3(5.868)	-476.168(577.895)
Currently registered as disabled					
Underestimate	0.061(0.605)	1.326(0.711)	-0.009(0.217)	0.011(0.458)	-2.325(0.977) †
Overestimate	0.368(0.430)	-0.359(0.642)	-0.026(0.012)*	0.003(0.013)	-1.232(0.436) **
Registered as employed					
Underestimate	-0.501(0.348)	-0.467(0.453)	0.007(0.008)	-0.009(0.009)	-1.720(0.301) ***
Overestimate	-0.318(0.563)	-0.283(0.673)	-0.001(0.014)	0.007(0.012)	-2.824(0.551) ***
Registered as unemployed					
Underestimate	-0.109(0.313)	0.060(0.256)	0.010(0.008)	-0.012(0.009)	-1.530(0.299) ***
Overestimate	-13.828(19.585)	0.939(0.838)	0.008(0.0411)	0.008(0.021)	-4.971(1.327) **
Length of unemployment	-40.379(37.608)	14.024(30.433)	-0.141(1.005)	-0.715(0.98)	305.678(35.08)
<i>Household characteristics</i>					
Has children in the benefit community					
Underestimate	0.713(1.110)	-7.103(14.730)	-0.036(0.030)	-0.011(0.046)	-2.907(1.255) *
Overestimate	0.795(0.581)	0.349(0.439)	-0.015(0.017)	0.008(0.013)	-2.644(0.536)

For CAPI respondents, however, the relationship between response propensity and measurement bias is negative for a few variables. For being registered as disabled, earlier respondents are less likely to underestimate their disability status than later respondents. Similarly, underestimates, overestimates, and absolute measurement bias is more likely among later CAPI respondents for being registered as employed.

6.2.4. Consent to link survey and record data

In order to estimate the measurement error bias for the above analyses, data from IEB was linked with PASS. However, a small number (see Table 28) of respondents did not consent to the link. This creates a missing data problem; measurement bias cannot be directly estimated for the nonconsenting respondents. This study used multiple imputation to reduce this missing data problem; all PASS survey data were imputed for nonconsenting respondents.

If consent is confounding any relationship between mode and measurement bias, then it should be added to the measurement bias models above. However, consent to link IEB and PASS data at Wave 2 was not related to mode of administration (Spearman's $\rho = 0.01$, $p > 0.5$).

6.3. Discussion

This study examined the use of a sequential mixed-mode design in a single wave of a panel survey. Nonresponse bias, measurement bias, and the relationship between nonresponse bias and measurement bias was estimated. As in Chapters 4 and 5, the findings here seemed to be good news for survey organizations who use sequential designs to increase response rates at a reasonable cost.

Table 27. Relationship between signed measurement bias and propensity scores: Estimates of β_1 .

Variable	Measurement bias type	Model (3.40): CATI respondents	Model (3.41): CAPI respondents	Model (3.39): All respondents
Gross monthly income	Signed	0.000(0.000)	0.000(0.000)	-0.000(0.000)
	Absolute	-0.000(0.000)	0.000(0.000)	0.000(0.000)
Currently registered as disabled	Signed:			
	Underestimate	-0.055(0.033)	-0.131(0.061) *	-0.014(0.009)
	Overestimate	0.013(0.021)	0.054(0.074)	0.011(0.009)
Registered as employed	Absolute	-0.029(0.017) †	-0.065(0.052)	-0.002(0.006)
	Signed:			
	Underestimate	-0.002(0.012)	-0.089(0.026) **	-0.007(0.004)
Registered as unemployed	Overestimate	0.035(0.024)	-0.162(0.066) *	-0.038(0.014) **
	Absolute	0.002(0.012)	-0.096(0.025) ***	-0.010(0.004) *
	Signed:			
Length of unemployment	Underestimate	0.010(0.014)	-0.035(0.033)	-0.03(0.009)
	Overestimate	-0.066(0.052)	0.104(0.092)	0.001(0.022)
	Absolute	0.005(0.014)	-0.025(0.032)	-0.003(0.008)
Has children in the benefit community/household	Signed	0.000(0.000) †	0.000(0.000)†	0.000(0.000)
	Absolute	0.000(0.000) †	0.000(0.000)†	0.000(0.000)
	Signed:			
Underestimate	Underestimate	0.049(0.012) ***	0.018(0.030)	0.007(0.006)
	Overestimate	0.009(0.027)	-0.025(0.067)	0.007(0.012)
	Absolute	0.021(0.069)	-0.012(0.089)	0.010(0.064)

†p<.1; *p<.05; **p<.01; ***p<.001.

Table 28. Consent to link IEB and PASS, Waves 1 and 2.

		No consent	Consent	Wave 2 Item nonresponse	Unit nonresponse	Total
Wave 1	No consent	116	231	18	346	711
	Consent	n/a	1706 [†]	n/a	1038	2744
	Item nonresponse	5	29	18	81	133
	<i>Total</i>	121	1966	36	1465	3588

Note. [†]This includes the 120 respondents who already provided consent in wave 1.

6.3.1. Nonresponse bias

First, nonresponse bias across all cases was effectively zero for all variables. If we did not attempt CAPI, however, we would have had some statistically significant nonresponse bias for being female, being a German national, and age. At the end of all CATI and CAPI attempts, all variables had no significant nonresponse bias associated with them.

Although the nonresponse bias of being female, being a German national, and age decreased from CATI to the full sample, the nonresponse bias estimates of the other variables were zero before and after the switch. CAPI was more successful at recruiting men, immigrants, and younger people than CATI.

It is unusual that the nonresponse bias of most of these variables was not significantly different from zero before the end of the field period. It is not uncommon to find that—for example—those who have children are more likely to participate than those who do not have children; those who work part-time are less likely to participate than those who do now work at all; and those who are married are more likely to participate than those who are not (Abraham *et al.*, 2006; Groves & Couper, 1998).

We also know that women tend to be more likely to participate than men; that older people are more likely to participate than younger (Abraham et al., 2006; Goyder, 1987; Groves & Couper, 1998); and that ethnic minorities are sometimes—but not always—less likely to participate than majority groups (Abraham et al., 2006; Goyder, 1987; Groves & Couper, 1998; Stoop, 2005). Before the switch to CAPI, the nonresponse bias of each of these variables (being female, age, being a German national) was significant and in the expected direction. Women, older people, and German nationals were overrepresented among respondents than in the full sample.

Why are these nonresponse bias estimates lower than expected? First, the nonresponse mechanisms may be different in this study, compared to most existing nonresponse studies. Nonresponse in this study is likely different because PASS is a panel survey. In a panel survey, the likelihood of participating in the second (or later) waves of a panel survey, conditional on response to Wave 1, is generally larger than the likelihood of participation in Wave 1. This tends to be due to higher rates of contact and cooperation in later waves (Lepkowski & Couper, 2002; Zabel, 1998). The 3588 Wave 2 panel respondents that make up the sample mean (\bar{y}_n), therefore, are generally more predisposed to participate than sample cases selected for the first wave of a panel or for a cross-sectional survey.

But this increase in the likelihood of response explains only the larger response rates in Wave 2, compared to Wave 1—not the nonresponse bias. The nonresponse bias may be small because the mean of all sample cases used in this study (\bar{y}_n) includes all existing Wave 1 nonresponse bias; being included in this sample mean necessitates that a gatekeeper is a Wave 1 respondent.

For example, if 50% of the original Wave 1 sample of gatekeepers were female, then \bar{y}_n in Wave 1=0.5. Imagine that 65% of Wave 1 respondents are women. The overall sample mean for gender in Wave 2 is $\bar{y}_n = 0.65$. Let us assume that, at Wave 2, women still have a 65% chance of participating. The *conditional* nonresponse bias—the nonresponse bias, conditional on participation in Wave 1—is $.65-.65=0$. This could account for the disappearing nonresponse bias. Future research could test this by examining the change in the nonresponse bias between Waves 1 and 2, much as Chapter 5 did with RDSL.

6.3.2. Measurement bias

Most estimates had some measurement bias associated with them. This study used administrative record data as a gold standard to estimate measurement bias. There are a number of reasons why the measurement bias estimates for nearly all variables in nearly all modes were significantly different from zero. First, this may have been true measurement bias; that is, respondents may have misreported income, length of unemployment, and so on, because of difficulties in the response process (Tourangeau *et al.*, 2000).

Alternative explanations for significant measurement bias include problems in the IEB data themselves. Some of the variables in IEB have varying degrees of data quality (Scioch & Bender, 2010). For example, gross income likely had high data quality and zero item nonresponse because employers are required, by law, to report this. Other variables, such type of profession (blue collar or white collar), are less consistently reported and have a large amount of item nonresponse.

However, this study was less concerned about the total amount of measurement bias and more concerned with the effect of the mode on the measurement bias. We have no reason to believe that the data quality of the IEB is better or worse for individuals who ended up responding in CATI or CAPI. Therefore, the quality of the IEB data are less of an issue, and instead, we consider differences in measurement bias between modes.

Measurement bias did not change across modes. This was a surprising finding because some of these estimates—especially gross monthly income and length of unemployment—are both sensitive and cognitively difficult (Holbrook *et al.*, 2003; Mathiowetz & Duncan, 1988; Moore *et al.*, 1999; Stauder & Hüning, 2003). Respondents tend to use more cognitive resources to respond in CAPI than CATI (Holbrook *et al.*, 2003), and they are also more likely to respond to sensitive questions (Holbrook *et al.*, 2003).

Two factors may have contributed to these null results. First, as noted in the nonresponse bias analysis, participation was dependent on response in Wave 1. These individuals had already answered these questions one year prior to the Wave 2 interview. The difficulty of these questions may have been mitigated by having been asked them before.

A second explanation for these null results is related to the first. The impact of the sensitive nature of these questions on measurement bias may have been attenuated by the time in sample. That is, respondents who have already participated in Wave 1 know that they will be asked about income and unemployment. If they agree to participate again, then perhaps the sensitivity of these questions is not a contributing factor to measurement bias. In other words, there may be nonresponse bias in the measurement

bias estimates, in which respondents are less likely to have a problem with the sensitive nature of these questions, compared to nonrespondents, had they participated. This post hoc explanation relates to the examination of nonresponse bias and its relationship to measurement bias.

6.3.3. Nonresponse bias and measurement bias

As in this second explanation for the lack of measurement bias differences across modes, earlier respondents may be more accurate than later respondents. Earlier respondents may be more motivated to answer accurately and may not modify answers because of the sensitivity of questions than later respondents.

For almost all variables, there was no relationship between the likelihood of response and measurement bias. Early respondents were just as inaccurate as later respondents, regardless of the mode of response. A few exceptions are being currently registered as disabled, being registered as employed, and having children. It is unclear why earlier respondents were less likely to have measurement bias than later respondents. Perhaps comparing a sequential design to a concurrent mixed-mode design and a single-mode design could answer this question. We could then parse out the mode of response from the latency of response; perhaps the relationship between nonresponse bias and measurement bias seen here is due entirely to the mode.

Like the findings in this study, the evidence for a relationship between response propensity and measurement bias is mixed; Olson (2007) found that some variables do have a relationship between the likelihood of response and measurement bias. But Kreuter et al. (2009) found that, for some variables, most of these relationships were

spurious, driven by the reference period dictated by the survey question. Once this reference period was controlled, most of these relationships disappeared.

6.3.4. Summary and conclusions

Sequential mixed-mode designs can therefore be used to raise response rates and contain costs, while maintaining or improving other aspects of data quality. The nonresponse bias either is unaffected or decreases when we add a second mode. Changing the mode also does not seem to affect the measurement error bias. And earlier respondents are just as accurate as later respondents.

Chapter 7

Conclusions

7.1. Introduction

This dissertation aimed to evaluate sequential mixed-mode designs with respect to response behavior, nonresponse bias, and measurement error. In Chapters 4 and 5, the sequential design of RDSL was evaluated with respect to response behavior and nonresponse bias across journal periods. In Chapter 6, nonresponse bias and measurement error were evaluated across modes.

7.2. Research findings

7.2.1. Response propensity, latency, and mode

In Chapter 4, we learned that past behavior has a strong influence on current behavior with respect to the likelihood of response, the latency of response, and the mode of response. Responding more often in the past—which resulted in a larger number of journal periods—was positively related with responding in the present. Similarly, being a late respondent in the past had a strong influence on being a late respondent in the present. And use of CATI in earlier journal periods is highly correlated to use of CATI in the present.

These findings may come as no surprise to survey researchers; it seems logical that gaining the participation of specific individuals will be difficult, even in a panel survey. What is surprising, however, is that even though some individuals may be perpetual nonrespondents, most (75%) return to the panel. Panel surveys that drop permanent attritors may be giving up too soon.

Furthermore, few of the respondent characteristics used in Chapter 4 were significant predictors of response, latency, or mode of response. We have no clear story with respect to the impact of privacy concerns, topic salience, or educational status. As noted at the end of Chapter 4, there are two main explanations for these null results: First, they are very weak indicators of these mechanisms. Like many methodologists, I was opportunistic in my selection of covariates.

A second explanation provided in Chapter 4 is that the relationship with the sponsor may be paramount. At baseline, perhaps the respondent characteristics are important predictors of response; but after agreeing to participate in a long-term panel, individuals' characteristics are less correlated with the likelihood of response. Conditional on agreement to participate in the panel, the nonresponse bias of those characteristics should be relatively low. But the nonresponse bias of the baseline interview may be quite high. Future research should investigate nonresponse bias of the baseline interview separately from subsequent waves of the panel.

7.2.2. Nonresponse bias

Chapter 5 investigated several aspects of nonresponse bias. First, is the overall nonresponse bias significantly different from zero? Does it change over time? Second,

does the sequential design affect nonresponse bias, and does any effect of the design on the bias change over journal periods?

This set of analyses did not produce significant findings. The nonresponse bias for all variables was effectively zero; the nonresponse bias did not change across journal periods; and the sequential design did not affect the nonresponse bias in any way.

The literature on nonresponse bias in sequential mixed-mode designs runs into two main problems: first, there are rarely values for nonrespondents, much like the literature on nonresponse bias in general. Second, the samples in sequential designs are not independent. Respondents in the second mode are necessarily nonrespondents to the first mode. Treating these as independent samples may be problematic. Using RDSL data, the first problem was still an issue. We had no values for nonrespondents. Multiple imputation is one of the ways we can estimate values for nonrespondents, but this is a clear weakness of this study. However, the new methods used in Chapter 5 deal with the dependent samples issue by simulating the two treatment groups.

7.2.3. Nonresponse bias and measurement error

In a sequential design, early respondents participate in a different mode than later respondents, and the mode can affect estimates with respect to nonresponse bias and/or measurement error. All of these moving parts can contribute to total survey error.

To unpack this, nonresponse bias was first estimated. In PASS, nonresponse bias before the switch was effectively zero for all variables except age, being a German national, and being female. After the switch, however, those nonresponse bias estimates dropped to a level indistinguishable from zero. That is, nonresponse bias was either

improved by the sequential design or was already equal to zero after the CATI portion of the design was completed.

Second, measurement error was estimated. Although some estimates had significant measurement error, the level of measurement error was consistent across modes. Time in sample may have helped reduce the measurement error for these estimates—but there may be nonresponse bias in the measurement error estimates.

For example, a question about income is both sensitive and cognitively challenging (Moore *et al.*, 2000). After responding to that question at Wave 1, the panel respondent may base the decision to participate at Wave 2, in part, on the knowledge that income will be asked. There is some evidence for this. Item nonresponse on the income question tends to lead to unit nonresponse in later waves (Bollinger & David, 2001; Loosveldt *et al.*, 2002). Although item nonresponse is not equivalent to measurement error, the causes of item missing data and measurement error in income are the same: sensitivity and cognitive difficulty (Moore *et al.*, 1999).

We could assert that measurement error at one wave is predictive of measurement error at a second wave. Panel respondents who misreport income at Wave 1 may misreport again at Wave 2. However, if these misreporters do not respond at Wave 2, then the measurement error estimates at Wave 2 could have some nonresponse bias.

This relationship between nonresponse and measurement error in PASS is unclear. For most variables, there is no relationship within either mode or in the full sequential design. However, being employed has a significant negative relationship between response propensity and measurement error within CAPI and in the full sequential

design—but not in CATI. This runs counter to the expectation that the relationship between propensity and measurement error should be higher in CATI than CAPI.

It is unclear why this is the case. In PASS, the mode is confounded with the latency of response. Early respondents always used CATI, and late respondents always used CAPI. This study was meant to be exploratory; however, a more clear comparison may be between two modes in a concurrent mixed-mode design.

7.2.4. Summary

The findings in this dissertation are generally good news for survey researchers. Sequential mixed-mode designs improve response rates—sometimes substantially so, at little risk of increased nonresponse bias in a single wave or in multiple waves of a panel survey. Mode differences in measurement error were minimal here. However, some important research questions remain.

7.3. Future research

7.3.1. Response behaviors in mixed-mode surveys

Incentives. In RDSL, incentives were used to encourage early response. In a strict sequential design, this incentive scheme could help increase participation in the less expensive first mode. Because the incentives provided in RDSL were provided to everyone (i.e., there was no control group), we cannot estimate the effect of the incentive on early response. But let us assume that they worked—that having the incentive increased the likelihood of early response, compared to not having that incentive. But in RDSL, early respondents can use web *or* CATI. Incentive schemes that successfully encourage the use of one particular mode may help decrease costs, relative to the existing

incentive scheme. Future research can investigate how—and if—incentives can be used in sequential mixed-mode designs.

Transitions and attrition. Chapter 4 investigated primarily contemporaneous proxies for topic interest, privacy concerns, and educational status. However, there is some evidence that transitions in a panel respondent's life can influence the likelihood of response (Lemay, 2009). When the panel respondent has a change in a variable of interest, then the likelihood of response may decrease. Future work should study this in a mixed-mode context. Do transitions affect the latency of response or the mode of response? Can we use respondent behaviors at earlier waves to predict transitions which can affect response?

7.3.2. Wave nonresponse bias

Linearity assumption. As an exploratory study, this dissertation assumed that any changes in nonresponse bias across journal periods in RDSL would be linear in nature. Linearity was also assumed for changes in δ_j across journal periods. In Chapter 5, we saw that neither nonresponse bias nor δ_j changed significantly across journal periods. The linearity assumption was likely not violated because both nonresponse bias and δ_j were zero across journal periods. But future work could develop methods to track nonlinear changes in bias and δ_j in other situations in which the bias and δ_j may be substantial.

7.3.2. Wave nonresponse bias and measurement error

Nonresponse bias of measurement error in panel surveys. A significant limitation of the analyses in RDSL is that there were no true values for our estimates of interest. Using

the IEB record data in Chapter 6 allowed us to separate measurement error from nonresponse bias.

The analyses in Chapter 6 used Wave 2 data instead of Wave 1. This allowed us to control for telephone coverage and use data from Wave 1 for predictive models. However, as noted above, there may be nonresponse bias in those measurement error estimates. Understanding how measurement error interacts with nonresponse bias across waves of the panel is largely not understood. This line of research can investigate longitudinal changes in a link between nonresponse bias and measurement error across waves of a panel.

