Journal of Official Statistics, Vol. 38, No. 4, 2022, pp. 963–986, http://dx.doi.org/10.2478/JOS-2022-0042

Testing a Planned Missing Design to Reduce Respondent Burden in Web and SMS Administrations of the CAHPS Clinician and Group Survey (CG-CAHPS)

Philip S. Brenner[1], *J. Lee Hargraves*[2], *and Carol Cosenza*[2]

We test a planned missing design to reduce respondent burden in Web and SMS administrations of the CAHPS Clinician and Group Survey (CG-CAHPS), a survey of patient experiences widely used by health care providers. Members of an online nonprobability panel were randomly assigned to one of three invitation and data collection mode protocols: email invitation to a Web survey, SMS invitation to a Web survey, or SMS invitation to an SMS survey. Within these three mode protocols, respondents were randomly assigned to a planned missing design, which shortened the survey by about 40%, or to a control group that received the survey in its entirety. We compare survey duration, breakoff and completion rates, and five key patient experience measures across conditions to assess the effect of the planned missing design across the three modes. We found that a planned missing design worked well with our Web survey, reducing survey duration and breakoff without changing estimates relative to the full-survey control condition. However, mixed findings in the SMS survey suggest that even shortened, 15-item surveys may be too long to substantially reduce respondent burden. We conclude with recommendations for future research.

Key words: Online data collection; text messaging; health.

1. Introduction

The high demand for cost-effective survey designs has been an impetus for methodological innovation (Schonlau and Couper 2017; De Leeuw 2018). Modes that contact, recruit, and measure electronically, such as Web surveys and text message surveys, fit this bill, allowing cost-effective designs that can leverage institutional resources (e.g., licenses for Web and text survey applications) to dramatically reduce costs (Dillman et al. 2014). Yet, taking advantage of these innovations to reduce costs may encounter an unintended consequence: high respondent burden (Crawford et al. 2001; Mavletova et al. 2018). Although respondent burden lacks a clear conceptualization, it likely results from multiple factors, including, but not limited to, survey length, complexity, and the effort required for completion (Crawford et al. 2001; Mavletova et al. 2018; Yan et al. 2020). These factors, however, do not determine respondent burden as they may be balanced or even outweighed by the positive characteristics of the survey, such as the respondent's interest in or perceived importance of the topic (Bradburn 1978; Sharp and Frankel 1983).

[1] Department of Methodology and Statistics, Padualaan 14, 3584 CH Utrecht, The Netherlands. Email: p.s.brenner@uu.nl.
[2] Center for Survey Research, University of Massachusetts Boston, 100 Morrissey Blvd. Boston, MA 02125 (617) 287-7200 U.S.A. Emails: lee.hargraves@umb.edu and carol.cosenza@umb.edu

High respondent burden generated by long and complex self-administered surveys such as those conducted on the Web or by short message service (SMS), may cause fatigue and resultant breakoffs that potentially harm data quality (Galesic and Bosnjak 2009; Revilla and Ochoa 2017). While cutting questions, condensing questionnaires, and curtailing complexity may help to reduce respondent burden in this case (Mavletova and Couper 2015; Toepoel and Lugtig 2018), this remedy may cause a side-effect. Even in a shortened and simplified survey questionnaire, important and complex concepts must be adequately measured for the survey to accomplish its goals (Keller et al. 2005). If too many questions are cut or if cuts are made haphazardly and without consideration of their effect on resulting data quality, reliability and validity, especially concurrent and discriminant validities, may be harmed (Ng et al. 2016). A promising solution – a planned missing design – randomly assigns respondents to answer only a subset of questions to shorten the survey and reduce respondent burden (Johnson et al. 2013).

Patient surveys provide an illustrative example of these cross-cutting demands that may be met with a planned missing design. Data from patient surveys are used by government agencies and insurance companies in reimbursement and pay-for-performance computations and are made available to consumers for their medical decision-making (Elliott et al. 2016; Holt 2019; Spranca et al. 2000). However, many healthcare providers and professionals view patient surveys as tertiary to providing care and see high data collection costs as an unnecessary burden on their already strained budgets (Alemi and Jasper 2014). As a result, many hospitals and clinics are demanding shorter self-administered patient surveys that take advantage of methodological and technological innovations to improve cost-effectiveness and allow increased flexibility and customizability (Keller et al. 2005; Lee et al. 2013; Stucky et al. 2016).

Therefore, we test the use of a planned missing design with the CAHPS Clinician and Group Survey (CG-CAHPS), a survey of patient experiences widely used by healthcare providers. We invite members of an online panel by email or SMS to complete a survey administered in one of two modes, by Web or SMS. Panelists who have had a visit with a physician in the past six months are assigned to one of three shortened modules or to the full version of the survey. Comparisons are made between the different invitation and data collection mode conditions and planned missing modules of the survey to assess the effects of these designs on respondent burden, operationalized as survey duration and completion and breakoff rates, and key patient experience outcomes.

2. Background

2.1. Respondent Burden and Web Surveys

Web surveys can be designed to reduce respondent burden (Mavletova and Couper 2015; Toepoel and Lugtig 2018). As a computerized data collection method, Web surveys can be programmed to simplify the respondent's task by streamlining their path through the questionnaire, using survey software to hide complex routing generated by filter questions and skip patterns (Dillman et al. 2014). However, respondent burden is not necessarily alleviated by computerization and the potential for problems remains (Crawford et al. 2001). If questions are too many or too complex, respondents may cope with the high cognitive demands by engaging in counterproductive behaviors that reduce data quality, such as straight-lining, satisficing, skipping questions, or breaking off from the survey before completion (Conrad et al. 2017; Kim et al. 2019; Zhang and Conrad 2018).

Thus, keeping Web surveys short and simple helps to reduce respondent burden and improve data quality (Mavletova and Couper 2015; Toepoel and Lugtig 2018). A planned missing design fits this need well as an effective way to reduce survey length and complexity (Peytchev and Peytcheva 2017). This approach reduces survey length by randomly assigning respondents to answer only subsets of questions (Enders 2010; Johnson et al. 2013; Graham et al. 2006). Given random assignment, unasked and unanswered questions generate nonresponse patterns in data sets that are known to be missing completely at random (Johnson et al. 2013). As a result, planned missing designs do not bias results, although they may reduce statistical efficiency for some types of analyses as they can shrink analytic sample sizes (Rhemtulla et al. 2016) – the price paid to shorten survey length, relieve respondent burden, and prevent resultant breakoffs. Thus, this study applies a planned missing design to reduce respondent burden in a Web survey.

2.2. Respondent Burden in SMS Surveys

An innovative extension to Web survey methods allows recruitment and data collection to be shifted to SMS. This approach takes advantage of many Web survey platforms' capabilities to recruit and measure using text messaging. SMS shows promise during recruitment to a Web survey, improving response when used as a pre-notification contact preceding an email invitation (Bosnjak et al. 2008) and yielding higher response than emailed invitations when used as a recruitment mode (Mavletova and Couper 2014). Importantly, using SMS as a recruitment or invitation mode in the United States requires an appropriate sampling frame that includes a cell number and permission to contact via text.

SMS also offers some potential benefits as a data collection mode including respondents' increased willingness to disclose sensitive information and a higher response rate compared to interviewer administration (Schober et al. 2015; West et al. 2015). However, SMS data collection also extends the duration of the survey due to the asynchronous nature and norms of texting, which may increase respondent burden and the potential for breakoffs (West et al. 2015). Much like for Web surveys, best practices suggest keeping SMS surveys short and simple to prevent breakoffs and errors resulting from respondent fatigue and high respondent burden (Lau et al. 2019). Therefore, SMS surveys may also benefit from a planned missing design to reduce survey length and respondent burden. Thus, this study novelly tests a planned missing design in an SMS survey.

2.3. Respondent Burden and Patient Surveys

Surveys of patients are well-positioned to benefit from innovations in recruitment and data collection, such as planned missing designs in Web and SMS surveys. One such survey, the CAHPS Clinician and Group Survey (CG-CAHPS), is used by general practitioners and specialists to assess patient experiences after office visits and outpatient procedures. CG-CAHPS collects information about patients' experiences with their provider and other medical and clinic staff to make comparisons across medical practices, giving providers a report card and patients the insights they need to make informed choices (Dyer et al. 2012;

Elliott et al. 2016; Holt 2019; Spranca et al. 2000). Data are commonly, though not exclusively, collected using mail surveys as a primary data collection mode, with telephone interviewers following up with nonrespondents (Elliott et al. 2009; Stucky et al. 2016). However, technological and methodological innovation may change how CG-CAHPS and other surveys of patients recruit and measure (Lee et al. 2013) in our increasingly online and cell-only society (Blumberg and Luke 2020; Pew Research Center 2019). Many medical practices request patients' email addresses and cell numbers to interact with them via medical portals and send them appointment reminders and confirmations (Garrido et al. 2016). Given the multiple types of contact information on patient lists, they provide rich sampling frames that offer multiple options for recruitment and measurement. Using this contact information to shift survey recruitment and data collection to Web and SMS may reduce costs and increase timeliness of response for CG-CAHPS surveys relative to mail and telephone data collection. Yet, without simplification and abbreviation of the questionnaire, respondent burden may yield relatively high rates of breakoffs, lower rates of completion, and other negative effects on data quality.

