Survey Research Methods (2015)
Vol. 9, No. 2, pp. 111-123
doi:10.18148/srm/2015.v9i2.6135

ISSN 1864-3361
http://www.surveymethods.org

Evaluating a Modular Design Approach to Collecting Survey Data Using Text Messages

Brady T. West
Survey Research Center
Institute for Social Research
University of Michigan-Ann Arbor, U.S.A.

Dirgha Ghimire
Population Studies Center
Institute for Social Research
University of Michigan-Ann Arbor, U.S.A.

William G. Axinn
Survey Research Center
Institute for Social Research
University of Michigan-Ann Arbor, U.S.A.

This article presents analyses of data from a pilot study in Nepal that was designed to provide an initial examination of the errors and costs associated with an innovative methodology for survey data collection. We embedded a randomized experiment within a long-standing panel survey, collecting data on a small number of items with varying sensitivity from a probability sample of 450 young Nepalese adults. Survey items ranged from simple demographics to indicators of substance abuse and mental health problems. Sampled adults were randomly assigned to one of three different modes of data collection: 1) a standard one-time telephone interview, 2) a "single sitting" back-and-forth interview with an interviewer using text messaging, and 3) an interview using text messages within a modular design framework (which generally involves breaking the survey response task into distinct parts over a short period of time). Respondents in the modular group were asked to respond (via text message exchanges with an interviewer) to only one question on a given day, rather than complete the entire survey. Both bivariate and multivariate analyses demonstrate that the two text messaging modes increased the probability of disclosing sensitive information relative to the telephone mode, and that respondents in the modular design group, while responding less frequently, found the survey to be significantly easier. Further, those who responded in the modular group were not unique in terms of available covariates, suggesting that the reduced item response rates only introduced limited nonresponse bias. Future research should consider enhancing this methodology, applying it with other modes of data collection (e. g., web surveys), and continuously evaluating its effectiveness from a total survey error perspective.

Keywords: modular survey design; survey methodology; survey nonresponse; text message surveys; total survey error; survey costs

1 Introduction

All common modes of survey data collection have been affected by increasing reluctance of the public to cooperate with survey requests (Abraham, Maitland, & Bianchi, 2006; Brick & Williams, 2013). Today's survey researchers are thus faced with a significant challenge of finding innovative methods for collecting high-quality survey data and improving the survey experience (Couper, 2013). Recent research in survey methodology suggests that the collection of survey data via text messages on mobile phones or Web browser applications is feasible and produces high-quality data (Brenner & DeLamater, 2014; Cocco & Tuzzi, 2013; Kuntsche & Robert, 2009; Schembre & Yuen, 2011; Schober et al., 2015). This is important, given the rapidly increasing proportions of people worldwide who use mobile phones but not landline phones. The rapid proliferation of mobile phones in the United States is an important example (Blumberg & Luke, 2013), but the rates of increase in mobile phone use are actually highest in some of the most isolated parts of the world where landline penetration has been low. Also important is the fact that most people with mobile phones send and receive text messages (73% of mobile phone owners in

Contact information: Brady T. West (corresponding author), Survey Research Center, Institute for Social Research, University of Michigan-Ann Arbor, 426 Thompson Street, Ann Arbor, MI 48106-1248, Phone: 734-647-4615, Fax: 734-647-2440, Email: bwest@umich.edu

the U.S., according to the Pew Research Center; A. Smith, 2011). The development of new methodologies for mobile data collection needs to be accompanied by rigorous studies of the errors and costs that can arise from using these methods.

This study presents one of the first comprehensive examinations of the errors and costs associated with what is known as a *modular design* approach to data collection among market researchers (Johnson, Kelly, & Stevens, 2012; Kelly, Johnson, & Stevens, 2013; R. Smith, Kotzev, Miller, & Kachhi, 2012). Modular design works under the assumption that the behaviors and attitudes a person will share on a survey do not change substantially over a short period of time. Although "modular" designs may refer to many different types of study designs in different fields, the specific survey designs that we focus on in this paper involve breaking the survey response task into distinct parts over a short period of time, rather than requiring participants to answer an entire survey in one sitting (which can lead to break-offs or lower-quality responses). This approach differs somewhat from split questionnaire or "matrix sampling" designs, where different respondents are only given random subsets of the items in the full survey (Raghunathan & Grizzle, 1995); in modular designs, all respondents eventually have the opportunity to answer all questions, just not in a single sitting.

Because survey researchers need to adapt to the increasing preferences of people to use mobile devices and the limited desire of people to spend long periods of time on surveys (Couper, 2013), modular designs may hold some potential to advance survey data collection. Unfortunately, relatively little is known about the measurement properties of modular design approaches. This limits our ability to predict how results from modular designs may differ from results from other designs. In this paper, we seek initial answers to three questions about the modular design approach:

1. How do modular survey responses on both sensitive and non-sensitive questions compare to those collected using other text messaging approaches and more standard data collection modes?
2. How do various paradata (e. g., item nonresponse indicators, time required to answer a question, etc.) compare between persons assigned to a modular approach and those assigned to alternative data collection approaches?
3. Do people responding to a modular survey using text messaging find the survey easier compared to alternative data collection modes?

2 Motivation for the Modular Design Approach

One can find motivation for the modular design approach to survey data collection via text messaging in multiple areas of the methodological literature. We review these areas here, beginning with motivations for attempting a modular approach and then assessing the potential measurement effects of using a modular approach with text messaging.

2.1 Negative Effects of Questionnaire Length

A vast inter-disciplinary literature has clearly demonstrated the negative effects of longer surveys (regardless of the mode of data collection) on both response rates and data quality. Several randomized studies have compared sample units given reduced length questionnaires to sample units given longer questionnaires in terms of response rates, and found higher response rates for the reduced length questionnaires. Heberlein and Baumgartner (1978) analyzed 98 mail survey experiments involving 240 design manipulations, including questionnaire length, and found a significant negative relationship between questionnaire length and response rate. This result was later replicated in meta-analyses reported by Goyder (1982) and Yammarino, Skinner, and Childers (1991). Analyzing data from a lengthy survey covering various topics regarding economic change and social life in Britain in 1987, Burchell and Marsh (1992) found that the length of the questionnaire was a significant disincentive to respond for many. Dillman, Sinclair, and Clark (1993) found evidence that shortening the U.S. decennial census questionnaire improved response rates in a large national study of U.S. households. Studying a mail survey, Biner and Kidd (1994) found that an incentive justified in a cover letter by an obligation on the part of the recipient to return a questionnaire was able to increase response rates (relative to a cover letter stating that the incentive was a gesture of appreciation) when a shorter questionnaire was used, suggesting that length may moderate other design effects. Kalantar and Talley (1999) found a response rate of 68.2% for a longer questionnaire and 75.6% for a shorter questionnaire in a randomized experiment involving a mailed health survey.