Appendix A. Supplemental tables and derivations.

Table A. 1. Mean and standard errors of baseline variables, RDSL.

Variable	Mean (std. err.)
Currently enrolled in school	
Not enrolled	0.284 (.0149)
Part-time enrollment	0.120 (.0108)
Full-time enrollment	0.596 (.0162)
Type of school (if enrolled)	
High school or less	0.183 (.0151)
2 year junior/community college	0.335 (.0185)
4 year college	0.402 (.0192)
Voc, tech, trade, or other school	0.080 (.0106)
Living with a parent	0.421 (.0163)
Having any sexual partners	0.755 (.0142)
Use of non-coital contraception	0.382 (.0161)
Pregnancy intentions [†]	0.095 (.0097)
Pregnancy avoidance [†]	0.899 (.0010)

[†] Proportion of panel respondents who have any intentions/avoidance.

Table A. 2. Means and standard errors of time-variant variables, RDSL.

Variable	Mean (std. err.)
Currently enrolled in school	
Not enrolled	0.246 (.0148)
Part-time enrollment	0.122 (.0102)
Full-time enrollment	0.632 (.0172)
Type of school (if enrolled)	
High school or less	0.117 (.0117)
2 year junior/community college	0.335 (.0207)
4 year college	0.479 (.0222)
Voc, tech, trade, or other school	0.069 (.0086)
Having any sexual partners	0.376 (0.146)
Use of non-coital contraception	0.326 (.0170)
Use of coital contraception	0.423 (.0223)
Pregnancy intentions [†]	0.077 (.0091)
Pregnancy avoidance [†]	0.912 (.0095)
Relationship change	0.072 (.0043)

[†] Proportion of panel respondents who have any intentions/avoidance.

Note. All standard errors are adjusted, using panel respondent as a cluster.

Equation 1. Derivations for contrasts in a concurrent mixed mode design.

The following model can be used to estimate nonresponse bias in a concurrent mixed mode design for the continuous variable y :

$$y_i = \beta_0 + \beta_1 r_i + \beta_2 M_i + \beta_3 r_i M_i + \varepsilon_i$$

where y_i =the value of some y variable (such as age), M_i =a dummy variable for the mode assigned—let us assume that the mode can be web ($M_i = 1$) or CATI ($M_i = 0$).
 r_i =a dummy variable indicating response ($r_i=1$) or nonresponse ($r_i=0$).

This regression model can be used to estimate the difference in nonresponse bias between two modes in a concurrent mixed-mode design. This difference is equal to:

$$(\bar{y}_{r,web} - \bar{y}_{n,web}) - (\bar{y}_{r,CATI} - \bar{y}_{n,CATI})$$

where $\bar{y}_{r,web}$ =the mean of y for respondents to web; $\bar{y}_{n,web}$ =the mean of the full sample—respondents and nonrespondents—assigned to the web condition; $\bar{y}_{r,CATI}$ =the mean of y for respondents to CATI; and $\bar{y}_{n,CATI}$ =the mean of the full sample assigned to CATI.

We know that

$$(\bar{y}_{r,web} - \bar{y}_{n,web}) - (\bar{y}_{r,CATI} - \bar{y}_{n,CATI}) = \frac{m_{web}}{n_{web}} (\bar{y}_{r,web} - \bar{y}_{m,web}) - \frac{m_{CATI}}{n_{CATI}} (\bar{y}_{r,CATI} - \bar{y}_{m,CATI})$$

where m_{web} =the number of nonrespondents in the web condition; n_{web} =the number of total sample cases assigned to web; m_{CATI} =the number of nonrespondents in the CATI condition; and n_{CATI} =the number of sample cases assigned to CATI.

The terms can be derived as follows:

$$\bar{y}_{r,web} = \beta_0 + \beta_1 + \beta_2 + \beta_3$$

$$\bar{y}_{m,web} = \beta_0 + \beta_2$$

$$\bar{y}_{r,CATI} = \beta_0 + \beta_1$$

$$\bar{y}_{m,CATI} = \beta_0$$

Therefore,

$$\begin{aligned} \frac{m_{web}}{n_{web}} (\bar{y}_{r,web} - \bar{y}_{m,web}) - \frac{m_{CATI}}{n_{CATI}} (\bar{y}_{r,CATI} - \bar{y}_{m,CATI}) &= \frac{m_{web}}{n_{web}} (\beta_0 + \beta_1 + \beta_2 + \beta_3 - \beta_0 - \beta_2) \\ - \frac{m_{CATI}}{n_{CATI}} (\beta_0 + \beta_1 - \beta_0) &= \frac{m_{web}}{n_{web}} (\beta_1 + \beta_3) - \frac{m_{CATI}}{n_{CATI}} (\beta_1) \end{aligned}$$

Equation 2. Derivations for the contrast to estimate $\bar{\delta}$ in sequential mixed-mode design.

Chapter 5 examines the change in nonresponse bias between web and the full sample in RDSL. This change within a single journal period j is known as δ_j :

$$\delta_j = (\bar{y}_{r,web,j} - \bar{y}_{n,web,j}) - (\bar{y}_{r.,j} - \bar{y}_{n.,j})$$

Per Chapter 3, an experiment was simulated in which the data were imputed twice: In Treatment 1, data for all CATI respondents and nonrespondents were imputed, along with item nonresponse for web respondents. In Treatment 2, data for nonrespondents and item nonresponse for all respondents were imputed. This simulates an experiment in which Treatment 1 was only attempted as a web study, and Treatment 2 was the full sequential design. Thus:

$$\delta_j = (\bar{y}_{r1j} - \bar{y}_{n1j}) - (\bar{y}_{r2j} - \bar{y}_{n2j})$$

$$\bar{\delta} = \frac{1}{J-1} \sum_{j=2}^J [(\bar{y}_{r1} - \bar{y}_{n1}) - (\bar{y}_{r2} - \bar{y}_{n2})] = \frac{1}{J-1} \sum_{j=2}^J D_j$$

The following logistic regression $\bar{\delta}$ model can be used to estimate whether $\bar{\delta} = 0$:

$$y_i = \beta_0 + \beta_1 T_i + \beta_2 G_{1ij} + \beta_3 G_{2ij} + \beta_4 JP_{ij} + \beta_5 T_i JP_{ij} + \beta_6 G_{1ij} JP_{ij} + \beta_7 G_{2ij} JP_{ij}$$

$$D_j = \frac{m_{j1}}{n_{j1}} D_{j1} - \frac{m_{j2}}{n_{j2}} D_{j2}$$

$$D_{j1} = \beta_0 + \beta_1 + \beta_2 + \beta_4 j + \beta_5 j + \beta_6 j - \beta_0 - \beta_1 - \beta_4 j - \beta_5 j = \beta_2 + \beta_6 j$$

$$D_{j2} = \beta_0 + \beta_3 + \beta_4 j + \beta_7 j - \beta_0 - \beta_4 j = \beta_3 + \beta_7 j$$

$$D_j = \frac{m_{j1}}{n_{j1}} (\beta_2 + \beta_6 j) - \frac{m_{j2}}{n_{j2}} (\beta_3 + \beta_7 j)$$

$$\bar{\delta} = \frac{1}{59} \sum_{j=2}^{60} \left[\frac{m_{j1}}{n_{j1}} (\beta_2 + \beta_6 j) - \frac{m_{j2}}{n_{j2}} (\beta_3 + \beta_7 j) \right]$$

$$\theta_D = \frac{1}{59} \left[\beta_2 \sum_{j=2}^{60} \left(\frac{m_{j1}}{n_{j1}} \right) + \beta_6 \sum_{j=2}^{60} \left(\frac{m_{j1} j}{n_{j1}} \right) - \beta_3 \sum_{j=2}^{60} \left(\frac{m_{j2}}{n_{j2}} \right) - \beta_7 \sum_{j=2}^{60} \left(\frac{m_{j2} j}{n_{j2}} \right) \right]$$

Table A. 3. Missing data: IEB variables.

IEB Variable	Wave 1		Wave 2	
	# Missing	% Missing	# Missing	% Missing
<i>Employment</i>				
Gross monthly income	0	0	0	0
Currently registered as disabled	554	15.44	750	20.9
Registered as employed	29	0.81	29	0.81
Employed full-time	29	0.81	29	0.81
Employed part-time	29	0.81	29	0.81
Registered as unemployed	29	0.81	29	0.81
Length of unemployment	55	2.14	71	3.19
<i>Household characteristics</i>				
Lives in former East German states	31	0.86	32	0.89
Has children in the benefit community	128	3.57	128	3.57
<i>Person characteristics</i>				
Marital status	115	3.21	115	3.21
Sex	29	0.81	29	0.81
German national	31	0.86	(No changes)	
Age	105	2.93%	105	2.93%

Table A. 4. Missing data: Paradata variables.

Variable	Wave 1		Wave 2	
	# Missing	% Missing	# Missing	% Missing
<i>Household interview:</i>				
Interviewer age	12	0.33	0	0
Interviewer education	12	0.33	0	0
Interviewer sex	12	0.33	0	0
<i>Person interview:</i>				
Interviewer age	65	1.81	64	3.01
Interviewer education	65	1.81	64	3.01
Interviewer sex	65	1.81	64	3.01

Table A. 5. Item nonresponse in PASS variables.

Variable	Wave 1		Wave 2	
	# Missing	% Missing	# Missing	% Missing
<i>Employment:</i>				
Gross monthly income	16	4.07	76	16.45
Currently registered as disabled	4	0.15	4	0.21
Employed full-time	0	0	5	1.08
Employed part-time	0	0	5	1.08
Registered as unemployed	19	0.69	1,922	100
Length of unemployment	272	15.29	103	9.03
<i>Household characteristics</i>				
Has partner in the house	0	0	0	0
German spoken in household	20	2.65	(not asked)	
Has children in the household	27	0.98	8	0.42
<i>Person characteristics</i>				
Marital status	47	1.71	12	0.62
Sex	0	0	(not asked)	
German national	3	0.11	(not asked)	
Age	0	0	0	0
Number of family and friends outside the household	29	1.13	16	0.9
Consent to link IEB and PASS	133	3.71	76	3.58

Table A. 6. Fraction of missing information for IEB variables.

Variable	Wave 1	Wave 2
<i>Employment</i>		
Gross monthly income	0.0000	0.0030
Currently registered as disabled	0.0050	0.0164
Registered as employed	0.0001	0.0000
Employed full-time	0.0000	0.0000
Employed part-time	0.0001	0.0000
Registered as unemployed	0.0000	0.0000
Length of unemployment	0.0000	0.0004
<i>Household characteristics</i>		
Lives in former East German states	0.0000	0.0000
Has children in the benefit community	0.0002	0.0002
<i>Person characteristics</i>		
Marital status	0.0000	0.0000
Sex	0.0000	0.0000
German national	0.0000	0.0000
Level of education	0.0099	0.0004
Age	0.0000	0.0000

Table A. 7. Fraction of missing information for paradata variables in PASS.

Variable	Wave 1	Wave 2
<i>Household interview:</i>		
Interviewer age	0.0000	0.0000
Interviewer education	0.0000	0.0000
Interviewer sex	0.0000	0.0000
<i>Person interview:</i>		
Interviewer age	0.0000	0.0007
Interviewer education	0.0000	0.0001
Interviewer sex	0.0000	0.0002

Table A. 8. Fraction of missing information for PASS survey variables.

Variable	Wave 1	Wave 2
<i>Employment:</i>		
Gross monthly income	0.0004	0.0709
Currently registered as disabled	0.0015	0.0023
Employed full-time	0.0000	0.0000
Employed part-time	0.0000	0.0000
Registered as unemployed	0.0253	0.0023
Length of unemployment	0.0000	0.0000
<i>Household characteristics</i>		
Has partner in the house	0.0082	0.0019
German spoken in household	0.0001	(not asked)
Has children in the household	0.0176	0.0009
<i>Person characteristics</i>		
Marital status	0.0214	0.0008
Sex	0.0001	(not asked)
German national	0.0127	(not asked)
Age	0.0011	(not asked)
Number of family and friends outside the household	0.0000	0.0000
Consent to link IEB and PASS	0.0000	0.0000

Table A. 9. Odds ratios and standard errors for response propensity models: Interaction models.

		Odds ratio (std. err.): Baseline	Odds ratio (std. err.): All journals
Having any sexual partners	Sex partner	0.601(0.210)	0.510(0.312)
	Mean # days in prev. journal periods	0.726(0.017)***	0.714(0.011)***
	JP	1.021(0.014)	1.014(0.009)
	JP * sex partner	1.001(0.016)	1.008(0.015)
Use of non-coital contraception	Use of non-coital contraception	1.226(0.363)	1.032(0.634)
	Mean # days in prev. journal periods	0.723(0.017)***	0.720(0.014)***
	JP	1.023(0.008)**	1.028(0.036)
	JP * use	1.003(0.016)	0.994(0.032)
Use of coital contraception	Use of coital contraception		.933(.255)
	Mean # days in prev. journal periods		.724(.021)***
	JP		1.022(.011)†
	JP * use		.975(.012)†
Living with a parent	Living with a parent	1.183(0.193)	
	Mean # days in prev. journal periods	0.713(0.011)***	
	JP	1.02(.006)**	
	JP * living with parent	.999(.008)	
Pregnancy intention	Pregnancy intention	.996(.188)	.097(.064)*
	Mean # days in prev. journal periods	.713(.011)***	.72(.012)***
	JP	1.024(.008)**	1.082(.045)
	JP * pregnancy intention	.994(.01)	.931(.041)
Pregnancy avoidance	Pregnancy avoidance	2.067(1.268)	2.314(2.295)
	Mean # days in prev. journal periods	.713(.011)***	.714(.011)***
	JP	1.033(.028)	.983(.028)
	JP * pregnancy avoidance	.987(.027)	1.04(.031)

Table A. 9. Odds ratios and standard errors for response propensity models (contd.)

		Odds ratio (std err.): Baseline	Odds ratio (std err.): All journals
Change in relationship status	Relationship change		.182(.097) *
	Mean # days in prev. journal periods		.731(.015) ***
	JP		1.03(.02)
	JP * relationship change		.976(.028)
Currently enrolled in school	Part-time enrollment	1.719(0.620)	0.267(0.115) *
	Full-time enrollment	1.632(0.432) †	6.798(7.495)
	Mean # days in prev. journal periods	0.723(0.017) ***	0.713(0.011) ***
	JP	1.026(0.013) *	1.019(0.015)
	Mean # days * Part-time	0.984(.018)	1.020(0.027)
	Mean # days * Full-time	0.996(.014)	0.967(0.035)
Type of school	2 year junior/community college	1.775(0.652)	1.459(1.268)
	4 year college	2.599(1.015) *	3.446(3.391)
	Voc, tech, trade, or other school	1.873(1.05)	0.150(0.141)
	Mean # days in prev. journal periods	0.721(0.019) ***	0.709(0.015) ***
	JP	1.037(0.020) †	0.979(0.024)
	JP * 2 year jr/comm.	0.983(0.022)	1.041(0.030)
	JP * 4 year college	0.978(0.023)	1.038(0.042)
	JP * Voc, tech, trade, other	0.957(0.029)	1.041(0.038)

Table A. 10. Odds ratios and standard errors from logistic regression models predicting latency of response: Panel respondent characteristics.

		Baseline		All journals	
Currently enrolled in school	Part-time enrollment	.834(0.173)	.824(0.152)	.856(0.116)	.881(0.089)
	Full-time enrollment	.654(0.148)**	.731(0.112)*	.646(0.087)***	.754(0.068)***
	LATEij-1		1.132(0.027)***		1.13(0.015)***
	JP		.997(0.005)		.997(0.003)
	MNLENij-1		1.197(0.020)***		1.213(0.015)***
	LATEij-1*JP		1.003(0.001)		1.002(0.001)**
Type of school	2 year junior/community college	1.01(0.194)	1.087(0.162)	.873(0.149)	.9(0.127)
	4 year college	.766(0.229)	.95(0.174)	.714(0.149)*	.815(0.122)
	Voc, tech, trade, or other school	1.423(0.298)	1.452(0.242)	.993(0.191)	1.079(0.163)
	LATEij-1		1.12(0.034)**		1.127(0.017)
	JP		1(0.006)		1.001(0.003)
	MNLENij-1		1.206(0.026)***		1.214(0.018)
Having any sexual partners	LATEij-1*JP		1.003(0.002)*		1.003(0.001)
	Sex partner	1.948(0.146)***	1.422(0.119)**	1.14(0.070)	1.108(0.058)
	LATEij-1		1.127(0.027)***		1.132(0.015)***
	JP		.997(0.005)		.997(0.003)
	MNLENij-1		1.196(0.020)***		1.214(0.015)***
	LATEij-1*JP		1.003(0.001)		1.002(0.001)**
Use of non-coital contraception	Use of non-coital contraception	1.027(0.173)	.978(0.112)	.795(0.084)**	.895(0.066)
	LATEij-1		1.133(0.028)***		1.13(0.016)***
	JP		.997(0.005)		.997(0.003)
	MNLENij-1		1.2(0.021)***		1.213(0.016)***
	LATEij-1*JP		1.003(0.001)		1.003(0.001)***

Table A. 9. Odds ratios and standard errors for response propensity models (contd.)

		Odds ratio (std err.): Baseline	Odds ratio (std err.): All journals
Change in relationship status	Relationship change		.182(.097) *
	Mean # days in prev. journal periods		.731(.015) ***
	JP		1.03(.02)
	JP * relationship change		.976(.028)
Currently enrolled in school	Part-time enrollment	1.719(0.620)	0.267(0.115) *
	Full-time enrollment	1.632(0.432) †	6.798(7.495)
	Mean # days in prev. journal periods	0.723(0.017) ***	0.713(0.011) ***
	JP	1.026(0.013) *	1.019(0.015)
	Mean # days * Part-time	0.984(.018)	1.020(0.027)
	Mean # days * Full-time	0.996(.014)	0.967(0.035)
Type of school	2 year junior/community college	1.775(0.652)	1.459(1.268)
	4 year college	2.599(1.015) *	3.446(3.391)
	Voc, tech, trade, or other school	1.873(1.05)	0.150(0.141)
	Mean # days in prev. journal periods	0.721(0.019) ***	0.709(0.015) ***
	JP	1.037(0.020) †	0.979(0.024)
	JP * 2 year jr/comm.	0.983(0.022)	1.041(0.030)
	JP * 4 year college	0.978(0.023)	1.038(0.042)
JP * Voc, tech, trade, other	0.957(0.029)	1.041(0.038)	

Table A. 10. Odds ratios and standard errors from logistic regression models predicting latency of response: Panel respondent characteristics. (contd.)

