Impact of question topics and filter question formats on web survey breakoffs

Zeming Chen, Alexandru Cernat and Natalie Shlomo
University of Manchester, UK

Stephanie Eckman
RTI International, Washington, DC, USA

Abstract
Web surveys have become increasingly popular over the last decade, but they tend to suffer from breakoffs, which take place when respondents start the survey but do not complete it. Many studies have investigated the factors impacting breakoffs, but they often ignored the breakoff timing and gave scant attention to two factors: question topics and filter question formats (grouped vs. interleaved as defined by whether filter questions are presented upfront or not). Using survival analysis, this study first identifies when breakoffs are more likely to happen and what are the time-varying predictors of breakoffs. Then, by using a web survey that experimentally manipulates the filter question format and randomly orders the question topic, this study investigates the effect of question topics and filter question formats on the breakoff event and its timing. We find that most breakoffs tend to happen at the beginning of the questionnaire and at the place where a new question topic is introduced. While item nonresponse is associated with more breakoffs, it is surprising to see that open-ended and long questions are associated with a lower breakoff risk. Additionally, we discover that grouping the filter questions leads to fewer breakoffs at the beginning compared to the interleaved counterpart, but the breakoff risk in the grouped format catches up quickly when respondents realise their previous answers will trigger more questions. This study also shows that questions about insurance have more breakoffs while questions on demographics and income have fewer breakoffs despite their sensitivity level.

Keywords
web surveys, breakoff timing, survival model, questionnaire design, experimental design

Corresponding author:
Zeming Chen, Department of Social Statistics, The University of Manchester, Oxford Road, Manchester M13 9PL, UK. Email: zeming.chen@manchester.ac.uk
Introduction

Surveys have been widely used in different fields, such as market research and political polling. Due to the cost concern and tight schedule, an increasingly number of surveys have been conducted online. However, running surveys on the web has some limitations, one of which is survey breakoff. Survey breakoffs happen when the respondent starts the survey but fails to complete it (Tourangeau et al., 2013). As a result, missing data are produced, causing subsequent analysis to have lower statistical power as well as potentially biasing results (Steinbrecher et al., 2015).

To address the breakoff, it is important to understand its determinants. Past studies (e.g. Peytchev, 2009) have identified many factors impacting breakoffs, but most of them studied breakoffs only as a binary outcome and ignored the breakoff timing. Investigating breakoff timing is important as survey practitioners want respondents not only to complete the survey but also to complete as many questions as possible before they break off (Sakshaug & Crawford, 2010).

The present study will apply survival analysis to an opt-in web survey to investigate when breakoffs are more likely to happen and what are the time-varying factors (factors whose value varies throughout the questionnaire) that explain breakoffs. Additionally, the study will investigate two other important factors that have received scant attention: question topic and the format of filter questions.

The content, sensitivity and placement of question topics can impact the breakoff and its timing. Topics that are relevant to the respondents can decrease or postpone breakoffs (Shropshire, Hawdon and Witte, 2009) while sensitive topics might have the opposite effect. When studying the effect of question topics on breakoff, randomising the topic order is important; otherwise, ignoring the ordering effect could confound the topic effect and cause spurious correlations with breakoffs. Nevertheless, this has not been done in prior research.

The format of filter questions is another factor that has received limited attention in the breakoff literature. Filter questions can produce a high degree of response burden as the positive answer to a filter question can lead to more questions. There are two main ways of presenting filter questions and their follow-ups. In the grouped format, all filter questions are asked before the follow-ups are displayed, whereas in the interleafed format every filter question immediately triggers its follow-ups (Kreuter, Eckman and Tourangeau, 2020). Although both filter formats can cause response burden and are prone to breakoffs, there is a difference between them in the timing when respondents learn about the burden. Respondents answering the interleafed format will quickly understand the response burden after giving affirmative answers to one or two filter questions. They could break off as early as the first pair of filter and follow-up questions. In the grouped format, respondents can only learn about the extra burden when they reach the follow-ups. They are therefore expected to break off later. However, no previous research has tested the relationship between breakoff timing and filter question formats.

This study uses a web survey that experimentally manipulated the filter question format and randomly ordered the question topic. Thus, we are able to causally investigate the impact of these two factors on the breakoff and its timing.

Background

Framework for studying breakoffs

Breakoffs are prevalent in web surveys. For example, Revilla (2017) reviewed 185 opt-in web surveys distributed through a Spanish survey company and found that the mean breakoff rate was
In an online probability survey about University of Michigan staff and students’ attitudes towards environmental issues, the breakoff rate was 13%, 14% and 17% in the year of 2014, 2015 and 2018, respectively (Mittereder, 2019).

Given the prevalence of survey breakoff, many researchers have been studying its causes. As a result, a framework has been developed to summarise different factors. According to Peytchev (2009) and Mittereder and West (2021), these factors can be grouped into four categories: (1) page/question characteristics, (2) survey design, (3) respondent factors and (4) paradata.

Page/question characteristics refer to the design features of survey pages and questions. Cognitively demanding questions such as matrix, open-ended questions and questions with more characters are associated with more breakoffs (Peytchev, 2009; Hoerger, 2010; Tijdens, 2014; Steinbrecher et al., 2015). These types of questions can impose extra burden on respondents when they engage in a series of actions required to answer the question such as comprehending the question and retrieving the relevant information (Tourangeau, 2018). To avoid the burden, respondents might choose to break off.

