Survey Research Methods (2016)
Vol. 10, No. 3, pp. 211-224
doi:10.18148/srm/2016.v10i3.6217

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

Are Final Comments in Web Survey Panels Associated with Next-Wave Attrition?

Cynthia McLauchlan
University of Waterloo
Waterloo, Canada

Matthias Schonlau
University of Waterloo
Waterloo, Canada

Near the end of a web survey respondents are often asked whether they have further comments. Such final comments are usually ignored, in part because open-ended questions are challenging to analyse. We explored whether final comments are associated with next-wave attrition in survey panels. We categorized a random sample of final comments in the Longitudinal Studies for the Social Sciences (LISS) panel and Dutch Immigrant panel into one of eight categories (neutral, positive, six subcategories of negative) and regressed the indicator of next-wave attrition on comment length, comment category and socio-demographic variables. In the Immigrant panel we found shorter final comments (< 30 words) are associated with increased next-wave attrition, and longer final comments (> 55 words) with decreased next-wave attrition relative to making no comment. Comments about unclear survey questions quadruple the odds of attrition and "other" (uncategorized) negative comments almost double the odds of attrition. In the LISS panel, making a comment (versus not) and comment length are not associated with attrition. However, when specifying individual comment categories, neutral comments are associated with half the odds of attrition relative to not making a comment.

Keywords:
Open-ended questions; final comments; next-wave attrition; web surveys; survey panels

1 Introduction

At the conclusion of surveys, respondents are often asked an open-ended question such as "Do you have any further comments?". We will refer to the respondents' answers to this open-ended question as "final comments". Final comments are often included in surveys, and the LISS and Dutch Immigrant panels routinely respond to questions posed in final comments to keep respondents engaged. However, final comments are generally not used in any apparent way in the analysis (Aldridge & Rowley, 1998; Bell & Tang, 1998; Hoekstra et al., 2011; Kingston, Carver, Evans, & Turton, 2000). When final comments are specifically mentioned in research, they often still directly relate to preceding subject matter questions rather than to the survey experience as a whole.

Final comments have been used after knowledge-based questions to ascertain the degree of certainty about the respondent's answer (Marshall, Mohammed, & Rouse, 2004). Belia, Fidler, Williams, and Cumming (2005) tested respondents on their understanding of confidence intervals and standard error bars and used final comments to gauge whether respondents who answered correctly were in fact not aware of the differences between the two concepts.

Similarly, final comments have been used after a multiple choice format to gain additional insight about a respondent's answer (Kendell & Pearce, 1997). For example, in an online survey regarding cyber bullying (Slonje & Smith, 2008) some final comments following multiple choice questions were included in the report. Such anecdotes were meant to convey some of the emotions participants shared in response to the multiple choice options provided.

When a respondent's desired answer is not permitted by the survey design, final comments provide participants with a chance to elaborate (T. Cook & Alexander, 2008; Stein et al., 2006; Webster, Merry, Gander, & Mann, 2004). This has been used by some researchers in explaining their survey results. While assessing the quality of life in narcoleptics, Daniels, King, Smith, and Shneerson (2001) found their survey respondents reported scenarios that were relevant to the research question, but not otherwise included in the survey, such as being unable to bathe while alone at home.

Similarly, final comments can be intentionally included in the analysis plan as a space for respondents to elaborate and not feel constrained only to the questions asked. In these cases, the final comments may need to be categorised and analysed separately. This approach was used by an education-based survey where only two questions were asked so that respondents would feel more inclined to elaborate on

Contact information: Cynthia McLauchlan, University of Waterloo, Bldg M3, 200 University Ave W, Waterloo, ON, Canada N2L 3G1 (email: cdmclauc@uwaterloo.ca)

whatever they felt was important using final comments (Martin & Edwards, 1998).

Final comments have been used to identify themes in the narrative (Brown et al., 2006; Hart & Macnee, 2007). For example, Hart and Macnee (2007) found four major themes related to a preceding question about improving nurse practitioner education. These four themes were reported in the study to provide a summary of respondents' opinions on the questionnaire topic.