		Baseline		All journals	
Living with a parent	Living with a parent	.874(0.086)	.887(0.058)		
	LATEij-1		1.132(0.015)		
	JP		.997(0.003)		
	MNLENij-1		1.214(0.015)		
	LATEij-1*JP		1.002(0.001)		
Pregnancy intention	Pregnancy intention	1.266(0.097)*	1.099(0.064)	1.426(0.102)***	1.157(0.082)
	LATEij-1		1.132(0.015)***		1.131(0.015)***
	JP		.997(0.003)		.997(0.003)
	MNLENij-1		1.215(0.015)***		1.214(0.015)***
	LATEij-1*JP		1.002(0.001)**		1.002(0.001)**
Pregnancy avoidance	Pregnancy avoidance	1.21(0.426)	1.095(0.309)	.727(0.240)	.93(0.205)
	LATEij-1		1.132(0.015)***		1.132(0.015)***
	JP		.997(0.003)		.997(0.003)
	MNLENij-1		1.215(0.015)***		1.215(0.015)***
	LATEij-1*JP		1.002(0.001)**		1.002(0.001)**
Change in relationship status	Relationship change			1.833(0.097)***	1.742(0.093)***
	LATEij-1				1.142(0.017)***
	JP				1.001(0.003)
	MNLENij-1				1.207(0.018)***
	LATEij-1*JP				1.001(0.001)

Table A. 11. Relative risk ratios and standard errors: Multinomial logistic regression models predicting use of a mode, with substantive predictors from baseline variables.

		Baseline			
		Bivariate model		Multivariate model	
		Inbound CATI	Outbound CATI	Inbound CATI	Outbound CATI
Currently enrolled in school	Part-time enrollment	.55(0.504)	.652(0.341)	.501(0.338)	.629(0.275)
	Full-time enrollment	.43(0.328)*	.432(0.247)**	.494(0.218)*	.53(0.177)**
	# Prev. inbound CATI journals			1.245(0.103)	1.15(0.090)
	# Prev. outbound CATI journals			1.289(0.049)***	1.326(0.045)***
	MNLENij-1			1.047(0.029)	1.07(0.020)**
Type of school	2 year junior/community college	.671(0.508)	.742(0.308)	.593(0.369)	.685(0.243)
	4 year college	.221(0.551)*	.287(0.350)**	.299(0.411)*	.35(0.288)**
	Voc, tech, trade, or other school	1.028(0.608)	1.576(0.620)	1.03(0.468)	1.405(0.475)
	# Prev. inbound CATI journals			1.269(0.133)	1.192(0.119)
	# Prev. outbound CATI journals			1.3(0.070)**	1.3(0.055)**
	MNLENij-1			1.052(0.034)	1.091(0.024)**
Having any sexual partners	Sex partner	1.527(0.377)	2.746(0.290)**	1.707(0.260)	2.411(0.236)**
	# Prev. inbound CATI journals			1.267(0.108)	1.178(0.096)
	# Prev. outbound CATI journals			1.276(0.049)***	1.302(0.046)***
	MNLENij-1			1.043(0.029)	1.061(0.020)*
Use of non-coital contraception	Use of non-coital contraception	.535(0.350)	.629(0.257)	.801(0.278)	.771(0.211)
	# Prev. inbound CATI journals			1.262(0.120)	1.184(0.108)
	# Prev. outbound CATI journals			1.262(0.068)**	1.27(0.062)**
	MNLENij-1			1.061(0.029)	1.088(0.020)**

Table A. 11. Relative risk ratios and standard errors: Multinomial logistic regression models predicting use of a mode, with substantive predictors from baseline variables. (contd).

		Baseline			
		Bivariate model		Multivariate model	
		Inbound CATI	Outbound CATI	Inbound CATI	Outbound CATI
	Living with a parent	.815(0.237)	.904(0.172)	.898(0.148)	.926(0.121)
Living with a parent	# Prev. inbound CATI journals			1.247(0.065)***	1.153(0.062)*
	# Prev. outbound CATI journals			1.276(0.035)***	1.315(0.032)***
	MNLENij-1			1.051(0.021)*	1.078(0.015)***
	Pregnancy intention	1.185(0.325)	1.454(0.241)	1.026(0.218)	1.195(0.191)
Pregnancy intention	# Prev. inbound CATI journals			1.249(0.065)*	1.156(0.061)
	# Prev. outbound CATI journals			1.275(0.035)***	1.311(0.032)***
	MNLENij-1			1.051(0.021)	1.078(0.015)**
	Pregnancy avoidance	1.07(0.729)	.765(0.650)	1.251(0.687)	1.111(0.485)
Pregnancy avoidance	# Prev. inbound CATI journals			1.249(0.065)**	1.154(0.061)*
	# Prev. outbound CATI journals			1.276(0.035)***	1.317(0.032)***
	MNLENij-1			1.051(0.020)*	1.078(0.015)***

Table A. 12. Relative risk ratios and standard errors: Multinomial logistic regression models predicting use of a mode, with substantive predictors across all journals.

		All journals				
		Inbound CATI	Outbound CATI	Inbound CATI	Outbound CATI	
168	Currently enrolled in school	Part-time enrollment	.774(0.264)	.705(0.190)	.794(0.231)	.748(0.185)
		Full-time enrollment	.395(0.210)***	.398(0.169)***	.545(0.159)***	.53(0.133)***
		# Prev. inbound CATI journals			1.244(0.064)***	1.149(0.060)*
		# Prev. outbound CATI journals			1.261(0.035)***	1.3(0.032)***
		MNLENij-1			1.045(0.021)*	1.074(0.015)***
	Type of school	2 year junior/community college	.733(0.316)	.592(0.225)*	.497(0.241)**	.524(0.183)***
		4 year college	.242(0.307)***	.224(0.238)***	.243(0.264)***	.25(0.193)***
		Voc, tech, trade, or other school	.823(0.369)	.717(0.312)	.718(0.422)	.688(0.323)
		# Prev. inbound CATI journals			1.301(0.087)**	1.213(0.082)*
		# Prev. outbound CATI journals			1.259(0.051)***	1.269(0.047)***
Having any sexual partners	MNLENij-1			1.044(0.024)	1.087(0.018)***	
	Sex partner	.729(0.174)	.859(0.122)	1.035(0.136)	.984(0.106)	
	# Prev. inbound CATI journals			1.25(0.065)***	1.154(0.061)*	
	# Prev. outbound CATI journals			1.275(0.035)***	1.315(0.032)***	
Use of non-coital contraception	MNLENij-1			1.051(0.020)*	1.078(0.015)***	
	Use of non-coital contraception	.446(0.206)***	.652(0.146)**	.72(0.156)*	.868(0.118)	
	# Prev. inbound CATI journals			1.245(0.067)**	1.161(0.063)*	
	# Prev. outbound CATI journals			1.255(0.035)***	1.289(0.032)***	
				1.05(0.020)*	1.089(0.015)***	

Table A. 12. Relative risk ratios and standard errors: Multinomial logistic regression models predicting use of a mode, with substantive predictors across all journals.(contd).

		All journals			
		Model X		Model y	
		Inbound CATI	Outbound CATI	Inbound CATI	Outbound CATI
		Model X		Model y	
169	Use of coital contraception	1.662(0.239)*	1.898(0.174)***	1.766(0.192)**	1.775(0.155)***
	Use of coital contraception # Prev. inbound CATI journals			1.649(0.083)***	1.471(0.080)***
	Use of coital contraception # Prev. outbound CATI journals			1.112(0.040)**	1.163(0.035)***
	MNLENij-1			1(0.039)	1.031(0.026)
	Pregnancy intention	1.079(0.213)	1.51(0.163)*	1.057(0.183)	1.331(0.166)
	Pregnancy intention # Prev. inbound CATI journals			1.25(0.065)***	1.156(0.061)*
	Pregnancy intention # Prev. outbound CATI journals			1.271(0.035)***	1.311(0.032)***
	MNLENij-1			1.051(0.021)*	1.077(0.015)***
	Pregnancy avoidance	1.208(0.358)	.969(0.307)	.922(0.380)	.905(0.320)
	Pregnancy avoidance # Prev. inbound CATI journals			1.249(0.065)***	1.155(0.061)*
	Pregnancy avoidance # Prev. outbound CATI journals			1.275(0.035)***	1.315(0.032)***
	MNLENij-1			1.051(0.021)*	1.078(0.015)***
Change in relationship status	1.155(0.215)	1.192(0.182)	.886(0.200)	1.024(0.177)	
Change in relationship status # Prev. inbound CATI journals			1.52(0.067)***	1.358(0.064)***	
Change in relationship status # Prev. outbound CATI journals			1.181(0.037)***	1.252(0.035)***	
MNLENij-1			1.022(0.027)	1.049(0.019)*	

Table A. 13. Change in overall nonresponse bias across journal periods: multinomial logistic regression coefficients.

	Baseline				All journals			
	complete	JP	complete*JP	_cons	complete	JP	complete*JP	_cons
Currently enrolled in school								
Part-time enrollment	.434(.424)	.017(.02)	-.005(.021)	-1.23(.403)**	-1.503(.327)**	-.028(.024)	.031(.025)	.742(.297)†
Full-time enrollment	.58(.268)*	.016(.014)	-.003(.014)	.221(.268)	1.778(1.178)	.031(.046)	-.026(.046)	-.915(1.174)
Type of school								
2 year junior/community college	.291(.395)	.002(.021)	.004(.022)	.322(.372)	.302(.685)	-.008(.028)	.058(.029)†	.008(.67)
4 year college	.673(.434)	.006(.023)	0(.024)	.155(.413)	1.515(.968)	.022(.039)	.034(.04)	-.971(.957)
Voc, tech, trade, or other school	1.098(.813)	.043(.032)	-.051(.037)	-1.887(.826)*	-1.728(.979)	.001(.036)	.04(.039)	.623(.956)
Any sexual partner	-.993(.412)*	-.017(.019)	.006(.019)	2.061(.426)***	-.603(.646)	-.017(.022)	.003(.022)	.351(.645)
Use of non-coital contraception	.174(.297)	-.007(.017)	.008(.018)	-.523(.31)	.194(.564)	.009(.033)	-.009(.033)	-.849(.56)
Use of coital contraception					.134(.308)	.018(.015)	-.038(.016)*	-.082(.292)
Living with a parent	.291(.205)	.015(.01)	-.006(.01)	-.668(.196)**				
Pregnancy intention	-.237(.158)	.006(.009)	.001(.009)	.313(.165)†	-2.275(.69)*	.073(.05)	-.081(.051)	.519(.681)
Pregnancy avoidance	.208(.949)	-.025(.036)	.006(.04)	4.702(.935)***	.837(.917)	-.023(.029)	.045(.032)	2.796(.909)*
Change in relationship status	-1.885(.529)*	.015(.029)	-.028(.029)	-.439(.525)				

†p<.1; *p<.05; **p<.01; ***p<.001.

Note: reference category for Currently enrolled in school is “Not enrolled”. The reference category for Type of school is “High school or less”.

Table A. 14. Coefficients and standard errors for nonresponse bias model in RDSL.

	Baseline			All journals		
	Complete	JP	Complete* JP	Complete	JP	Complete* JP
Currently enrolled in school						
Part-time enrollment	1.578(.692)	1.017(.02)	.994(.021)	.226(.074)**	.973(.023)	1.031(.025)
Full-time enrollment	1.826(.504)*	1.017(.014)	.997(.015)	6.022(7.098)	1.031(.047)	.973(.045)
Type of school						
2 year junior/community college	1.345(.553)	1.002(.021)	1.004(.022)	1.295(.889)	.992(.027)	1.061(.031)†
4 year college	1.986(.887)	1.006(.023)	.999(.024)	4.389(4.254)	1.022(.04)	1.036(.042)
Voc, tech, trade, or other school	3.031(2.488)	1.044(.033)	.95(.036)	.17(.167)	1.001(.036)	1.043(.041)
Any sexual partner	.364(.15)*	.984(.019)	1.007(.019)	.456(.294)	.983(.021)	1.01(.022)
Use of non-coital contraception	1.198(.358)	.993(.017)	1.008(.018)	1.168(.659)	1.009(.033)	.993(.033)
Use of coital contraception				1.33(.277)	1.016(.01)	.995(.01)
Living with a parent	1.33(.277)	1.016(.01)	.995(.01)			
Pregnancy intentions	.972(.155)	1.006(.009)	.994(.009)	.104(.072)*	1.075(.054)	.922(.047)
Pregnancy avoidance	1.386(1.339)	.975(.035)	1.002(.04)	2.157(1.985)	.978(.029)	1.049(.034)
Change in relationship status				.152(.08)*	1.015(.029)	.973(.028)

Table A. 15. Coefficients and standard errors for of change in nonresponse bias between treatments: Baseline variables.

	T	G1	G2	JP	T*JP	G1*JP	G2*JP	cons.
Currently enrolled in school								
Part-time enrollment	0.280(0.433)	0.265(0.383)	0.456(0.438)	0.017(0.020)	-0.009(0.025)	0.003(0.022)	-0.006(0.021)	-1.23(0.403)**
Full-time enrollment	0.037(0.265)	0.650(0.225)**	0.602(0.276)*	0.016(0.014)	-0.004(0.018)	0.001(0.014)	-0.003(0.015)	0.221(0.268)
Type of school								
2 year junior/community college	0.075(0.381)	0.210(0.337)	0.296(0.411)	0.002(0.021)	-0.002(0.027)	0.003(0.023)	0.004(0.022)	0.322(0.372)
4 year college	-0.113(0.387)	0.927(0.343)**	0.686(0.447)	0.006(0.023)	-0.013(0.028)	0.009(0.019)	-0.001(0.024)	0.155(0.413)
Voc, tech, trade, or other school	1.150(0.747)	-0.179(0.523)	1.109(0.821)	0.043(0.032)	-0.043(0.037)	-0.009(0.026)	-0.051(0.038)	-1.887(0.826)*
Any sexual partner	-0.186(0.401)	0.907(0.271)**	-1.01(0.412)*	-0.017(0.019)	-0.008(0.024)	0.015(0.016)	0.007(0.019)	2.061(0.426)**
Use of non-coital contraception	-0.078(0.342)	0.255(0.244)	0.181(0.299)	-0.007(0.017)	-0.005(0.019)	0.011(0.012)	0.008(0.018)	-0.523(0.310)
Living with a parent	0.161(0.174)	0.115(0.154)	0.285(0.209)	0.015(0.010)	-0.006(0.012)	0.002(0.010)	-0.005(0.010)	0.668(0.196)**
Pregnancy intentions	0.097(0.226)	-0.153(0.124)	-0.029(0.159)	0.006(0.009)	0.001(0.014)	-0.007(0.009)	-0.006(0.009)	0.313(0.165)†
Pregnancy avoidance	-0.795(0.993)	1.038(0.509)*	0.326(0.966)	-0.025(0.036)	0.051(0.033)	-0.037(0.032)	0.002(0.040)	4.702(0.935)**

172

Table A. 16. Coefficients and standard errors for change in nonresponse bias between treatments: Variables across all journals.

	T	G1	G2	JP	T*JP	G1*JP	G2*JP	cons.
Currently enrolled in school								
Part-time enrollment	0.073(0.993)	-1.222(0.738)	1.489(0.329)**	-0.028(0.024)	0.003(0.046)	0.019(0.025)	0.031(0.025)	0.742(0.297)†
Full-time enrollment	0.455(1.307)	1.447(0.803)	1.795(1.179)	0.031(0.046)	-0.015(0.047)	-0.013(0.022)	-0.027(0.046)	-0.915(1.174)
Type of school								
2 year junior/community college	0.300(0.736)	-0.030(0.649)	0.258(0.687)	-0.008(0.028)	-0.010(0.026)	0.067(0.032)†	0.060(0.030)†	0.008(0.670)
4 year college	1.376(1.544)	0.188(0.987)	1.479(0.969)	0.022(0.039)	-0.053(0.065)	0.086(0.033)*	0.035(0.040)	-0.971(0.957)
Voc, tech, trade, or other school	0.231(0.642)	-1.718(0.644)*	-1.77(0.982)	0.001(0.036)	-0.012(0.024)	0.048(0.026)	0.042(0.039)	0.623(0.956)
Any sexual partner	-0.558(0.790)	-0.190(0.550)	-0.786(0.646)	-0.017(0.022)	0.027(0.025)	-0.016(0.021)	0.010(0.022)	0.351(0.645)
Use of non-coital contraception	-0.331(0.636)	0.537(0.317)	0.156(0.564)	0.009(0.033)	0.020(0.026)	-0.027(0.013)†	-0.007(0.033)	-0.849(0.560)
Use of coital contraception	0.011(0.305)	0.087(0.232)	0.134(0.308)	0.018(0.015)	-0.022(0.018)	-0.015(0.014)	-0.038(0.016)*	-0.082(0.292)
Pregnancy intentions	0.000(0.598)	1.944(0.299)**	-2.264(0.691)*	0.073(0.050)	-0.030(0.046)	-0.058(0.025)†	-0.082(0.051)	0.519(0.681)
Pregnancy avoidance	-0.590(0.955)	1.269(0.454)*	0.769(0.920)	-0.023(0.029)	0.018(0.033)	0.028(0.017)	0.047(0.033)	2.796(0.909)*
Change in relationship status	0.202(0.992)	-1.762(0.559)*	-1.885(0.529)*	0.015(0.029)	-0.017(0.052)	-0.017(0.026)	-0.028(0.029)	-0.439(0.525)

Table A. 17. Coefficients and standard errors for nonresponse bias model: PASS.

	T	G1	G2
<i>Employment</i>			
Gross monthly income	-11.034(15.998)	15.584(71.093)	-3.733(71.164)
Current profession: blue collar	.983(.046)	.636(.158)†	.64(.172)
Current profession: white collar	1.05(.051)	1.683(.382)*	1.754(.439)*
Currently registered as disabled	.979(.053)	.987(.262)	.954(.272)
Registered as employed	.947(.019)**	.924(.075)	.85(.07)†
Employed full-time	.954(.021)*	.848(.078)†	.795(.074)*
Employed part-time	.946(.025)*	.982(.113)	.897(.104)
Registered as unemployed	1.043(.026)†	1.014(.108)	1.085(.117)
Length of unemployment	-4.651(3.177)	6.342(12.984)	-1.869(13.351)
<i>Household characteristics</i>			
Lives in former East German states	.988(.022)	1.034(.089)	1.01(.089)
Has children in the benefit community	1.023(.022)	.865(.084)	.911(.088)
<i>Person characteristics</i>			
Married (vs unmarried)	.948(.02)*	.81(.097)†	.753(.092)*
Female	1.009(.021)	1.314(.104)**	1.3(.104)**
German national	1.098(.023)***	1.274(.195)	1.475(.225)*
Age	-0.354(.132)**	2.829(.507)***	1.972(.522)***

Appendix B. Question wording in Relationship Dynamics and Social Life

Note: Question wording for selected predictors and confounding variables are presented here. The full questionnaire is available upon request.

Currently enrolled in school (at baseline):

Are you going to school at all now?

Yes

No

Currently enrolled in school (journals):

Which of the following describes your current enrollment in school?

Not enrolled in school

Attending school part-time

Attending school full-time

Type of school (baseline and journals):

What kind of school do you attend?