The second factor that is related to breakoff is the survey design. Examples of this are providing incentives unconditional on survey completion (Silber, Lischewski and Leibold, 2013), using a lengthy questionnaire (Hoerger, 2010) and displaying the progress bar alongside the questionnaire (Villar et al., 2013).

The third group of factors – respondent factors – refers to the characteristics of the respondents, and they are included in the study of breakoffs for different reasons. On the one hand, respondent characteristics are the proxy indicator of their ability to handle the survey burden. Respondents at a cognitive disadvantage are expected to be less capable of coping with the burden and more likely to break off. Such respondents were found to be older and have a lower education degree (Peytchev, 2009; Blumenberg et al., 2018). On the other hand, some demographic variables are associated with respondents’ general level of cooperation with the survey request. The findings regarding the relationship between breakoffs and these variables are often mixed, including gender, race, marital status, student and income (Galesic, 2006; Peytchev, 2009; 2011; Klein et al., 2011; Mittereder & West, 2021), but the general trend is that male, non-white, student and more affluent respondents are more likely to break off.

Paradata, the final category in the framework, refer to the information collected during the response process (Kreuter, 2013). This type of data is believed to reflect the change in the response burden and respondents’ motivation throughout the questionnaire (Mittereder & West, 2021), thereby being useful for predicting the imminent breakoff. Some paradata that have been associated with breakoffs are the proportion of questions that are not answered (Mittereder & West, 2021) and using mobile devices to answer the survey (Wenz, 2017).

The above four categories form a comprehensive framework. However, previous research has found that web survey breakoffs were preceded by an accumulated respondent burden (Galesic, 2006). This means that the burden caused by the factors in the framework takes some time before it can actually exert its influence on breakoffs. An example of this can be seen in the study conducted by Mittereder and West (2021). By allowing the effects of the responding device to vary in time (measured as the cumulative number of questions answered), they found that there was no difference in the breakoff between the non-mobile and mobile devices at the beginning of the survey. However, when mobile device respondents answered more questions, they were more likely to break off.

Based on the review so far, we argue that the timing dimension is a necessary factor to consider in the study of breakoffs and the effect of time-varying factors on the breakoff needs further research. In the present study, we derive the time-varying factors from question characteristics (e.g. question word count) and paradata (e.g. item nonresponse rate) and investigate how they impact breakoffs,
after controlling for the difference in respondents’ cognitive ability and survey cooperation using their demographic information such as gender, age, ethnicity and education. The first two research questions are:

RQ1. When is the breakoff more likely to happen in the web survey?

RQ2. What are the timing-varying predictors of the web survey breakoff?

In addition to examining breakoff timing, this study will contribute to the literature by focussing on two specific factors: question topic and the filter question format.

Question topics and breakoffs

Many studies have identified the survey topic as an important factor for unit nonresponse (not answering the survey at all) (Groves, Singer and Corning, 2000) and item nonresponse (not answering some of the questions) (Tourangeau & Yan, 2007). Survey breakoff, as a special type of nonresponse, is also impacted by respondents’ topic interest. For instance, when analysing the data from a web survey that covered the topic of conservation, Shropshire, Hawdon and Witte (2009) documented that respondents who scored higher in their conservation support were less likely to break off.

In addition to the interest in the topic, the perceived sensitivity of the topic can also impact survey nonresponse. When facing a sensitive topic such as income or sexual orientation, respondents might feel uncomfortable with its intrusiveness or worried about the potential threat of disclosing personal information (Tourangeau & Yan, 2007). As a result, respondents will skip those sensitive questions, leading to item nonresponse. In fact, the behaviour of skipping sensitive questions was found to be more frequent in the interviewer-administered survey mode compared to a self-completion mode (Kreuter et al., 2008).

As an alternative to not answering the sensitive question, respondents might terminate their survey participation. However, among those papers that investigated the relationship between question topics and breakoffs (see McGonagle, 2013; Mittereder & West, 2021), the order of the topics was not randomised. For example, McGonagle (2013) analysed a telephone survey about the U.S. families’ economic status, but the topic about respondents’ housing was always followed by their employment history, income and so on. Previous research has noted that the order of the topic could affect the rates of item nonresponse (Teclaw, Price and Osatuke, 2012).

Without the order randomisation, prior work failed to separate the impact of the topic content from that of the topic order.

An ideal design to investigate the impact of question topic on breakoff timing is to randomise the order of questions and include questions with different levels of sensitivity. We used such a design to answer our third research question:

RQ3. Does the topic of the questions impact the breakoff and its timing?

Filter question formats and breakoffs

The use of filter questions can also impact web survey breakoff and its timing. Many surveys use filter questions which trigger some follow-up questions when answered positively. For example, if the respondent chooses ‘yes’ to the filter question ‘Have you held a full-time job during the past
12 months’, then more questions will follow (e.g. ‘From when and until when did you hold this job’). The grouped and interleaved formats are two main ways to present filter and follow-up questions. A visual example of both formats is shown in Figure 1.

One advantage of the grouped format is that the connection between a ‘yes’ answer to the filter questions and the activation of follow-ups is not immediately apparent, so respondents facing the grouped filter questions would choose more ‘yes’ answers compared to the interleaved format. This was found in Eckman et al. (2014) after randomly assigning the respondents of a probability-based telephone survey to either grouped or interleaved formats and comparing the number of ‘yes’ in filter questions between the two formats.