Lastly, final comments can be used as an ad hoc means to understand a result that is contrary to what was expected. For example, Beekes (2006) expected a student evaluation method to improve students' confidence. However, after surveying students who used the evaluation method, the question asking whether the student's confidence was increased was met with a low rating. The final comments made by respondents explained that the evaluation method was useful for improving their understanding and thus indirectly improved their confidence, but the students did not perceive this association while completing the questionnaire.

Final comments have never, to our knowledge, been considered when investigating factors relevant to next-wave attrition. In this paper, we investigate whether final comments inform next-wave attrition in web surveys.

Two research questions guide this paper. First, is making a final comment, or the length of final comments associated with next-wave attrition? This is of interest since the length of final comments can be easily measured. Second, are specific comment categories associated with next-wave attrition above and beyond comment length? We would like to consider the content of final comments, rather than their length, to draw conclusions regarding the reasons respondents attrite.

The remainder of this paper is as follows: Section 2 gives a brief overview of the socio-demographic correlates of attrition in survey panels. Section 3 describes the data and methodology used. Section 4 gives the results. Section 5 concludes with an interpretation of the findings and important implications.

2 Socio-Demographic Correlates of Attrition in Panels

This paper investigates whether final comments are associated with attrition. Why do respondents attrite? Several authors (e.g. Laurie, Smith, and Scott, 1999; Lugtig, 2014) give four reasons for attrition: absence of commitment, habit to respond (leading to decreased attrition), panel fatigue, and shock. Shock may relate to a life event or a particularly unpleasant survey experience. This important theoretical distinction is hard to measure in practice (but see a structural equation-based approach by Lugtig, 2014). Most studies rely on socio-demographic correlates of attrition.

In addition to the question "Why do respondents attrite?" it is also important to note at what stage in the contact/response process they attrite: failure to locate the sample members, non-contact, or refusal (Lepkowski & Couper, 2002). This typology is more relevant to interviewer administered surveys (face-to-face or phone) than to self-administered surveys. Respondents of the LISS and Immigrant panels are contacted by email making refusal the most prominent type of attrition.

We now give a brief overview of this literature with a focus on panel surveys. Panels mentioned below include in alphabetical order: BHPS (Great Britain), ECHP (many European countries), HILDA (Australia), PSID (USA), LISS (Netherlands) and the SHP (Switzerland).

Gender: When respondents are contacted at home, many studies have found response rates to be higher for women than for men, in part because women are more likely to be at home (Watson & Wooden, 2009). This does not apply to Internet surveys with email as the contact mode.

Conditional on contact, there is no evidence of differential attrition by gender in the HILDA survey (Watson & Wooden, 2004, 2009), BHPS (Uhrig, 2008), and the LISS panel (De Vos, 2009). In the ECHP panel, men are a little more likely to attrite than women (Behr, Bellgardt, & Rendtel, 2005), but differences disappear conditional on contact (Nicoletti & Peracchi, 2005). Other studies find men attrite slightly more than women even conditional on contact (Lepkowski & Couper, 2002).

Age: Overall, younger respondents (e.g. 18-25) and elderly respondents are more likely to attrite than respondents in a middle age, though some panels only find increased attrition for one of these two groups. Increased attrition for both young and old are found in the HILDA panel (Watson & Wooden, 2009) and the LISS panel (De Vos, 2009). For the SHP and ECHP only the younger age is associated with attrition (Behr et al., 2005; Lipps, 2009) as did an earlier analysis for wave 2 of the HILDA panel (Watson & Wooden, 2004). In the PSID, older respondents are more likely to attrite (Fitzgerald, Gottschalk, & Moffitt, 1998). Conditional on contact, in the BHPS, older respondents are more likely to attrite (Uhrig, 2008).

Education: Attrition is generally concentrated among respondents with lower education (De Vos, 2009; Fitzgerald et al., 1998; Uhrig, 2008; Watson & Wooden, 2004, 2009). For the ECHP the results are mixed (Behr et al., 2005) and, when conditioning on contact, no differences are found (Nicoletti & Peracchi, 2005).

Income: The evidence for income is inconclusive. The HILDA panel finds no correlation between income and attrition (Watson & Wooden, 2009). In the PSID, attriters tend to be poorer (Fitzgerald et al., 1998). For the BHPS, higher income is correlated with attrition, but this effect is due to non-contact and not refusal (Uhrig, 2008). In the ECHP, evidence is mixed (Behr et al., 2005) but when conditional on contact, a separate analysis finds no differences (Nicoletti &

Peracchi, 2005).