High school or less

2 year junior or community college

4 year college

Vocational, technical, or trade school

Other

Having any sexual partners (baseline):

Have you ever had sexual intercourse? Sexual intercourse is when a man inserts his penis into a woman's vagina.

Yes

No

Having any sexual partners (journals)

In the past ----- days (since -----), did you have sexual intercourse with -----? By sexual intercourse, we mean when a man puts his penis into a woman's vagina.

Yes

No

In the past ----- days (since -----), did you have sexual intercourse with anyone other than -----?

Yes

No

Use of noncoital contraception (baseline and journals)

Did you use birth control pills (for any reason)?

Yes

No

[IF NO OR DK/RF]

Did you use the birth control patch (for any reason)?

Yes

No

[IF NO OR DK/RF]

Did you use the NuvaRing?

Yes

No

[IF NO OR DK/RF]

Did you use Depo-Provera or any other type of contraceptive shot?

Yes

No

[IF NO OR DK/RF]

Did you have an implant such as Implanon™ or another contraceptive implant?

Yes

No

[IF NO OR DK/RF]

Did you have an IUD?

Yes

No

Use of coital contraception

In the past ----- days (since -----), did you use a condom?

Yes

No

In the past ----- days (since -----), did you use a diaphragm or cervical cap?

Yes

No

In the past ----- days (since -----), did you use spermicide?

Yes

No

In the past ----- days (since -----), did you use a female condom?

Yes

No

In the past ----- days (since -----), did you use the morning after pill?

Yes

No

In the past ----- days (since -----), did you use a condom?

Yes
No

Living with a parent (baseline only)

Who do you currently live with? Please select one or more from the list.

Biological mother
Biological father
Adoptive mother
Adoptive father
Step-mother
Step-father
Grandmother
Grandfather
Husband
Other relative(s)
Foster mother
Foster father
Romantic partner
Roommate(s)
No one/Live alone

Pregnancy intention (baseline)

You know, getting pregnant and having a baby is a big event, one that has a lot of consequences. Most people your age have some positive and some negative feelings about getting pregnant and having a child. For this reason we are going to ask you first how much you want to get pregnant, using a scale from 0 to 5. Then we are going to ask you how much you want to avoid getting pregnant, using a scale from 0 to 5.

First, how much do you want to get pregnant during the next month? Please give me a number between 0 and 5, where 0 means you don't at all want to get pregnant and 5 means you really want to get pregnant.

Pregnancy intention (journals)

How much do you want to get pregnant during the next month? Please give a number between 0 and 5, where 0 means you don't at all want to get pregnant and 5 means you really want to get pregnant.

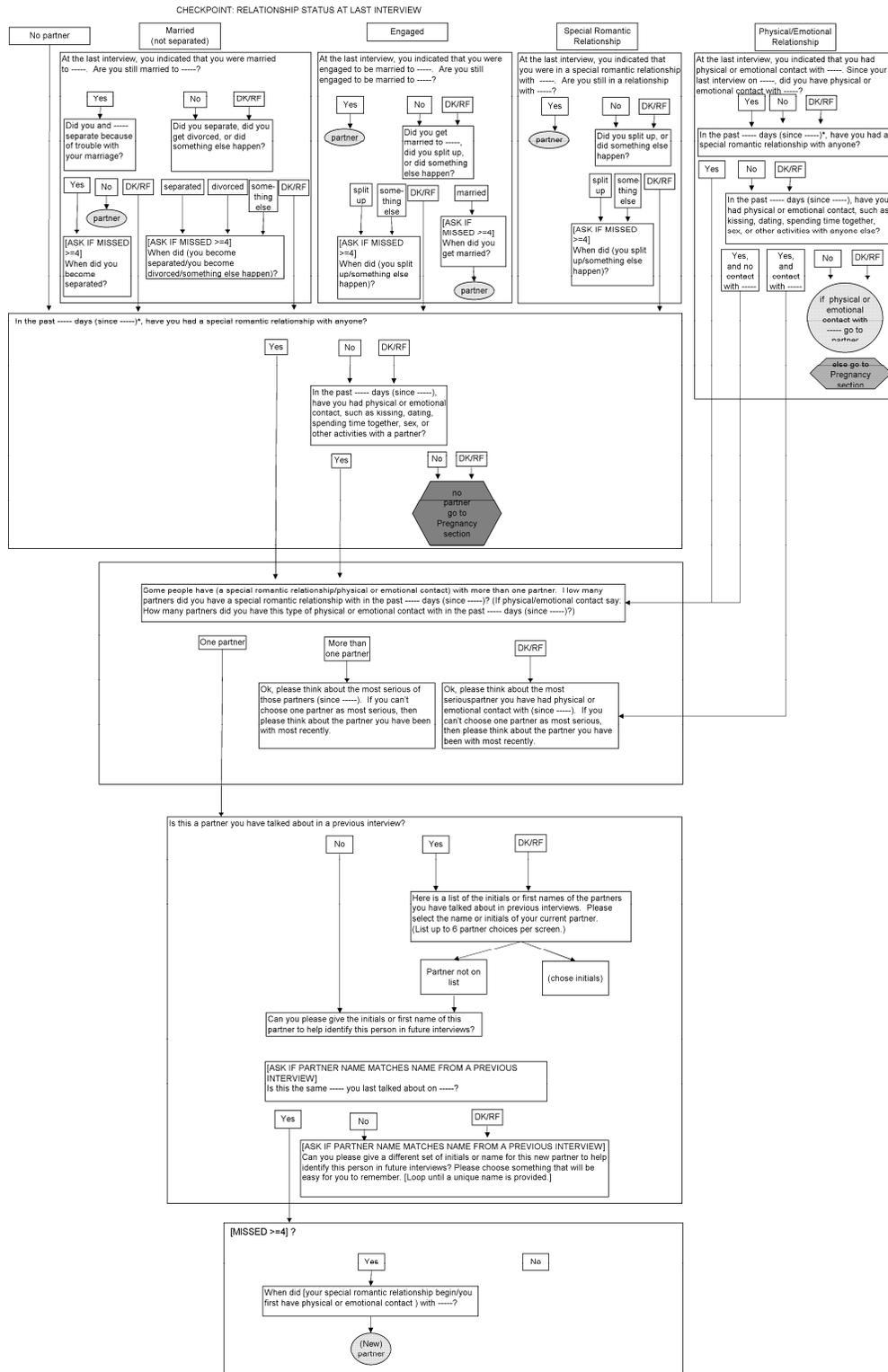
Pregnancy avoidance (baseline)

And next, how much do you want to avoid getting pregnant during the next month? Please give me a number between 0 and 5, where 0 means you don't at all want to avoid getting pregnant and 5 means you really want to avoid getting pregnant.

Pregnancy avoidance (journals)

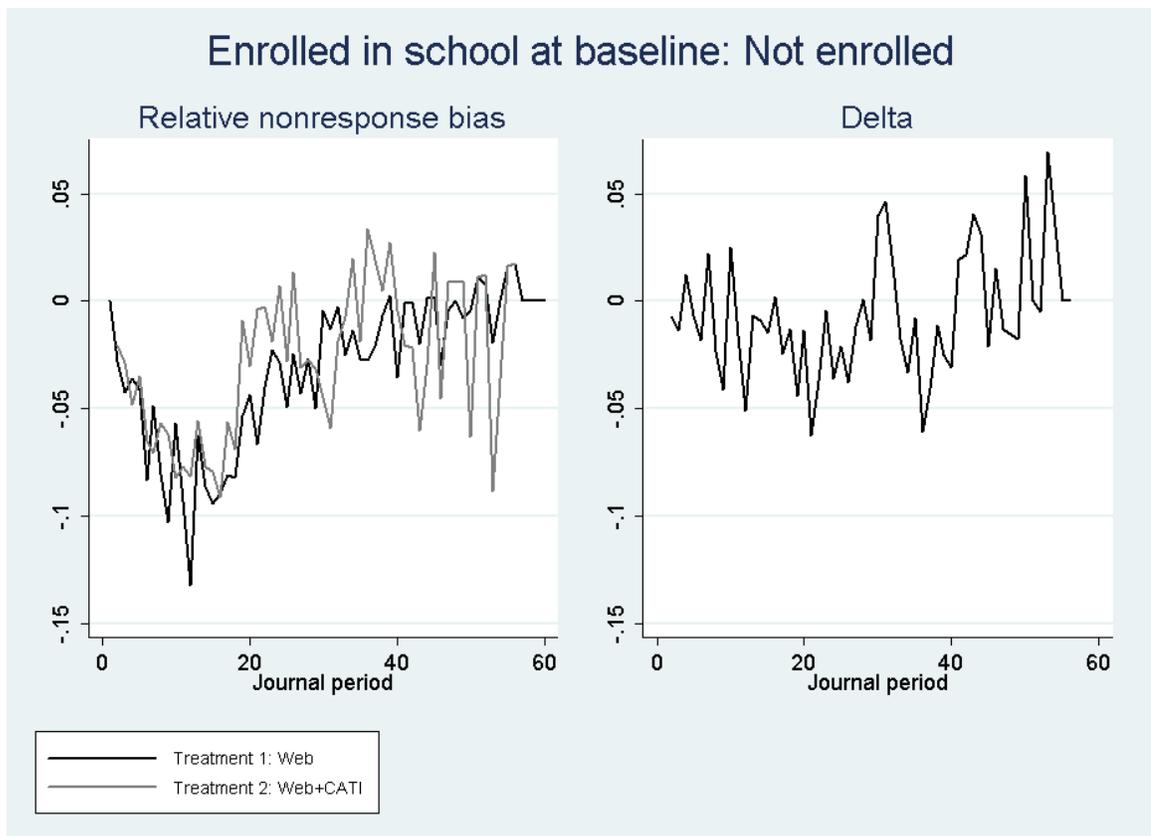
How much do you want to avoid getting pregnant during the next month? Please give a number between 0 and 5, where 0 means you don't at all want to avoid getting pregnant and 5 means you really want to avoid getting pregnant.

Figure B. 1. Question series for having a partner.

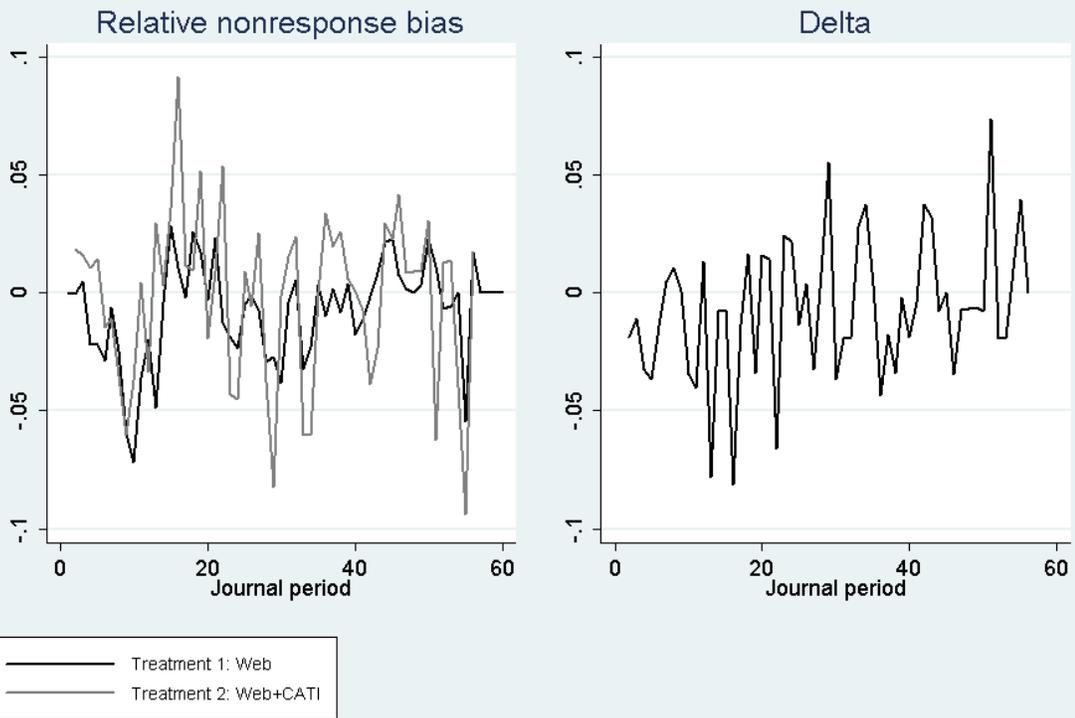


* For all Identify Partner Measures that begin with "In the past x days (since y)": If 14 or more days since the last interview then fill with "In the past 7 days, (since [Today's Date minus 7 days])"

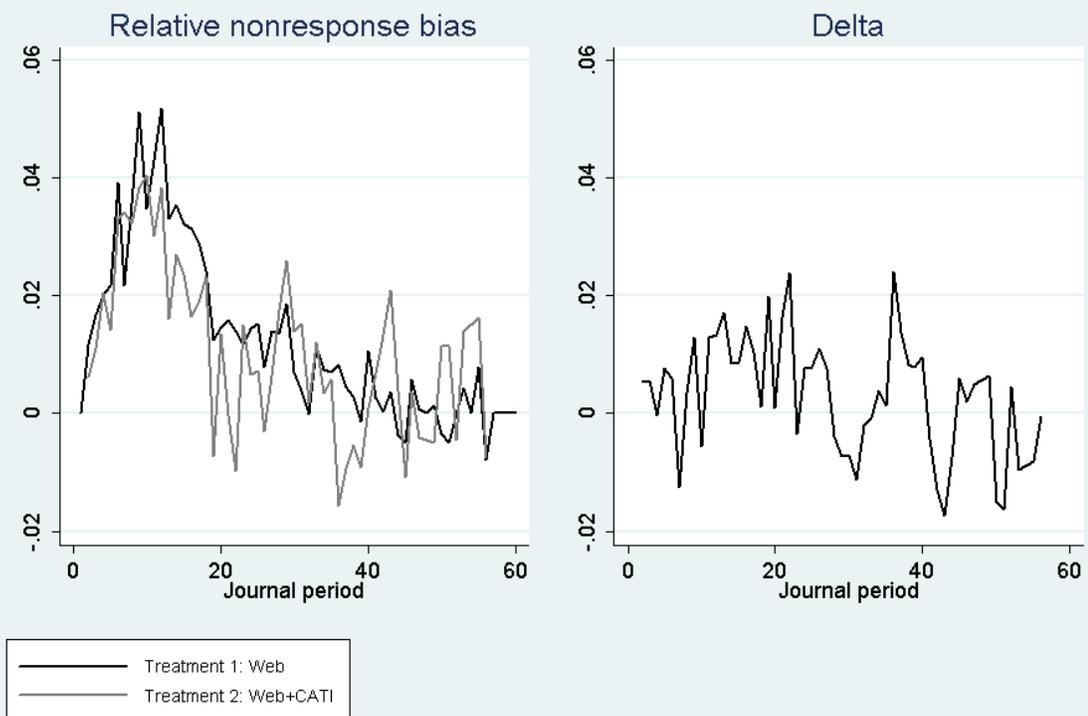
Appendix C. Nonresponse bias in Panel Arbeitsmarkt und Soziale Sicherung



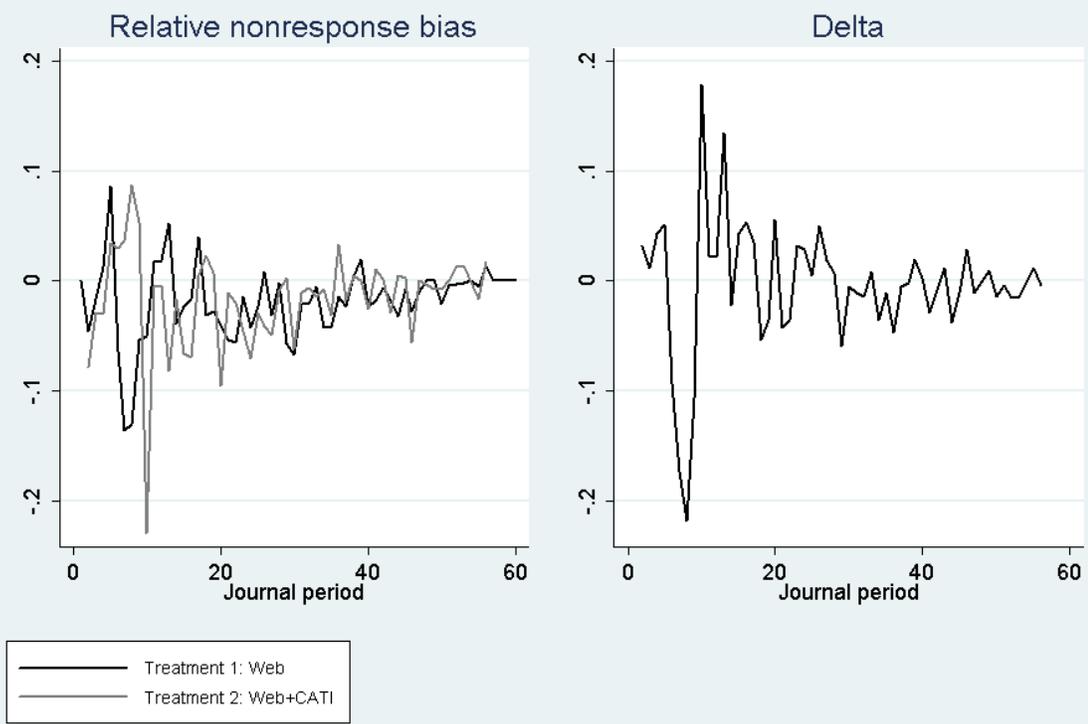
Enrolled in school at baseline: Part-time enrollment



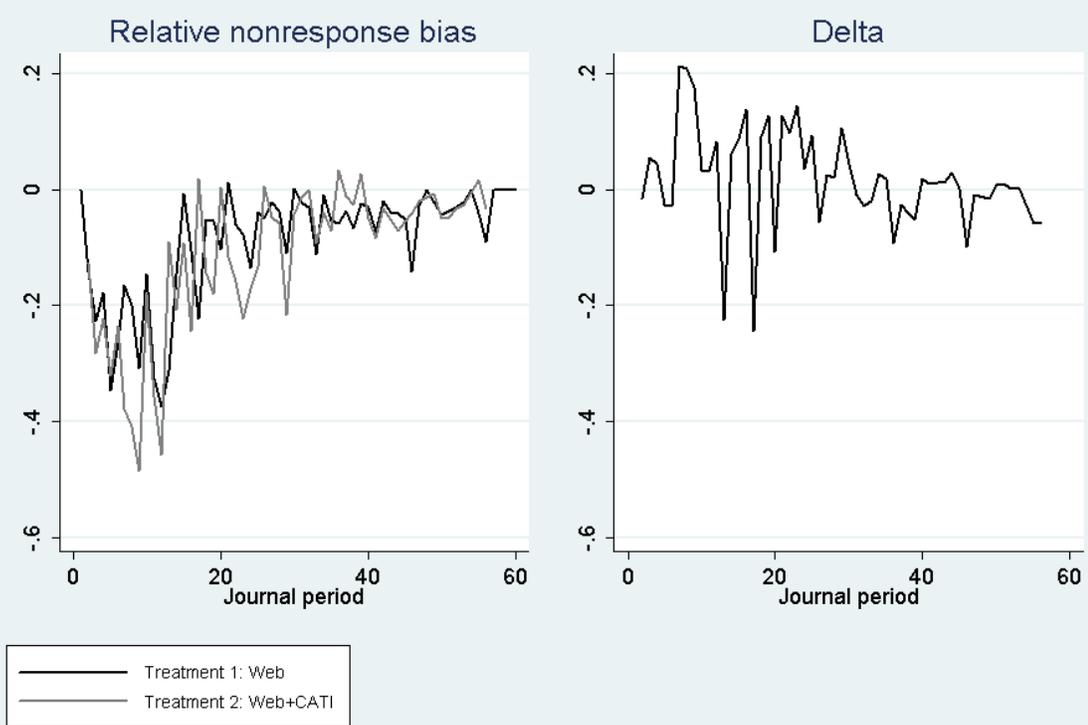
Enrolled in school at baseline: Full-time enrollment