However, in the grouped version, the follow-up questions are far away from the corresponding filter questions, so respondents have to recall the relevant information from their memory again, which causes recall difficulties and hampers the cognitive processing (Clark-Fobia, Kephart and Nelson, 2018). Kreuter et al. (2011) also randomly allocated sample members of a different telephone survey to the grouped or interleaved format and noted that respondents in the grouped format chose more non-substantive answers (e.g. ‘Don’t know’) for the follow-up questions.

---

**Figure 1.** Example of grouped and interleaved formats in the web survey analyzed in this study (filter questions are highlighted).
Unlike the grouped format, the interleaved version puts together questions that are of the same topic, serving as a recall aid (Kreuter, Eckman and Tourangeau, 2020). Yet, in the interleaved version, respondents can quickly learn that a positive answer to the filter question will trigger more questions. They then are more likely to deliberately choose a ‘no’ to shorten the questionnaire, which was documented in the Eckman et al. (2014) and Kreuter et al. (2011) studies mentioned above.

The review of the grouped and interleaved formats highlights that both formats impose a burden on the response process and respondents who do not want to or cannot handle this burden will provide lower data quality. Rather than giving incorrect answers to reduce the length of the survey, respondents could break off. In addition to comparing the effect of grouped and interleaved formats on measurement error, Kreuter et al. (2011) and Eckman and Kreuter (2018) also looked at the influence of the filter question format on breakoffs. Both studies found that the format was not associated with breakoffs.

However, previous studies did not investigate whether there is a difference in the breakoff timing between filter question formats. As we argued previously, it is important to consider the timing in breakoff studies as it might produce new insights regarding mitigating breakoffs. Thus, the final research question in this study is:

**RQ4.** Does the filter question format affect the timing of the survey breakoff?

**Data**

The data used in this study come from a web survey conducted between September and October 2019. The web survey was administered to members of the Lightspeed Panel, an opt-in web panel in the United States. Upon completing the survey, the respondents received reward points which could be accrued and redeemed later. Given the opt-in nature, it is impossible to calculate the response rate. The survey analysed here is dominated by white respondents (74%) and females (66%). These deviations from the US population make the subsequent findings less generalisable.

Nevertheless, we consider that this web survey is appropriate to be analysed for three reasons. First, it records the outcome of interest – breakoffs. After removing two individuals with an unknown response status, the final sample size for analysis is 3128. Out of these, 520 respondents accessed but did not complete the survey, resulting in a breakoff rate of approximately 17%. This breakoff rate is slightly higher than other surveys reviewed in the previous section even though individuals in our survey voluntarily participated in the survey and could only receive the reward upon survey completion. Meanwhile, the survey recorded the last question respondents saw, enabling the investigation of breakoff timing.

Second, the web survey includes six different topics with varying levels of sensitivity. Questions of the same topic are organised into a single block, resulting in six blocks in the survey. As shown in Table 1, the topic of Block 1, 2 and 6 always remains the same, namely, respondents’ demographic information, housing unit and household income. In contrast, the topics of the three remaining blocks (Block 3, 4 and 5) are randomised among respondents’ clothing purchase, utilities payment and non-health insurances. This randomisation leads to six possible orders among the blocks (see Table 1). Respondents were randomly assigned to one of the six orders upon seeing the first randomised blocks (i.e. Block 3). For 317 respondents who broke off at Block 1 or 2, their assigned order is unknown. Within each block, the order of the questions is fixed. Different topics in the survey help answer the research question regarding how question topics affect breakoffs, especially for the three topics whose order is randomised. Also, Demographics and Income (Block 1 and 6) are
considered sensitive as they are either found to suffer from more item nonresponse (Tourangeau & Yan, 2007) or recommended by survey practitioners to be placed towards the end of the questionnaire so respondents feel more comfortable to share such information (Allen, 2017). This varying sensitivity allows us to study the relationship between topic sensitivity and the survey breakoff.

Lastly, the survey embedded an experiment about filter question formats in the three randomly ordered blocks (i.e. Block 3, 4 and 5). The respondents were randomly assigned to either the grouped (49% of the sample) or the interleaved format (51% of the sample). Depending on the block, there are five or six filter questions, each of which can trigger five or six follow-ups.

In total, the survey analysed here has 196 question items, and approximately 80% of them are in the three randomised blocks (see Supplemental Appendix A for the number of questions in each block). At the beginning of nearly every question block, there is an introduction statement which informs respondents of the upcoming block’s topic and encourages respondents to give accurate answers. Respondents can either click a radio button to show their acknowledgement or skip to the next question. Among the 196 total items in the questionnaire, six items are introduction statements, which we code as the reference category for the question topic. On average, the respondent who broke off saw 16 questions (standard deviation = 19), much lower compared to those completed the survey (85 questions, standard deviation = 21). The descriptive summary for all variables used in this study along with how they are coded are provided in Supplemental Appendix A.