Marital status: Not being married is associated with greater attrition in the PSID, HILDA, BHPS and ECHP (Fitzgerald et al., 1998; Lugtig, 2014; Uhrig, 2008; Watson & Wooden, 2009). Conditional on contact, this effect disappears in the BHPS (Uhrig, 2008) but not in the ECHP (Nicoletti & Peracchi, 2005). By contrast, in the LISS panel, being single is associated with lower attrition (De Vos, 2009).

Ethnicity: The influence of ethnicity on attrition is often conflated with whether the respondent speaks the relevant language well (Lugtig, 2014). These two issues are disentangled in the HILDA panel. There are no differences in attrition for Australian-born respondents and respondents born in other English speaking countries. Attrition does increase for respondents born in non-English speaking countries and is worse if respondents do not speak English well (Watson & Wooden, 2009). In the BHPS, non-white is associated with increased attrition, but this difference disappears conditional on contact (Uhrig, 2008). In the LISS panel, differences between households with and without migrants are minor (De Vos, 2009). For the SHP, respondents from ethnic minorities attrite more often (Lipps, 2009). In PSID panel, being non-white is associated with greater attrition (Fitzgerald et al., 1998).

3 Methods

We first describe the two survey panels used throughout the analysis. We then explain the methods used in addressing the two research questions followed by an explanation of how final comments are categorized for our second research question. Finally, we describe two modeling approaches: GEE logistic regression, which accounts for within-respondent correlation, and SIMEX, a method that accounts for measurement error from different ratings of comment categories.

3.1 The LISS and Dutch Immigrant survey panels

We use the Longitudinal Internet Studies for the Social sciences (LISS) panel, and Dutch Immigrant panel (CentERdata, n. d.). These probability-based survey panels were sampled from the Statistics Netherlands' population register. Panel members are paid proportional to an expected survey completion time of either 15 or 30 minutes per survey. The LISS panel was created in 2007 with around 8000 members in the panel. The Immigrant panel was created in 2010 with around 2400 panel members, most of whom (1700) are non-Dutch.

For each panel, at least one online survey was sent each month to every panel member and additional surveys were sent to subsets of the panel. Since information about who was invited to which survey was not available, we restricted our analysis to the largest monthly survey sent to all panel members. In the LISS panel, we analyzed data from November 2007 until October 2013. Over this time, 378,505 responses were made to 71 surveys. In the Immigrant panel, we analyzed data from March 2010 until October 2013. Over this time, 55,841 responses were made to 44 surveys.

A respondent was defined to have next-wave attrition for any given month if they failed to complete the following month's survey sent to all respondents. The final month of the dataset was only used to define next-wave attrition on the previous month. Next-wave attrition is used instead of permanent attrition because any association between final comments and attrition may be more pronounced in the following month rather than several months later.

In each survey, a final comment section was included at the survey's conclusion asking the following open-ended question: "Do you have any remarks about the questionnaire?" (in Dutch: "Hebt u nog opmerkingen over deze vragenlijst?" This question could be answered with yes or no. If the respondent chose yes, he/she could then write a comment. There are no differences in terms of how the question was asked in either panel.

3.2 Research questions

To answer the first research question of whether final comment length is associated with next-wave attrition, we utilise a robust (Huber, 1976; White, 1982) GEE logistic regression model to regress next-wave attrition on covariates. Covariates include an indicator variable for whether or not a comment was made, comment length (number of words), and socio-demographic variables. To avoid undue influence of very long final comments in the regression analyses, final comment lengths were capped at the 95th percentile. This corresponds to 74 words for the LISS panel, and 88 words for the Immigrant panel.

To answer the second research question of whether final comment categories are associated with next-wave attrition when already adjusting for comment length, we regress next-wave attrition on covariates: indicator variables for each of eight comment categories, the comment length, and socio-demographic variables. The SIMEX method is used for this regression to account for measurement error in the rating of comment categories.

