Investigating the Effects of Questionnaire Design and Question Characteristics on Respondent Fatigue

Elias Markstedt, the SOM Institute, University of Gothenburg,
elias.markstedt@som.gu.se

Frida Vernersdotter, the SOM Institute, University of Gothenburg,
frida.vernersdotter@som.gu.se

Prepared for the AAPOR Annual Conference 2013, Boston, May 16-19

Introduction

The contexts in which breakoffs occur in mail-back surveys have not been thoroughly investigated. In this study we examine how questionnaire design, question characteristics, and respondent characteristics affect breakoff propensity in the case of self-administered mail surveys. We draw on 13 years of consecutive self-administered mail surveys in Sweden, annually conducted by the SOM Institute at the University of Gothenburg, with a total of over 52 000 respondents and 31 different questionnaires.

The SOM study is an annually repeated general social survey that covers a wide range of topics on society, media and politics and is used for academic research on attitudes, values, self-reported behavior, and socio-economic status. The questionnaires are on average 20 pages long and have a mean response rate of 55 percent (RR5), 58 percent (RR6). The SOM studies have been fielded as a mail-back questionnaire since the beginning in 1986. They use a random sample of the adult population of Sweden and are characterized by high quality field work and representativeness of their samples. The SOM-institute has always focused on high response rates and included many reminders and other efforts to reduce unit nonresponse. However, just as other surveys, they have faced increasing problems with declining response rates, despite their increased efforts with reminders, both via mail and telephone. The SOM surveys do not use any economic incentives for participation.

Breakoff and partial breakoff

Survey noncompletion, breakoff, is often overlooked in the discussion on survey response rates as a measure of data quality. It is urgent to differentiate the item nonresponse due to breakoffs from the selective item nonresponse, as the questions following the breakoff have not been seen and assessed by the respondent. This paper replicates the study of Andy Peytchev published in POQ issue 73 2009 "Survey Breakoff" (Peytchev 2009) analyzing breakoffs in mail-back surveys instead of the web surveys. Defining breakoffs in paper surveys is problematic as the condition of not having seen a question is difficult to verify. One can argue that a true breakoff in a paper survey is one that is never mailed back to the researcher, making it impossible to study. It is however still of interest to study the respondent behavior of skipping several pages in a

questionnaire, which is why we proceed in our analysis calling this form of breakoff a partial breakoff.

We define partial breakoff as the point after which a respondent has omitted to answer a sequence of at least 30 items. It should be noted that most questions in the SOM study are multiple item grid questions and that the item – not the question – is our unit of analysis meaning that a breakoff can occur in the middle of a question. We argue that a sequence of 30 items is a reasonable limit for partial breakoff since the average survey page contains 20 items, meaning that a sequence of 30 unanswered questions is equivalent to a full survey page. The median length of a nonresponse sequence is 49 items (mean 75). Our definition rely on the assumption that the respondent who decides to skip a full page does not read the questions on the skipped page and only resumes reading when he or she reaches the next page. In accordance with the concept of partial breakoff it should be noted that our definition allow the respondents to take up answering questions after the sequence of 30 unanswered questions and still count as a partial breakoff.