### Method

We use the survival model to answer the research questions. The survival model is useful in explaining whether, and if so when, the event of interest happens (Singer & Willett, 2003). Following previous studies on survey breakoffs (Peytchev, 2009; Mittereder & West, 2021), time is measured by the cumulative number of questions respondents saw and treated as discrete. As Willett and Singer (1993) emphasised, when the time metric is discrete, the likelihood function for estimating the discrete-time survival model and the standard logistic regression is algebraically equivalent. We therefore use the standard logistic regression to fit the discrete-time survival model in this study. The model is estimated using the `glm` command in R 4.0.2 (R Core Team, 2020) and takes the following form

\[
\ln\left(\frac{P_{1q}}{1 - P_{1q}}\right) = \alpha_q + \beta_1 X_{1q} + \beta_2 X_{2q}(q)
\]

| Order | Block 1   | Block 2   | Block 3   | Block 4   | Block 5   | Block 6   | Breakoff/Total |
|-------|-----------|-----------|-----------|-----------|-----------|-----------|----------------|
| 1     | Demographics | Housing   | Clothing  | Utilities | Insurance | Income    | 21/461         |
| 2     | Demographics | Housing   | Clothing  | Insurance | Utilities | Income    | 32/470         |
| 3     | Demographics | Housing   | Utilities | Clothing  | Insurance | Income    | 27/476         |
| 4     | Demographics | Housing   | Utilities | Insurance | Clothing  | Income    | 36/454         |
| 5     | Demographics | Housing   | Insurance | Clothing  | Utilities | Income    | 48/478         |
| 6     | Demographics | Housing   | Insurance | Utilities | Clothing  | Income    | 39/472         |
| 7     | Demographics | Housing   | Unknown   | Unknown   | Unknown   | Income    | 317/317        |
| Total |           |           |           |           |           |           | **520/3128**   |

Table 1. All possible orders of the question blocks in the web survey analysed in this study.
\( P_{iq} \) represents the probability of person \( i \) breaking off at question \( q \) given that this person has not broken off at any question prior to \( q \). This conditional probability is called hazard in the survival literature (Singer & Willett, 2003). The equation shows that the logit transformed hazard is a linear function of three terms. The first term, \( \alpha_q \), is the baseline hazard, which quantifies the hazard of breaking off at question \( q \) when all covariates in the model are zero. The other two terms, \( \beta_1 X_{i1} \) and \( \beta_2 X_{i2}(q) \) represent a set of different covariates \( X \) and their impact \( \beta \) on the logit hazard. The difference between them is that \( X_{i1} \) represents the time-constant covariates (e.g. ethnicity) and \( X_{i2}(q) \) represents the time-varying counterparts (e.g. question word count).

Four logistic models are developed to answer the research questions in this study. Model 1 involves only time represented as the number of questions seen and the respondents’ demographic characteristics while Model 2 adds in the time-varying factors. These two models will together address RQ 1 and 2 (i.e. when breakoffs are likely to happen and what are the time-varying predictors).

To answer RQ 3 and 4 (i.e. how question topics and filter question formats affect breakoff and its timing), the analysis sample is restricted to only Blocks 3, 4 and 5. As mentioned earlier, the experiment of topic orders and filter question formats only exists in these three blocks. The sample restriction enables us to only investigate the breakoffs happening under the experimental design and measure the effect of both factors on breakoffs more directly.

Model 3 is derived by applying Model 2 to the restricted sample but with two changes. Firstly, given that some topics are discarded in the restricted sample, the variable about question topics now includes only four categories, namely, Clothing (the reference category), Utilities, Insurance and Introduction Statement. Using Clothing as the reference category (rather than the Introduction Statement as in Model 1 and 2) helps investigate how the topics of other two randomised blocks (Utilities and Insurance) impact breakoffs in comparison to Clothing. Secondly, the variable representing the matrix questions is excluded as these questions only exist in Block 2, which is eliminated from the restricted sample. In Model 4, we add in two interaction terms between time (i.e. number of questions seen) and the grouped/interleafed format as well as the question topics. Model 3 will investigate whether the question topics and filter question formats impact the breakoff risk, and Model 4 will answer whether their impact on breakoffs changes throughout time.

Due to a sizable number of breakoffs prior to the demographic-related questions, demographic variables suffer from missing data (ranging from 2% to 8% as shown in Supplemental Appendix A). To minimise the sample loss throughout the model development, we use multiple imputation for the missing demographics. Following Enders (2010), we included all variables in the substantive model in the imputation (i.e. breakoff status, time, demographics, question characteristics and paradata) as well as the order of question blocks respondents were assigned to (as shown in Table 1). We created 10 imputed datasets, each of which was obtained after 50 iterations. Parameters of all substantive models were separately estimated on these 10 datasets and then pooled together by the combining rule of Rubin (1987). We also conducted a sensitivity analysis by coding the missingness in demographics variables explicitly as a level in the model, but the conclusion regarding our research questions does not change (see these results in Supplemental Appendix B); therefore, models built upon the imputed datasets are reported here. The imputation was performed in R 4.0.2 using the mice package (van Buuren & Groothuis-Oudshoorn, 2011). For the univariate description of the variables before and after imputation, see Supplemental Appendix A.
Results

Change in breakoff hazard over time

Figure 2 plots time (i.e. the number of questions seen) on the x axis and the breakoff hazard on the y axis. A larger hazard indicates a higher breakoff risk. Figure 2 illustrates that the largest breakoff hazard is at the beginning of the survey. The second peak lies between the 15th and 20th questions. Because questions in the range of the second peak either involve sensitive topics (i.e. rent/mortgage for the dwelling), belong to matrix questions or introduce a new series of topics, we speculate that the second peak is more likely attributed to the question characteristics rather than time. After the second peak, the breakoff hazard tapers off. All peaks after 100 questions are mainly due to the rare breakoff event and decreasing number of respondents included in the denominator for calculating the breakoff hazard (For instance, at the 115th question, only 206 respondents remained in the survey, and a single breakoff event among this small denominator is causing large peaks in the tail of the distribution).