The socio-demographic variables included in all models are: gender, age (grouped in 10 year spans), marital status (married, single, divorced/separated/widowed), monthly net income, and highest level of education achieved. The demographic indicator variables for education are based on the Dutch school acronyms: primary school and VMBO comprise Secondary Education or Less, HAVO and VWO comprise Selective Secondary Education, MBO comprises Vocational Education, and HBO and WO comprise University or more.

3.3 Categorization and length of final comments

The second research question concerns the association between different categories of final comments and next-wave attrition.

The final comments were categorized into nine mutually exclusive categories: trivial comments, positive comments, neutral comments and six categories for negative comments (Schonlau, 2015). Positive comments ("positive") include comments such as "this is interesting" and "it made me think about the topic". Neutral comments ("neutral") include comments related to the survey topic, and personal information such as "I was on vacation". Six categories are negative in nature: comments regarding whether the respondent perceived any part of the survey to be difficult ("difficult question"), mentioning an html technical error ("technical error"), complaints regarding the length of the survey ("too long"), comments on how the survey in whole or part was not applicable to the respondent ("not applicable"), and comments regarding a question being unclear ("question unclear"). All other negative comments were contained within a sixth category ("other negative"). When respondents complained about the survey being too long, this was usually in reference to the promised length/payment ratio. Additionally, a category was created for trivial comments such as "I have no comment" or "aaah". Figure 1 illustrates the percentage of final comments contained within each of the nine categories.

All observations with a trivial comment are hereafter fully omitted from the analysis due to the infrequent occurrence of trivial comments and the lack of substance as a final comment. This effectively results in eight mutually exclusive comment categories.

For the LISS panel, a random sample of 1,250 final comments were categorised by two individuals. Of these final comments, 450 were categorised by both individuals as well as an expert with an inter-rater agreement of Kappa = 0.48. The Kappa coefficient of agreement accounts for agreement by chance (Cohen, 1968). Our observed Kappa value corresponds to moderate agreement between the three sets of classifications (Landis & Koch, 1977). For the Immigrant panel, 850 final comments were categorised by one individual, 450 of which were also categorised by a second individual. The inter-rater agreement Kappa was 0.50. This Kappa value also corresponds to moderate agreement between the two sets. While moderate values of kappa were observed, the regression analyses based on the SIMEX method took these measurement errors into account.

Since not all final comments were classified, and those that were classified were randomly selected, we eliminated any observations that included an unclassified final comment from all further analysis to create a consistent dataset across both research questions. This resulted in a dataset of 378,505 responses for the LISS panel from 11,959 panel members. Each panel member responded to at least 1 survey, and at most 70 surveys, with a mean of 31.65 surveys per panel member. In the Dutch Immigrant panel, the dataset consists of 55,841 responses from 2,867 different panel members. Each panel member responded to at least 1 survey, and at most 40 surveys, with an average of 19.48 surveys.

3.4 Regression Models: GEE and SIMEX

Here the use of standard logistic regression has two drawbacks: (1) observations are assumed to be independent, and (2) the variance is a function of the mean and cannot accommodate overdispersion (i.e. when more variance is observed in the data than would be expected from the model).

Generalized estimating equations (GEEs) solve both these issues. First, GEEs model correlation between repeated measurements from the same subject (Kohler & Kreuter, 2005; Liang & Zeger, 1986; Zeger & Liang, 1986). Since panel members are surveyed repeatedly, multiple observations (surveys) from the same respondent may be correlated. We assume an exchangeable correlation structure for within-respondent correlation and still assume independence among observations from different respondents. Second, GEEs allow the mean and variance to be specified independently through a multiplicative overdispersion parameter. For example, when more variance is observed than would be expected by a logistic regression model, the logistic regression model (with underestimated variance) may lead, incorrectly, to significant findings. The GEE logistic regression addresses this problem by introducing an overdispersion parameter that allows for larger variances where appropriate. In GEE models, standard errors are estimated consistently if the model (including the within-respondent correlation matrix) is specified correctly. It is common practice to use robust (Huber, 1976; White, 1982) or sandwich estimators instead of the "naïve" standard errors that do not require a correct working correlation structure.