Therefore, we test the use of a planned missing design in a survey of patient experiences. Members of an online panel were invited by email or text message to complete a Web survey or an SMS survey. We assess and compare evidence of respondent burden across four versions of the questionnaire: three abbreviated questionnaires applying a planned missing design and a control group that includes the full survey instrument. We operationalize respondent burden in three ways: more burdensome modules will take longer to complete, yield more breakoffs, and, accordingly, lower completion rates. We also compare survey progress at interim stages of the recruitment process – click (on URL), screening, and eligibility rates – to assess differences in recruitment and survey progress between modes and modules given the negative effect of respondent burden (Galesic and Bosnjak 2009; Revilla and Ochoa 2017). We then assess the effect of respondent burden and the planned missing design on patient experience outcomes in the CG-CAHPS survey in both Web and SMS survey modes. A lack of differences in patient experience outcomes from the shortened questionnaires and the full-survey control condition would provide evidence for the effectiveness and appropriateness of using a planned missing design to reduce respondent burden with CG-CAHPS in these modes. To our knowledge, this is the first study using a planned missing design to examine the differential effects of respondent burden in Web and SMS modes.

3. Methods

3.1. *Sample Design and Justification*

Although the intended outcome of the planned missing design is higher quality data (Peytchev and Peytcheva 2017), removing questions may potentially cause unintended errors. Omitted questions may change the context in which subsequent questions are asked, introducing artifactual differences between questionnaire versions (Morgan and Poppe 2015; Swain 2015). These artifactual differences generated by methodological experiments could expose healthcare providers to negative outcomes by invalidating comparisons between providers and reducing the usefulness of CG-CAHPS data which are used by government agencies and insurance companies in reimbursement and pay-for-performance computations and by consumers for their medical decision-making (Elliott et al. 2016; Holt 2019; Spranca et al. 2000). Thus, it is critical that if alternative methods of data collection are employed, such as planned missing designs in Web and SMS surveys, they not differentially affect CG-CAHPS estimates. Therefore, we test this design using a sample from a nonprobability online panel rather than nesting it within a production survey using a clinic's patient list as a sampling frame. Fifty-thousand members of Qualtrics' online nonprobability panel were invited to participate in a survey using their standard, generic invitation message. Panelists were offered an incentive given their pre-stated preferences (e.g., frequent flyer program miles, hotel brand loyalty points).

3.2. Invitation and Data Collection Modes

Combining the mode of invitation (email or SMS) with the mode of data collection (Web or SMS) resulted in three conditions: email invitation to a Web survey, SMS invitation to a Web survey, and SMS invitation to an SMS survey. Note that a fourth combination, an email invitation to an SMS survey, is possible. However, this design would require an additional step not required by the other designs. An invitation to a Web survey, either by email or by SMS, only requires clicking a link to open a browser. An SMS invitation to an SMS survey can simply be responded to. These approaches seem to us to be relatively seamless. An email invitation to an SMS survey would require a more active switch on the part of the respondent, moving from an email application to a texting application, entering a telephone number or shortcode, and sending a message to this new number. While shifting from email recruitment to SMS data collection is possible and has been used previously (Brenner and DeLamater 2013), we did not foresee it being a useful design for a CAHPS survey and did not test it in this study.

Email invitations were sent to 20,000 panelists that included a URL that forwarded to the survey. Another 20,000 panelists were invited by SMS to complete a Web survey and received a link identical to those sent via email. The remaining 10,000 panelists were invited by SMS to complete an SMS survey. These panelists received questions sequentially by SMS after agreeing to participate by responding to the initial message. The difference in sample sizes is partially due to the number of panelists available and partially due to cost. SMS surveys encounter extra costs in addition to the Web survey platform license, including an additional cost for SMS survey capabilities and per-message transmission costs for both outgoing questions and incoming answers. Data collection began in August 2019 and was completed in September 2019.

3.3. Planned Missing Experiment and Respondent Burden

In each mode, panelists were randomly assigned to one of four survey modules. Three of these modules used a planned missing design to reduce the length of the survey. Each of the first three survey modules included a subset of questions. Module A included questions assessing access to the provider; module B assessed communication with the provider; and module C assessed coordination of care. All three of these survey modules also included two questions about their provider's office staff, an overall provider rating, patient demographic questions, and self-rated health. These shortened modules included 14, 15, or

16 questions; shorter if the respondent answered "no" to one or more filter questions. The final module, module D, served as a control group and included all questions asked in each of the first three modules; a total of 25 questions, shorter if the respondent answered "no" to one or more filter questions.

The planned missing design is used to manipulate respondent burden as the shortened questionnaire modules are hypothesized to relieve respondent burden relative to the full-length questionnaire module. Following previous theory and research, respondent burden is operationalized as survey duration, and reflected in breakoff rates and, accordingly, completion rates (Antoun and Cernat 2020; Bradburn 1978; Mavletova and Couper 2015; Peytchev 2009; Sharp and Frankel 1983; Steinbrecher et al. 2015). Breakoff rates are computed as the percentage of eligible respondents who started the survey but did not complete it. Completion rates are computed as the percent of invited panelists who complete the survey.

3.4. Dependent Variables

Panelists were first screened for a doctor's visit in the past six months to be eligible for the survey. Subsequent questions refer to the provider identified in the screening question. Five dependent variables, created by combining question responses into composites, are measured and analyzed. The first is a composite combining three items, each measuring the patient's perception of the accessibility of their provider. The second combines four items, assessing how often the provider communicated effectively with the patient. The third combines three items, each assessing coordination of patient care. The fourth combines two items, measuring the patient's assessment of the courteousness of the medical office staff. The final dependent variable is a single item measure of the patient's assessment of his or her provider.

The four composites use scales demonstrated to have good internal consistency (Dyer et al. 2012). Each measured variable was recoded following common practice with CAHPS measures (never (0), sometimes (3.33), usually (6.67), and always (10)). For each scale, these values were then averaged and rounded to the nearest integer to create 11-point (0–10) composite measures (AHRQ n.d.). For scales with embedded filter questions, values for unasked follow-up items were missing if respondents responded "no" to the filter question. Scales values were computed using remaining answered scale items. This procedure puts these first four dependent variables on a scale similar to that of the fifth dependent variable, a 0 to 10 overall rating of the provider. See Table 1 for the full question text, response options, question order, and use across planned missing modules.

The survey was optimized for mobile devices, including only item specific scales and using no question grids, drop down menus, or sliders. Nearly all of the questions were successfully adapted for SMS, shortened to 160 characters or fewer, a US-based limit for text messages. Unfortunately, five questions could not be shortened enough without changing their meaning or excluding important information and, thus, exceeded this character limit. Note that most smartphones and carriers reassemble messages exceeding this limit before they appear on the device, and therefore we do not expect that this small subset of longer questions caused a problem for SMS respondents (Ayers et al. 2014).

Table 1. Question order, wordings, and response options for each CAHPS scale, by module.

Scale	Question	Response options	In modules			
Screening question	In the last six months, did you get medical care from a health care provider (physician, nurse practitioner, or physician's assistant)? Do not include dental care or overnight stays in a hospital.	YN	A	B	C	D
	Is this the provider you usually see if you need medical care? If you saw more than one provider, think about the one you saw the most.	YN	A	B	C	D
Provider accessibility	In the last six months, did you contact this provider's office to get an appointment for an illness, injury, or condition that needed care right away?	YN	A			D
Provider accessibility	In the last six months, when you contacted this provider's office to get an appointment for care you needed right away, how often did you get an appointment as soon as you needed?	NSUA	A			D
Provider accessibility	In the last six months, did you make any appointments for a check-up or routine care with this provider?	YN	A			D
Provider accessibility	In the last six months, when you made an appointment for a check-up or routine care with this provider, how often did you get an appointment as soon as you needed?	NSUA	A			D
Provider accessibility	In the last six months, did you contact this provider's office with a medical question during regular office hours?	YN	A			D
Provider accessibility	In the last six months, when you contacted this provider's office during regular office hours, how often did you get an answer to your medical question that same day?	NSUA	A			D
Effective communication	In the last six months, how often did this provider explain things in a way that was easy to understand?	NSUA		B		D
Effective communication	In the last six months, how often did this provider listen carefully to you?	NSUA		B		D
Care coordination	In the last six months, how often did this provider seem to know the important information about your medical history?	NSUA			C	D
Effective communication	In the last six months, how often did this provider show respect for what you had to say?	NSUA		B		D
Effective communication	In the last six months, how often did this provider spend enough time with you?	NSUA		B		D
Care coordination	In the last six months, did this provider order a blood test, x-ray, or other test for you?	YN			C	D

Table 1. Continued

Scale	Question	Response options	In modules			
Care coordination	In the last six months, when this provider ordered a blood test, x-ray, or other test for you, how often did someone from this provider's office follow up to give you those results?	NSUA			C	D
Provider rating	Using any number from 0 to 10, where 0 is the worst provider possible and 10 is the best provider possible, what number would you use to rate this provider?	0-10	A	B	C	D
Care coordination	In the last six months, did you take any prescription medicine?	YN			C	D
Care coordination	In the last six months, how often did you and someone from this provider's office talk about all the prescription medicines you were taking?	NSUA			C	D
Office staff rating	In the last six months, how often were clerks and receptionists at this provider's office as helpful as you thought they should be?	NSUA	A	B	C	D
Office staff rating	In the last six months, how often did clerks and receptionists at this provider's office treat you with courtesy and respect?	NSUA	A	B	C	D

Note: NSUA stands for "Never, Sometimes, Usually, Always"; YN stands for "Yes, No."

