Hello…Good-bye.: Hang-ups and Breakoffs by Mode, Topic, and Geography in Oregon

Debbie Krug Mangipudi[1], Matt Jans[2], Robynne Locke[1],
Stephen Haas[2], John Boyle[2], Randy ZuWallack[1], Lizzie Remrey[2], Heather
Driscoll[1], Siobhan McAlister[3], Samantha McCoy[3], Kelly Officer[3],
Ken Sanchagrin[3], Michael Weinerman[3], Zoe Padgett[2], Melinda Scott[2]

[1] ICF, 126 College St Suite 2, Burlington, VT 05401
[2] ICF, 530 Gaither Road, Rockville, MD 20850
[3] Oregon Criminal Justice Commission, 885 Summer St NE, Salem, OR 97301

Abstract
Survey topic and mode can influence response propensity, even if the topic only becomes truly salient to respondents once questions are read. We usually design interviews and questionnaires to begin with topic-relevant, simple, nonsensitive questions. However, perceived sensitivity can vary widely by respondent experience and background, producing disproportionate unit and item nonresponse. Breakoffs and hang-ups are interesting types of nonresponse because the act of not answering a specific question (i.e., item nonresponse) results in item nonresponse in the resulting data if the respondent completes the interview or questionnaire, but results in unit nonresponse (i.e., breakoffs and hang-ups as a final disposition) if they do not. Data came from the Oregon Crime Victimization Survey (OCVS), which used dual-frame random-digit-dial (RDD) and address-based samples (ABS). In both samples, adults living in Oregon for at least the past 12 months were eligible. The questionnaire was identical in both modes and included the following sections: 1) eligibility screening, 2) consent, 3) non-crime (i.e., quality of life), 4) core demographics, 5) index crimes, 6) non-index crimes, 7) crime follow-up, and 8) additional demographics. Focusing on breakoffs and hang-ups that were final dispositions, this paper addresses the following questions: 1) Do breakoffs differ between phone and web modes, 2) do specific topics, questions, or screens lead to more hang-ups and breakoffs than others, and 3) do breakoffs vary by respondent geographic and demographic characteristics of where they live? To address these research questions, we compared the percentage that hung-up or broke off by instrument section, and geographic stratum. Exploratory results show that breakoffs and hang-ups among respondents who began the instrument were three times higher in web than phone. However, once reaching consent screens/scripts, the web questionnaire had about half of the breakoff rate per question as the phone interview. Further, phone respondents may be more prone to hang-up at consent than web respondents, but web respondents may be more prone to breakoff at crime follow-up questions than phone respondents. Finally, overall breakoffs and hang-ups varied significantly across region, underscoring the importance of social context in nonresponse. Results are interpreted in the context of questionnaire design decisions, question sensitivity, mode differences, and overall nonresponse rates.

Key Words: nonresponse, paradata, breakoffs, hang-ups, web survey, phone survey

1. Introduction

Nonresponse is a large problem in survey research, with breakoffs being a relatively understudied type of nonresponse. Breakoffs (often called hang-ups in phone surveys) are situations in which a respondent terminates participation before reaching the end of the questionnaire or interview. They lead to item nonresponse or unit nonresponse (depending on the definition of partially complete questionnaires). In both cases, they reduce the amount of reported data, which increases the need for imputation or weighting to address nonresponse.

Several reasons for breakoffs include question interview length being too long, questions being too sensitive or personal, not having time to complete the survey, distrust of the government (for government surveys), simply not wanting to participate, a lack of English fluency, a survey instrument problem, or a household member not allowing participation (Stussman, Taylor, and Riddick, 2003). With telephone and face-to-face interviews, the presence of other people may increase breakoff. For crime victimization specifically, nonresponse may occur when respondents in telephone interviews are "gagged" in reporting crimes such as rape and domestic violence when a spouse is present (Yu, Stasny, and Li, 2008). Web survey breakoffs have also been associated with respondent characteristics (race/ethnicity, age, and education) with a lower breakoff likelihood overall among white, older, and more educated respondents, however not by question or questionnaire characteristics (Peytchev, 2006). Although race/ethnicity has been associated with breakoffs, it was to a much lower magnitude than nonresponse, suggesting these characteristics are not as important to predicting breakoff behavior as they are to predicting nonresponse behavior.