In a clinical trial setting, Edwards, Roberts, Sandercock, and Frost (2004) used a meta-analysis to review the effect of questionnaire length on response rates in 38 randomized controlled trials, and these authors found evidence of a benefit to reducing the length of questionnaires to improve response rates, along with evidence that moderate changes to shorter questionnaires will be more effective than the same changes to longer questionnaires. Similar results were reported in a meta-analysis by VanGeest, Johnson, and Welch (2007).

Some of this work can also be found in the literature on nonresponse in panel surveys. Bogen (1996) provides a review of literature on the effect of questionnaire length on response rates for the first and subsequent waves in panel surveys. Based on this review of the literature, there seems to be some evidence that attrition rates could be reduced by decreasing interview length (Zabel, 1994). Not all findings have been in the same direction, however: Kantorowitz (1998) found no relationship between interview length and

response rate when the survey questionnaire was shortened.

The effects of questionnaire length on motivation to respond may also be moderated by the mode of data collection used. A recent study (Kelly et al., 2013) suggests that only 25% of smartphone users are willing to spend more than five minutes completing surveys. The widespread use of mobile technologies globally has required survey researchers to explore the possibility of collecting survey data using mobile devices, and this means that findings related to survey length and mobile devices need to be weighed carefully. In theory, modular designs could also effectively combat problems with break-offs in longer web and mobile phone surveys, given that each additional question on a given page in a web survey and the presence of a progress indicator have been shown to increase the likelihood of breakoffs (Peytchev, 2009), but this theory needs to be tested before the method can become widely used. Collectively, the findings in the literature related to questionnaire length make it clear that reducing the size of the task for a survey respondent may have positive benefits in terms of overall data quality. We thus hypothesize that modular designs will be more appealing to the survey participant and reduce the burden associated with participating in longer surveys in a single sitting.

2.2 Improved Reporting in Text Message Interviews

A rich literature also exists demonstrating the positive effects of self-administration on data quality and reporting related to sensitive survey items such as drug use (see Groves et al., 2009, for a review). Initial research conducted to date has suggested that interviews conducted via text messaging on mobile devices essentially mimic self-administration, in terms of higher reporting of sensitive behaviors and higher data quality in general (Brenner & DeLamater, 2014; Cocco & Tuzzi, 2013; Schober et al., 2015). Interviews using text messages also remove the potential of interviewer effects on responses due to observable characteristics of the interviewers, on which there is a rich history of research (Schaeffer, Dykema, & Maynard, 2010). In theory, one can therefore hypothesize that modular designs involving interviewers and mobile phones will simultaneously reduce respondent burden (due to decreased questionnaire length) and increase the quality of reporting for sensitive survey items, and we intend to test this hypothesis with the present study.

2.3 Positive Benefits of Brief Assessment

Modular designs, which involve "brief" measurement of variables of interest, also share similar features with ecological momentary assessment (EMA) methods (e. g. Moskowitz & Young, 2006; Shiffman, Stone, & Hufford, 2008; Stone & Shiffman, 1994) and day reconstruction methods (DRM; Kahneman, Krueger, Schkade, Schwarz, and Stone, 2004) in epidemiology and psychology. These methods collect brief measures of behaviors and attitudes in real time and in natural environments, limiting recall bias and enabling detailed monitoring of behaviors over time. Modular designs using mobile phones therefore also have the potential to increase data quality from the point of view of eliminating context effects and generating real-time responses in more natural settings.

Although there is reasonably strong theoretical support for the success of the modular design approach, no studies to date have demonstrated differences in survey reporting between this approach and related approaches to data collection. Our objective in this study is to test the hypotheses outlined above, focusing specifically on potential measurement error related to the modular design approach and respondent burden.

3 Methods

Nepal provides an ideal setting for testing the measurement properties of a modular design approach, specifically because survey response rates in Nepal are extremely high across alternative data collection modes (Axinn, Ghimire, & Williams, 2012). High response rates across modes mean that observed measurement differences between modes cannot be attributed to differential nonresponse. Even though the modular design may be most beneficial in settings characterized by high nonresponse, this special setting of low nonresponse allows us to focus our investigation on the measurement differences across modes. In Nepal, subscriptions for mobile phones increased 5% in 2006 and 75% in 2011 (MOHP, New ERA, & ICF International Inc., 2007, 2012). This rapid upward trajectory in mobile phone use means that large portions of the population now have access to these devices, making survey data collection using mobile phones feasible for a representative sample of the general population.

3.1 The Chitwan Valley Family Study (CVFS)

We tested the modular design approach using a subsample of the Chitwan Valley Family Study (CVFS) in south central Nepal. The CVFS is an ongoing, 18-year panel study of a stratified, systematic probability sample of 151 communities, 2,671 households, and over 15,000 individuals in Nepal, and features a case-control comparison design at the community level (Axinn & Yabiku, 2001; Barber, Shivakoti, Axinn, & Gajurel, 1997). This unusual panel study provides an outstanding environment for testing the proposed modular design approach, where relatively high cooperation rates among panel cases (a 92% response rate among originally sampled individuals to date, using AAPOR RR5) enable a focus on the measurement properties of the approach (Axinn et al., 2012). Data are collected from a scientific, representative probability sample, and respondent contact information, including mobile telephone numbers and substantial survey

information collected from previous waves, is readily available from the panel.

3.2 Randomized Assignment

We first prepared a complete list of 685 sampled CVFS household members between the ages of 18 and 24 (67% of all sample members in this age range) who had a mobile phone. We then randomly assigned the 685 individuals into three groups, and randomly selected 150 individuals from each group for a total sample of 450 individuals. The first group (*voice interview*) was recruited and asked all of the survey questions during a single telephone call. The second group (*text message interview*) was asked all of the survey questions in a single (one-time) series of text message exchanges with a human interviewer. The third group (*modular design interview*) was sent one survey question per day via text message by an interviewer. Respondents in the second and third groups were reimbursed for all expenses related to the text messages. While the target sample size of the proposed study is relatively small (450 people), we expected a 95 - 97% response rate in each group, which would yield more than 80% power to detect small-to-moderate effect sizes at the 0.05 significance level when using chi-square tests. A 50% response rate on a given survey item within a given group would still enable us to detect moderate effects with 80% power at the 0.05 level.