3.5. Control Variables

Self-rated health was included as a standard question that asked "In general, how would you rate your overall health?" using a five-point scale: excellent, very good, good, fair, or poor. Self-rated health is included in these analyses as it is predictive of patient experiences and healthcare utilization and is used as a patient-mix adjustment variable for comparing healthcare organizations (DeSalvo et al. 2004; Elliott et al. 2009; Paddison et al. 2013). Mental health was also measured using a question that asked: "In general, how would you rate your overall mental or emotional health?" and used the same response scale as the general health question. This measure of mental health is included as it is predictive of patient experiences and healthcare utilization (Ahmad et al. 2014). Both health variables have been recoded so that higher values reflect better health.

A series of demographic controls, including education (less than a high school degree, high school degree or GED, some college or two-year degree, four-year college degree, or more than a college degree), sex (male or female), and age (18–24, 25–34, 35–44, 45–54, 55–64, and 65 or older) are also included as categorical variables. Descriptive statistics for these independent variables and significance tests comparing across modes and modules are available in Table 2.

Table 2. Independent variables descriptive statistics and comparsion by mode and module

Mode	Email/Web					SMS/Web					SMS/SMS					Mode diff.
Module	Overall	A	B	C	D	Overall	A	B	C	D	Overall	A	B	C	D	
Self-rated health (%)																$X^2(8)=47.4$ ***
Excellent	14.0	14.1	14.3	13.3	14.4	12.1	13.0	11.4	12.8	11.1	7.8	8.3	5.9	8.1	8.9	
Very good	30.8	35.6	27.9	29.6	30.2	25.2	27.9	22.7	24.5	25.7	23.3	22.9	19.3	25.9	25.2	
Good	33.2	30.4	34.8	31.1	36.9	36.1	33.0	39.0	37.2	35.3	36.9	34.9	32.6	43.7	35.8	
Fair	18.3	17.3	18.8	21.6	15.1	21.3	20.7	20.3	21.3	25.7	26.9	28.4	35.6	17.8	26.0	
Poor	3.7	2.6	4.2	4.4	3.5	5.3	5.3	6.6	4.3	11.1	5.2	5.5	6.7	4.4	4.1	
Module diff.					$X^2(12)=15.9$					$X^2(12)=7.8$					$X^2(12)=14.1$	
Mental health (%)																$X^2(8)=53.2$ ***
Excellent	19.7	18.6	19.8	20.1	20.5	15.2	16.2	14.6	14.7	15.2	15.6	17.6	10.4	16.4	18.7	
Very good	27.1	29.4	28.9	23.9	26.5	23.1	24.7	23.0	21.6	22.9	20.2	21.3	19.3	26.1	13.8	
Good	27.8	26.8	29.1	28.6	26.7	28.3	28.2	30.0	27.7	27.4	27.2	30.6	28.9	23.9	26.0	
Fair	20.2	19.8	17.3	21.9	21.8	24.5	22.9	24.1	26.1	25.0	26.8	21.3	28.1	27.6	29.3	
Poor	5.1	5.4	4.9	5.6	4.5	9.0	8.0	8.5	9.9	9.6	10.2	9.3	13.3	6.0	12.2	
Module diff.					$X^2(12)=7.7$					$X^2(12)=3.6$					$X^2(12)=15.4$	
Sex (%)																$X^2(2)=127.9$ ***
Female	63.8	67.7	58.7	64.8	63.6	78.2	76.1	77.9	80.5	78.4	84.6	82.2	84.3	84.1	86.7	
Module diff.					$X^2(3)=7.6$					$X^2(3)=2.2$					$X^2(3)=1.0$	
Age (%)																$X^2(10)=69.2$ ***
18-24	11.2	11.3	10.4	11.6	11.4	11.6	8.2	10.6	16.2	11.2	14.0	8.3	12.6	10.5	21.7	
25-34	25.9	23.3	22.8	28.3	29.2	25.5	25.8	26.5	24.2	25.5	29.6	35.2	29.6	30.1	25.5	
35-44	25.1	27.1	24.3	24.7	24.3	26.4	27.1	29.4	24.2	25.0	27.2	30.6	28.1	24.1	26.7	
45-54	16.0	16.3	16.8	15.6	15.3	19.7	19.7	19.0	17.0	22.9	14.9	11.1	14.8	16.5	16.1	
55-64	11.5	13.7	13.4	9.4	9.7	12.8	15.2	10.0	14.1	12.0	10.2	9.3	13.3	9.8	8.7	
65 and older	10.3	8.3	12.4	10.5	10.2	4.1	4.0	4.5	4.3	3.5	4.1	5.6	1.5	9.0	1.2	
Module diff.					$X^2(15)=16.3$					$X^2(15)=22.2$					$X^2(15)=30.9$ **	
Education (%)																$X^2(8)=49.4$ ***
Less than HS	3.3	4.9	1.5	4.2	2.2	3.7	4.5	2.4	4.8	3.2	4.0	1.0	1.5	5.4	7.0	
HS grad./GED	23.4	21.2	23.8	23.6	25.0	22.8	22.9	19.0	23.9	25.3	21.6	21.9	20.1	16.9	26.6	
Some coll., 2yr deg.	35.9	36.0	33.7	36.1	37.9	44.0	41.1	47.6	47.2	40.3	47.8	45.7	58.2	43.8	43.7	
4-yr coll. degree	26.9	28.2	28.2	25.4	26.0	19.3	18.1	21.2	17.4	20.3	17.8	21.0	13.4	17.7	19.6	
More than 4yr deg.	10.5	9.6	12.9	10.7	8.9	10.2	13.3	9.8	6.7	10.9	8.7	10.5	6.7	16.2	3.2	
Module diff.					$X^2(12)=17.0$					$X^2(12)=21.6$ *					$X^2(12)=33.2$ ***	
Race and ethnicity (%)																$X^2(6)=10.7$
White	68.1	69.9	69.4	67.0	66.1	69.9	72.0	70.1	68.3	69.1	74.8	76.2	75.9	76.2	71.6	
Black	14.4	12.7	12.9	15.8	16.1	13.4	11.7	14.0	13.1	14.9	9.6	7.6	7.5	11.1	11.6	
Latino/a, any race	12.0	11.8	13.7	11.4	11.1	11.0	12.3	8.7	13.1	10.1	10.2	10.5	11.3	9.5	9.7	
Asian or other race	5.5	5.6	4.0	5.8	6.7	5.7	4.0	7.1	5.6	5.9	5.4	5.7	5.3	3.2	7.1	
Module diff.					$X^2(9)=7.7$					$X^2(9)=9.3$					$X^2(9)=4.5$	

Note: *$p<.05$; **$p<.01$; ***$p<.001$

3.6. Analysis

We first compare screening and eligibility rates by predicting a series of nested logistic regression models. The first models are estimated with dummy variables for respondents' randomly assigned invitation and data collection mode (email/Web, SMS/Web, SMS/SMS) as the sole predictor. Model fit is compared and assessed using likelihood ratio X^2 tests. Given significant improvement in fit, pairwise comparisons between categorical predictors are presented. The second models add dummy variables for planned missing design module (Modules A, B, C, and D) as a predictor. Given that screening for eligibility takes place before the respondent is exposed to their randomly assigned questionnaire module, we expect that module will have no effect on screening and eligibility rates. Finally, the interaction between mode and module is tested.

We similarly assess measures of respondent burden by estimating nested logistic regression models that predict the propensity of a case to be completed or to breakoff using mode and module as predictors. We then assess survey duration, the final measure of respondent burden, by estimating a median regression predicting survey duration using mode and module as predictors. Given our expectation of a large difference in completion times between Web and SMS surveys, as suggested in the existing research (Schober et al. 2015) they are analyzed separately.