Partial breakoffs occur throughout the length of the questionnaire as shown in figure 1. The y-axis represents the longest spell of unanswered questions and x-axis represents the point in the survey at which the partial breakoff occurred. Partial breakoffs in paper surveys cannot be explained as a function of the length of the questionnaire. Question characteristics, survey design and respondent characteristics are more important in the understanding of partial breakoffs than questionnaire length.
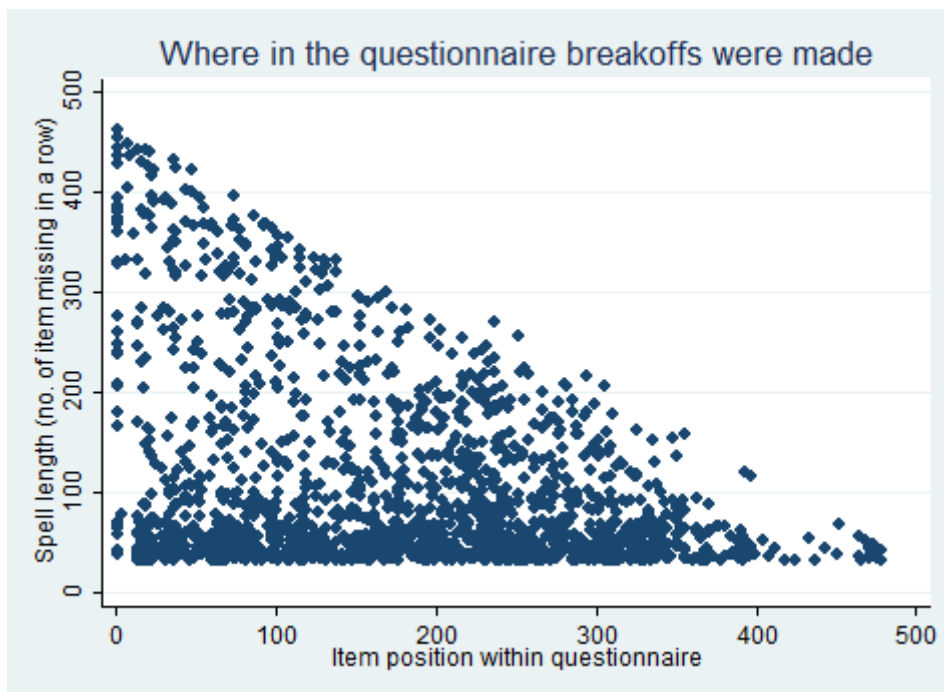

Figure 1. Partial breakoff scatter plot

Method and data

We use aggregated data from the SOM study collected over thirteen years through 31 questionnaires containing almost 7 000 items. Each questionnaire was sent out to a sample of 3000 people with an average response rate of 55 percent (RR5), 58 percent (RR6), resulting in a total of   52 551 respondents, out of which 4 103 were identified as partial breakoffs.

Investigating the propensity of breakoff we use the method of survival analysis where a breakoff is considered a failure. Each item was coded for its position in the questionnaire, constituting the time axis of the analysis. Our model includes independent variables on question characteristics, questionnaire design and respondent characteristics that have been shown by Peytchev (2009) to have an effect on breakoffs.

Questionnaire design

The questionnaires used in the SOM study are on average 20 pages long and A4 sized. The study is carried out annually using one to three parallel but differently profiled questionnaires in which approximately one fifth of the questions are identical. The questions are thematically clustered in sections introduced by a banner. There are approximately ten sections in each questionnaire, each ending with a section of background questions. Respondents who breakoff in the middle of the survey sometimes find their way back at the end of the survey, filling out the final section of background questions. The questionnaire design characteristics included in our model are:

Number of items per page – Item intense pages are cognitively challenging. This is a continuous variable denoting the total number of items on each page. (continuous)

Question position – Each item was coded for the position of the question in which it appears in the questionnaire. This is a proxy for questionnaire length. (continuous)

Question characteristics

The predominant question type is multiple item grid questions with verbal rating scales measuring attitudes and habits. Few questions are open-ended requiring a simple number or year or a more elaborate written answer. These are the question characteristics included in the analysis.

Battery length – Most questions in the SOM study have multiple items. Each question was coded for the total number of item it contains. (continuous)

Open ended – Open ended questions vary in their design. Some are completely open requiring a free text answer, some questions have one open ended item where the respondent can add additional information using the same scale as the close ended items in the question. Each item was coded for being open ended or not. 4 percent of the questions were coded as open ended. (binary)

Long question – The SOM institute strives to keep questions as short and simple. They make an effort to phrase questions using only simple sentence constructions and terms. Each question and item is coded for its total number of characters. In our model a long question is a defined as the top quintile of the distribution of question length in the 25 questionnaires setting the limit to 119 characters. 20 percent of the questions are longer than 119 characters. (binary)

Sensitive question – What constitutes a sensitive question is subject to personal predisposition. Every question has been coded for sensitivity by the authors with an undeniable subjective bias to personal sense of integrity. The authors have been restricted in the coding of sensitivity trying to limit the scope to what is perceived as a lowest common denominator, including questions on health, sexuality and income. 3 percent of the questions were coded for sensitivity. (binary)

Question type – Each question was coded as a judgment or a retrieval question. A retrieval question requires the retrieval of objective information about habits or behavior. A judgment

question requires a subjective opinion or evaluation of something. Every question was categorized as one of the two. 56 percent of the questions were coded as judgment questions. (binary)

Verbal scale points – Some scales are numeric, others have verbal scale points. Each question was coded for the presence of verbal scale points. Numeric scales with verbal end points were coded as non-verbal scales. 90 percent of the questions have verbal scale points. (binary)

No opinion – In recent years it has become increasingly popular to offer the respondents an escape route by adding a "no opinion"/"don't know" option at the end of the scale. Each question has been coded for having or not having such an escape option. 20 percent of the questions have the escape route of "no opinion". (binary)

Respondent characteristics

Respondent characteristics included in the model are age, gender, education, income, labor market status, and political interest.