Three trained Nepali interviewers then contacted all 450 sampled individuals by telephone. These three interviewers explained the study, requested participation in the study, completed the informed consent process, and clearly explained the data collection protocol randomly assigned to each respondent. Interviewers encouraged respondents in the two text messaging groups to delete text messages after they responded to the survey questions. Individuals selected to complete the survey by telephone were asked to complete the survey during the recruitment call. Any individual who declined to participate in the study was re-contacted once by a supervisor-level staff person in Nepal who also encouraged participation in the study.

Finally, we administered a relatively brief survey interview in three different ways. For the *voice interview* group, a trained interviewer asked 15 closed-ended and multiple-choice type questions. For the *text message interview* group, we arranged a one-hour time frame during which the respondent could exchange text messages with the human interviewer. In this group, the survey questions were first transmitted by the interviewers, and the interviewer then waited for a response to each question before transmitting the next question. After one hour, the interview was considered "complete", and any break-offs or item-missing data were noted and recorded. For the *modular design interview* group, we sent one text message with one question per day for 15 consecutive days after the initial telephone contact. No reminders or follow-up contacts were made for missing responses on each day, and respondents were allowed to respond to a daily question on a later date. In each of the three groups, a final question assessed respondent burden from the survey process using a Likert-type question, with possible response options "very easy", "somewhat easy", "somewhat hard", and "very hard".

3.3 Measures

The 15-item questionnaire (see the online Appendix) collected a series of simple yes/no and numeric responses to demographic questions (e. g., age, marital status), drug use behavior questions (e. g., "Have you ever smoked marijuana?"), mental health questions (e. g., "Have you ever in your life had a period lasting several days or longer when most of the day you felt sad, empty or depressed?"), and social media behaviors (e. g., "Do you have a Facebook account?"). These questions were adapted from other surveys, and chosen specifically because they represent behaviors or lifetime histories where respondent answers are unlikely to change over a short period of time (15 days). IRB approval was obtained for the use of the questionnaire and the entire pilot study.

3.4 Data Analysis

Consistent with our three research questions, we initially used chi-square tests and one-way analysis of variance to compare the three groups in terms of: 1) response distributions for key survey measures; 2) item nonresponse rates for each of the 15 survey items, along with other paradata describing response behavior; and 3) respondent burden. For each survey item, we also compared respondents in the modular design group with the full sample assigned to the modular group in terms of distributions on selected auxiliary variables, to estimate the potential for nonresponse bias in the estimates from the modular design group respondents.

For those survey measures found to have different response distributions among the three groups in the bivariate analyses, we fitted linear and logistic regression models to further explore the effects of the alternative modes on the distributions when adjusting for other relevant covariates. In these multivariate models, we integrated measures of both local community context (proximity to schools, markets, and health posts, and percentage of individuals enrolled in school) and key individual characteristics (age, gender, marital status, caste / ethnicity) that may explain differences among the groups despite the randomization. This multivariate approach provided the means to test interactions between the three-category group variable (modular design, text interview, phone interview) and these important individual and community characteristics. Interactions were tested for significance using multi-parameter Wald tests, and dropped from the models if they were not found to significantly improve model fit. All analyses were performed using

commands in the Stata software (Version 13.1), and a significance level of 0.05 was used for testing null hypotheses (e. g., a given regression coefficient is equal to zero, or the three groups have equal response distributions on a given item).

4 Results

4.1 Unit Nonresponse

Overall, unit nonresponse was not a critical issue in any of the three study groups (as expected). The rates at which people agreed to participate in the study were 98.7% (148/150), 94.7% (142/150) and 94.7% (142/150) in the voice interview, text message, and modular design groups, respectively. As expected, the highly cooperative character of the Nepalese population allows us to compare the measurement properties of the different interviewing modes, all with a low level of unit nonresponse. Thus, we focus on features of the measurement process with our analysis, and compare item-missing data rates across the three modes for each survey item.

4.2 Marginal Differences in Response Distributions

Table 1 presents comparisons of descriptive statistics (means and percentages) for the various survey items between the three modes, in addition to the number of sampled persons responding to each item in each mode. Table 1 also shows, for the modular design group, the percentage of valid responses that arrived the same day that the question was originally sent via text message.

The results in Table 1 suggest that the modes did not tend to differ in terms of reports on the more factual survey items (e. g., age, marital status). Interestingly, we found evidence of differences between the groups in reporting on the more sensitive survey items, which in some cases emerged to be significant (e. g., age of drinking onset, ever smoking marijuana, prolonged period of losing interest in things usually enjoyed). In general, we found more frequent reporting of less socially desirable behaviors in the text message groups. We also found that respondents in the modular design group found the survey to be significantly easier than respondents in the other two groups, and that the single-sitting text message group found the survey to be marginally easier than the telephone group.

4.3 Item Nonresponse Bias in the Modular Design Group

Although the vast majority of respondents in the modular design group responded to the survey questions on the same day that they were initially sent via text message, this group had higher item nonresponse rates than the other two groups (see the item-specific respondent counts in Table 1). This is a measurement problem that introduces an increased risk of nonresponse bias for *individual items*; we distinguish this problem from one of unit nonresponse, where an individual chooses not to participate in the entire survey. To assess this bias risk, we analyzed covariates that were fully available for all panel members, and examined differences in distributions on these variables between those responding to each item and the full sample assigned to the modular design group (Table 2). Large differences in these distributions between respondents and the full sample indicate potential nonresponse bias for any survey items correlated with these measures.

Overall, the differences between respondents and the full sample in the modular group were slight across the survey items. There is some evidence of a higher percentage of respondents to the mental health items being female, suggesting possible nonresponse bias if gender is predictive of mental health reports. We found that gender was in fact not significantly associated with reports on Q11, Q12, or Q13 (questions related to mental health), suggesting that the increased reporting on these items was not simply due to a higher proportion of females responding than expected. We also see evidence of respondents to the ease of survey question being more likely to be female; importantly, gender was not found to be associated with responses to this question either, both across groups and for the modular design group only, suggesting that the earlier finding of increased ease of survey participation was not simply due to more females responding to this question. Finally, we see some evidence of respondents to the self-reported health question tending to live farther away from markets and health posts, but these differences were also slight.