We then examine the distributions of the demographic and self-reported health variables between modes and modules using chi-square tests. These demographic and health variables are used as covariates in the final set of models assessing the effect of the planned missing design on patient experience outcomes, the key measures of the CG-CAHPS survey. A set of nested ordinary least squares regression models are estimated and adjusted means are presented for each outcome: provider accessibility, communication with provider, coordination of care, assessment of the provider's office staff, and the overall provider assessment. The baseline model adjusts for demographic and health variables. We then estimate two additional sets of models. First, we add dummy variables for mode of invitation and data collection, comparing this model to the prior baseline model using an F-test to test the statistical significance of including mode as a predictor of the dependent variable. Where the F-test indicates by-mode differences, predicted values of the dependent variables are compared by mode. We then add dummy variables for the four questionnaire modules (A, B, C, and D), comparing this model to the prior model including mode. Where the F-test indicates by-module differences, predicted values for the dependent variable are compared by module. Finally, we test the interaction of mode and module by comparing a model with this interaction term to the prior model using an F-test.

4. Results

4.1. URL Click, Screening, Eligibility, Breakoff, and Completion Rates

About 14% of those receiving email invitations to a Web survey (2,788 panelists) started the survey by clicking on the URL. Of those panelists who clicked the link to the survey, 82% (2,272 panelists) completed the first question, a screener for a recent (in the past six months) doctor visit. Three-quarters (1,721 panelists) of those answering the screening question were eligible, and nearly all (97%) of these eligible cases completed the survey

(1,665 respondents), yielding a 3.3% breakoff rate and an 8.3% completion rate in the email/Web condition. Sample sizes, eligibility, screening, breakoff, and completion rates by survey mode and module are available in Table 3.

In the condition using an SMS invitation to a Web survey, 2,131 (11%) of those receiving a text invitation started the survey by clicking the link in their SMS messaging application. Of those panelists who clicked the link they received by SMS, 1,991 (93%) answered the screening question. Over 80% (1,635 panelists) of those answering the screening question were eligible, and 90% (1,473 respondents) of these eligible panelists completed the survey, yielding a 9.9% breakoff rate and a 7.4% completion rate in the SMS/Web condition.

Nearly 10% (977) of those invited by SMS to the SMS survey started the survey by replying to the screener question. Note that this condition lacks a URL to click. Thus, the survey begins with screening: an initial text message that asks a question to determine eligibility. Responding to this initial question functions both as the initial "URL click" of a sort and an eligibility screen. Of the panelists who answered the screening question, 787 (81%) reported an eligible doctor's visit. Of the eligible cases, 70% or 550 respondents, completed the survey yielding a 30% breakoff rate and a completion rate of 5.5% in the SMS/SMS condition.

We estimated logistic regression models predicting screening and eligibility using mode as an independent variable. Differences emerge in screening ($X^2 = 27.6$, $p < .001$) and eligibility rates ($X^2 = 27.6$, $p < .001$) between modes during recruitment (see Table 4). Given the significant X^2 statistics, we compared modes pairwise. When predicting screening, significant differences emerge between email/Web and both of the other modes, SMS/Web ($z = 4.55$; $p < .001$) and SMS/SMS ($z = 4.27$; $p < .001$). The small difference between the screening rate in email/Web (11%) and that in SMS/Web and SMS/SMS (10%) is likely significant given the very large sample size, all 50,000 panelists. Predicting eligibility, significant differences also emerge between email/Web (76%) and both of the other modes, SMS/Web (82%; $z = 5.12$; $p < .001$) and SMS/SMS (81%, $z = 3.09$; $p < .01$). We then added module to these models. For neither screening ($X^2 = 1.1$) nor eligibility ($X^2 = 5.0$) does adding module to these models improve fit. Finally, we tested the interaction of mode and module in these models. For neither model does adding the interaction improve model fit.

4.2. *Modeling Respondent Burden: Completion Rates*

Next, we estimate logistic regression models to examine differences in respondent burden between components of the study design. Survey completion is used as a first (negative) indictor of respondent burden (Bradburn 1978; Sharp and Frankel 1983). First, we predict survey completion using mode as the sole predictor, then add module as a predictor, and finally their interaction. The first model suggests that completion rates differ significantly between the three modes ($X^2 = 77.9$, $p < .001$). Pairwise comparisons show significant differences between these three rates. Completion is highest in email/Web (8.3%) compared with SMS/Web ($z = 3.5$, $p < .001$) and SMS/SMS ($z = 9.4$, $p < .001$) and lowest in SMS/SMS (5.5%), compared with SMS/Web (7.4%; $z = 6.4$, $p < .001$) which has a completion rate that falls in between the other two modes.

Table 3. Samples sizes, screening, eligibility, breakoff, and completion rates, by condition

Mode		All Modules		Module A		Module B		Module C		Module D	
		#	rate	#	rate	#	rate	#	rate	#	rate
Email/Web	# Invited	20000		5000		5000		5000		5000	
	# Clicked URL (% of invited)	2788	13.9%	733	14.7%	622	12.4%	805	16.1%	628	12.6%
	# Screened (% of invited)	2272	11.4%	594	11.9%	548	11.0%	583	11.7%	547	10.9%
	(% of URL clicked)		81.5%		81.0%		88.1%		72.4%		87.1%
	# Eligible (% of screened)	1721	75.7%	436	73.4%	413	75.4%	455	78.0%	417	76.2%
	# Breakoff (% of eligible)	56	3.3%	19	4.4%	9	2.2%	6	1.3%	22	5.3%
	# Completed / (% of invited)	1665	8.3%	417	8.3%	404	8.1%	449	9.0%	395	7.9%
SMS/Web	# Invited	20000		5000		5000		5000		5000	
	# Clicked URL (% of invited)	2131	10.7%	551	11.0%	570	11.4%	488	9.8%	522	10.4%
	# Screened (% of invited)	1991	10.0%	510	10.2%	494	9.9%	478	9.6%	509	10.2%
	(% of URL clicked)		93.4%		92.6%		86.7%		98.0%		97.5%
	# Eligible (% of screened)	1635	82.1%	411	80.6%	413	83.6%	402	84.1%	409	80.4%
	# Breakoff (% of eligible)	162	9.9%	50	12.2%	32	7.7%	27	6.7%	53	13.0%
	# Completed / (% of invited)	1473	7.4%	361	7.2%	381	7.6%	375	7.5%	356	7.1%
SMS/SMS	# Invited	10000		2500		2500		2500		2500	
	# Screened (% of invited)	977	9.8%	224	9.0%	242	9.7%	242	9.7%	269	10.8%
	# Eligible (% of screened)	787	80.6%	186	83.0%	194	80.2%	197	81.4%	210	78.1%
	# Breakoff (% of eligible)	237	30.1%	63	33.9%	50	25.8%	45	22.8%	79	37.6%
	# Completed / (% of invited)	550	5.5%	123	4.9%	144	5.8%	152	6.1%	131	5.2%

Table 4. Predicted recruitment (screening and eligibility rates) and respondent burden (completion and breakoff rates and survey duration), modes and modules

		Overall		Adding mode			Adding module									Testing interaction
							Module A		Module B		Module C		Module D			
		%/*M*	SE	%/*M*	SE	LR X^2 / F z	%/*M*	SE	%/*M*	SE	%/*M*	SE	%/*M*	SE	LR X^2 / F z	LR X^2 / F
Recruitment																
Screened	Overall	0.10	0.00			$X^2(2)$= 27.57***									$X^2(3)$= 1.09	$X^2(6)$=8.55
	Email/Web (1)			0.11	0.00	$z_{1\text{-}2}$=4.55***	0.12	0.00	0.11	0.00	0.11	0.00	0.11	0.00		
	SMS/Web (2)			0.10	0.00	$z_{1\text{-}3}$=4.27***	0.10	0.00	0.10	0.00	0.10	0.00	0.10	0.00		
	SMS/SMS (3)			0.10	0.00	$z_{2\text{-}3}$=0.51	0.10	0.00	0.10	0.00	0.10	0.00	0.10	0.00		
Eligible	Overall	0.79	0.01			$X^2(2)$= 27.55***									$X^2(3)$= 4.99	$X^2(6)$=4.58
	Email/Web (1)			0.76	0.01	$z_{1\text{-}2}$=5.12***	0.74	0.01	0.76	0.01	0.78	0.01	0.75	0.01		
	SMS/Web (2)			0.82	0.01	$z_{1\text{-}3}$=3.09**	0.81	0.01	0.82	0.01	0.84	0.01	0.81	0.01		
	SMS/SMS (3)			0.81	0.01	$z_{2\text{-}3}$=1.02	0.79	0.02	0.81	0.02	0.82	0.01	0.80	0.02		
Respondent burden																
Complete	Overall	0.07	0.00			$X^2(2)$= 77.88***									$X^2(3)$= 5.63	$X^2(6)$=3.55
	Email/Web (1)			0.08	0.00	$z_{1\text{-}2}$=3.50***	0.08	0.00	0.08	0.00	0.09	0.00	0.08	0.00		
	SMS/Web (2)			0.07	0.00	$z_{1\text{-}3}$=9.35***	0.07	0.00	0.07	0.00	0.08	0.00	0.07	0.00		
	SMS/SMS (3)			0.06	0.00	$z_{2\text{-}3}$=6.36***	0.05	0.00	0.06	0.00	0.06	0.00	0.05	0.00		
Breakoff	Overall	0.11	0.00			$X^2(2)$=344.80***									$X^2(3)$=37.22***	$X^2(6)$=1.34
	Email/Web (1)			0.03	0.00	$z_{1\text{-}2}$=7.46***	0.04	0.01	0.03	0.00	0.02	0.00	0.05	0.01	$z_{A\text{-}B}$=3.19***	
	SMS/Web (2)			0.10	0.01	$z_{1\text{-}3}$=15.72***	0.12	0.01	0.08	0.01	0.07	0.01	0.13	0.01	$z_{A\text{-}C}$=4.26***	
	SMS/SMS (3)			0.30	0.02	$z_{2\text{-}3}$=11.26***	0.35	0.03	0.25	0.02	0.22	0.02	0.38	0.03	$z_{B\text{-}D}$=4.13***	
															$z_{C\text{-}D}$=5.22***	
Survey duration	Overall (Web)	137	1.54			$F_{(1,3177)}$=0.11	141	3.03	116	3.07	113	2.99	186	3.07	$F_{(3,3175)}$=121.0***	$F_{(3,3171)}$=0.88
	Email/Web (1)			138	2.07										$z_{A\text{-}B}$=5.72***	
	SMS/Web (2)			137	2.19										$z_{A\text{-}C}$=6.73***	
															$z_{A\text{-}D}$=10.30***	
															$z_{B\text{-}D}$=15.93***	
															$z_{C\text{-}D}$=17.05***	
	SMS/SMS (3)	1845	191				1835	430	1975	366	1684	391	1884	390	$F_{(3,482)}$=0.2	