Survey characteristics may be better predictors of breakoff than respondent characteristics (Peytchev, 2011). For web surveys specifically, design elements (e.g., large grids) and transitions between topics have been blamed for increased breakoff rates. Other questionnaire characteristics can also increase breakoff, including questions with greater cognitive burden (e.g., long questions, attitudinal questions, and open-ended questions; Peytchev, 2009). With web and paper questionnaires, breakoffs increase with the number of questions on a single page, but not the cumulative number of questions in the instrument (Peytchev, 2006). Section introductions without questions have been found to induce breakoff in telephone (Groves and Kahn, 1979) and web-based surveys (Peytchev, 2009), perhaps because they are an opportunity to reassess the desire to participate.

This paper addresses breakoffs in a multiframe, multimode crime victimization survey, providing an interesting context in which to assess the effect of mode on breakoffs related to potentially sensitive content. To that end, the paper addresses the following research questions:

> 1) Do breakoffs differ between phone and web modes?

> 2) Do specific topics, questions, or screens lead to more hang-ups and breakoffs than others?

> 3) Do breakoffs vary by respondent geographic and demographic characteristics of where they live?

2. Data and Methods

2.1 Survey Design

Data came from the Oregon Crime Victimization Survey (OCVS). This multiframe (address-based sample [ABS] and random digit dial [RDD]) and multimode (web and computer-assisted telephone interviewing [CATI]) survey was conducted in late 2020 through early 2021 and was designed to estimate the incidence of various crime victimization experiences and perceptions of public safety, police, and victim services in the state of Oregon. Eligible respondents were adults (18 years and older) who were current Oregon residents and had lived in the state for at least 12 months. The survey instrument had approximately 120 questions and was offered in both English and Spanish. A pre-paid incentive of $2 was sent with all ABS sample mailed invitations to the web questionnaire. No pre-paid incentive was offered in the phone mode, and no post-incentives were offered in either mode.

Both ABS and RDD samples were stratified geographically to allow regional analysis and oversampling of Black and Hispanic populations. Oregon was divided by county into five geographic regions: Metro, North Coast, Central Valley, and East.

In both modes, the instrument began with eligibility screening questions and informed consent, followed by non-crime (quality of life) questions, and core demographics. Questions on index crimes, defined by the Federal Bureau of Investigation (FBI) as physical violence, robbery, burglary, theft, sexual assault, and rape came next. Following that, questions on non-index crimes (including nonphysical abuse, mental and emotional abuse, hate crimes, fraud, phishing, and stalking) were then asked. The final sections were crime follow-up questions about index crimes reported earlier, additional demographics, and the close of the survey. The average completion time for phone and web were 17 and 12 minutes respectively.

The ABS sample was only able to respond by web, and the majority of the RDD sample responded by phone. However, a small portion of the RDD frame responded by web. For the purposes of this analysis, respondents are categorized by their response mode.

2.2 Breakoff Analysis

The analysis focused on breakoff rates by question, questionnaire section, and geographic region. Breakoffs (i.e., web questionnaire breakoffs and phone interview hang-ups) were calculated separately for each mode. In both modes, only sample units that ended up as nonrespondents were coded as a "breakoff" (i.e., respondents who broke off or hung up at some point during the field period, but who eventually completed the instrument were not counted as a breakoff). For simplicity, the overall breakoff rate for each mode was calculated as the number of respondents who broke off before the end of the instrument over the total initial sample. Section-specific breakoff rates were calculated as the number of respondents who broke off at any point during a specific section, over the total number of respondents that reaching that section.