4.4 Regression Models

Next, we investigate multivariate models of those outcomes with the largest marginal mode differences. To do this, we combined measures indicating a history of depression (ever having a long period of feeling sad, empty or depressed; ever having a long period of feeling very discouraged about life; ever having a long period of losing interesting in things usually enjoyed) and a history of drug use (smoking marijuana; smoking other narcotics) into single indicators of any history of depression or drug use. We then fitted logistic regression models to these two indicators and the indicator of rating the survey as "very easy", including the aforementioned control variables and two-way interaction terms. We also fitted a linear regression model to the variable measuring age of drinking onset (for those indicating a history of drinking only). Table 3 presents the estimated coefficients in these four regression models.

The bivariate results were largely supported for the indicator of ever smoking marijuana or other narcotics, with the single-sitting text interview group found to have significantly higher odds of reporting this behavior than the telephone group when controlling for the other covariates. The same was true for the indicator of any depression history, with the modular design group having the highest odds of

Table 1
Comparisons of descriptive estimates and item response rates among the three study groups

	Voice Interview Group		Single-Sitting Text Message Group		Modular Design Group			*F*-Statistic / Chi-square (χ^2) Statistic, *p*-value	
Item	Est.	Resp. (%)	Est.	Resp. (%)	Est.	Resp. (%)	Same Day Resp. (%)	Statistic	*p*
Mean Age	21.2	100.0	20.7	100.0	21.0	73.9	90.5	F(2,392) = 1.6	0.21*
Married (%)	41.2	100.0	36.4	98.6	38.8	69.0	89.8	$\chi^2(2) = 0.7$	0.71
Mean Age First Married (If married)	19.2	41.2	18.4	35.9	19.4	15.5	86.4	F(2,131) = 1.6	0.21*
Traveled Outside of Nepal (%)	13.5	100.0	22.9	98.6	13.9	55.6	89.9	$\chi^2(2) = 5.2$	0.08
Mean Age First Drink (If drinker) (%)	17.1	28.4	15.3	30.3	10.8	7.0	60.0	**F(2,92) = 7.2**	**0.01**
Having Drink In Past Week (If drinker) (%)	11.9	28.4	21.6	26.1	16.0	17.6	88.0	$\chi^2(2) = 1.4$	0.51
Fair / Poor Self-Reported Health (%)	46.0	100.0	41.5	95.1	32.2	57.0	88.5	$\chi^2(2) = 2.8$	0.25
Mean Age First Smoking (If smoker)	16.0	12.2	13.7	19.7	11.6	4.9	71.4	F(2,50) = 2.4	0.10*
Ever Smoked Marijuana (%)	5.4	100.0	12.4	96.5	5.4	64.8	84.8	**$\chi^2(2) = 5.8$**	**0.05**
Ever Smoked Other Narcotic (%)	2.7	100.0	4.3	97.9	3.5	60.6	80.2	$\chi^2(2) = 0.6$	0.76
Ever Long Depression Period (%)	21.0	100.0	24.3	98.6	26.8	57.7	80.5	$\chi^2(2) = 1.1$	0.58
Ever Discouraged with Life (%)	18.2	100.0	24.1	99.3	27.0	62.7	88.8	$\chi^2(2) = 2.8$	0.25
Lose Interest In Things Enjoyed(%)	10.1	100.0	20.4	100.0	20.8	54.2	85.7	**$\chi^2(2) = 7.0$**	**0.03**
Use The Internet (%)	62.8	100.0	66.9	100.0	60.2	58.5	83.1	$\chi^2(2) = 1.1$	0.58
With Facebook Account (%)	66.2	100.0	69.0	100.0	62.8	54.9	84.6	$\chi^2(2) = 0.9$	0.64
Perceived Burden							84.9	**$\chi^2(2) = 19.9$**	**0.01**
Very Easy (%)	51.4	100.0	69.7	100.0	80.3	46.5			

* Bartlet's test of equal variance among the three groups was satisfied at the 0.01 level.

Table 2
Differences between respondents providing valid measures and the full sample of persons assigned to the modular design group in terms of means or percentages on selected auxiliary variables, by survey item

	Difference between Modular Design Respondents and the full Modular Design Sample in terms of ...									
Survey Item	Age[a]	Female[b]	School < 10 minutes walk[b]	Market < 10 minutes walk[b]	Health Post < 10 min. walk[b]	Currently Enrolled in School[a]	Months Enrolled in School since 2008[a]	Months Recored in CVFS Sample[a]	Support Groups[a]	Current Household Size[a]
Q1 (Age)	−0.10	−0.35	1.79	−1.94	−1.26	2.91	1.88	0.78	0.04	−0.29
Q2 (Marital Status)	−0.04	4.82	0.16	−1.40	−1.19	3.05	0.91	−0.51	0.20	−0.06
Q4 (Outside Travel)	0.04	2.31	0.60	0.02	0.50	0.47	0.04	−0.51	0.40	−0.31
Q7 (Self-Rep Health)	−0.19	0.17	−2.94	**−6.28**	−4.84	3.14	1.20	−0.52	0.25	−0.19
Q9 (Marijuana)	−0.05	2.42	0.74	−3.82	−0.50	0.23	0.50	−0.49	0.11	−0.25
Q10 (Other narcotic)	−0.16	−0.31	−0.93	−3.08	3.77	−1.82	0.26	−0.07	0.06	−0.39
Q11 (Long Depress)	−0.10	2.53	2.70	−0.50	1.27	−2.13	−0.15	−0.23	0.17	−0.50
Q12 (Discouraged)	−0.08	4.47	−0.07	−1.21	−1.25	1.49	0.05	−0.37	0.18	−0.20
Q13 (Lose Interest)	< 0.01	**5.09**	**5.17**	−1.77	0.38	−4.19	−1.29	−0.03	0.28	0.06
Q14 (Use Internet)	−0.08	0.59	2.96	−1.47	1.92	0.66	1.77	0.33	0.21	−0.19
Q15 (Facebook)	−0.18	−0.76	−0.99	−4.07	−1.48	−1.19	−0.60	−0.53	−0.22	−0.54
Q16 (Ease of Survey)	−0.01	**8.22**	**−5.65**	−3.72	0.38	0.79	0.59	0.70	−0.24	−0.53

Differences above five percentage points in boldface.
[a] Mean of auxiliary variable [b] Percentage of auxiliary variable