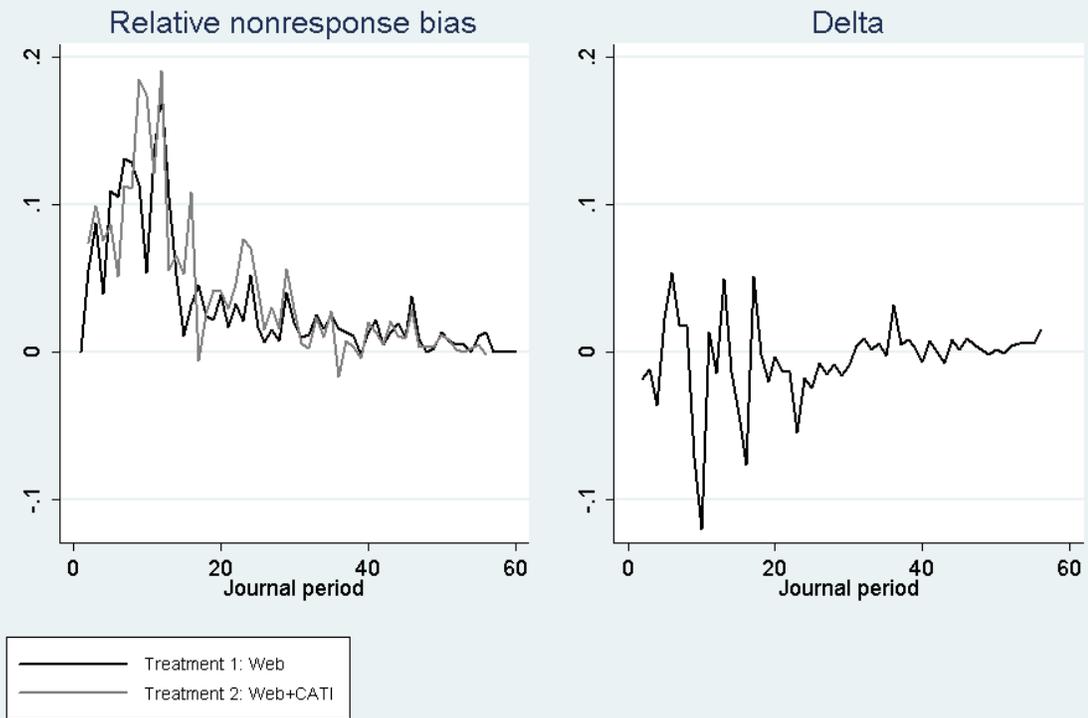
Currently enrolled in school: Not enrolled



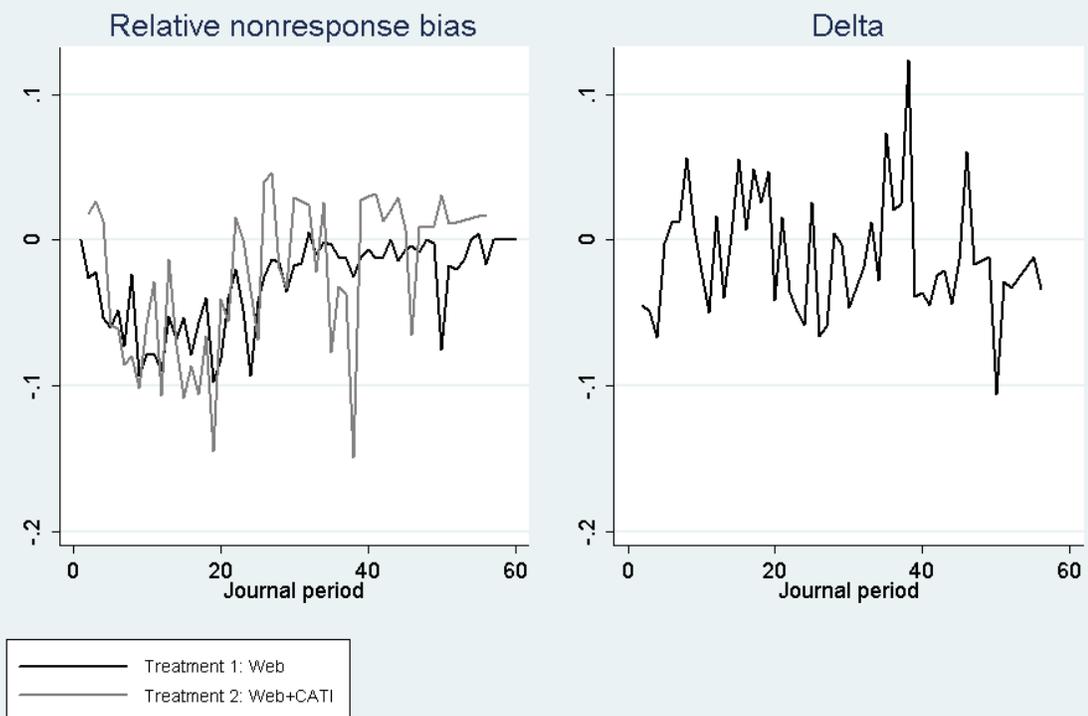
Currently enrolled in school: Part-time enrollment



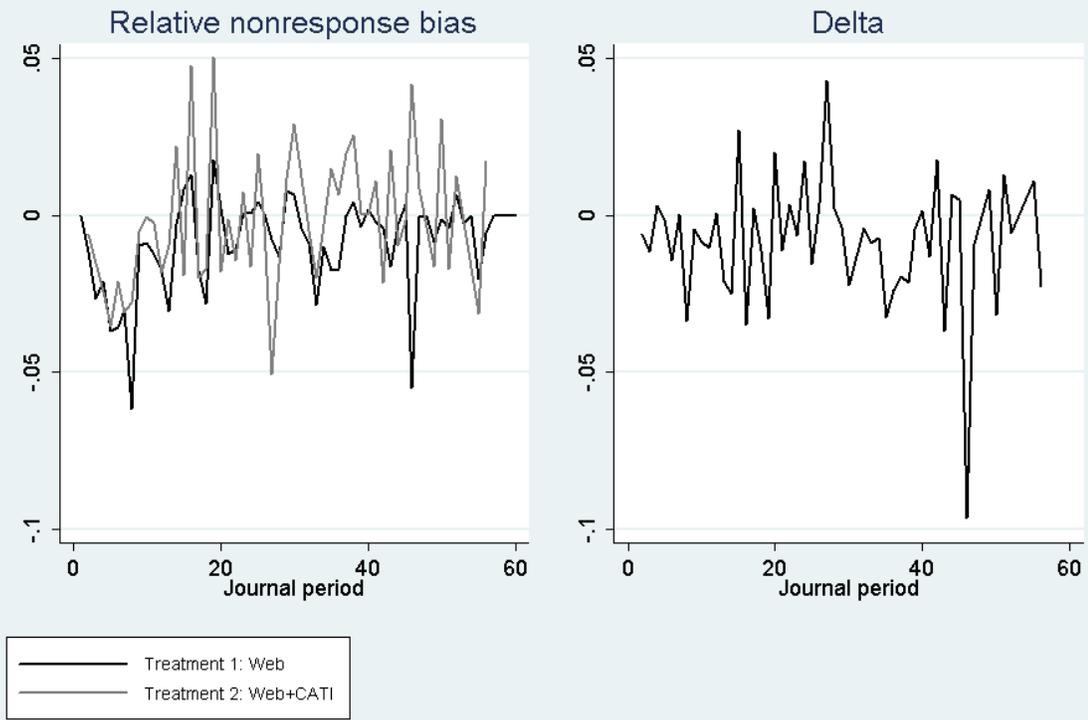
Currently enrolled in school: Full-time enrollment



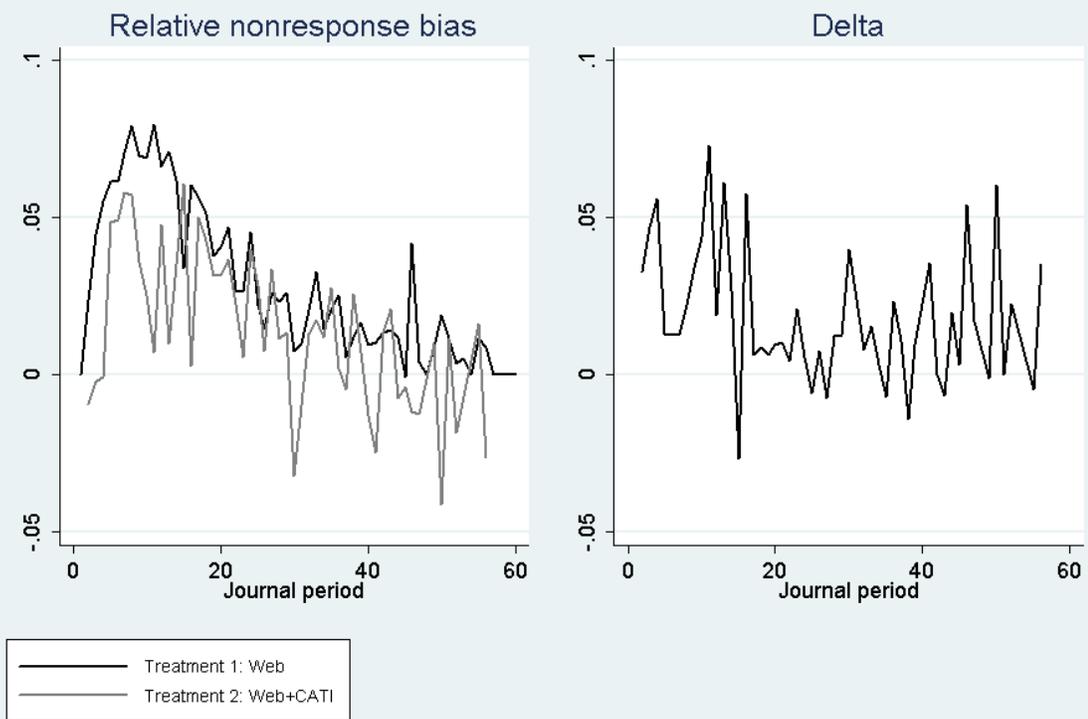
Type of school at baseline: High school or less



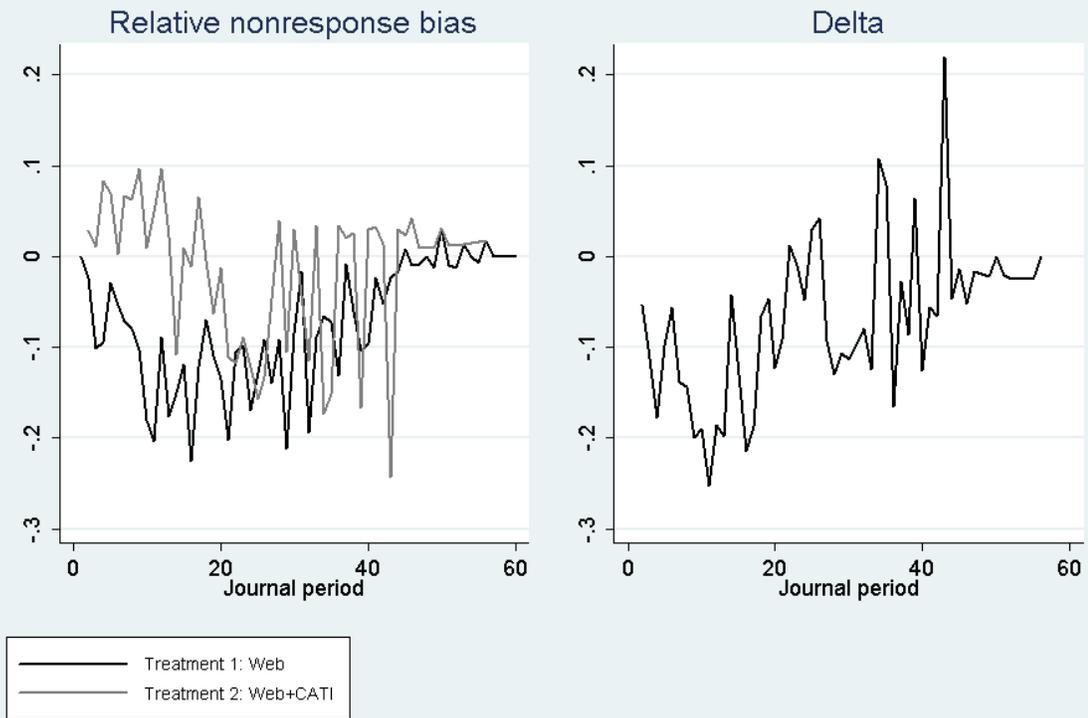
Type of school at baseline: 2 year junior/community college



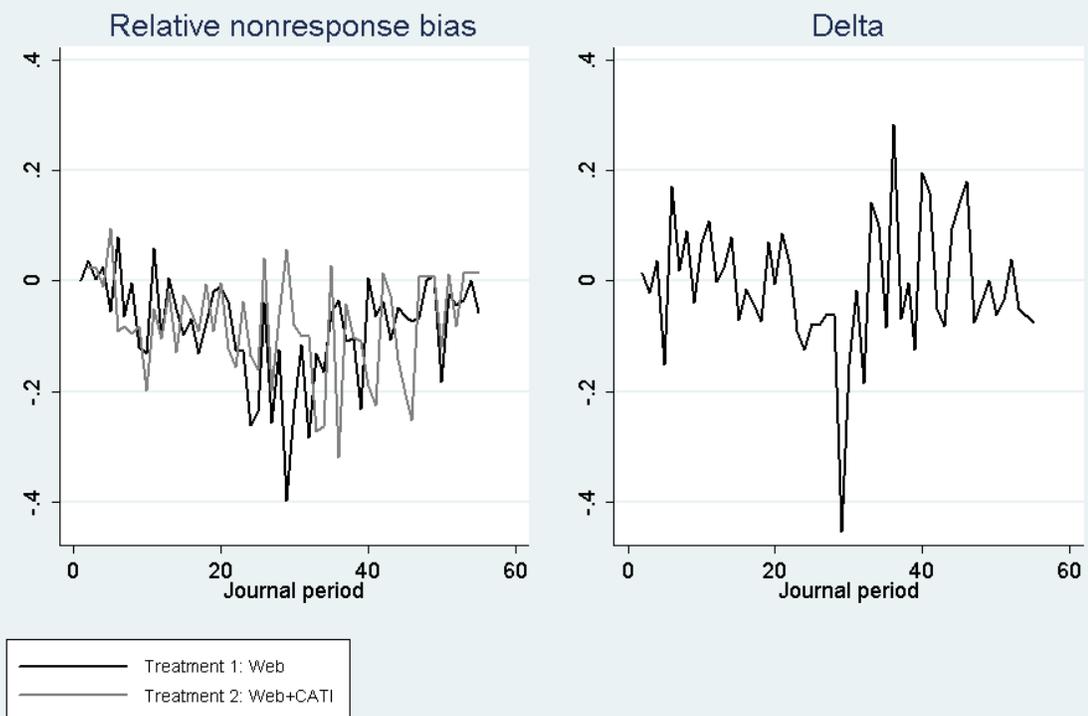
Type of school at baseline: 4 year college



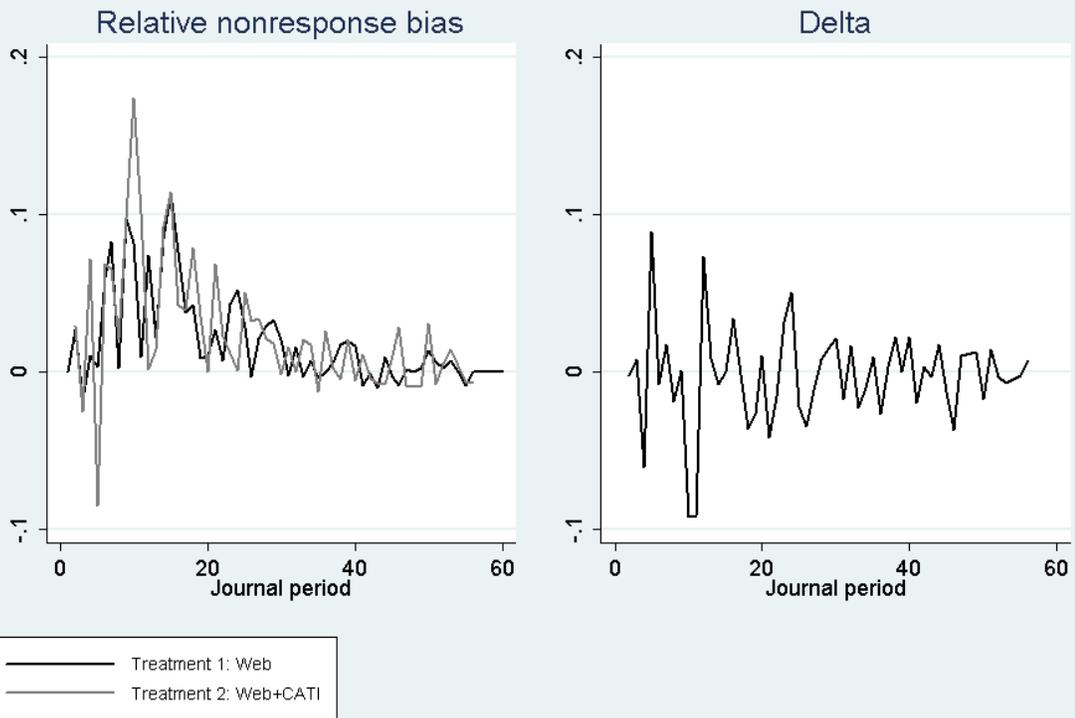
Type of school at baseline: Voc, tech, trade, or other school