As shown in Figure 2, the breakoff hazard is non-linearly associated with time, and there is only one change in the direction of breakoff hazard that is genuinely related to time. We therefore decided to fit all our survival models using linear and quadratic forms of time. We also conducted a sensitivity analysis by coding the time differently (see Supplemental Appendix C), but the quadratic time model conforms to the trend in Figure 2 and strikes a good balance between model interpretation, goodness-of-fit and parsimony. Thus, the quadratic time model will be reported in the following section.

Factors impacting breakoff and its timing

Models 1 and 2 investigate what factors impact the breakoff on the full sample. As can be seen in Table 2, the odds ratio of linear time (i.e. number of questions seen) of Model 1 is smaller than one, indicating that the more questions a respondent answers, the less likely she is to break off. This trend does not remain constant. The odds ratio corresponding to the quadratic time is greater than 1 (exact
value not shown due to rounding), so the downward breakoff likelihood flattens out to some extent as time passes by.

Model 1 also estimates the impact of different respondent demographics on survey breakoffs. Non-white respondents are 19% less likely to break off than the white peers. Students have a five-fold increase in the breakoff risk. Meanwhile, compared to respondents with a degree at the high school level or below, holders of a degree at the college level or above are about 80% less likely to break off.

Adding question characteristics and paradata to the model (i.e. Model 2) improves the overall model fit given the large AIC decrease. Compared to Model 1, the impact of student status and

### Table 2. Odds ratio of logistic regression predicting breakoff (based on the full sample).

| Variable                                | Model 1 | Model 2 |
|-----------------------------------------|---------|---------|
| Intercept                               | .01***  | .28***  |
| Number of questions seen (linear)       | .92***  | .91***  |
| Number of questions seen (quadratic)    | 1.00*** | 1.00*** |
| Married (ref: no)                       | .72     | .92     |
| Male (ref: female)                      | 1.10    | 1.14    |
| Age                                     | 1.01    | 1.01*** |
| Non-white (ref: white)                  | .81*    | .93     |
| Current student (ref: no)               | 5.41*** | 2.27*** |
| Education (ref: high school or below)   |         |         |
| College                                 | .24***  | .46***  |
| Bachelor or above                       | .20***  | .44***  |
| Household income (ref: low)             |         |         |
| Middle                                  | .93     | 1.12    |
| High                                    | 1.56    | 1.64*** |
| Topic (ref: Introduction statement)     |         |         |
| Demographics                            |         | .05***  |
| Housing                                 | .27***  |         |
| Clothing                                | .35***  |         |
| Utilities                               | .34***  |         |
| Insurance                               | .54***  |         |
| Income                                  | .07***  |         |
| Matrix question (ref: no)               |         | 1.33    |
| Open-ended question (ref: no)           | .87     |         |
| Question stem word count                | .98***  |         |
| Item nonresponse rate                   | 1.03*** |         |
| Grouped (ref: Interleafed)              | 1.15    |         |
| Mobile device (ref: non-mobile)         | 1.26**  |         |
| Multiple sessions (ref: one session)    | 1.08    |         |
| Survey duration (min)                   | .76***  |         |
| N of respondents                        | 3125    | 3125    |
| N of observations                       | 229,816 | 229,816 |
| Log likelihood                          | −3042.74| −2380.26|
| AIC                                     | 6109.48 | 4812.53 |

* p < .1, ** p < .05, *** p < .01.
education are attenuated but still significant. While the odds associated with ethnicity become insignificant, age and household income become positively related to breakoffs. The odds of breakoffs for an individual who is 10 years older are 10% higher. The odds of breakoff for respondents from the high household income group are 64% higher than that of peers from the low-income household.

Some of the estimates in Model 2 conform to expectations. The introduction statement gives the respondent a chance to re-evaluate whether they want to continue the survey and thus is expected to associate with more breakoffs. Indeed, compared to the introduction statement, the odds of breakoffs in other topics are lower. More interestingly, when facing sensitive topics about demographics and income, respondents are approximately 95% less likely to break off compared to the introduction statement. Additionally, item nonresponse rate and mobile device are positively associated with breakoffs as expected. For every unit increase in the item nonresponse rate, the breakoff odds increase by 3%, and mobile device users would have 26% higher odds of breakoffs. Also, the more time respondents spend in the questionnaire, the less likely they will break off.

In contrast to prior studies, questions with more words are found to be associated with fewer breakoffs. More specifically, each additional word in the question stem leads to a decrease of 2% in the breakoff risk.

Impact of question topic and filter question format on breakoff timing

Model 3 in Table 3 is the result of fitting Model 2 to the restricted sample. After the restriction, the number of remaining respondents reduces from 3128 to 2797, of whom 188 break off. As a result, the breakoff rate declines to 6.7%.

As before, the Introduction Statement is still associated with higher breakoff odds. However, when comparing to Clothing, Insurance has a higher breakoff risk. In total, there are 73 questions in the clothing block and 55 in the insurance block. The fewer questions in the insurance block and randomisation of question blocks together demonstrate that the insurance topic is genuinely associated with more breakoffs. The utilities block does not differ from the clothing block in terms of breakoff. The less-than-one odds ratio of open-ended questions in Model 4 is a surprising finding because nearly all prior studies documented that open-ended question is positively linked with survey breakoffs.