Table 3
Estimated Coefficients in Logistic and Linear Regression Models for Respondent Burden and Selected Sensitive Measures

	Survey Rated as Very Easy		Ever Smoked Marijuana or Other Narcotics		Age First Drinking Alcohol (Among Drinkers)		Any Evidence of Depression History	
Regression Model Predictor	b	S.E.	b	S.E.	b	S.E.	b	S.E.
Study Group (*Ref.:* Phone Interview)								
Modular Design	2.67	1.09**	−0.03	0.71	−4.44	2.18**	0.78	0.35**
Text Interview	0.34	0.42	0.91	0.43**	−2.07	1.05*	0.50	0.25**
Health Post < 20-minute Walk	−0.21	0.30	−0.40	0.47	0.62	1.32	0.15	0.28
School < 10-minute Walk	−0.05	0.31	0.26	0.49	−1.17	1.28	−0.19	0.29
Number of Support Groups	−0.10	0.05*	−0.02	0.09	−0.07	0.25	0.01	0.05
Months Enr. in School Since 2008	0.74	0.34**	−0.45	0.52	0.48	1.48	−0.36	0.33
Current Household Size	−0.12	0.05***	0.01	0.08	0.10	0.23	0.03	0.04
Caste/Ethnicity *(Ref.:* Brahmin)								
Hill Janajati	−0.37	0.38	1.29	0.57**	0.34	1.52	0.10	0.35
Dalit	−0.28	0.44	0.92	0.67	2.09	2.06	0.72	0.41*
Newar	0.48	0.61	-[a]	-[a]	3.08	2.78	−0.38	0.54
Terai Janajati	−0.05	0.35	0.94	0.56*	−0.98	1.44	−0.38	0.33
Male	0.96	0.41**	1.80	0.48***	2.69	1.24**	0.04	0.26
Age (Age - 20)	0.08	0.07	0.20	0.12	0.24	0.30	−0.03	0.07
Not Married	−0.91	0.46**	0.13	0.53	0.60	1.47	0.36	0.32
Intercept	1.22	0.62**	−4.02	1.00***	14.26	2.60***	−1.03	0.56*
Interactions								
Male × Modular Design	−1.51	0.98						
Not Married × Modular Design	−0.76	1.26						
Male × Text Interview	−1.45	0.61**						
Not Married × Text Interview	1.86	0.63***						
(Pseudo) R-squared	0.131		0.164		0.213		0.038	
Observations[b]	332		319		91		338	

[a] All respondents in this category had a value of 0 on the dependent variable, meaning that a coefficient could not be estimated.
[b] Sample sizes vary due to item-missing data on the variables analyzed.
* $p < 0.10$. ** $p < 0.05$, *** $p < 0.01$,

reporting any depressive symptoms and both text message groups having higher odds in general of reporting depressive symptoms compared to the telephone interview group. The modular design group was found to have a significantly ($p < 0.05$) lower mean age of drinking onset than the telephone group when controlling for the other covariates, and the single-sitting group was found to have a marginally lower mean age of drinking onset. No significant interactions were found in the models for these three variables.

We did find significant interactions in the model for the indicator of rating the survey as "very easy". First, among those who were married, the modular design group had the highest marginal probability of rating the survey as very easy, and the single-sitting text interview group had the *lowest* marginal probability of rating the survey as very easy (see Figure 1). Second, the positive effects of the text message modes on the probability of rating the survey as very easy (relative to the telephone mode) were much stronger for females relative to males, with the change in the effect of the single-sitting mode for females relative to males found to be significant (see Figure 2). Interestingly, among individuals assigned to the modular design group, females also had a significantly higher probability than males of responding to the burden question on the same day that it was sent ($p < 0.05$), which could be another indicator of increased convenience of the text message approaches for females. No other significant differences between socio-demographic groups in terms of same-day response behavior (in the modular design group) were found. These interactions suggest that the perceived convenience of the modular design approach is not only higher overall in a marginal sense, but may vary more substantially across demographic subgroups.

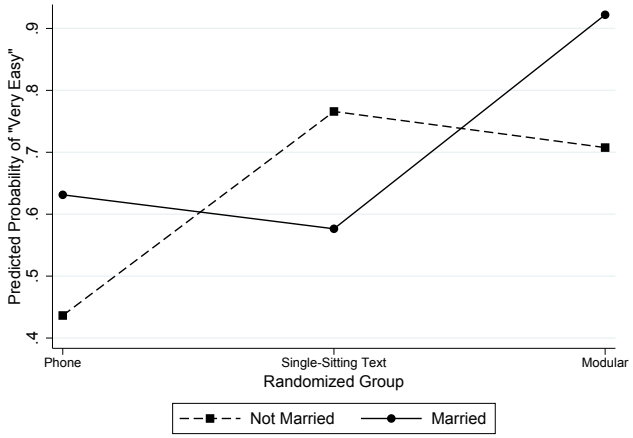

Figure 1. Predicted Probabilities by experimental setting and marital status

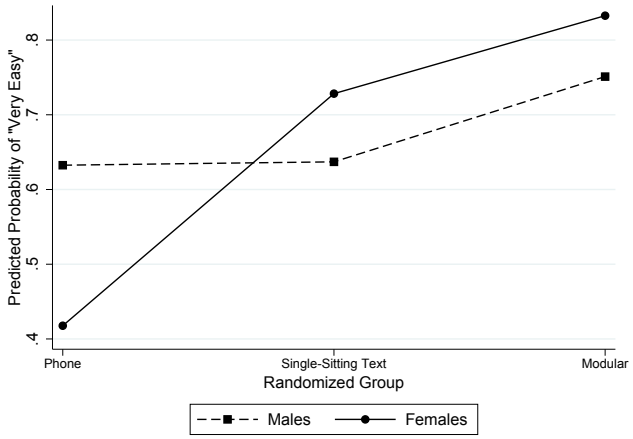

Figure 2. Predicted Probabilities by experimental setting and gender

4.5 Data Collection Costs

For the telephone interview group, the average interview time was 7.46 minutes, enabling easy calculation of the cost per interview. Given this average interview time, the cost per interview in this group was U.S. $6.87.