Models that do not account for measurement error require a single categorization. However, multiple raters may categorize comments differently. Analysis methods can account for such measurement error when repeated measurements are available. One popular approach for accounting for measurement error is the simulation extrapolation (SIMEX) method (Carroll, Küchenhoff, Lombard, & Stefanski, 1996; J. Cook & Stefanski, 1994; Stefanski & J. Cook, 1995). The SIMEX method simulates different versions of the data, each time adding varying amounts of measurement error. A regression is conducted for each version of the data, establishing a trend in the estimated coefficients of the error-prone variable as a function of measurement error. The trend can be extrapolated back to the case of no measurement error. The bootstrap method (Efron, 1979) can then be used to calculate the standard errors of these estimates. In this particular case, each final comment is associated with either two or three raters' categorisation. The variability between the ratings reflects

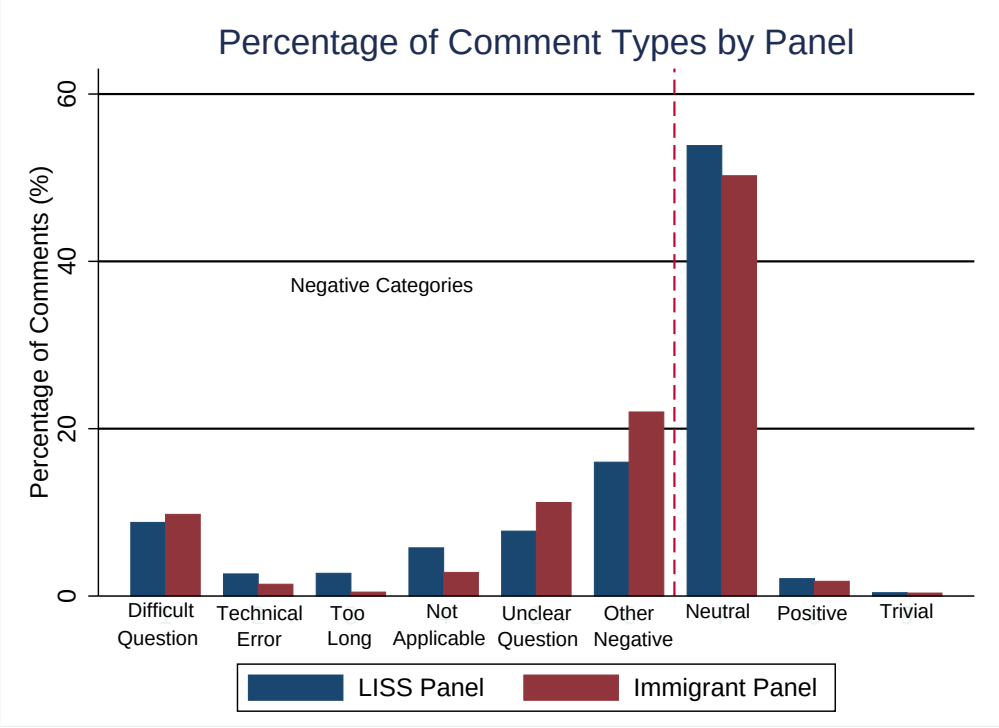

Figure 1. Percentage of final comments contained within each category by survey panel. The six categories to the left are negative in nature, the 3 remaining categories to the right are neutral, positive, and trivial comments.

the measurement error. Consequently, the estimated standard errors of the coefficients in the regression model tend to increase. This appropriately reflects the increased variability due to measurement error.

4 Results

Before investigating the associations between final comments and next-wave attrition we give some descriptive statistics. In the LISS panel, 3.7% of responses contained final comments with a comment length ranging from 1 to 196 words, a mean of 30.11 and a median of 24 words. On average, each respondent completed 31.65 surveys. In the Immigrant panel, 5.5% of responses contained final comments with a comment length ranging from 1 to 299 words, a mean of 33.24 and a median of 25 words. On average, each respondent completed 19.48 surveys. For the purposes of answering the second research question, 1,250 randomly selected final comments were manually categorised in the LISS panel, and 850 were categorised in the Immigrant panel.

Because we consider next-wave attrition rather than permanent attrition, the same individual can next-wave attrite and return into the panel at a later time. The LISS panel had a next-wave attrition rate of 13.3%. On average, a panel member next-wave attrited 4.22 times. In total, 50,410 next-wave attritions were observed. The Immigrant panel had a next-wave attrition rate of 21.4%. On average, a panel member next-wave attrited 4.17 times. In total, 11,953 next-wave attritions were observed.