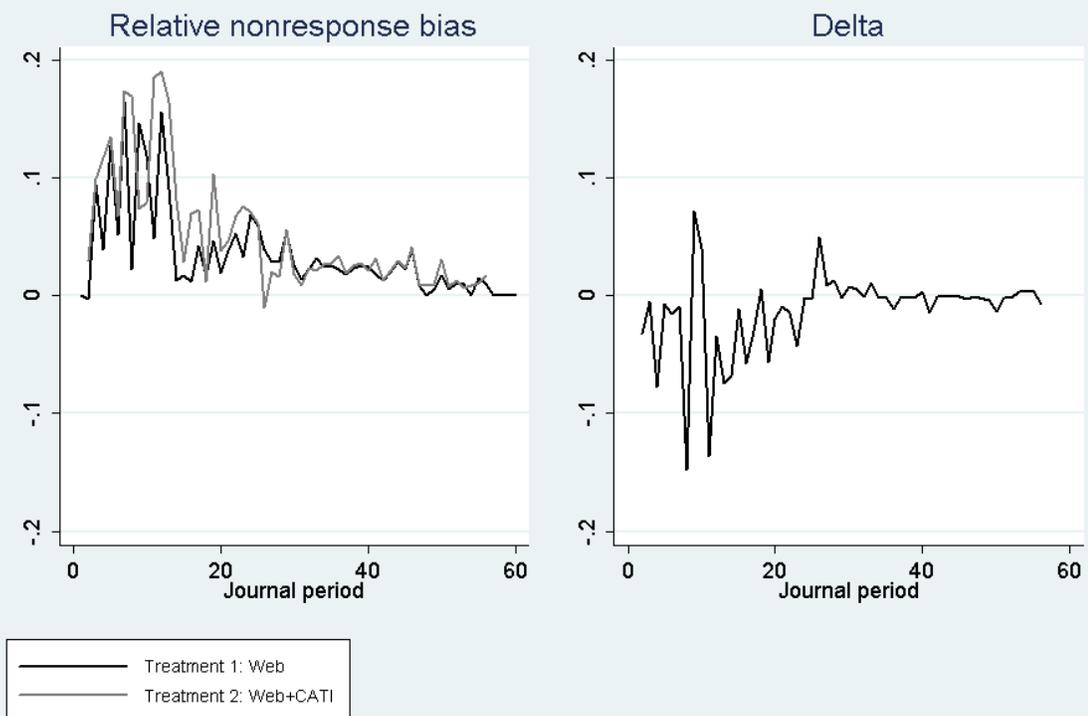
Type of school: High school or less



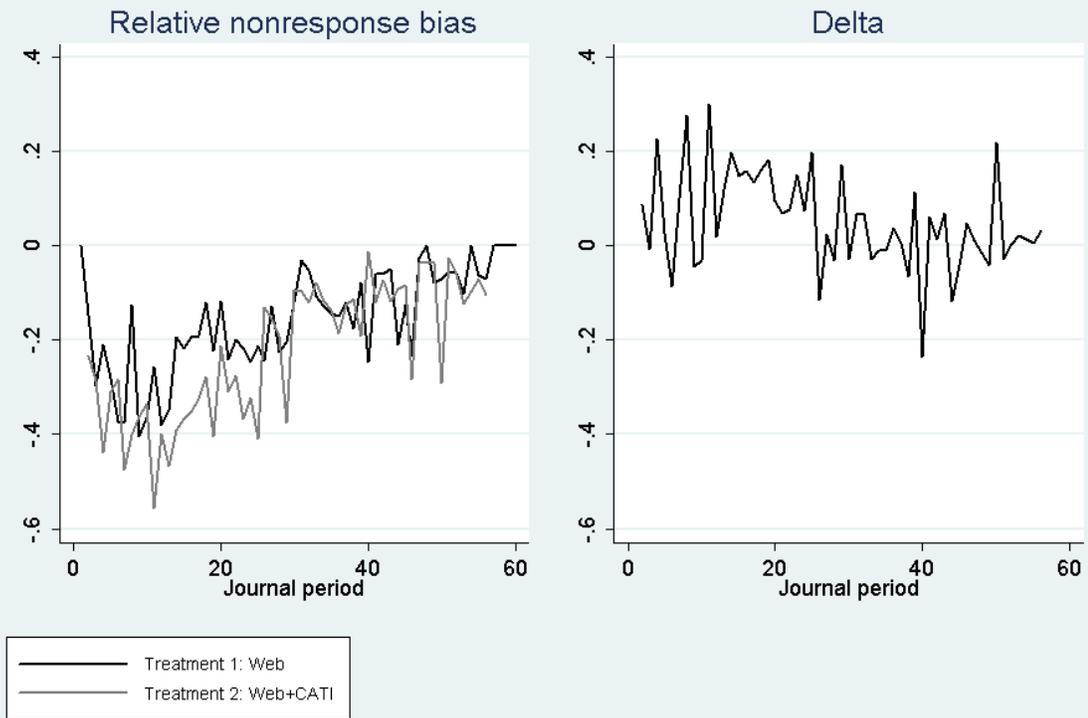
Type of school: 2 year junior/community college



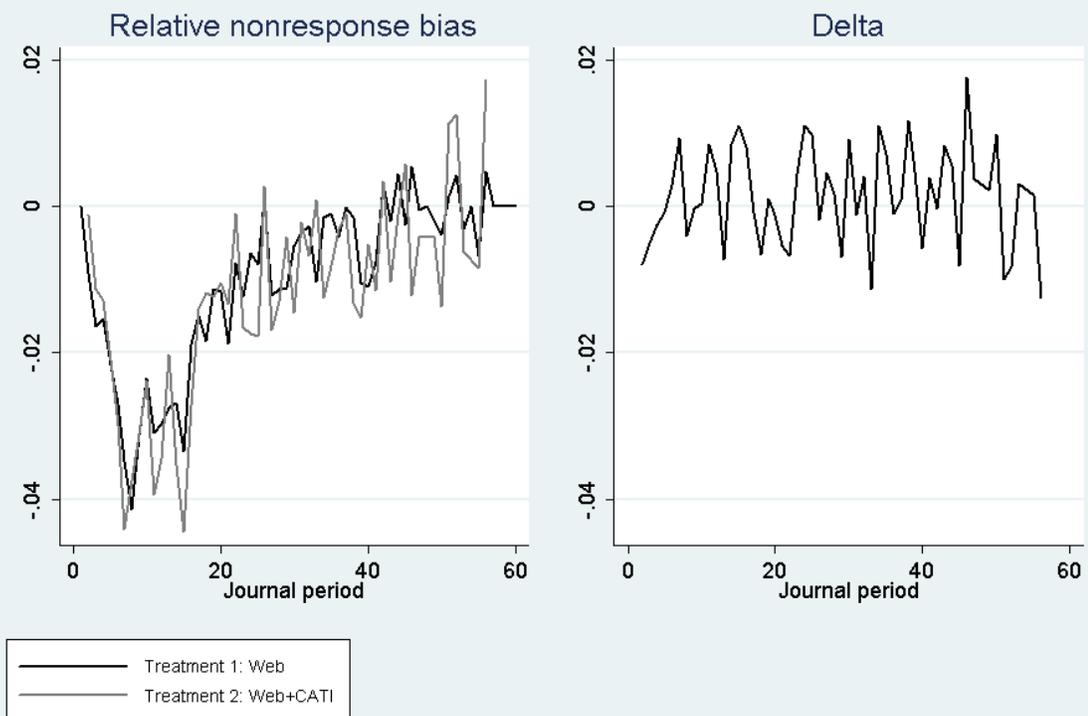
Type of school: 4 year college



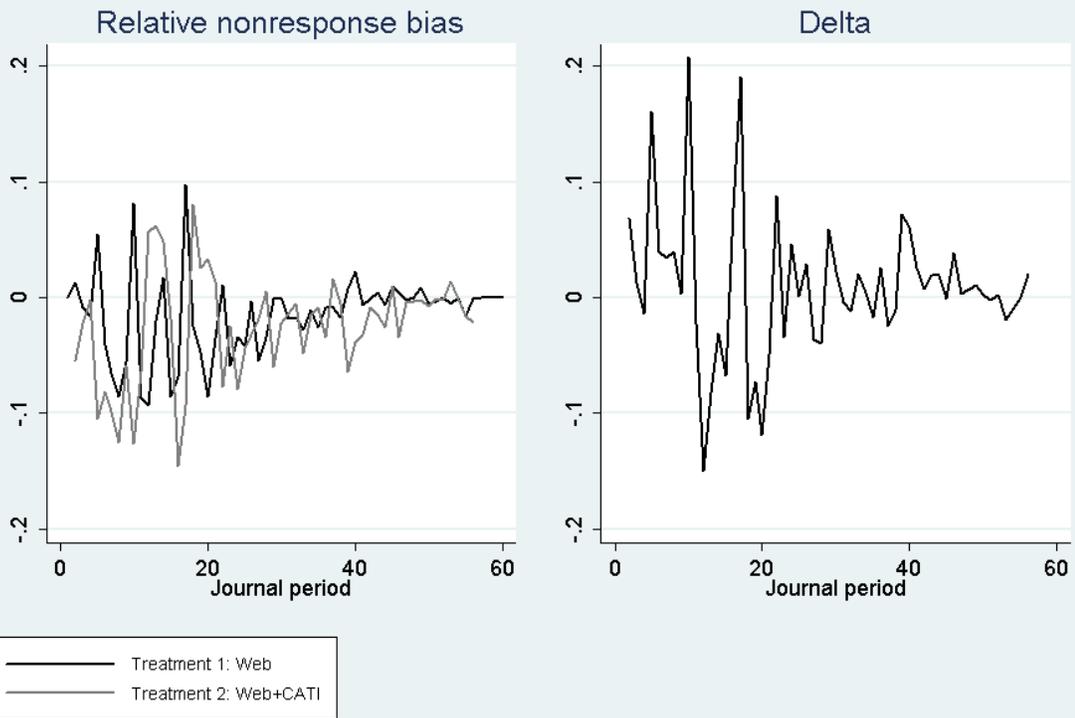
Type of school: Voc, tech, trade, or other school



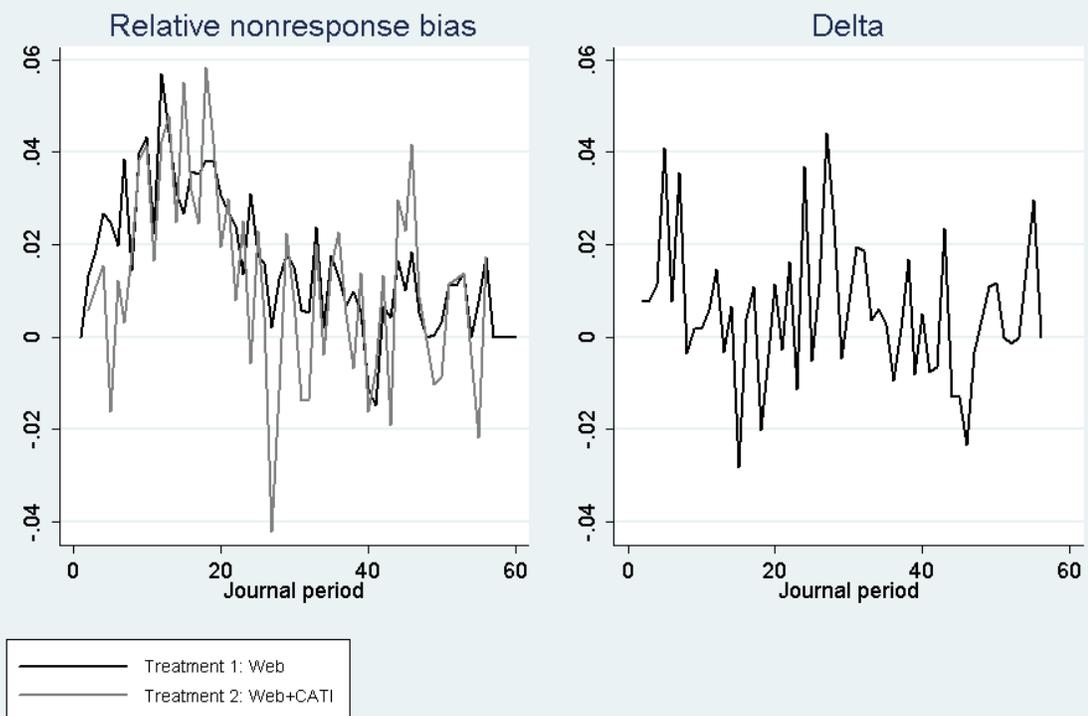
Any sexual partner (baseline)



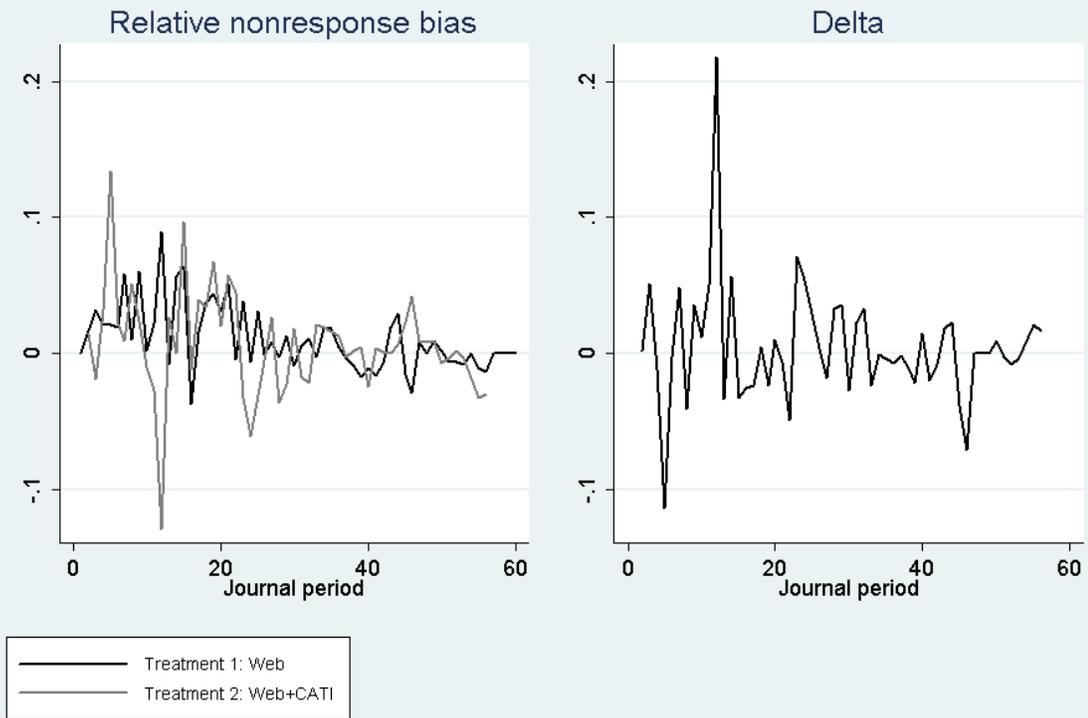
Any sexual partner



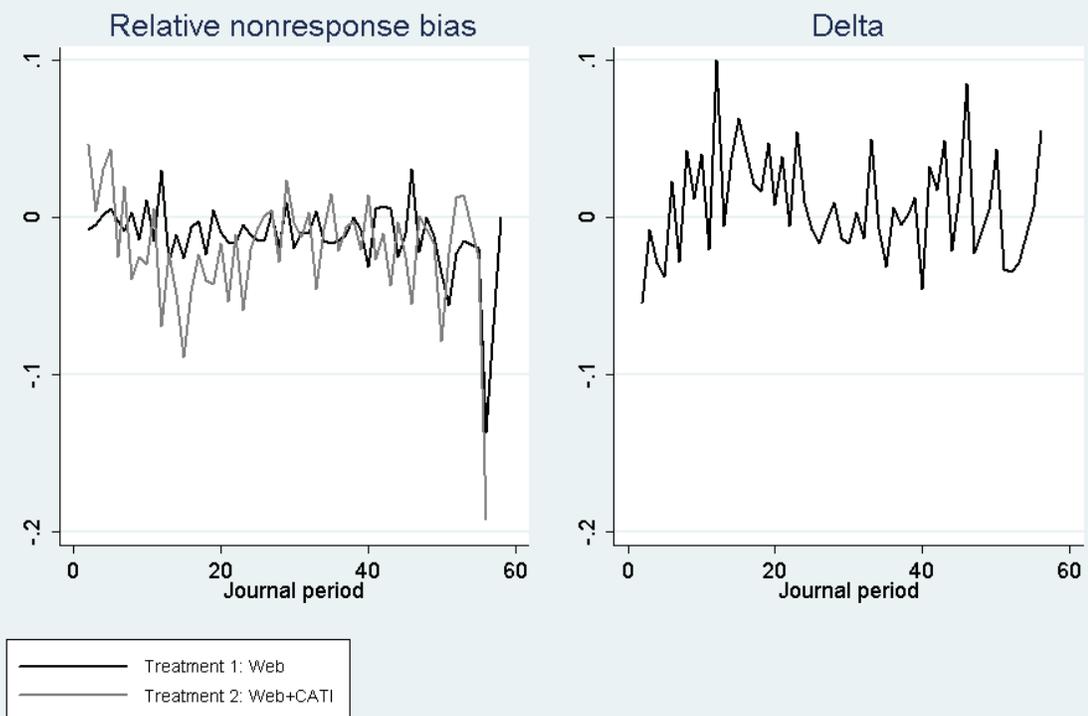
Use of non-coital contraception (baseline)