Education – Each respondent was coded for education level. A respondent who has attended college or university was coded for high education. 32 percent of the respondents have a high education. (binary)

Income – Information on income was collected from a question on combined household income. The scale is composed by intervals that have been adjusted over the years to represent the normal distribution of the income in the country. The top third share was coded as high income. 23 percent of the respondents live in high income households. (binary)

Labor market status – Each respondent was coded for employment. No employment includes not only unemployed but students, retirees and people with permanent disabilities excluding them from the workforce. 41 percent of the respondents were unemployed (binary)

Political interest – There is a question in the SOM surveys on self-evaluated political interest on a four point scale from "not at all interested" to "very interested". The response categories were pooled and dichotomized. 52 percent of the respondents reports being interested in politics. (binary)
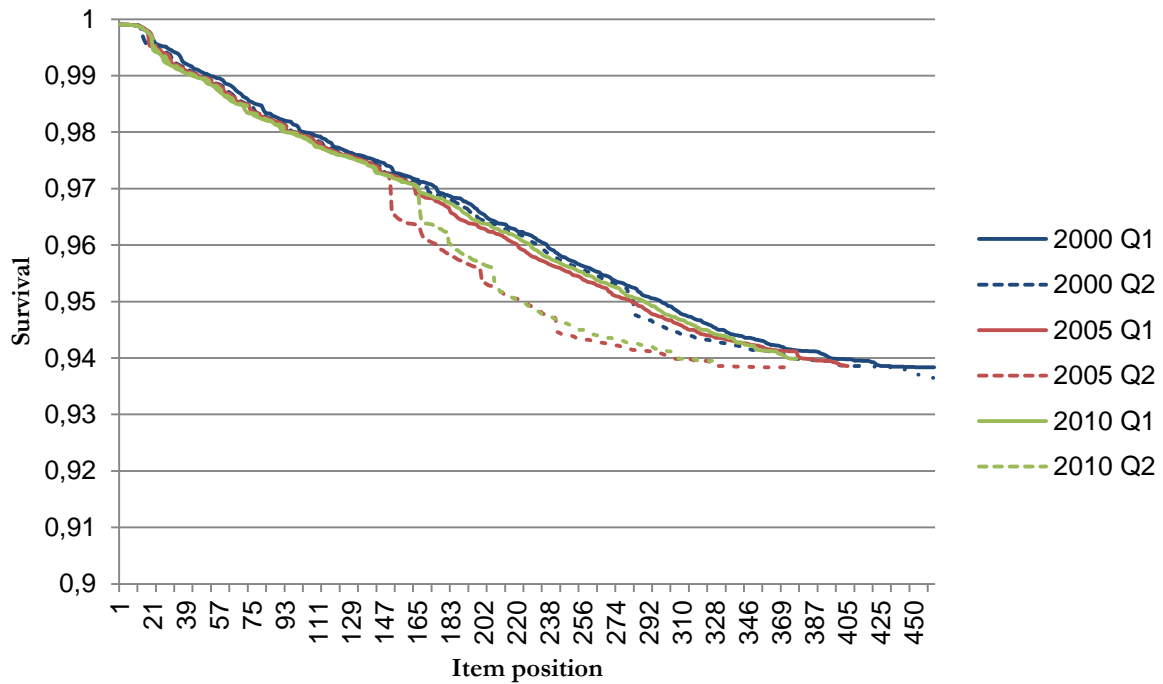
Figure 2. Kaplan-Meier survival estimate for six questionnaires from three years