Note: *M* is used to represent the median. ***p<.001; **p<.01

In the second model, we add planned missing modules as a predictor. Accounting for module does not significantly improve the fit of the model ($X^2 = 5.6$). A subsequent test including the interaction of mode and module also unsurprisingly fails to significantly improve the fit of the model ($X^2 = 3.55$).

4.2.1. Modeling Respondent Burden: Breakoff Rates

Survey breakoff is used as a second indictor of respondent burden (Peytchev 2009; Steinbrecher et al. 2015). We estimate logistic regression models to further examine differences in breakoff rates between components of the study design, first including only mode as a predictor, then adding module, and finally testing their interaction. The first model suggests that mode is a significant predictor of breakoffs ($X^2 = 344$, $p < .001$). Given the significant X^2 statistic, we compared modes pairwise, finding significant differences between all three modes, with email/Web (3.3%) prompting a lower breakoff rate than SMS/Web (9.9%; $z = 7.5$; $p < .001$) and SMS/SMS (30%) resulting in a higher breakoff rate than email/Web ($z = 15.7$; $p < .001$) and SMS/Web ($z = 11.3$; $p < .001$).

In the second model, we add planned missing modules as a predictor. Adding module significantly improves the fit of the model ($X^2 = 37.2$; $p < .001$). Given the significant X^2 statistic, we compared modules pairwise. Full length module D increases breakoffs compared to modules B ($z = 4.1$; $p < .001$) and C ($z = 5.2$; $p < .001$). The longest of the shortened modules, module A, also shows a higher breakoff rate than modules B ($z = 3.2$; $p < .001$) and C ($z = 4.3$; $p < .001$). Notably, modules A and D and modules B and C do not differ in their breakoff rates. Finally, the interaction of mode and module was tested but was not statistically significant ($X^2 = 1.3$).

4.2.2. Modeling Respondent Burden: Survey Duration

Survey duration is used here as the final measure of respondent burden (Antoun and Cernat 2020; Mavletova and Couper 2015). We examined differences in survey duration between modes and modules by estimating quantile (median) regression models. These models predicted survey duration in seconds, first introducing mode as a predictor, then adding module to the model. Models are estimated separately for Web (email/Web and SMS/Web) and SMS surveys given our expectation of a large difference in survey duration between data collection modes (Schober et al. 2015).

Median duration did not differ between the two Web surveys ($F_{1,3177} = 0.1$): 138 seconds in email/Web and 137 seconds in SMS/Web. As no significant difference emerged between these two modes, we pool the Email/Web and SMS/Web data to test differences across modules. Median survey duration differed between the four modules ($F_{3,3175} = 121.0$, $p < .001$). Given the significant F statistic, we compared modules pairwise. Median duration for the pooled web-based surveys differs between each pair of modules except for modules B and C for which duration fails to differ. Time to complete Module D exceeds all other modules: A ($z = 10.3$, $p < .001$), B ($z = 15.9$, $p < .001$) and C ($z = 17.1$, $p < .001$) and time to complete module A exceeds modules B ($z = 5.7$, $p < .001$) and C ($z = 6.7$, $p < .001$). Finally, the interaction of mode and module was tested but was not statistically significant ($F_{1,3171} = 0.9$).

In the SMS/SMS survey, median duration was over 30 minutes (1,845 seconds), ranging from 1,668 seconds for Module C to 1975 for Module B. However, survey duration did not significantly differ between modules in the SMS/SMS condition ($F_{3,482} = 0.2$).

4.3. *Demographic and Self-Reported Health Covariates*

A number of significant differences emerge in the distributions of the four demographic variables and two self-reported health variables between the three modes (see Table 2). Respondents answering the survey in the SMS/SMS condition report relatively poorer health and mental health. SMS respondents are also more likely female and are relatively younger than their counterparts completing Web surveys. Those invited by email to complete a Web survey, however, are relatively more educated than those invited by SMS to take either a Web or SMS survey and report better mental health. Relatively few differences emerge across modules within modes. Differences emerge in the SMS/Web survey on education, and in the SMS/SMS survey on age and education. However, no clear and consistent patterns emerge from these comparisons across modules.

4.4. *Predicting Patient Experience Outcomes*

The first set of regression models provide a baseline restricted model predicting the five outcomes controlling only for demographic and health covariates. The adjusted means from these models for the five outcomes range from 7.0 for provider accessibility to 8.1 for effective communication and the overall provider rating (see Table 5).

The next set of regression models include mode as an independent variable. *F*-tests comparing each model to the corresponding previously estimated baseline model show statistically significant effects for mode in models for three of the five outcomes, effective communication ($F_{2,1794} = 10.5$, $p < .001$), office staff rating ($F_{(2,3640)} = 13.2$, $p < .001$), and provider rating ($F_{(2,3635)} = 7.8$, $p < .001$), and was marginally significant in a fourth model for provider accessibility ($F_{(2,1712)} = 2.6$, $p < .10$).

For each model with a significant *F*-statistic, we compare adjusted means for these outcomes by mode. Email/Web responses differ from SMS/SMS for four outcomes, provider accessibility (6.8 and 7.3; $t = 2.14$, $p < .05$), effective communication (7.8 and 8.5; $t = 3.96$, $p < .001$), office staff rating (7.7 and 8.3; $t = 5.03$, $p < .001$), and provider rating (8.0 and 8.4; $t = 3.50$, $p < .001$), in each case the SMS survey yielding the higher value. Email/Web responses differ from SMS/Web for three outcomes, effective communication (7.8 and 8.3; $t = 3.64$, $p < .001$), office staff rating (7.7 and 7.9; $t = 2.79$, $p < .01$), and provider rating (8.0 and 8.2; $t = 2.81$, $p < .01$), in each case the SMS/Web condition yielding the higher value. For only one outcome, the office staff rating, do SMS/Web (7.9) and SMS/SMS (8.3) responses differ ($t = 3.12$, $p < .01$).

The next set of models include module as an independent variable. *F*-tests comparing each model to the corresponding previously estimated model show statistically significant effects for module for just one of the five outcomes, care coordination ($F_{2,1794} = 10.5$, $p < .001$), with module C yielding lower values of the outcome than module D.

The final set of models include the interaction of mode and module. *F*-tests comparing each model to the corresponding previously estimated main-effect only models show no statistically significant effects for the interaction terms.