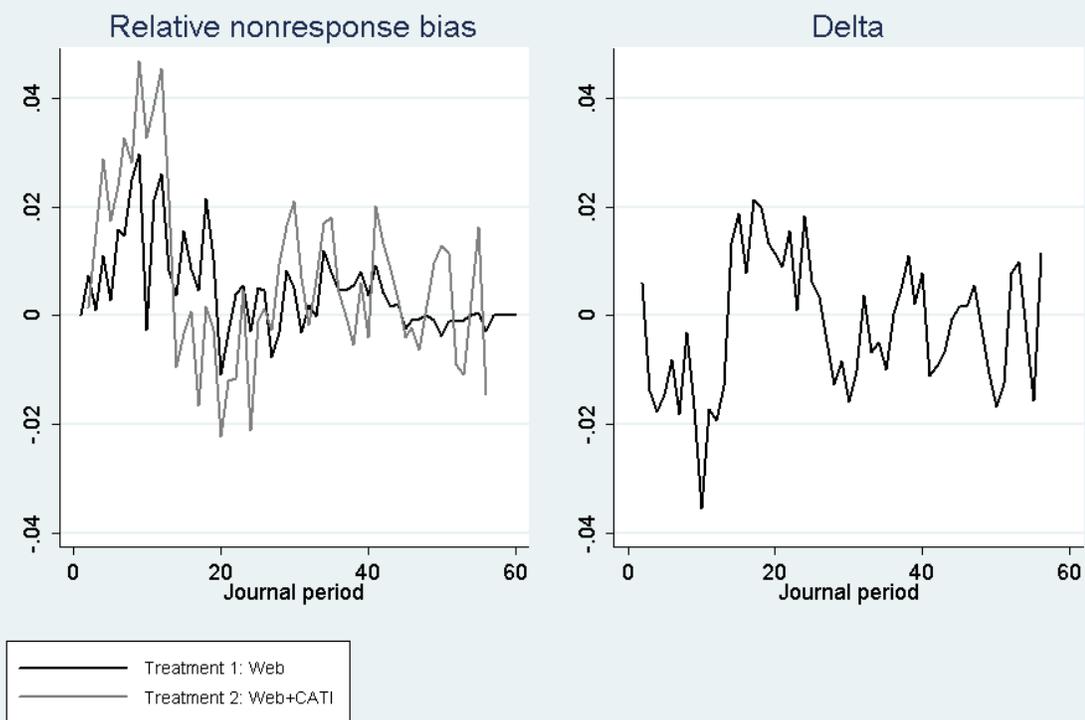
Use of non-coital contraception



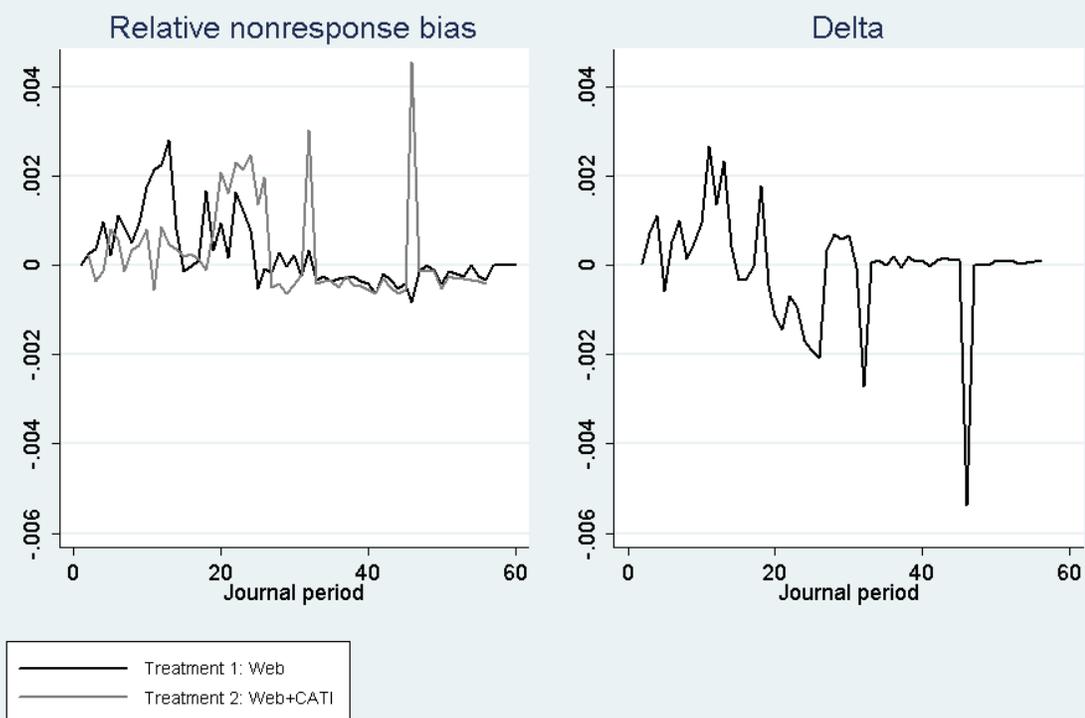
Use of coital contraception



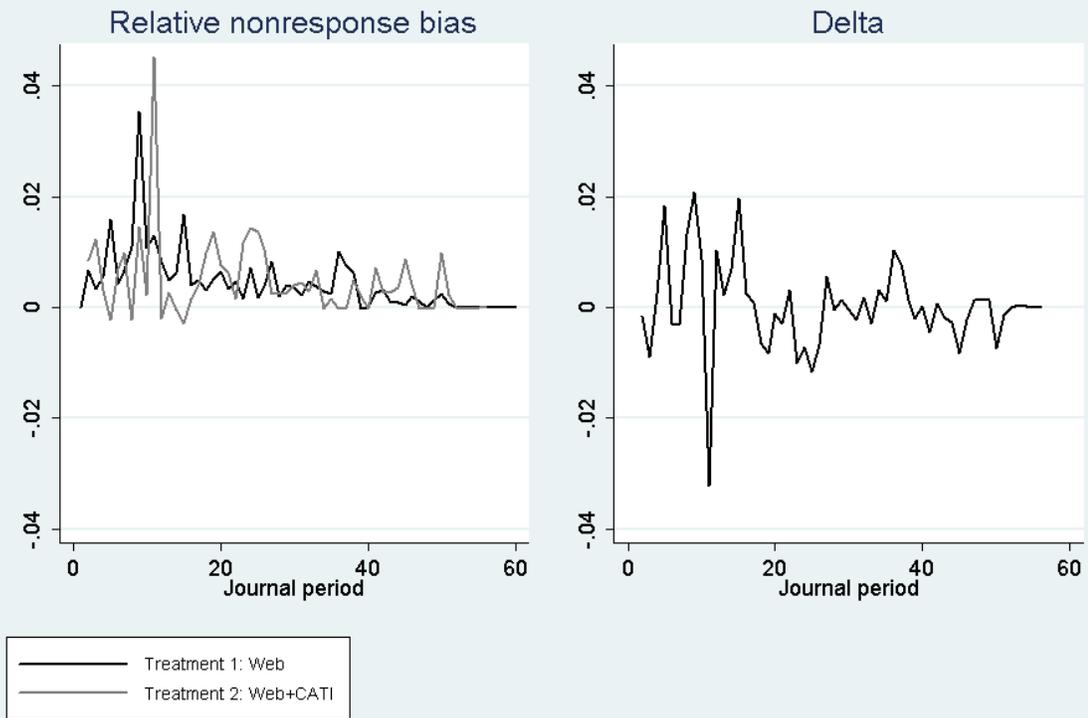
Living with a parent (baseline)



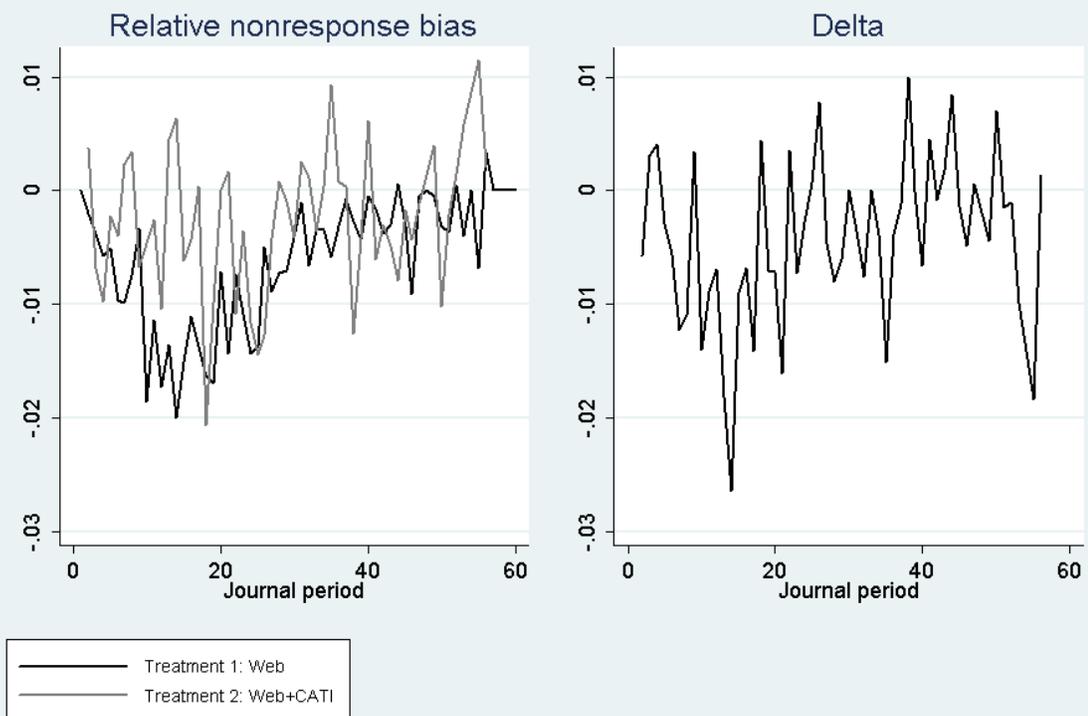
Pregnancy avoidance (baseline)



Pregnancy avoidance

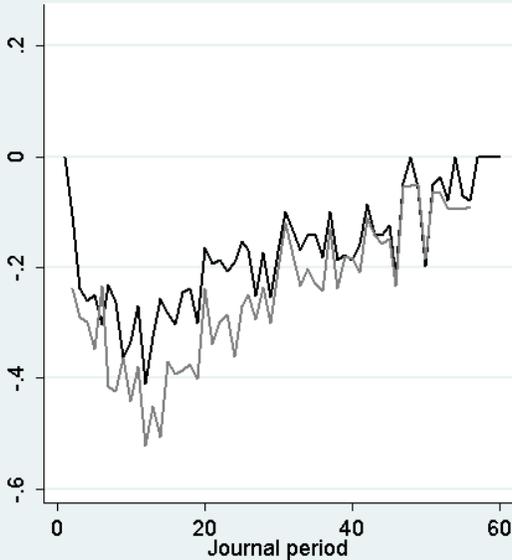


Pregnancy intentions (baseline)

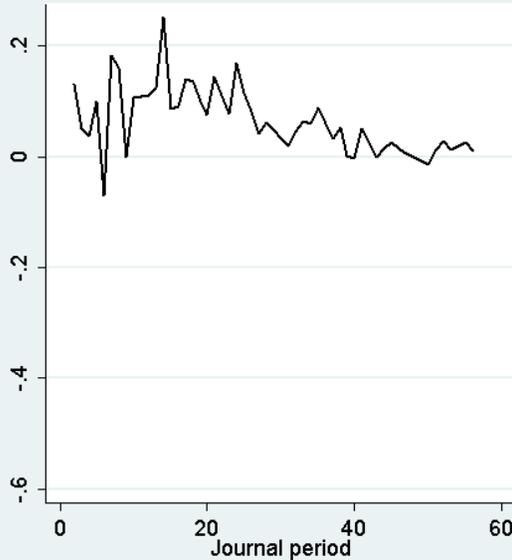


Pregnancy intentions

Relative nonresponse bias



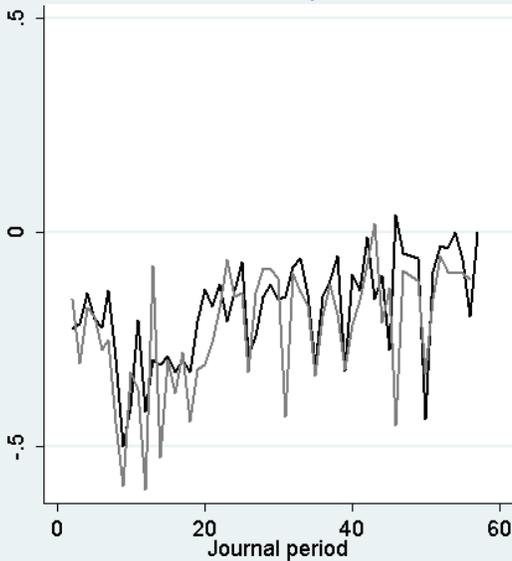
Delta



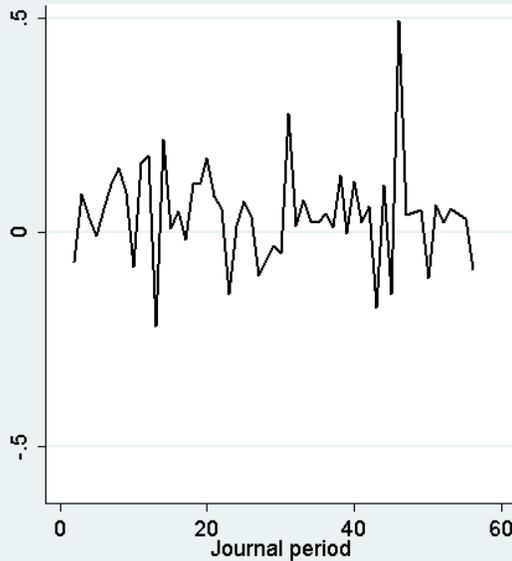
— Treatment 1: Web
— Treatment 2: Web+CATI

Change in relationship status

Relative nonresponse bias



Delta



— Treatment 1: Web
— Treatment 2: Web+CATI

References

- U.S. Census Bureau. (2009). Design and methodology of the american community survey. Washington, DC: U.S. Government Printing Office.
- Abraham, K. G., Maitland, A., & Bianchi, S. M. (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.
- Abraham, K. G., S., H., & Presser, S. (2009). How social processes distort measurement: The impact of survey nonresponse on estimates of volunteer work in the united states. *American Journal Of Sociology*, 114(4), 1129-1165.
- Aquilino, W. S., & Lo Sciuto, L. A. (1990). Effects of interview mode on self-reported drug use. *Public Opinion Quarterly*, 54(3), 362.
- Aquilino, W. S., & LoSciuto, L. A. (1990). Effects of interview mode on self-reported drug use. *Public Opinion Quarterly*, 54(3), 362-395.
- Archer, K. J., & Lemeshow, S. (2006). Goodness-of-fit test for a logistic regression model fitted using survey sample data. *The Stata Journal*, 6(1), 97-105.
- Atrostic, B. K., Bates, N., Burt, G., & Silbertein, A. (2001). Nonresponse in US government household surveys: Consistent measures, recent trends, and new insights. *Journal of Official Statistics*, 17(2), 209-226.
- Barber, J. S., Gatny, H., & Kusunoki, Y. (2008). Relationship dynamics and social life: Design and implementation of a mixed mode, mixed method study of young women: Survey Research Center, University of Michigan.
- Bates, N., Dahlhamer, J. M., & Singer, E. (2008). Privacy concerns, too busy, or just not interested: Using doorstep concerns to predict survey nonresponse. *Journal of Official Statistics*, 24(4), 591-612.
- Beebe, T. J., Locke, G. R., Barnes, S., Davern, M., & Anderson, K. J. (2007). Mixing web and mail methods in a survey of physicians. *Health Services Research*, 42(3), 1219.

- Beebe, T. J., McRae, J. A., Jr., Harrison, P. A., Davern, M. E., & Quinlan, K. B. (2005). Mail surveys resulted in more reports of substance abuse than telephone surveys. *Journal of Clinical Epidemiology*, *58*(4), 421.
- Behr, A., Bellgardt, E., & Rendtel, U. (2005). Extent and determinants of panel attrition in the european community household panel. *European Sociological Review*, *21*(5), 489-512.
- Bergbauer, H., Lehnert, C., Schneider, A., Seysen, C., Thomsen, U., & Zimmermann, R. (2010). *Benutzerhandbuch: Integrierte erwerbsbiografien* (No. V08.01.00-100222). Nuremberg: Institut für Arbeitsmarkt- und Berufsforschung.
- Bethlehem, J. (2002). Weighting nonresponse adjustments based on auxiliary information. In R. M. Groves, D. A. Dillman, J. L. Eltinge & R. J. A. Little (Eds.), *Survey nonresponse* (pp. 275-288). New York: Wiley.
- Blank, T., & Schmidt, P. (2003). National identity in a united germany: Nationalism or patriotism? An empirical test with representative data. *Political Psychology*, *24*(2), 289-312.
- Bollinger, C. R., & David, M. H. (2001). Estimation with response error and nonresponse: Food-stamp participation in the sipp. *Journal of Business and Economic Statistics*, *19*(2), 129-141.
- Bradburn, N., & Sudman, S. (1979). *Improving interview method and questionnaire design*. San Francisco: Jossey-Bass.
- Brault, M. (2008). Americans with disabilities: 2005. *U.S. Census Bureau*.
- Brick, J. M., Dipko, S., Presser, S., Tucker, C., & Yuan, Y. (2006). Nonresponse bias in a dual frame sample of cell and landline numbers. *Public Opinion Quarterly*, *70*(5), 780-793.
- Brown, N. R., & Sinclair, R. C. (1999). Estimating number of lifetime sexual partners: Men and women do it differently. *Journal of Sex Research*, *36*(3), 292.
- Campbell, A. (1954). *The voter decides*. Evanston, IL: Row & Peterson.
- Campbell, A., Converse, P. E., Miller, W. E., & Stokes, D. E. (1960). *The american voter*. Chicago: University of Chicago Press.
- Cannell, C. F., & Fowler, F. J. (1963). Comparison of a self-enumerative procedure and a personal interview: A validity study. *Public Opinion Quarterly*, *27*(2), 250-264.
- Catania, J. A., Gibson, D., Coates, D. D., Chitwood, D. D., & Coats, T. J. (1990). Methodological problems in aids behavioral research: Influences of measurement error and participation bias in studies of sexual behavior. *Psychological Bulletin*, *108*, 339-362.

- Conrad, F. G., Park, H., Singer, E., Couper, M. P., Hubbard, F., & Groves, R. M. (2006). Impact of disclosure risk on survey participation decisions, *American Association of Public Opinion Research*. Montreal.
- Couper, M. P. (1998). *Measuring survey quality in a casic environment*. Proceedings of the Joint Statistical Meetings of the American Statistical Association.
- Couper, M. P. (2000). Web surveys: A review of issues and approaches. *Public Opinion Quarterly*, 64, 464.
- Couper, M. P. (2005). Technology trends in survey data collection. *Social Science Computer Review*, 23(4), 1.
- Couper, M. P., Hansen, S. E., & Sadosky, S. A. (1997). Evaluating interviewer performance in a capi survey. In L. E. Lyberg, P. P. Biemer, M. Collins, E. D. De Leeuw, C. Dippo, N. Schwarz & D. Trewin (Eds.), *Survey measurement and process quality*. New York: Wiley.
- Couper, M. P., Singer, E., Conrad, F. G., & Groves, R. M. (2008). Risk of disclosure, perceptions of risk, and concerns about privacy and confidentiality as factors in survey participation. *Journal of Official Statistics*, 24(2), 255-275.
- Couper, M. P., Singer, E., & Kulka, R. A. (1998). Participation in the 1990 decennial census: Politics, privacy, pressures. *American Politics Research*, 26(1), 59-80.
- de Leeuw, E. D. (1992). *Data quality in mail, telephone, and face to face surveys*. Unpublished Dissertation/Thesis.
- de Leeuw, E. D. (2005). To mix or not to mix data collection modes in surveys. *Journal of Official Statistics*, 21(2), 233-255.
- de Leeuw, E. D. (2008). Choosing the mode of data collection. In E. D. de Leeuw, J. Hox & D. A. Dillman (Eds.), *International handbook of survey methodology* (pp. 113-135). New York: Lawrence Erlbaum.
- de Leeuw, E. D., & de Heer, W. (2002). Trends in household survey nonresponse: A longitudinal and international comparison. In R. M. Groves, D. A. Dillman, J. L. Eltinge & R. J. A. Little (Eds.), *Survey nonresponse* (pp. 41). New York: Wiley.
- de Leeuw, E. D., & van der Zouwen, J. (1988). Data quality in telephone and face to face surveys: A comparative meta-analysis. In R. M. Groves, P. P. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nicholls, II & J. Waksberg (Eds.), *Telephone survey methodology* (pp. 283). New York: Wiley.
- Dillman, D. A., Brown, T. L., Carlson, J., Carpenter, E. H., Lorenz, F. O., Mason, R., et al. (1995). Effects of category order on answers to mail and telephone surveys. *Rural Sociology*, 60, 674-687.