Calculation of the cost per interview in the two text messaging groups was much more complex. Because we did not automate the text messaging in the text message interview group (where interviews were completed in a single sitting), the interviewer had to wait for the response to a given question before they could send the respondent the next question. This resulted in additional waiting time for the interviewers, adding to the costs. The minimum interview time in this group was 14.0 minutes, and the mean interview time was 36.4 minutes. In addition to the interviewer's time, the cost of text messaging, both from the interviewer to the respondent and the respondent's response to the interviewer, also substantially contributed to the cost of each interview. This all resulted in a total estimated cost of U.S. $16.58 per interview.

After obtaining consent, interviewers in the modular design group had to spend about one minute to send one question per day for a total of 16 days, so the total time to complete one interview turned out to be 16 minutes plus 3 minutes for obtaining consent, resulting in approximately 19 minutes per interview. Because some of the respondents did not get all questions (as some were inappropriate and skipped), the minimum time was 12 minutes, with a maximum of 37 minutes and a mean of 21 minutes. Although the interview length was lower in this mode than in the "single sitting" mode, the additional cost of text messaging resulted in a total cost per interview that was still much higher than the telephone interview. This equated to a total estimated cost of U.S. $16.10 per interview.

The two text messaging modes therefore introduced a higher cost per interview in the absence of automation. Even so, this increased cost should be viewed in light of the potential data quality benefits associated with these two modes (discussed above). Clearly, simple automation of the text receipt re-text process could reduce these costs, and text messaging costs may be higher or lower in other countries. Future work on the costs of using text messaging for survey data collection should focus on the application of automation to this process, as this strategy has the potential to significantly drive down the text messaging costs.

5 Discussion

5.1 Summary of Results

This study presents evidence of the potential of modular survey designs to collect similar or higher-quality data on a variety of survey items relative to other survey modes, in a manner that is easier for survey respondents. Whether the increased reporting of sensitive events and behaviors in the texting modes (including the modular design group) is due to decreased time pressure or increased privacy in these modes relative to telephone (voice) interviewing remains a topic for future research.

In this special setting of high respondent cooperation and high response rates overall, we did find evidence of higher rates of item-nonresponse in the modular design group. However, respondents in this mode did not systematically vary from the full sample assigned to that design in terms of distributions on a variety of key characteristics, suggesting minimal nonresponse bias in estimates derived using variables that are correlated with these characteristics. In addition, we did not find evidence of significant differences between the groups in terms of response distributions for more factual survey items. Collectively, we view these results as

good news for the use of the text approaches in collecting both objective and sensitive information, but more work is still needed to reduce item nonresponse when using the modular design approach.

Our results are largely consistent with recent work in this new area (Schober et al., 2015), suggesting that the text message interviewing approaches offer benefits similar to that of self-administration. We add to the new literature in this area by demonstrating the potential of the modular design approach to reduce respondent burden while also generating response distributions on sensitive items that are largely consistent with self-administration. Furthermore, we find evidence of increased convenience of the modular design approach for specific subsamples; in this case, married individuals and females found the modular design approach to be much easier than the competing modes. These results suggest that specific population subgroups may prefer the modular approach, depending on the topic of the survey and the population being studied, and this increased convenience may ultimately improve the engagement of certain subgroups of interest in survey data collections.

Finally, our results also demonstrate that interviewing by text message is entirely feasible in remote Asian settings such as Nepal, and that the benefits in terms of higher-quality responses to sensitive items appear to hold up in this radically different setting, not just in wealthier European settings. This is important because the vast majority of the world population lives under conditions more similar to Nepal than to Europe and the United States. The large and more rapidly growing populations of Asia and Africa, in particular, create substantial new demand for survey data collection. As governments and non-government agencies struggle to provide services to these enormous populations, data about both life circumstances and human needs are crucial to targeting scarce resources. But the infrastructure for interviewing in these settings is different than in the United States and Europe. Rapid recent proliferation of cellular infrastructure into poor populations in both rural and urban areas in Asia and Africa has now made it possible to contact large fractions of these populations by phone and text messaging. Our study demonstrates that administering surveys by these modes is entirely feasible, that data from such measures are robust, and that these modes of contact are not an obstacle to reporting on potentially sensitive health and wellbeing issues.

5.2 Directions for Future Work

One of the major advantages of using mobile technologies for survey data collection is the ability to monitor behaviors and attitudes in respondents' everyday lives (similar to EMA and journal keeping methods). Although the results of this study suggest that assessing behaviors and attitudes that are unlikely to change substantially over short periods of time via text messaging may be superior to using telephone (voice) interviewing, possibly due to decreased time pressure, decreased burden, and/or increased privacy, one could argue that the same would be true when using a simple self-administered (paper-and-pencil or online) questionnaire. In addition, these self-administered questionnaires tend to be much less time- and cost-intensive than text messaging, and would thus seem to be the method of choice for one-time assessments of past behaviors and current attitudes that are not posited to change substantially over a short period of time. However, respondent convenience is an important issue facing modern survey researchers (Couper, 2013), and survey researchers need to continue seeking ways to keep survey respondents engaged in survey data collections (McCutcheon, 2015). Given that respondents found the modular design approach using text messaging to be much easier (which could be viewed as a measure of increased convenience) and provided responses similar to those that might be expected using self-administration, this mode may be as effective as simple mail or web questionnaires in practice.

The recruitment of underrepresented and/or hard-to-reach groups using brief text message surveys may also be easier than attempting to convince these groups to take longer and more complex web-based or paper surveys. If this proves to be true, then findings from this study would be more generalizable, and perhaps the methods could even be used to target specific groups (e. g., high risk groups such as injection drug users) in the future. This is certainly a testable hypothesis for future work. Future research should continue to compare costs (including those associated with automating the text message process), respondent burden / ratings of convenience, item nonresponse rates, and breakoff rates between modular designs and mail / web questionnaires to see if the modular approach offers additional advantages that could not be assessed in the present study. For example, modular design approaches may prove to be valuable in longitudinal studies employing continuous intensive monitoring of selected measures, where respondent burden is a critical issue (Bolger & Laurenceau, 2013).

Importantly, this study was conducted with a probability sample of young adults (ages 18-24) known to have mobile phones in a setting of high respondent cooperation (Western Chitwan in Nepal), enabling assessment of the measurement consequences of using the two text messaging modes. That assessment demonstrated that these modes were at least as good if not better than the telephone mode from a measurement error perspective. The sample itself is representative of both urban and rural South Asian populations experiencing dramatic increases in cellular phone and text message use, and therefore similar to much of Asia and the majority of the world's population. Nevertheless, these results clearly are not representative of all populations, and future research regarding the modular design approach certainly needs to examine the feasibility of this approach in other populations.