Results

The Kaplan–Meier survival estimates from three different years are presented in figure 2. The selected years from the beginning, middle and end of the investigated period, are representative for the same. The Q1 and Q2 denoting the curves refer to the questionnaire edition. As above mentioned the SOM study uses parallel and somewhat differently profiled questionnaires for data collection. The Q1 has a political profile and contains more judgment intense questions than Q2. Q2 has a media profile with more retrieval intense questions. The curves are declining smoothly showing that the overall breakoff propensity is similar across the editions. There is however a systematic difference between the editions as the retrieval intense media editions show a somewhat sharper decline at an earlier point in the questionnaire than do the political questionnaire. This decline is due to a general trend of higher item nonresponse in the media questionnaires. From previous studies on the SOM surveys we know that the retrieval intense Q2 is perceived as repetitive resulting in a higher item non-response as the respondents, halfway through the questionnaire have the impression of having already provided the information asked in a previous question.

Table 1. Discrete Hazard Survival Models with Page-Varying Covariates Predicting Breakoff in thirteen years of SOM surveys

Predictor	Relative hazard	p-value
Questionnaire design		
Number of items per page	1. 06	0.000
Question position	0.99	0.000
Question characteristics		
Battery length	0.93	0.000
Long question (yes=1)	1.31	0.000
Open question (yes=1)	1.83	0.000
Sensitive question (yes=1)	1.46	0.000
Questions type (judgment question=1)	0.90	0.004
Verbal scale points (yes=1)	0.65	0.000
"No opinion" (yes=1)	1.28	0.000
Respondent characteristics		
Sex (female=1)	1.06	0.065
Age	1.01	0.000
Income (high income=1)	0.67	0.000
Labor market status (Non-working=1)	0.80	0.000
Political interest (Interested=1)	0.57	0.000

Number at risk/number of breakoffs 52 551/3 930

The length of the questionnaire is known to have an effect on overall response rates. Among those who made the decision to take the survey however, the length of the questionnaire is of less importance to whether they complete it or not. The cognitive burden is more important to survey completion. Crammed pages and questions induce breakoffs. Every additional item on a page increases the relative risk of breakoff with 6 percent and long questions increase the relative risk with one third. The total number of items in a question however has no effect on the relative risk of breakoffs. This can be explained in the context of the SOM study as a function of a uniform battery design. The respondent is introduced to the battery format early in the questionnaire which them accustomed to the format and less likely to intimidated by long batteries later on. Another explanation could be that the longest batteries, in the case of the SOM studies, concerns everyday lifestyle habits.

Other cognitively burdensome question characteristics that affect the breakoff propensity negatively are open questions and non-verbal scales. Open questions increase the relative risk with 80 percent. Sensitive questions, that generally known to increase item-non response, also have an effect on breakoffs. Sensitive questions increase the relative risk of breakoff with 46 percent.

Contrary to our expectations retrieval questions appear to be slightly more burdensome than judgment questions. This could be the result of a too inclusive categorization. Instead of rating the intensity of the retrieval or judgment requirement of each question, we have chosen to categorize each question as one of the two, resulting in very little effect of question type. It is however possible interpret the "no opinion" characteristics as a proxy for intensity. Questions that the SOM Institute expects to be difficult to answer due to the intensity of retrieval or

judgment are equipped with a "no option" or "don't know" alternative. Interpreting the "no opinion" characteristic this way, the retrieval and judgment intense questions increase the relative risk of breakoff with one third.

Respondent characteristics such as age and gender seem to have less importance in predicting breakoffs, even though older respondents are somewhat more prone to breaking off. High income, high education and high political interest however decrease the relative risk of breakoff.

Conclusion

In accordance with the results of Andy Peytchev we conclude that cognitively burdensome questions and questionnaire characteristics increase the relative risk of breakoff and partial breakoff in paper surveys as well as web surveys. The relative risk of partial breakoff in paper surveys generally lower than the risk of (complete) breakoff in web surveys due to the fact that the *true breakoffs* in paper surveys never make it to the researcher. Respondents who are better equipped cognitively living in high income households are less likely to break off than their counterpart.

References

AAPOR Standard Definitions - Final Dispositions of Case Codes and Outcome Rates for Surveys. 2011. Available online: "http://aapor.org/Content/NavigationMenu/AboutAAPOR/StandardsampEthics/StandardDefinitions/StandardDefinitions2011.pdf".

Peytchev, Andy. 2009. "Survey Breakoff". *Public Opinion Quarterly*. 73(1): 74–97.