To investigate the change in time of the breakoff by question topics and filter formats, their interaction with time is included in Model 4. None of the interaction terms involving topics is significant. In contrast to Model 3, the model estimate of the grouped format on breakoffs in Model 4 becomes significant. The odds of breakoffs for respondents seeing the grouped format are only 19% of that of those answering the interleaved version. Yet, this difference varies by the number of questions respondents see. According to its interaction with the linear time, for every additional question, respondents of the grouped format see their breakoff odds increase by a factor of 8% compared to the interleaved format.

A more intuitive interpretation of the interaction effect of the grouped format and time (number of questions seen) can be seen in Figure 3 where the fitted hazard of breakoffs generated by Model 4 is plotted against time for both grouped and interleaved formats. When respondents see only a few questions, those receiving the interleaved format are more likely to break off. However, after approximately the 26th question, this trend is reversed; the grouped format starts to experience a higher breakoff risk. As respondents see more questions, the breakoff hazard between the two formats eventually converges. Again, the fluctuation in both curves after the 120th question is mainly due to the small denominator in the hazard calculation.
Table 3. Odds ratio of logistic regression predicting breakoff (based on the restricted sample).

| Variable                                                                 | Model 3   | Model 4   |
|-------------------------------------------------------------------------|-----------|-----------|
| Intercept                                                               | .12***    | .19*      |
| Number of questions seen (linear)                                       | .91***    | .89***    |
| Number of questions seen (quadratic)                                    | 1.00***   | 1.00***   |
| Married (ref: no)                                                       | 1.10      | 1.10      |
| Male (ref: female)                                                      | .97       | .97       |
| Age                                                                     | 1.00      | 1.00      |
| Non-white (ref: white)                                                  | .96       | .97       |
| Current student (ref: no)                                               | .94       | .94       |
| Education (ref: high school or below)                                   |           |           |
| College                                                                 | .96       | .96       |
| Bachelor or above                                                       | .76       | .77       |
| Household income (ref: low)                                             |           |           |
| Middle                                                                  | .96       | .96       |
| High                                                                    | 1.63**    | 1.63**    |
| Topic (ref: Clothing)                                                   |           |           |
| Utilities                                                               | 1.07      | .88       |
| Insurance                                                               | 1.74***   | 4.39*     |
| Introduction statement                                                  | 2.99***   | 4.34      |
| Open-ended question (ref: no)                                           | .42***    | .42***    |
| Question stem word count                                                | .98**     | .98**     |
| Item nonresponse rate                                                   | 1.00      | 1.00      |
| Grouped (ref: Interleafed)                                              | 1.19      | .19****   |
| Mobile device (ref: non-mobile)                                         | 1.38**    | 1.39**    |
| Multiple sessions (ref: one session)                                    | .95       | .95       |
| Survey duration (min)                                                   | .84***    | .85***    |
| Grouped × Questions seen (linear)                                       |           |           |
| Grouped × Questions seen (quadratic)                                    |           |           |
| Utilities × Questions seen (linear)                                     | 1.00**    | 1.00      |
| Utilities × Questions seen (quadratic)                                  | 1.00      | 1.00      |
| Insurance × Questions seen (linear)                                     | .96       | .96       |
| Insurance × Questions seen (quadratic)                                  | 1.00      | 1.00      |
| Introduction statement × Questions seen (linear)                        | .98       | .98       |
| Introduction statement × Questions seen (quadratic)                     | 1.00      | 1.00      |
| N of respondents                                                        | 2797      | 2797      |
| N of observations                                                       | 149,154   | 149,154   |
| Log likelihood                                                          | −1269.67  | −1262.49  |
| AIC                                                                     | 2583.35   | 2584.98   |

* p < .1, ** p < .05, *** p < .01

Discussion

The prevalence of the survey breakoff and the damage it can cause has led to a growing body of research into the factors causing it. This study extends this line of research by investigating two aspects of questions in particular: question topic and filter question format.
Our analysis finds two specific time points when breakoffs are more likely to happen (RQ 1). The first one is at the beginning of the survey. This finding is in accordance with previous research (Mittereder & West, 2021; Peytchev, 2009; Vehovar & Cehovin, 2014). The second timing is at the beginning of each question block where an introduction statement indicates a new set of questions. To further investigate when breakoffs are more likely to happen, two additional time-related variables are used in the analysis, namely, the number of questions respondents see and survey duration. We find that the more questions respondents see the less likely they will break off. However, we remain cautious about this finding because of the possible confounding with breakoffs.

In terms of the impact of different time-varying factors on breakoffs (RQ 2), some factors are in line with prior studies. Respondents who use a mobile device to answer the survey and have a higher item nonresponse rate are more likely to break off (Mittereder & West, 2021; Wenz, 2017).

However, some predictors affect the breakoff risk in an unexpected direction. The first is the negative relationship between word count in the question stem and breakoffs. We notice that the questions in our survey that have more words tend to be filter and follow-up questions. Most of the words in those questions are in fact repetitive. For example, every question about the price of different clothing items has the same instruction ‘Round to the nearest dollar. Please include any shipping and handling charges with the cost of any item that was shipped’. When facing the repetitive content, respondents might ignore them and only read the changing parts in the question. In comparison, for questions that are shorter but not repetitive, respondents have to read every word to understand it. This in fact makes questions with more words ‘shorter’ while questions with fewer words ‘longer’. Another surprising finding is that the breakoff risk for open-ended questions is lower than that of closed ones. We suspect that this is perhaps because our survey has a large number of open-ended questions (36% of the total questions are open-ended). The extensive use of open-ended questions might make respondents accustomed to this challenging question type.