Table 5. Adjusted means for outcomes (CAHPS composites), by module and mode types

		Demographic controls			Adding mode				Adding module									Adding interaction
									Module A		Module B		Module C		Module D			
Outcomes		mean	SE	N	mean	SE	N	F (mode) t (mode diff.)	mean	SE	mean	SE	mean	SE	mean	SE	F (module)	F(mode*module)
Provider accessibility	Overall	6.96	0.06	1730				$F_{(2,1712)}=2.6^{\dagger}$	6.95	0.09					6.98	0.09	$F_{(1,1711)}=0.1$	$F_{(2,1709)}=1.2$
	Email/Web (1)				6.83	0.10	806	$t_{1\text{-}3}=2.14^{*}$	6.81	0.11					6.84	0.12		
	SMS/Web (2)				7.03	0.10	710		7.01	0.12					7.04	0.12		
	SMS/SMS (3)				7.28	0.19	214		7.26	0.20					7.29	0.20		
Effective communication	Overall	8.09	0.05	1812				$F_{(2,1794)}=10.5^{***}$			8.04	0.07			8.14	0.07	$F_{(1,1793)}=0.9$	$F_{(2,1791)}=0.4$
	Email/Web (1)				7.83	0.08	806	$t_{1\text{-}3}=3.96^{***}$			7.78	0.09			7.88	0.09		
	SMS/Web (2)				8.25	0.08	751	$t_{1\text{-}2}=3.64^{***}$			8.20	0.10			8.29	0.10		
	SMS/SMS (3)				8.48	0.14	255				8.43	0.15			8.52	0.15		
Care coordination	Overall	7.30	0.06	1846				$F_{(2,1828)}=0.7$					7.12	0.08	7.49	0.09	$F_{(1,1827)}=9.2^{**}$	$F_{(2,1825)}=2.0$
	Email/Web (1)				7.24	0.09	851						7.06	0.11	7.42	0.11		
	SMS/Web (2)				7.33	0.09	747						7.16	0.11	7.52	0.11		
	SMS/SMS (3)				7.44	0.16	248						7.26	0.18	7.62	0.18		
Provider rating	Overall	8.12	0.03	3653				$F_{(2,3635)}=7.8^{***}$	8.07	0.06	8.21	0.06	8.05	0.06	8.17	0.06	$F_{(3,3632)}=1.4$	$F_{(6,3626)}=1.0$
	Email/Web (1)				8.00	0.05	1673	$t_{1\text{-}3}=3.50^{***}$	7.95	0.07	8.08	0.07	7.93	0.07	8.05	0.07		
	SMS/Web (2)				8.20	0.05	1494	$t_{1\text{-}2}=2.81^{**}$	8.14	0.07	8.28	0.07	8.12	0.07	8.24	0.07		
	SMS/SMS (3)				8.35	0.09	486		8.30	0.10	8.43	0.10	8.28	0.10	8.40	0.10		
Office staff rating	Overall	7.88	0.04	3658				$F_{(2,3640)}=13.2^{***}$	7.87	0.08	7.86	0.08	7.93	0.08	7.88	0.08	$F_{(3,3637)}=0.2$	$F_{(6,3631)}=1.1$
	Email/Web (1)				7.70	0.06	1675	$t_{1\text{-}3}=5.03^{***}$	7.69	0.09	7.68	0.09	7.75	0.09	7.70	0.09		
	SMS/Web (2)				7.94	0.06	1497	$t_{1\text{-}2}=2.79^{**}$	7.93	0.09	7.92	0.09	7.99	0.09	7.93	0.09		
	SMS/SMS (3)				8.33	0.11	486	$t_{2\text{-}3}=3.12^{**}$	8.31	0.13	8.30	0.13	8.37	0.13	8.32	0.13		

Note: $^{***}p<.001$; $^{**}p<.01$; $^{*}p<.05$; $^{\dagger}p<.10$;

5. Discussion

5.1. Respondent Burden and Planned Missing Modules

That shorter surveys reduce respondent burden (Johnson et al. 2013; Toepoel and Lugtig 2018; West et al. 2015) motivated our novel test of a planned missing design in Web and SMS patient surveys. We operationalized respondent burden as (longer) survey duration, (lower) completion rates, and (higher) breakoff rates in line with prior theory and research (Antoun and Cernat 2020; Bradburn 1978; Mavletova and Couper 2015; Peytchev 2009; Sharp and Frankel 1983; Steinbrecher et al. 2015). Taken together, findings suggest that the planned missing design did reduce respondent burden, especially in both Web (email/Web and SMS/Web) surveys. Module D, the full-length version of the web survey questionnaire, took significantly longer for Web respondents to answer compared with the three shortened questionnaire modules. Of the planned missing modules, Module A asked more questions than did the others (Modules B and C) and took respondents longer to answer than the other shortened versions. In the SMS/SMS mode, however, no significant differences in survey duration emerged between modules. This is likely due to two important factors: (1) reduced statistical power given the smaller number of completed SMS surveys, and (2) the relative difference between the very brief amount of time needed to answer questions and the longer spans of time between active question answering. Finally, although completion rates did not significantly vary between modules, Modules B and C did yield significantly lower breakoff rates compared to Modules A and D.

5.2. Respondent Burden and Invitation and Data Collection Modes

The three invitation and data collection modes clearly varied by respondent burden as operationalized in the current study. SMS as a data collection mode appeared to increase the burden placed on respondents relative to the two Web surveys. Panelists assigned to SMS were less likely to complete the survey and those who do start an SMS survey are more likely to breakoff, at a rate in line with prior research (Lau et al. 2019). Moreover, those SMS panelists who persist through the survey spend an order of magnitude more time completing it compared to the two groups of Web survey respondents regardless of how they were invited (i.e., via email or SMS). In comparison with Web surveys, SMS surveys may also present respondent burden that differs in kind as well as degree. In the current study, Web surveys are completed by most respondents in a single session, answering the first question and the last within a few minutes of each other. Conversely, SMS respondents stretch out the question-answering process, distributing the burden of responding over a longer period of time (Schober et al. 2015; West et al. 2005) and increasing the potential for breakoffs (Lee et al. 2013).

Respondent burden in the SMS/Web condition fell in between the SMS survey with the SMS invitation and the other Web survey using an email invitation. This finding follows extant research demonstrating that SMS invitations can lead to higher respondent burden, including higher breakoffs and reduced completion rates (Mavletova and Couper 2014).

5.3. Effect of Respondent Burden on Outcomes

Although the planned missing design appears to reduce respondent burden, few differences emerge in patient experience outcomes. This finding suggests that partitioning the survey instrument as part of the planned missing design did not change how respondents answered questions. For only one outcome, coordination of care, did including module in the model significantly improve model fit. Respondents in full-length module D report better experiences with care coordination than respondents in shortened module C, the only shortened module with this question included. This single difference between modules is likely a context effect (Swain 2015). In module D, the full survey, care coordination questions were interleaved with questions about the effectiveness of the provider's communication and the overall provider rating (see Table 1). Questions about the provider explaining things well, listening carefully, showing respect, and spending adequate time may encourage module D respondents to feel more positively about their provider. Thus, respondents answering those questions may give higher ratings of care coordination (knowing the patient's medical history, providing test results) than module C respondents for whom care coordination questions are the first set answered after the initial screening items. In sum, while the longer survey modules arguably increased respondent burden, illustrated in longer survey duration and higher breakoff rates, respondent burden seemingly did not alter measured outcomes. This finding provides some initial evidence supporting the use of the planned missing design with CG-CAHPS.

5.4. *Effect of Invitation and Data Collection Mode on Outcomes*

Invitation and data collection modes also influenced how respondents answered in four of the five outcomes. Email/Web respondents reported poorer experiences with their provider than did respondents invited by SMS. For only one outcome did SMS/Web respondents differ from SMS/SMS respondents, the former reporting poorer experiences with their provider's office staff than the latter. While the source of these by-mode differences in outcomes is uncertain, the pattern of findings suggests a potential cause. The consistent differences in four of five outcomes between email/Web and SMS/SMS, the consistent lack of differences in four of five outcomes between SMS/Web and SMS/SMS, and that outcomes from SMS/Web consistently fell in between those from the other two mode protocols, suggest a mode of invitation effect generating these differences. Panelists who respond to an invitation by SMS, or are even willing to receive an invitation by SMS, clearly differ from panelists who receive and respond to email invitations.

5.5. *Limitations and Future Directions*

In summary, the planned missing design was successful for the Web surveys, reducing respondent burden – shorter survey duration and fewer breakoffs – without changing patient experience estimates. While we are optimistic about the use of a planned missing design with CG-CAHPS surveys in the future, further study is needed, specifically with a sample of patients from a provider or healthcare organization. We used a sample from an online non-probability panel as a first test. It is possible that online panelists may differ from a sample of patients from a provider list in key ways that could alter response patterns and patient experience estimates. Given their greater experience with answering surveys, panelists may be more adept at survey response than the average patient (Toepoel et al. 2008). Thus, the applicability of these findings may be limited if panelists answer these questions in ways that differ systematically from a sample from a clinic, hospital, or physician patient list; a concern that is not unique to this study (Hillygus et al. 2014). However, given that we screened panelists for a physician visit in the last six months, and this is not a rare population, these findings are arguably likely to be similar to what would be found using a patient sample. Future research should validate these findings with a sample generated from a list of patients with a recent visit to a healthcare provider.

The planned missing design yielded relatively more mixed results in the SMS survey, reducing breakoff rates but failing to shorten survey duration. This finding suggests that respondent burden may be linked to qualities of the planned missing design other than time to completion, such as the sheer number of questions asked. Respondents may fear that the SMS survey in particular is never-ending as every answer sent leads to a new question received Indeed, SMS achieved the lowest completion rate and highest breakoff rate in comparison with both Web survey modes. While SMS also yielded far and away the longest survey duration, our measure of duration lacks detail, an important limitation to be addressed in future research. Ideally, researchers would record the time that a question is sent, the message is opened, and the answer is received. With these additional details, future research may better understand how respondents interact with SMS surveys and whether and how the number of questions influences respondent burden.