Table 1 provides some descriptive statistics of the socio-demographic variables used in the analysis. Overall, the two panels have similar demographic representations. As compared to the LISS panel, the immigrant panel contains more panel members with less than a secondary education (36.7% vs. 27.8%) and fewer married panel members (58.3% vs. 51.9%).

Looking at the first research question of whether final comment length is associated with next-wave attrition, we fit a robust GEE logistic regression model with next-wave attrition as the dependent variable in both the LISS panel and Dutch Immigrant panel. Table 2 gives the results of this regression for the LISS panel, and Table 3 for the Immigrant panel.

The coefficients in Tables 2 and 3 represent the log odds ratio of next-wave attrition for the covariate level versus the reference level. Usually the exponentiated coefficient – an odds ratio – is interpreted rather than the coefficient itself.

In the LISS panel, neither making a final comment nor the comment length is associated with next-wave attrition. The indicator variable for making a final comment was still not significant if the variable for final comment length is removed from the model.

In the Immigrant panel, after adjusting for covariates, the odds of next-wave attrition for those who made a comment of

Table 1
Socio-demographic characteristics of respondents. Due to rounding, percentages may not add to 100%

	Percentage of respondents	
	LISS panel	Immigrant Panel
Gender		
Female	53.7	54.0
Male	46.3	46.0
Age		
Aged 15-24	11.3	11.4
Aged 25-34	12.2	15.6
Aged 35-44	17.2	22.8
Aged 45-54	19.3	20.0
Aged 55-64	21.1	17.0
Aged 65 or older	19.0	13.3
Marital Status		
Married	58.3	51.9
Divorced/Separated/Widowed	13.5	16.0
Single	28.2	32.1
Monthly Net Income		
No income	11.4	12.0
EUR 500 or less	7.5	7.0
EUR 501-1000	17.8	16.3
EUR 1001-1500	20.4	20.6
EUR 1501-2000	21.5	21.5
EUR 2001-2500	11.1	10.5
EUR 2501-3000	5.3	5.8
EUR 3001-3500	2.4	2.7
EUR 3001 or more	2.6	3.6
Highest Level of Education		
Secondary Education or Less	36.7	27.8
Selective Secondary Education	11.0	13.7
Intermediate Vocational	22.7	22.5
University	29.6	35.9

median length (25 words) increase by a factor of 1.15 relative to those who do not make a comment ($p < 0.001$):

$$1.15 = e^{0.64-0.02\cdot 25} \quad (1)$$

To put this into context, we translate the change in odds to a change in probabilities for a respondent. A respondent with all values set to their reference levels (45-54 year old, married female respondent with a monthly income of 1501-2000 Euros and selective secondary education) who did not leave a comment has a probability of 0.22 for next-wave attrition. The probability of attrition for the same respondent with a 25-word comment increases to 0.27.

For shorter comments, the positive coefficient for making a final comment seen in Table 3 suggests that making a final comment results in higher attrition relative to not making a final comment as with equation (1). But for longer final comments, the negative coefficient for comment length outweighs the positive coefficient for making the comment, and results in lower attrition relative to not making a final comment. This can be seen from equation (2) for a comment with 60 words.

$$0.57 = e^{0.64-0.02\cdot 60} \quad (2)$$

The probability of attrition for the above respondent (with values set to their reference levels) with a 60-word comment decreases to 0.17.

A final comment with 56 or more words is significantly associated with decreased next-wave attrition relative to not making a comment, while making a final comment that is 29 words or shorter is significantly associated with increased

Table 2
The association between final comment length and next-wave attrition in the LISS panel with sample size of 378,505 from 11,959 panel members. Results are reported using a GEE logistic regression model with robust standard errors.