The third research question (RQ 3) investigated whether the topic of the question impacts the breakoff and its timing. Compared to the topic of clothing, the insurance-related topic suffers from a higher breakoff risk while there is no difference in the breakoff risk between clothing and utilities.
Meanwhile, in comparison to the introduction statement, topics on income and demographics are relatively more sensitive but have a lower breakoff risk. Yet, the position of both blocks was not randomised, so the finding could be confounded by question order. Although some topics are associated with a higher/lower breakoff risk, we find that the relative difference between topics’ effects on breakoffs remains constant regardless of how many questions respondents have seen.

The final question (RQ 4) was whether the filter question format impacts the breakoff timing. We find that the grouped format can delay the breakoff but only until respondents realise the relationship between filter question and follow-ups and gain a sense of the extra burden.

The present study also has some limitations. Firstly, the web survey analysed here has a limited amount of paradata from the question level (e.g. question response time). Lacking such information prohibits a more detailed analysis on the process leading to breakoffs. Also, given that some respondents broke off at Block 1 and 2, there is a possibility that these early breakoff cases might differ from those reaching Block 3 (i.e. the first of the three randomised blocks). However, our analysis includes respondents’ demographic background, so we expect that this issue could be resolved to a large degree. Another limitation is that the survey analysed here is a non-probability survey and skewed towards female and white individuals, making the findings less applicable to the general population. Furthermore, the topics in the survey are not fully randomly ordered, so we can only test the effect of those randomised topics on breakoffs. Moreover, respondents answering the interleaved format might learn to reduce burden by deliberately under-reporting in the filter questions. In this case, they are not shown the follow-up questions and consequently break off less often at the later stage compared to the grouped counterpart. Future research is needed to answer whether under-reporting could explain the difference in the breakoff timing between grouped and interleaved formats. Lastly, researchers can also investigate whether our finding about filter questions still hold when all filter questions are presented in a matrix format (as opposed to showing them on separate pages).

In spite of these limitations, we believe findings in this paper will be useful to survey practitioners. For example, given the fact that a large number of breakoffs happen at the introduction statement, questionnaire designers should think about ways to keep respondents engaged (e.g. placing this type of statement on the same page with a substantive question or replacing this long statement with a short title about the topic). Meanwhile, findings about the insurance topic demonstrates that some question topics can impact breakoffs. Survey designers should place those topics towards the end of the questionnaire or give some motivations to the respondents in those topic blocks (e.g. emphasising the anonymity of the response). Additionally, the finding about the effect of filter question formats on breakoff timing is helpful for surveys that use filter questions extensively. For example, if the interest is in the prevalence of instances (e.g. purchase of different clothing items), the survey designer might prefer the grouped format as the postponing effect of this format would expose respondents to more filter questions. On the other hand, if the researcher cares more about the detail of the reported instance, it would be helpful to put the most important pair of filter and follow-up questions at the beginning of the interleaved format.

Declaration of conflicting interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) received no financial support for the research, authorship, and/or publication of this article.
Notes

1. We also fitted the continuous-time survival model using R’s survival package (Therneau, 2021), but both survival and logistic models gave the same result (not shown). We report the logistic regression in this study as it is the model used in many fields for discrete time events.

2. We also ran a model using Insurance as the reference level for the question topic and found that both Clothing and Utilities have a lower breakoff risk compared to Insurance (see Supplemental Appendix D).

References

Allen, M. (2017). The SAGE encyclopedia of communication research methods. Sage Publications. https://doi.org/10.4135/9781483381411

Blumenberg, C., Zugna, D., Popovic, M., Pizzi, C., Barros, A. J. D., & Richiardi, L. (2018). Questionnaire breakoff and item nonresponse in web-based questionnaires: multilevel analysis of person-level and item design factors in a birth cohort. Journal of Medical Internet Research, 20(12), Article e11046. https://doi.org/10.2196/11046

Clark-Fobia, A., Kephart, K., & Nelson, D. V. (2018). A qualitative study on the effects of grouped versus interleaved filter questions. Survey Practice, 11(2), 1–11. https://doi.org/10.29115/sp-2018-0009

Eckman, S., & Kreuter, F. (2018). Misreporting to looping questions in surveys: Recall, motivation and burden. Survey Research Methods, 12(1), 59–74. https://doi.org/10.18148/srm/2018.v12i1.7168

Eckman, S., Kreuter, F., Kirchner, A., Jäckle, A., Tourangeau, R., & Presser, S. (2014). Assessing the mechanisms of misreporting to filter questions in surveys. Public Opinion Quarterly, 78(3), 721–733. https://doi.org/10.1093/poq/nfu030

Enders, C. K. (2010). Applied missing data analysis (2nd ed.). Guilford Press.

Galesic, M. (2006). Dropouts on the web: Effects of interest and burden experienced during an online survey. Journal of Official Statistics, 22(2), 313–328.