Brief indicators of progress (e.g., "question 10 of 15" or "66% complete") sent by text in combination with or in addition to the survey questions may help motivate SMS survey respondents to continue with the survey and resist the desire to breakoff. However, these progress indictors may counterproductively extend the length of the text message. Moreover, progress indicators may reduce rather than increase response rates in Web surveys (Villar et al. 2013). Therefore, future research should carefully evaluate their use in SMS surveys.

6. Conclusion

This article contributes to the literature on respondent burden by testing a planned missing design – randomly assigning respondents to answer only a subset of questions to shorten the survey and reduce respondent burden (Johnson et al. 2013) – in Web and SMS surveys. To our knowledge, this is the first test of this approach to reduce respondent burden in an SMS survey.

Three shortened versions of the CG-CAHPS patient experience survey applying a planned missing design reduced the number of questions by about 40% in comparison to the full survey module. We invited members of an online panel using one of two invitation modes, email or SMS, to complete a survey in one of two data collection modes, Web or SMS surveys. We found that a planned missing design reduced respondent burden for Web survey respondents, yielding fewer breakoffs and shorter survey duration. Most importantly, we reduced respondent burden without significantly altering the way that respondents answered most of the questions about patient experience. While findings should be investigated further using samples from clinic patient lists, the findings from the Web survey are encouraging: a planned missing design can make our respondents' jobs easier while keeping survey statistics consistent.

The findings for the SMS survey were less clear. Although the shortened versions of the SMS surveys tended to reduce the breakoff rate relative to the full-length control version, they did not yield a reduction in the time it took respondents to complete the survey.

Arguably, both the full 25-item CG-CAHPS survey and the shortened 15-item surveys may simply be too long to expect respondents to complete by text message. Moreover, the piecemeal way questions are sent after previous questions are answered may prevent the planned missing design from perceptibly reducing the burden of the task. Thus, the nature of the mode combined with the number of questions may simply exhaust even those respondents initially willing to complete the survey, as reflected in a lower completion rate among eligible respondents and a breakoff rate that is 3-10 times higher than that in the Web surveys. The solution may be to further reduce the number of questions asked in an SMS survey, perhaps to around five or so, and make more explicit ways for respondents skip questions they do not wish to answer (Lau et al. 2019). Moreover, sample members who are capable of and willing to complete an SMS survey may be very different from those without the ability and willingness to respond via SMS. While this may change over time, this hurdle to completing SMS surveys with general populations may present a significant barrier to moving surveys into this mode for data collection.

We also found evidence for a mode of invitation effect. Respondents able and willing to even receive an SMS invitation may be substantially different from those without the capability and willingness. Those respondents invited by SMS to participate in Web survey responded to the survey in a way that split the difference between those who were invited by email to complete a Web survey and those invited by SMS to an SMS survey. Future research should further investigate how using SMS as an invitation mode may alter who responds.

In conclusion, the planned missing design can clearly be applied successfully in Web surveys. Our original full-length survey was already relatively brief, about 25 questions. Thus, other survey researchers may find additional reductions in response burden yielding even larger increases in completion or response rates and reductions in breakoffs and survey duration. However, work remains to determine if and how a planned missing design can be successfully adapted to reduce respondent burden in SMS surveys. The next step in applying a planned missing design in SMS may be to further reduce respondent burden by further shortening the survey, or in other words, planning for more missing.

7. References

Ahmad, F., A.K. Jhajj, D.E. Stewart, M. Burghardt, and A.S. Bierman. 2014. "Single Item Measures of Self-Rated Mental Health: A Scoping Review." *BMC Health Services Research* 14: 398. DOI: https://doi.org/10.1186/1472-6963-14-398.

AHRQ, Agency for Healthcare Research and Quality. n.d. "How to Report Results of the CAHPS Clinician & Group Survey." Available at: https://www.ahrq.gov/sites/default/-files/wysiwyg/cahps/surveys-guidance/cg/cgkit/HowtoReportResultsofCGC-AHPS080610FINAL.pdf (accessed September 2021).

Alemi, F., and H.M. Jasper. 2014. "An Alternative to Satisfaction Surveys: Let the Patients Talk." *Quality Management in Health Care* 23(1): 10–19. DOI: https://doi.org/10.1097/QMH.0000000000000014.

Antoun, C., and A. Cernat. 2020. "Factors Affecting Completion Times: A Comparative Analysis of Smartphone and PC Web Surveys." *Social Science Computer Review* 38(4): 477–489. DOI: https://doi.org/10.1177/0894439318823703.

Ayers, R., S. Brothers, and W. Jansen. 2014. "Guidelines on Mobile Device Forensics." National Institute of Standards and Technology, Special Publication 800-101r1. Washington, DC: US Department of Commerce. DOI: http://dx.doi.org/10.6028/NIST.Sp.800-101r1.

Blumberg, S.J., and J.V. Luke. 2020. "Wireless substitution: Early release of estimates from the National Health Interview Survey, July-December 2019." National Center for Health Statistics. Available at: https://www.cdc.gov/nchs/nhis.htm. (accessed September 2021).

Bosnjak, M., W. Neibarth, M.P. Couper, W. Bandilla, and L. Kaczmirek. 2008. "Prenotification in Web-Based Access Panel Surveys." *Social Science Computer Review* 26(2): 213–223. DOI: https://doi.org/10.1177/0894439307305895.

Bradburn, N. 1978. "Respondent Burden." In Proceedings of the Survey Research Methods Section of the American Statistical Association: 35–40. Alexandria, VA: American Statistical Association. Available at: http://www.asasrms.org/Proceedings/y1978f.html (accessed September 2021).

Brenner, P.S., and J.D. DeLamater. 2013. "Paradata Correlates of Data Quality in an SMS Time Use Study–Evidence from a Validation Study." *Electronic International Journal of Time Use Research* 10(1): 38–54. DOI: dx.doi.org/10.13085/eIJTUR.10.1.38-54.

Conrad, FG., M.P. Couper, R. Tourangeau, and C. Zhang. 2017. "Reducing Speeding in Web Surveys by Providing Immediate Feedback." *Survey Research Methods* 11(1): 45–61. DOI: https://doi.org/10.18148/srm/2017.v11i1.6304.

Crawford, S.D., M.P. Couper, and M.J. Lamias. 2001. "Web Surveys: Perception of Burden." *Social Science Computer Review* 19(2): 146–162. DOI: https://doi.org/10.1177/089443930101900202.

De Leeuw, E.D. 2018. "Mixed-Mode: Past, Present, and Future." *Survey Research Methods* 12(2): 75–89. DOI: doi:10.18148/srm/2018.v12i2.7402.

DeSalvo, K.B., V.S. Fan, M.B. McDonell, and S.D. Fihn. 2004. "Predicting Mortality and Healthcare Utilization with a Single Question." *Health Services Research* 40(4): 1234–1246. DOI: https://doi.org/10.1111/j.1475-6773.2005.00404.x.

Dillman, D.A., J.D. Smyth, and L.M. Christian. 2014. *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method.* New York: Wiley.

Dyer, N., J.S. Sorra, S.A. Smith, P. Cleary, and R. Hays. 2012. "Psychometric Properties of the Consumer Assessment of Healthcare Providers and Systems (CAHPS) Clinician and Group Adult Visit Survey." *Medical Care* 50: S28. DOI: https://doi.org/10.1097/MLR.0b013e31826cbc0d.

Elliott M.N., M.K. Beckett, W.G. Lehrman, P. Cleary, C.W. Cohea, L.A. Giordano, E.H. Goldstein, and C.L. Damberg. 2016. "Understanding The Role Played By Medicare's Patient Experience Points System In Hospital Reimbursement." *Health Affairs* 35(9): 1673–1680. DOI: https://doi.org/10.1377/hlthaff.2015.0691.

Elliott, M.N., A.M. Zaslavsky, E. Goldstein, W. Lehrman, K. Hambarsoomians, M.K. Beckett, and L. Giordano. 2009. Effects of survey mode, patient mix, and nonresponse on CAHPS® hospital survey scores. *Health Services Research*, 44 (2p1): 501–518. DOI: https://doi.org/10.1111/j.1475-6773.2008.00914.x.

Enders, C.K. 2010. *Applied Missing Data Analysis*. New York: Guilford.

Galesic, M., and M. Bosnjak. 2009. "Effects of Questionnaire Length on Participation and Indicators of Response Quality in a Web Survey." *Public Opinion Quarterly* 73(2): 349–360. DOI: https://doi.org/10.1093/poq/nfp031.

Garrido, T., B. Raymond, and B. Wheatley. 2016. "Lessons From More than a Decade In patient Portals." *Health Affairs Blog*. DOI: https://doi.org/10.1377/hblog20160407.054362.