All analyses were unweighted because the purpose was not to create population estimates. Results should be considered exploratory due to the complexity of breakoff analyses. To assess the association between design features (e.g., mode and breakoff rate), statistical testing was conducted with Pearson chi-square tests of association at a 0.05 alpha level. Whenever testing was done, the results of that testing are mentioned, but

some results presented in this paper were not tested for statistical significance underscoring the importance of treating results as exploratory.

3. Results

3.1 Overall Breakoff Rates
About 3 percent of respondents who accessed the web survey and 1 percent who started the phone interview broke off. However, between the consent screens and end of the substantive questionnaire, the web questionnaire had only 4 breakoffs per question on average, while the phone interview had 7. Breakoffs occurred at almost every question on the web and CATI instruments.

3.2 Breakoffs by Section and Topic
Figure 1 shows the overall breakoff trends by instrument section for both modes. While the overall breakoff rates are relatively low, and most sections show breakoff rates under half a percent, two things jump out. First, breakoffs were higher in CATI at the consent screens. This may have been due to the amount of text that interviewers had to read to the respondent on those screens. Web respondents could read that information quicker than it could be read to them and could move though the screens at their own pace, even skipping the information if they wished. Second, and conversely, a higher percentage of web respondents broke off at the crime follow-up questions compared to CATI respondents. This may be due to the presence of an interviewer who can motivate the respondent to continue answering. No statistical testing was done on these results.
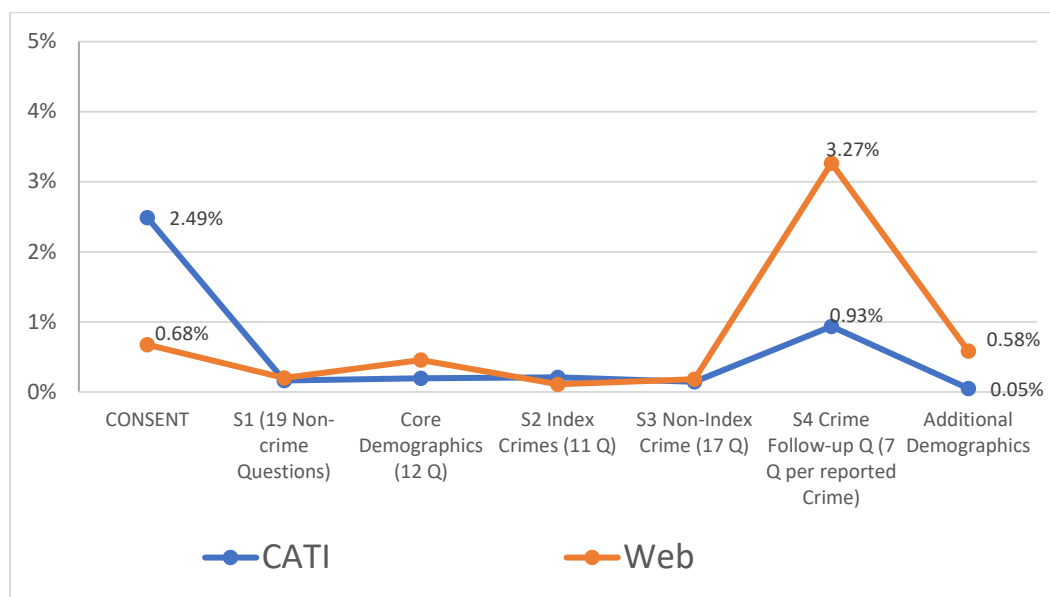

Figure 1: Breakoff rates by mode and topic

3.3 Breakoffs by Geography
While there was a significant association between breakoffs by mode and region (results not shown), there was no association between section and region within each mode. That is, no specific sections appeared to be more prone to breakoff in one region than another.