Groves, R. M., Singer, E., & Corning, A. (2000). Leverage-saliency theory of survey participation: Description and an illustration. Public Opinion Quarterly, 64(3), 299–308. https://doi.org/10.1086/317990

Hoerger, M. (2010). Participant dropout as a function of survey length in internet-mediated university studies: Implications for study design and voluntary participation in psychological research. Cyberpsychology, Behavior and Social Networking, 13(6), 697–700. https://doi.org/10.1089/cyber.2009.0445

Klein, D. J., Elliott, M. N., Haviland, A. M., Saliba, D., Burkhart, Q., Edwards, C., & Zaslavsky, A. M. (2011). Understanding nonresponse to the 2007 Medicare CAHPS survey. Gerontologist, 51(6), 843–855. https://doi.org/10.1093/geront/gnr046

Kreuter, F. (2013). Improving surveys with paradata: Introduction. In F. Kreuter (Ed), Improving surveys with paradata: Analytic uses of process information (pp. 1–9). John Wiley & Sons. https://doi.org/10.1002/9781118596869.ch1

Kreuter, F., Eckman, S., & Tourangeau, R. (2020). The salience of survey burden and its effect on response behavior to skip questions: Experimental results from telephone and web surveys. In P. C. Beatty, D. Collins, L. Kaye, J.-L. Padilla, G. B. Willis, & A. Wilmot (Eds.), Advances in questionnaire design,
development, evaluation and testing (pp. 213–227). John Wiley & Sons. https://doi.org/10.1002/9781119263685.ch9

Kreuter, F., McCulloch, S., Presser, S., & Tourangeau, R. (2011). The effects of asking filter questions in interleaved versus grouped format. *Sociological Methods and Research, 40*(1), 88–104. https://doi.org/10.1177/0049124110392342

Kreuter, F., Presser, S., & Tourangeau, R. (2008). Social desirability bias in CATI, IVR, and web surveys: The effects of mode and question sensitivity. *Public Opinion Quarterly, 72*(5), 847–865. https://doi.org/10.1093/poq/nfn063

McGonagle, K. A. (2013). Survey breakoffs in a computer-assisted telephone interview. *Survey Research Methods, 7*(2), 79–90. https://doi.org/10.18148/srm/2013.v7i2.5126

Mittereder, F. K. (2019). *Predicting and preventing breakoff in web surveys*. University of Michigan.

Mittereder, F., &West, B. T. (2021). A dynamic survival modeling approach to the prediction of web survey breakoff. *Journal of Survey Statistics and Methodology*, 1-34. https://doi.org/10.1093/jssam/smab015.

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

Peytchev, A. (2011). Breakoff and unit nonresponse across web surveys. *Journal of Official Statistics, 27*(1), 33–47.

R Core Team (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Revilla, M. (2017). Analyzing survey characteristics, participation, and evaluation across 186 surveys in an online opt-in panel in Spain. *Methods, Data, Analyses, 11*(2), 135–162. https://doi.org/10.12758/mda.2017.02

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. John Wiley & Sons.

Sakshaug, J. W., & Crawford, S. D. (2010). The impact of textual messages of encouragement on web survey breakoffs: an experiment. *International Journal of Internet Science, 4*(1), 50–60.

Shropshire, K. O., Hawdon, J. E., & Witte, J. C. (2009). Web survey design: Balancing measurement, response, and topical interest. *Sociological Methods and Research, 37*(3), 344–370. https://doi.org/10.1177/0049124108327130

Silber, H., Lischewski, J., & Leibold, J. (2013). Comparing different types of web surveys: Examining dropouts, non-response and social desirability. *Metodoloski Zvezki, 10*(2), 121–143. https://doi.org/10.51936/luih3006

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford University Press.

Steinbrecher, M., Roßmann, J., & Blumenstiel, J. E. (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. https://doi.org/10.1093/ijpor/edu025

Teclaw, R., Price, M. C., & Osatuke, K. (2012). Demographic question placement: Effect on item response rates and means of a veterans health administration survey. *Journal of Business and Psychology, 27*(3), 281–290. https://doi.org/10.1007/s10869-011-9249-y

Therneau, M. T. (2021). A package for survival analysis in R. https://github.com/therneau/survival

Tijdens, K. (2014). Dropout rates and response times of an occupation search tree in a web survey. *Journal of Official Statistics, 30*(1), 23–43. https://doi.org/10.2478/jos-2014-0002

Tourangeau, R. (2018). The survey response process from a cognitive viewpoint. *Quality Assurance in Education, 26*(2), 169–181. https://doi.org/10.1108/qae-06-2017-0034

Tourangeau, R., Conrad, F. G., & Couper, M. P. (2013). *The science of web surveys*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199747047.001.0001
Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin, 133*(5), 859–883. https://doi.org/10.1037/0033-2909.133.5.859

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software, 45*(3), 1–67. https://doi.org/10.18637/jss.v045.i03

Vehovar, V., & Cehovin, G. (2014) Questionnaire length and breakoffs in web surveys: A meta study. 7th Internet Survey Methodology Workshop 2014, Bozen-Bolzano, Italy, 01 December, 2014

Villar, A., Callegaro, M., & Yang, Y. (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. https://doi.org/10.1177/0894439313497468

Wenz, A. (2017). *Sources of error in mobile survey data collection*. Institute for Social and Economic Research, University of Essex.

Willett, J. B., & Singer, J. D. (1993). It’s about time: Using discrete-time survival analysis to study duration and the timing of events. *Journal of Educational Statistics, 18*(2), 155–195.