Graham, J.W., B.J. Taylor, A.E. Olchowski, and P.E. Cumsille. 2006. "Planned Missing Data Designs in Psychological Research." *Psychological Methods* 11(4): 323–343. DOI: https://doi.org/10.1037/1082-989X.11.4.323.

Hillygus, D.S., N. Jackson, and M. Young. 2014. "Professional respondents in non-probability online panels." In *Online panel research: A data quality perspective,* edited by M. Callegaro, R. Baker, J. Bethlehem, A.S. Góritz, J.A. Krosnick and P.J. Lavrakas: 219–237. New York: Wiley.

Holt, J.M. 2019. "Patient Experience in Primary Care: A Systematic Review of CG-CAHPS Surveys." *Journal of Patient Experience* 6(2): 93–102. DOI: https://doi.org/10.1177/2374373518793143.

Johnson, D.R., V. Roth, and R. Young. 2013. "Planned Missing Data Designs in Health Surveys." In Proceedings of the Tenth Conference on Health Survey Research Methods, November 2013: 43–53. Hyattsville, MD: Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics. Available at: https://www.cdc.gov/nchs/data/hsrmc/hsrmc_10th_proceedings_2011.pdf (accessed February 2020).

Keller, S., A.J. O'Malley, R.D. Hays, R.A. Matthew, A.M. Zaslavsky, K.A. Hepner, and P.D. Cleary. 2005. "Methods Used to Streamline the CAHPS® Hospital Survey." *Health Services Research*, 40: 2057–2077. DOI: https://doi.org/10.1111/j.1475-6773.2005.00478.x.

Kim, Y., J. Dykema, J. Stevenson, P. Black, and D.P. Moberg. 2019. "Straightlining: Overview of Measurement, Comparison of Indicators, and Effects in Mail–Web Mixed-Mode Surveys." *Social Science Computer Review* 37(2): 214–233. DOI: https://doi.org/10.1177/0894439317752406.

Lau, C.Q., H. Sanders, and A. Lombaard. 2019. "Questionnaire Design in Short Message Service (SMS) Surveys." *Field Methods* 31(3): 214–229. DOI: https://doi.org/10.1177/1525822X19862506.

Lee, S.S.S., X. Xin, W.P. Lee, E.J. Sim, B. Tan, M.P.G. Bien, A.S.T. Lau, and J. Thumboo. 2013. "The Feasibility of Using SMS as a Health Survey Tool: An Exploratory Study in Patients with Rheumatoid Arthritis." *International Journal of Medical Informatics* 82(5): 427–434. DOI: https://doi.org/10.1016/j.ijmedinf.2012.12.003.

Mavletova, A. and M.P. Couper. 2014. "Mobile Web Survey Design: Scrolling versus Paging, SMS versus E-mail Invitations." *Journal of Survey Statistics and Methodology* 2: 498–518. DOI: https://doi.org/10.1093/jssam/smu015.

Mavletova, A., and M.P. Couper. 2015. "A Meta-Analysis of Breakoff Rates in Mobile Web Surveys." In *Mobile Research Methods: Opportunities and Challenges of Mobile Research Methodologies*, edited by D. Toninelli, R. Pinter, and P. de Pedraza: 81–98. London: Ubiquity Press.

Mavletova, A., M.P. Couper, and D. Lebedev. 2018. "Grid and Item-by-Item Formats in PC and Mobile Web Surveys." *Social Science Computer Review* 36(6):647-668. DOI: https://doi.org/10.1177/0894439317735307.

Morgan, S.L., and E.S. Taylor Poppe. 2015. "A Design and a Model For Investigating the Heterogeneity of Context Effects in Public Opinion Surveys." *Sociological Methodology* 45(1): 184–222. DOI: https://doi.org/10.1177/0081175015570095.

Ng, J.H., E. Henry, T. Oberlander, P. Shi, S.H. Scholle. 2016. "Shortening a Patient Experiences Survey for Medical Homes." *Healthcare* 4:1. DOI: https://doi.org/10.3390/healthcare4010001.

Paddison, C.A.M., G.A. Abel, M.O. Roland, M.N. Elliott, G. Lyratzopoulos, and J.L. Campbell. 2013. "Drivers of Overall Satisfaction with Primary Care: Evidence from the English General Practice Patient Survey." *Health Expectations* 18: 1081-1092. DOI: https://doi.org/10.1111/hex.12081.

Pew Research Center. 2019. *Mobile Technology and Home Broadband.* Available at: https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2019/06/PI_2019.06.13_Mobile-Technology-and-Home-Broadband_FINAL2.pdf. (accessed September 2021).

Peytchev, A. 2009. "Survey Breakoff." *Public Opinion Quarterly* 73(1): 74–97. DOI: https://doi.org/10.1093/poq/nfp014.

Peytchev, A. and E. Peytcheva. 2017. "Reduction of Measurement Error due to Survey Length: Evaluation of the Split Questionnaire Design Approach." *Survey Research Methods* 11(4): 361–368. DOI: https://doi.org/10.18148/srm/2017.v11i4.7145.

Revilla, M. and C. Ochoa. 2017. "Ideal and Maximum Length for a Web Survey." *International Journal of Market Research* 59(5): 557–565. DOI: https://doi.org/10.2501/IJMR-2017-039.

Rhemtulla, M., V. Savalei, and T.D. Little. 2016. "On the Asymptotic Relative Efficiency of Planned Missingness Designs." *Psychometrika* 81(1): 60–89. DOI: https://doi.org/10.1007/s11336-014-9422-0.

Schober, M.F., F.G. Conrad, C. Antoun, P. Ehlen, S. Fail, A.L. Hupp, M. Johnston, L. Vickers, H.Y. Yan, and C. Zhang. 2015. "Precision and Disclosure in Text and Voice Interviews on Smartphones." *PLoSONE* 10. DOI: https://doi.org/10.1371/journal.pone.0128337.

Schonlau, M. and M. Couper. 2017. "Options for Conducting Web Surveys." *Statistical Science* 32(2): 279–292. DOI: https://doi.org/10.1214/16-STS597.

Sharp, L.M., and J. Frankel. 1983. "Respondent Burden: A Test of Some Common Assumptions." *Public Opinion Quarterly* 47: 36–53. DOI: https://doi.org/10.1086/268765.

Spranca, M., D.E. Kanouse, M. Elliott, P.F. Short, D.O. Farley, and R.D. Hays. 2000. "Do Consumer Reports of Health Plan Quality Affect Health Plan Selection?" *Health Services Review* 35(5): 933–947. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1089177. (accessed September 2021).

Steinbrecher, M, J. Roßmann, and J.E. Blumenstiel. 2015. "Why Do Respondents Break Off Web Surveys and Does it Matter? Results from Four Follow-up Surveys." *International Journal of Public Opinion Research* 27(2): 289–302. DOI: https://doi.org/10.1093/ijpor/edu025.

Stucky, B.D., R.D. Hays, M.O. Edelen, J. Gurvey, and J.A. Brown. 2016. "Possibilities for Shortening the CAHPS Clinician and Group Survey." *Medical Care* 54(1): 32–37. DOI: https://doi.org/10.1097/MLR.0000000000000452.

Swain, M.S. 2015. "The Effects of a Planned Missingness Design on Examinee Motivation and Psychometric Quality." PhD Dissertation, James Madison University. Available at: https://www.proquest.com/docview/1685390254 (accessed September 2021).

Toepoel, V., and P. Lugtig. 2018. "Modularization in an Era of Mobile Web: Investigating the Effects of Cutting a Survey Into Smaller Pieces on Data Quality." *Social Science Computer Review*: 1–15. DOI: https://doi.org/10.1177/0894439318784882.

Toepoel, V., M. Das, and A. van Soest. 2008. "Effects of Design in Web Surveys: Comparing Trained and Fresh Respondents." *Public Opinion Quarterly* 72(5): 985–1007. DOI: https://doi.org/10.1093/poq/nfn060.

Villar, A., M. Callegaro, and Y. Yang. 2013. "Where am I? A meta-analysis of experiments on the effects of progress indicators for web surveys." *Social Science Computer Review* 31(6): 744–762. DOI: https://doi.org/10.1177/0894439313497468.

West, B.T., D. Ghimire, and W.G. Axinn. 2015. "Evaluating a Modular Design Approach to Collecting Survey Data Using Text Messages." *Survey Research Methods* 9:111–123. DOI: https://doi.org/10.18148/srm/2015.v9i2.6135.

Yan, T., S. Fricker, and S. Tsai. 2020. "Response Burden: What Is It and What Predicts It?" In Advances in Questionnaire Design, Development, Evaluation, and Testing, edited by P.C. Beatty, D. Collins, L. Kaye, J.L. Padilla, G. B. Willis, and A. Wilmot: 193–212. New York: Wiley.

Zhang, C., and F.G. Conrad. 2018. "Intervening to Reduce Satisficing Behaviors in Web Surveys: Evidence From Two Experiments on How it Works." *Social Science Computer Review* 36(1): 57–81. DOI: https://doi.org/10.1177/0894439316683923

Received May 2020
Revised July 2021
Accepted December 2021