4. Discussion

While exploratory, this study found that overall breakoff rates were low, but phone interviews were more prone to breakoff than web questionnaires (about 3 percentage points higher breakoff rate). Consent was more prone to telephone hang-ups than web breakoffs, possibly due to the amount of text read or the pacing of it. Conversely, follow-up questions to reported index crimes were more likely to lead to breakoff in web than a hang-up in phone interviews. This finding underscores the importance of interviewers for motivating response to follow-up questions like this (see e.g., de Leeuw, Hox, & Huisman, 2003).

In short, there is no "winning" mode on the breakoff metric. Web respondents overall produced about 3 times the breakoff rate of phone respondents, but phone interviews led to a higher breakoff rate on substantive questions (i.e., questions after screening and consent). Further, section-specific breakoffs differed by mode as discussed above. Low breakoff rates and small sample sizes made it difficult to determine whether item-level differences were real or meaningful. For example, while the crime follow-up section showed a potentially higher overall breakoff rate in web questionnaires than phone interviews, with as low as 11 respondents to some questions (and breakoff rates as low as 0.02%) it was difficult to identify which specific questions are most prone to breakoff. Sample sizes were particularly small in the crime follow-up questions, due to filtering and earlier breakoff. Thus, making inferences about differences between individual questions becomes speculative, particularly in the absence of significance testing. In such situations, in-depth qualitative review of interviewer-respondent interactions (for phone interviews) or screen-level respondent paradata (for web questionnaires) often provide more helpful information than statistical tests. Such observations can help clarify the respondents' response processes, as well as any interaction difficulties with the question, whether the difficulties are related to the question, interviewer, or web survey interface.

Further, this paper assessed breakoff that led to unit nonresponse. Breakoff can also be conceptualized as the presence of questionnaire or interview termination that resolves to a complete or partial by the end of the field period (i.e., "ever breakoff" versus the "final breakoff" definition used here). Differing definitions can produce differing results that shed light on the relationship between item nonresponse and unit nonresponse. Additional research with these data, and certainly future research on this topic should combine multiple breakoff definitions with simple item nonresponse (i.e., missing data among completed and partial interviews and questionnaires) to better understand which items are problematic and to what degree.

Despite these limitations, some insights for questionnaire development and future research can be gleaned from these results.

1) Review phone consent scripts and any available phone interview recordings to identify situations that increase breakoff risk at those points in the interview
2) Evaluate the specific index crime follow-up questions that are contributing the most to breakoff, particularly in web questionnaires; Review interview recordings to see whether there are any interviewer behaviors that could be programmed into the web instrument as encouragement/motivation messages on these screens
3) For items with the highest breakoff rates, experimentally test alternative instructions, item wording, or response options to reduce breakoff risk

Placing these results in the context of other crime victimization surveys is also important. To the degree they are available, it would be insightful to compare OCVS breakoff rates with rates from other crime victimization surveys conducted in the same modes. The findings, like all breakoff results, need to be interpreted in the context of overall unit nonresponse (including noncontact and refusal) which is a much larger nonresponse problem and threat to generalizability.

References

de Leeuw, E. D., Hox, J., & Huisman, M. (2003). Prevention and treatment of item nonresponse. *Journal of Official Statistics, 19*(2), 153-176.

Groves, R. and R. L. Kahn (1979). Surveys by Telephone. New York, Academic Press.

Peytchev, A. (2006). A Framework for Survey Breakoffs. AAPOR - ASA Section on Survey Research Methods, 4205-4212.

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

Peytchev, A. (2011). Breakoff and Unit Nonresponse Across Web Surveys. Journal of Official Statistics. 27. 33-47.

Stussman, B., Taylor, B., & Riddick, H. (2003). Partials and Break-offs in the National Health Interview Survey, 2002 1.

Yu, Q., Stasny, E., & Li, B. (2008). Bayesian models to adjust for response bias in survey data for estimating rape and domestic violence rates from the NCVS. The Annals of Applied Statistics. 2. 10.1214/08-AOAS160.