These crucial initial results open the door to more large-scale applications of text message and modular approaches to potentially reduce unit nonresponse, and bring a larger proportion of selected respondents into surveys. Given that the measurement differences found here were not harmful, the greatest benefits of these tools are likely to come in settings with rapidly declining response rates, such as the United States and Western Europe, and replications of this study in these cultures are certainly needed.

Although the results of this work are promising, effective reproduction of this research in other populations and settings with lower cooperation rates (e. g., the U.S. and Western Europe) will require survey researchers to address many remaining unanswered questions. First, what is the best way to convince representative samples of persons in general populations (possibly contacted using address-based sampling; e. g., Iannacchione, Staab, & Redden, 2003; Link, Battaglia, Frankel, Osborn, & Mokdad, 2008) to provide mobile phone or email contact information, and participate in surveys using modular designs? Second, can we develop applications for cooperating persons with smartphones, alerting them when "daily" questions are available for answering? Nearly two-thirds of Americans now own smartphones, and this number continues to rise (A. Smith, 2015). Data collections implementing the convenient modular design approach via smart phone applications therefore have the potential to reach a rapidly increasing fraction of the U.S. population. Future research in this area needs to examine differences in response behaviors depending on the type of mobile phone used (i. e., smartphone vs. feature phone); we did not have the resources to do this in the current study. Third, what strategies (e. g., cumulative incentives for each module answered) can be used to increase item response rates if only a small number of questions will be asked per day? Fourth, will there be selection bias for those who agree to participate using a given mode? Fifth, are there interviewer effects associated with modes of data collection that involve text messaging with human interviewers, and will automation eliminate these effects? All of these questions need to be addressed with future work in this area. Innovative studies designed to answer these questions will identify methodologies enabling more widespread implementation of modular design approaches.

Future work also needs to consider the feasibility of modular design approaches for web surveys and longer surveys with more content, using more diverse groups of individuals across a wider range of settings. The effort required to start a survey online may be higher, and engaging in this effort just to answer a single question on a given day may not be appealing to respondents. When administering longer surveys with more content, different "doses" of questions in modular designs need more exploration. For example, in a survey with 60 questions from four modules, should one module be asked per day? Or should only a fixed number of questions be asked? Modular designs open up many measurement possibilities, and the measurement quality associated with each design needs careful consideration.

The stability of response distributions generated using modular designs over short periods of time also requires further examination. Future studies using modular designs should assign *random* orders of survey questions to respondents to study this stability in more detail. The present study, which allowed respondents as much time as needed to respond to a daily question, did not employ this randomization and always asked questions in the same order to the modular design group. If the response distribution for a given survey item changes substantially over time during the course of a given survey, the modular design approach may not be as effective, instead representing a series of surveys on time-sensitive measures rather than the desired "one-time" survey. Ideally, during the time period of a given survey, the response distribution for a given survey item should not change substantially, and one could accumulate the responses collected on different days to form a final estimate. We do note that in repeated measures designs, where many measures of the *same* variable are desired over a short time period, this "rapid measurement" approach may work well (Axinn, Jennings, & Couper, 2015).

Although the results of the present study suggest that modular survey designs could simplify the survey response process for people using mobile technologies in their daily lives, this is one of the first steps in an exciting area that is ripe for future research.

Acknowledgements

This research was jointly supported by internal funds from the Population Studies Center at the University of Michigan Institute for Social Research, and a series of grants from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (R01HD032912, R01HD033551, R37HD039425, and R24HD041028) for the panel study. We thank Armani Hawes and Jennifer Mamer for their assistance with preparing this manuscript; the staff of the Institute for Social and Environmental Research–Nepal (ISER-N) for assistance with data collection; and the residents of the Western Chitwan Valley for their contributions to the research reported here. All errors and omissions in this manuscript are the sole responsibility of the authors.

Conflict of Interest. Dr. Ghimire is also the Director of the Institute for Social and Environmental Research in Nepal (ISER-N) that collected the data for the research reported here. Dr. Ghimire's conflict of interest management plan is approved and monitored by the Regents of the University of Michigan.

References

Abraham, K. G., Maitland, M., & 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*, 676–703.

Axinn, W. G., Ghimire, D. J., & Williams, N. E. (2012). Collecting survey data during armed conflict. *Journal of Official Statistics*, *28(2)*, 153–171.

Axinn, W. G., Jennings, E. A., & Couper, M. P. (2015). Response of sensitive behaviors to frequent measurement. *Social Science Research*, *49*, 1–15.

Axinn, W. G. & Yabiku, S. (2001). Social change, the social organization of families, and fertility limitation. *American Journal of Sociology*, *106(5)*, 1219–1261.

Barber, J. S., Shivakoti, G. P., Axinn, W. G., & Gajurel, K. (1997). Sampling strategies for rural settings: a detailed example from chitwan valley family study, nepal. *Nepal Population Journal*, *6(5)*, 193–203.

Biner, P. M. & Kidd, H. J. (1994). The interactive effects of monetary incentive justification and questionnaire length on mail survey response rates. *Psychology and Marketing*, *11(5)*, 483–492.

Blumberg, S. J. & Luke, J. V. (2013). Wireless substitution: early release of estimates from the national health interview survey, july - december 2012. National Center for Health Statistics. Retrieved from http://www.cdc.gov/nchs/nhis.htm

Bogen, K. (1996). The effect of questionnaire length on response rates - a review of the literature. 1020–1025. Proceedings of the Survey Research Methods Section of the American Statistical Association. Retrieved from http : / / www . amstat . org / sections / srms / proceedings

Bolger, N. & Laurenceau, J.-P. (2013). *Intensive longitudinal methods: an introduction to diary and experience sampling research*. New York: The Guilford Press.

Brenner, P. S. & DeLamater, J. D. (2014). Social desirability bias in self-reports of physical activity: is an exercise identity the culprit? *Social Indicators Research*, *117*, 489–504.

Brick, J. M. & Williams, D. (2013). Explaining rising nonresponse rates in cross-sectional surveys. *The ANNALS of the American Academy of Political and Social Science*, *645*, 36–59.

Burchell, B. & Marsh, C. (1992). The effect of questionnaire length on survey response. *Quality and Quantity*, *26*, 233–244.

Cocco, M. & Tuzzi, A. (2013). New data collection modes for surveys: a comparative analysis of the influence of survey mode on question-wording effects. *Quality and Quantit*, *47*, 3135–3152.