- Dillman, D. A., Phelps, G., Tortora, R., Swift, K., Kohrell, J., & Berck, J. (2008). Response rate and measurement differences in mixed mode surveys using mail, telephone, interactive voice response and the internet. *Social Science Research*, 38(1), 1-18.
- Dillman, D. A., Sinclair, M. D., & Clark, J. R. (1993). Effects of questionnaire length, respondent-friendly design, and a difficult question on response rates for occupant-addressed census mail surveys. *Public Opinion Quarterly*, 57(3), 289-304.
- Dillman, D. A., Smyth, J., & Christian, L. M. (2009a). *Internet, mail, and mixed-mode designs*. Hoboken, NJ: Wiley.
- Dillman, D. A., Smyth, J., & Christian, L. M. (2009b). *Internet, mail, and mixed-mode surveys: The tailored design method* (3rd ed.). Hoboken: Wiley.
- Dillman, D. A., West, K. K., & Clark, J. R. (1994). Influence of an invitation to answer by telephone on response to census questionnaires. *Public Opinion Quarterly*, 58(4), 557.
- Duncan, G. J., & Hill, M. S. (1985). Conceptions of longitudinal households: Fertile or futile? *Journal of Economic and Social Measurement*, 13(3/4), 361-375.
- Durrant, G. B., Groves, R. M., Staetsky, S., & Steele, F. (2010). Effects of interviewer attitudes and behaviors on refusal in household surveys. *Public Opinion Quarterly*, 74(1), 1-36.
- Egerton, M. (2003). Higher education and civic engagement. *British Journal of Sociology*, 53(4), 603-620.
- Fitzgerald, J., Gottschalk, P., & Moffitt, R. (1998). An analysis of sample attrition in panel data: The michigan panel study of income dynamics. *Journal of Human Resources*, 33, 251-299.
- Fortson, B. L., Scotti, J. R., Chen, Y., Malone, J., & Del Ben, K. S. (2007). Internet use, abuse, and dependence among students at a southeastern regional university. *Journal of American College Health*, 56(2), 137-144.
- Gfroerer, J., Lessler, J., & Parsley, T. (1997). Studies of nonresponse and measurement error in the national household survey on drug abuse. In L. Harrison & A. L. Hughes (Eds.), *The validity of self-reported drug use: Improving the accuracy of survey estimates* (pp. 273-295). Rockville, MD: US Department of Health and Human Services.
- Goyder, J. (1986). Surveys on surveys: Limitations and potentialities. *Public Opinion Quarterly*, 50(1), 27-41.

- Goyder, J. (1987). *The silent minority: Nonrespondents on sample surveys*. Boulder, CO: Westview Press.
- Green, K. E. (1996). Sociodemographic factors and mail survey response. *Psychology and Marketing*, 13(2), 171-184.
- Greenberg, B. S., & Stokes, S. L. (1990). Developing an optimal call scheduling strategy for a telephone survey. *Journal of Official Statistics*, 6(4), 421-435.
- Gribble, J. N., Miller, H. G., Rogers, S. M., & Turner, C. F. (1999). Interview mode and measurement of sexual behaviors: Methodological issues. *Journal of Sex Research*, 36(1), 16-24.
- Griffin, D. H. (2009). What the American Community Survey can tell us about mixed mode surveys: Presented at the Survey Research Center, Ann Arbor, MI.
- Groves, R. M. (1989). *Survey errors and survey costs*. New York: Wiley.
- Groves, R. M. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly*, 70(5), 646-675.
- Groves, R. M., Cialdini, R. B., & Couper, M. P. (1992). Understanding the decision to participate in a survey. *Public Opinion Quarterly*, 56(4), 475-495.
- Groves, R. M., & Couper, M. P. (1998). *Nonresponse in household interview surveys*. New York: Wiley.
- Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J., Singer, E., & Tourangeau, R. (2009). *Survey methodology* (2nd ed.). Hoboken, NJ: Wiley.
- Groves, R. M., & 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 (Statistics in Society)*, 169(3), 439.
- Groves, R. M., & Kahn, R. L. (1979). *Surveys by telephone: A national comparison with personal interviews*. New York: Academic Press.
- Groves, R. M., & McGonagle, K. A. (2001). A theory-guided interview training protocol regarding survey participation. *Journal of Official Statistics*, 17(2), 249-265.
- Groves, R. M., & Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias: A meta-analysis. *Public Opinion Quarterly*, 72(2), 167-189.
- Groves, R. M., Presser, S., & Dipko, S. (2004). The role of topic interest in survey participation decisions. *Public Opinion Quarterly*, 68(1), 2-31.
- Groves, R. M., Singer, E., & Corning, A. (2000). Leverage-salience theory of survey participation. *Public Opinion Quarterly*, 64(3), 299.

- Heeringa, S., Hill, D. H., & Howell, D. A. (1993). *Unfolding brackets for reducing item nonresponse in economic surveys* (No. 94-029, AHEAD/HRS Report).
- Hill, D. H., & Willis, R. J. (2001). Reducing panel attrition: A search for effective policy instruments. *Journal of Human Resources*, 36(3), 416-438.
- Hill, M. S. (1992). *The panel study of income dynamics: A user's guide*. Newbury Park, CA: Sage Publications.
- Hochstim, J. R. (1967). A critical comparison of three strategies of collecting data from households. *Journal of the American Statistical Association*, 62, 976.
- Holbrook, A. L., Green, M. C., & Krosnick, J. A. (2003). Telephone versus face-to-face interviewing of national probability samples with long questionnaires: Comparisons of respondent satisficing and social desirability bias. *Public Opinion Quarterly*, 67(1), 79-125.
- Jäckle, A., Roberts, C., & Lynn, P. (2010). Assessing the effect of data collection mode on measurement. *International Statistical Review*, 78(1), 3-20.
- James, T. L. (1997). Results of the wave 1 incentive experiment in the 1996 survey of income and program participation, *Proceedings of the American Statistical Association, Survey Research Methods Section* (pp. 834-839).
- Jones, E. F., & Forrest, J. D. (1992). Underreporting of abortion in surveys of u.S. Women: 1976 to 1988. *Demography*, 29(1), 113-126.
- Jones, S. (2002). *The internet goes to college: How students are living in the future with today's technology*: Pew Internet & American Life Project.
- Jones, S., Johnson-Yale, C., Millermaier, S., & Pérez, F. S. (2009). U.S. College students' internet use: Race, gender, and digital divides. *Journal of Computer-Mediated Communication*, 14(2), 244-264.
- Kalsbeek, W. D., Botman, S. L., Massey, J. T., & Liu, P.-W. (1994). Cost-efficiency and the number of allowable call attempts in the national health interview survey. *Journal of Official Statistics*, 10(2), 133-152.
- Kalton, G., Winglee, M., Rizzo, L., Jabine, T., Levine, D., Huggins, V., et al. (1998). *Sipp quality profile*.
- Keeter, S., Zukin, C., Andolina, M., & Jenkins, K. (2002). *The civic and political health of the nation: A generational portrait*. College Park, MD: Center for Information and Research on Civic Learning and Engagement.
- Knack, S. (1992). Civic norms, social sanctions, and voter turnout. *Rationality and Society*, 4, 133-156.

- Kreuter, F., Mueller, G., & Trappmann, M. (2009a). Nonresponse and measurement error in employment research, *American Association of Public Opinion Research*. Hollywood, FL.
- Kreuter, F., Presser, S., & Tourangeau, R. (2009b). Social desirability bias in cati, ivr, and web surveys. *Public Opinion Quarterly*, 72(5), 847-865.
- Kronthaler, F. (2005). Economic capability of east german regions: Results of a cluster analysis. *Regional Studies*, 39(6), 739-750.
- Krosnick, J. A., & Alwin, D. (1987). An evaluation of a cognitive theory of response order effects in survey measurement. *Public Opinion Quarterly*, 51(2), 201.
- Kuusela, V., Callegaro, M., & Vehovar, V. (2008). The influence of mobile telephones on telephone surveys. In J. M. Lepkowski, C. Tucker, J. M. Brick, E. D. De Leeuw, L. Japac, P. J. Lavrakas, M. W. Link & R. Sangster (Eds.), *Advances in telephone survey methodology* (pp. 87-112). Hoboken, NJ: Wiley.
- Laurie, H., Smith, R., & Scott, L. (1999). Strategies for reducing nonresponse in a longitudinal panel survey. *Journal of Official Statistics*, 15(2), 269-282.
- Lemay, M. (2009). *Understanding the mechanism of panel attrition*. University of Maryland, College Park, MD.
- Lepkowski, J. M., & Couper, M. P. (2002). Nonresponse in longitudinal household surveys. In R. M. Groves, D. A. Dillman, J. L. Eltinge & R. J. A. Little (Eds.), *Survey nonresponse*. New York: Wiley.
- Lessler, J. T., & Kalsbeek, W. D. (1992). *Nonsampling error in surveys*. New York: Wiley.
- Li, K. H., Meng, X. L., Raghunathan, T. E., & Rubin, D. B. (1991a). Significance levels from repeated p values from multiply-imputed data. *Statistica Sinica*, 1, 65-92.
- Li, K. H., Raghunathan, T. E., & Rubin, D. B. (1991b). Large sample significance levels from multiply imputed data using moment-based statistics and an f reference distribution. *Journal of the American Statistical Association*, 86, 1065-1073.
- Liebig, T. (2007). *The labour market integration of immigrants in germany* (No. 47): Organization for Economic Cooperation and Development.
- Lillard, L. A., & Panis, C. W. A. (1998). Panel attrition from the panel study of income dynamics: Household income, marital status, and mortality. *Journal of Human Resources*, 33(2), 437-457.
- Loosveldt, G., Pickery, J., & Billiet, J. (2002). Item nonresponse as a predictor of unit nonresponse in panel surveys. *Journal of Official Statistics*, 18(4), 545-557.

- Lozar Manfreda, K., Bosnjak, M., Berzelak, J., Haas, I., & Vehovar, V. (2008). Web surveys versus other survey modes: A meta-analysis comparing response rates. *International Journal of Market Research*, 50(1), 79-104.
- Mack, S., Huggins, V., Keathley, D., & Sundukchi, M. (1998). Do monetary incentives improve response rates in the survey of income and program participation? *Proceedings of the American Statistical Association, Survey Research Methods Section* (pp. 529-534).
- Mathiowetz, N. A., & Duncan, G. J. (1988). Out of work, out of mind: Response errors in retrospective reports of unemployment. *Journal of Business and Economic Statistics*, 6(2), 221-229.
- McMillen, D. B., & Herriot, R. (1985). Toward a longitudinal definition of households. *Journal of Economic and Social Measurement*, 13(3/4), 349-360.
- McPherson, J., & Rotolo, T. (1996). Diversity and change in voluntary groups. *American Sociological Review*, 61, 179-202.
- Meng, X. L., & Rubin, D. B. (1992). Performing likelihood ratio tests with multiply imputed data sets. *Biometrika*, 79, 103-111.
- Miller, P. V., & Groves, R. M. (1985). Matching survey responses to official records: An exploration of validity in victimization reporting. *Public Opinion Quarterly*, 49(3), 366.
- Moore, J. C., Stinson, L. L., & Welniak, E. J., Jr. (1999). Income reporting in surveys: Cognitive issues and measurement error. In M. G. Sirken, D. J. Herrmann, S. Schechter, N. Schwarz, J. M. Tanur & R. Tourangeau (Eds.), *Cognition and survey research* (pp. 155-174). New York: Wiley.
- Moore, J. C., Stinson, L. L., & Welniak, E. J., Jr. (2000). Income measurement error in surveys: A review. *Journal of Official Statistics*, 6(4), 331-361.
- Neter, J., & Waksberg, J. (1964). A study of response errors in expenditures data from household interviews. *Journal of the American Statistical Association*, 59, 17-55.
- Okun, M., & Eisenberg, N. (1992). A comparison of office and adult day-care center older volunteers. *International Journal of Aging and Human Development*, 35(3), 219-233.
- Olson, K. (2006). Survey participation, nonresponse bias, measurement error bias, and total bias. *Public Opinion Quarterly*, 70(5), 737-758.
- Price, V., & Stroud, N. F. (2005). Public attitudes toward polls: Evidence from the 2000 u.S. Presidential election. *International Journal of Public Opinion Research*, 18(4), 393-421.

- Putnam, R. D. (2000). *Bowling alone: The collapse and revival of american community*. New York: Simon & Schuster.
- Raghunathan, T. E. (2004). What do we do with missing data? Some options for analysis of incomplete data. *Annual Review of Public Health, 25*, 99-117.
- Raghunathan, T. E., Lepkowski, J., Van Hoewyk, J., & Solenberger, P. (2001). A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey Methodology, 27*(1), 85-95.
- Rogelberg, S. G., Fisher, G. G., Maynard, D. C., Hakel, M. D., & Horvath, M. (2001). Attitudes toward surveys: Development of a measure and its relationship to respondent behavior. *Organizational Research Methods, 4*(1), 3-25.
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley.
- Sanchez, T. W., Lang, R. E., & Dhavale, D. M. (2005). Security versus status? A first look at the census's gated community data. *Journal of Planning Education and Research, 24*(3), 281-291.
- Schwarz, N., Hippler, H. J., & Noelle-Neumann, E. (1992). A cognitive model of response-order effects in survey measurement. In N. Schwarz & S. Sudman (Eds.), *Context effects in social and psychological research*. New York: Springer-Verlag.
- Scioch, P., & Bender, S. (2010). *Quality and quantity: Using administrative data for scientific purposes in labor market research*. Proceedings of the American Association of Public Opinion Research, Chicago, IL.
- Selfa, L., & Sederstrom, S. (2006). *Respondent mode preference in a multi-mode survey*. Proceedings of the American Association for Public Opinion Research.
- Singer, E. (2002). The use of incentives to reduce nonresponse in household surveys. In R. M. Groves, D. A. Dillman, J. L. Eltinge & R. J. A. Little (Eds.), (pp. 163). New York: Wiley.
- Spiess, M., & Pannenberg, M. (2003). *Documentation of sample sizes and panel attrition in the german socio economic panel (gsoep) (1984 until 2002)*. Berlin.
- Stauder, J., & Hüning, W. (2003). *Die messung von äquivalenzeinkommen und armutsquoten auf der basis des mikrozensus*. Unpublished manuscript.
- Steeh, C., & Piekarski, L. (2008). Accommodating new technologies: Mobile and voip communication. In J. M. Lepkowski, C. Tucker, J. M. Brick, E. D. de Leeuw, L. Japac, P. J. Lavrakas, M. W. Link & R. L. Sangster (Eds.), *Advances in telephone survey methodology* (pp. 423-446). Hoboken, NJ: Wiley.

- Stoop, I. (2005). *The hunt for the last respondent: Nonresponse in sample surveys*. The Hague, Netherlands: Social and Cultural Planning Office.
- Stratton, L. S., O'Toole, D. M., & Wetzel, J. N. (2008). A multinomial logit model of college stopout and dropout behavior. *Economics of Education Review*, 27(3), 319-331.
- Sykes, W., & Collins, M. (1988). Effects of mode of interview: Experiments in the uk. In R. M. Groves, P. P. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nicholls, II & J. Waksberg (Eds.), *Telephone survey methodology* (pp. 301). New York: Wiley.
- Tourangeau, R., Rasinski, K. A., Jobe, J. B., Smith, T. W., & Pratt, W. F. (1997). Sources of error in a survey on sexual behavior. *Journal of Official Statistics*, 13(4), 341.
- Tourangeau, R., Rips, L. J., & Rasinski, K. A. (2000). *The psychology of survey response*. Cambridge: Cambridge University Press.
- Tourangeau, R., & Smith, T. W. (1996). Asking sensitive questions: The impact of data collection mode, question format, and question context. *Public Opinion Quarterly*, 60, 275-304.
- Trappmann, M., Christoph, B., Achatz, J., Wenzig, C., Mueller, G., & Gebhardt, D. (2009). *Design and stratification of pass: A new panel study for research on long term unemployment* (No. 5/2009). Nuremberg: Institut für Arbeitsmarkt- und Berufsforschung.
- Uhlig, H. (2006). Regional labor markets, network externalities and migration: The case of german reunification. *American Economic Review*, 96(2), 383-387.
- Uhrig, S. C. N. (2008). *The nature and causes of attrition in the british household panel survey* (No. 2008-05). Colchester: University of Essex.
- van Campen, C., Sixma, H., Kerssens, J. J., & Peters, L. (1998). Comparisons of the costs and quality of patient data collection by mail versus telephone versus in-person interviews. *European Journal of Public Health*, 8, 66-70.
- Voogt, R. J. J., & Saris, W. E. (2005). Mixed mode designs: Finding the balance between nonresponse bias and mode effects. *Journal of Official Statistics*, 21(3), 367-387.
- Wagner, J. (2010). The fraction of missing information as a tool for monitoring the quality of survey data. *Public Opinion Quarterly*, 74(2), 223-243.
- Walsh, J. (2005). Experiencing part-time work: Temporal tensions, social relations and the work-family interface. *British Journal of Industrial Relations*, 45(1), 155-177.
- Watson, D. (2003). Sample attrition between waves 1 and 5 in the european community household panel. *European Sociological Review*, 19(4), 361-378.

- Watson, N., & Wooden, M. (2009). Identifying factors affecting longitudinal survey response. In P. Lynn (Ed.), *Methodology of longitudinal surveys* (pp. 157-182). Chichester, UK: Wiley.
- Wilke, R. A. (2005). New estimates of the duration and risk of unemployment for west-germany. *Journal of Applied Social Science Studies*, 128(2), 207-237.
- Wilson, J. (2000). Volunteering. *Annual Review of Sociology*, 26, 215-240.
- Winkels, J. W., & Withers, S. D. (2000). Panel attrition. In D. Rose (Ed.), *Researching social and economic change: The uses of household panel studies* (pp. 79-95). New York: Routledge.
- Zabel, J. E. (1998). An analysis of attrition in the panel study of income dynamics and the survey of income and program participation with an application to a model of labor market behavior. *Journal of Human Resources*, 33(2), 479.
- Zimmerman, M. A., & Rappaport, J. (1988). Citizen participation, perceived control, and psychological empowerment. *American Journal of Community Psychology*, 16(5), 725-750.