Couper, M. P. (2013). Is the sky falling? New technology, changing media, and the future of surveys. *Survey Research Methods*, *7*, 145–156.

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.

Edwards, P., Roberts, I., Sandercock, P., & Frost, C. (2004). Follow-up by mail in clinical trials: does questionnaire length matter? *Controlled Clinical Trials*, *25(1)*, 31–52.

Goyder, J. C. (1982). Further evidence on factors affecting response rates to mailed questionnaires. *American Sociological Review*, *47(4)*, 550–553.

Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey methodology* (2nd ed.). Wiley.

Heberlein, T. A. & Baumgartner, R. (1978). Factors affecting response rates to mailed questionnaires: a quantitative analysis of the published literature. *American Sociological Review*, *43(4)*, 447–462.

Iannacchione, V. G., Staab, J. M., & Redden, D. T. (2003). Evaluating the use of residential mailing addresses in a metropolitan household survey. *Public Opinion Quarterly*, *67(2)*, 202–210.

Johnson, A., Kelly, F., & Stevens, S. (2012). Modular survey designs for mobile devices. Retrieved from https://c.ymcdn.com/sites/www.casro.org/resource/collection/E270CC91 - 6B72 - 4C37 - BCC0 - 5503CBB66C55 / Paper_ - _Frank _ Kelly _ and _ Alex _ Johnson_ - _Lightspeed_Research_and_Kantar_Operations.pdf

Kahneman, D., Krueger, A. B., Schkade, D. A., Schwarz, N., & Stone, A. A. (2004). A survey method for characterizing daily life experience: the day reconstruction method. *Science*, *306(5702)*, 1776–1780.

Kalantar, J. S. & Talley, N. J. (1999). The effects of lottery incentive and length of questionnaire on health survey response rates: a randomized study. *Journal of Clinical Epidemiology*, *52(11)*, 1117–1122.

Kantorowitz, M. (1998). Is it true that nonresponse rates in a panel survey increase when supplement surveys are annexed? In A. Koch & R. Porst (Eds.), *Nonresponse in survey research* (pp. 121–137). Mannheim: ZUMA Nachrichten Spezial 4.

Kelly, F., Johnson, A., & Stevens, S. (2013). Modular survey design: bite sized chunks 2. Retrieved from http://c.ymcdn.com/sites/www.casro.org/resource/resmgr/conference/paper_-_frank_kelly_-_lights.pdf

Kuntsche, E. & Robert, B. (2009). Short message service (SMS) technology in alcohol research - a feasibility study. *Alcohol and Alcoholism*, *44(4)*, 423–428.

Link, M. W., Battaglia, M. P., Frankel, M. R., Osborn, L., & Mokdad, A. H. (2008). A comparison of address-based sampling (ABS) versus random-digit dialing (RDD) for general population surveys. *Public Opinion Quarterly*, *72(1)*, 6–27.

McCutcheon, A. L. (2015). Web surveys, online panels, and paradata: automating responsive design. The Joint Program in Survey Methodology Distinguished Lecture. Retrieved from http://jpsm.umd.edu

MOHP, New ERA, & ICF International Inc. (2007). Nepal demographic and health survey 2006. Kathmandu, Nepal. Retrieved from http://dhsprogram.com/pubs/pdf/FR191/FR191.pdf

MOHP, New ERA, & ICF International Inc. (2012). Nepal demographic and health survey 2011. Kathmandu, Nepal. Retrieved from http://dhsprogram.com/pubs/pdf/FR257/FR257[13April2012].pdf

Moskowitz, D. S. & Young, S. N. (2006). Ecological momentary assessment: what it is and why it is a method of the future in clinical psychopharmacology. *Journal of Psychiatry and Neuroscience*, *31(1)*, 13–20.

Peytchev, A. (2009). Survey breakoff. *Public Opinion Quarterly*, *73*, 74–97.

Raghunathan, T. E. & Grizzle, J. E. (1995). A split questionnaire survey design. *Journal of the American Statistical Association*, *90(429)*, 54–63.

Schaeffer, N. C., Dykema, J., & Maynard, D. W. (2010). Interviewers and interviewing. In J. D. Wright & P. V. Marsden (Eds.), *Handbook of survey research* (2nd ed., pp. 437–470). Bingley, UK: Emerald Group Publishing Limited.

Schembre, S. M. & Yuen, J. (2011). Project TwEATs: a feasibility study testing the use of automated text messaging to monitor appetite ratings in a free-living population. *Appetite*, *56(2)*, 465–468.

Schober, M. F., Conrad, F. G., Antoun, C., Ehlen, P., Fail, S., Hupp, A. L., … Zhang, C. (2015). Precision and disclosure in text and voice interviews on smartphones. *PLOS ONE*, *10*(6), e0128337. Retrieved from http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0128337

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology*, *4*, 1–32.

Smith, A. (2011). Americans and text messaging. Pew Research Center. Retrieved from http://www.pewinternet.org/2011/09/19/americans-and-text-messaging/

Smith, A. (2015). The smartphone difference. Pew Research Center. Retrieved from http://www.pewinternet.org/2015/04/01/us-smartphone-use-in-2015/

Smith, R., Kotzev, I., Miller, C., & Kachhi, D. (2012). Modularizing surveys to meet today's respondent challenges. Retrieved from https://c.ymcdn.com/sites/www.casro.org/resource/collection/E270CC91-6B72-4C37-BCC0-5503CBB66C55/Paper_-_Renee_Smith_and_Chuck_Miller_-_Kantar_and_uSamp.pdf

Stone, A. A. & Shiffman, S. (1994). Ecological momentary assessment (EMA) in behavioral medicine. *Annals of Behavioral Medicine*, *16(3)*, 199–202.

VanGeest, J. B., Johnson, T. P., & Welch, V. L. (2007). Methodologies for improving response rates in surveys of physicians: a systematic review. *Evaluation and the Health Professions*, *30(4)*, 303–321.

Yammarino, F. J., Skinner, S. J., & Childers, T. L. (1991). Understanding mail survey response behavior: a meta-analysis. *Public Opinion Quarterly*, *55*, 613–639.

Zabel, J. E. (1994). *A comparison of attrition in the panel study of income dynamics and the survey of income and program participation (no. 189)*. United States Bureau of the Census. Retrieved from http://www2.census.gov/prod2/sipp/wp/SIPP_WP_189.pdf