	Coef.	exp(coeff)	S.E.
Comment Indicator			
No final comment made^{+}	-	-	-
Left final comment	−0.11	0.89	0.12
Comment Length			
Number of words	0.00	1.00	0.00
Gender			
Female^{+}	-	-	-
Male	−0.02	0.98	0.02
Age			
Aged 15-24	0.24***	1.27	0.04
Aged 25-34	0.25***	1.28	0.03
Aged 35-44	0.18***	1.19	0.03
Aged 45-54^{+}	-	-	-
Aged 55-64	−0.33***	0.72	0.03
Aged 65 or older	−0.47***	0.62	0.03
Marital Status			
Married^{+}	-	-	-
Divorced/Separated/Widowed	−0.01	0.99	0.03
Single	0.05	1.05	0.03
Monthly Net Income			
No income	−0.11***	0.89	0.03
EUR 500 or less	−0.19***	0.83	0.04
EUR 501-1000	−0.05***	0.95	0.03
EUR 1001-1500	−0.08***	0.92	0.02
EUR 1501-2000^{+}	-	-	-
EUR 2001-2500	0.06***	1.06	0.03
EUR 2501-3000	0.00	1.00	0.03
EUR 3001-3500	0.13**	1.13	0.05
EUR 3001 or more	0.14**	1.15	0.05
Highest Level of Education			
Secondary Education or Less	−0.06*	0.94	0.03
Selective Secondary Education^{+}	-	-	-
Intermediate Vocational	0.04	1.04	0.03
University	0.06*	1.06	0.03
Constant	−1.47***	0.23	0.04

$^{+}$ reference level * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

next-wave attrition. These boundaries were computed based on hypothesis tests about linear combinations from the model (Table 3): For example, for a comment with 29 words (or fewer words) we could reject the corresponding hypothesis $H_0 : (\beta_1 + 29\beta_2) = 0$ where β_1 is the coefficient for the comment indicator and β_2 the coefficient for the number of words. In the Immigrant panel, 29 words corresponds to the 57$^{\text{th}}$ percentile for comment length, and 56 words corresponds to the 85$^{\text{th}}$ percentile.

The GEE logistic model (Table 3) showed that longer comments are associated with decreased next-wave attrition, whereas shorter comments are associated with increased

Table 3
The association between final comment length and next-wave attrition in the Dutch Immigrant panel with sample size of 55,841 from 2,867 panel members. Results are reported using a GEE logistic regression model with robust standard errors.

	Coef.	exp(coeff)	S.E.
Comment Indicator			
No final comment made^{+}	-	-	-
Left final comment	0.64^{***}	1.90	0.14
Comment Length			
Number of words	-0.02^{***}	0.98	0.00
Gender			
Female^{+}	-	-	-
Male	0.08^{*}	1.08	0.04
Age			
Aged 15-24	0.35^{***}	1.42	0.07
Aged 25-34	0.36^{***}	1.43	0.06
Aged 35-44	0.19^{***}	1.21	0.05
Aged 45-54^{+}	-	-	-
Aged 55-64	-0.32^{***}	0.73	0.05
Aged 65 or older	-0.52^{***}	0.59	0.06
Marital Status			
Married $^{+}$	-	-	-
Divorced/Separated/Widowed	0.07	1.07	0.05
Single	−0.02	0.98	0.04
Monthly Net Income			
No income	0.02	1.02	0.06
EUR 500 or less	0.07	1.08	0.07
EUR 501-1000	0.04	1.04	0.05
EUR 1001-1500	0.05	1.05	0.04
EUR 1501-2000 $^{+}$	-	-	-
EUR 2001-2500	0.01	1.01	0.06
EUR 2501-3000	−0.07	0.94	0.08
EUR 3001-3500	0.04	1.04	0.09
EUR 3001 or more	0.10	1.11	0.09
Highest Level of Education			
Secondary Education or Less	0.08	1.08	0.05
Selective Secondary Education $^{+}$	-	-	-
Intermediate Vocational	0.04	1.04	0.06
University	0.02	1.02	0.05
Constant	-1.25^{***}	0.29	0.07

$^{+}$ reference level * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

next-wave attrition. We next checked this finding graphically from the data directly to confirm the finding was real. Because next-wave attrition has only two values (attrite vs. not attrite), we used a lowess smoother (Cleveland, 1981) to plot next-wave attrition as a function of final comment length (Figure 2). Figure 2 confirms that relative to attrition among responses without a comment (horizontal line), shorter comments are associated with increased attrition, and longer comments with decreased attrition. Comments of middling lengths are not associated with significantly different attrition.

Adjusted for socio-demographic variables, both regres-

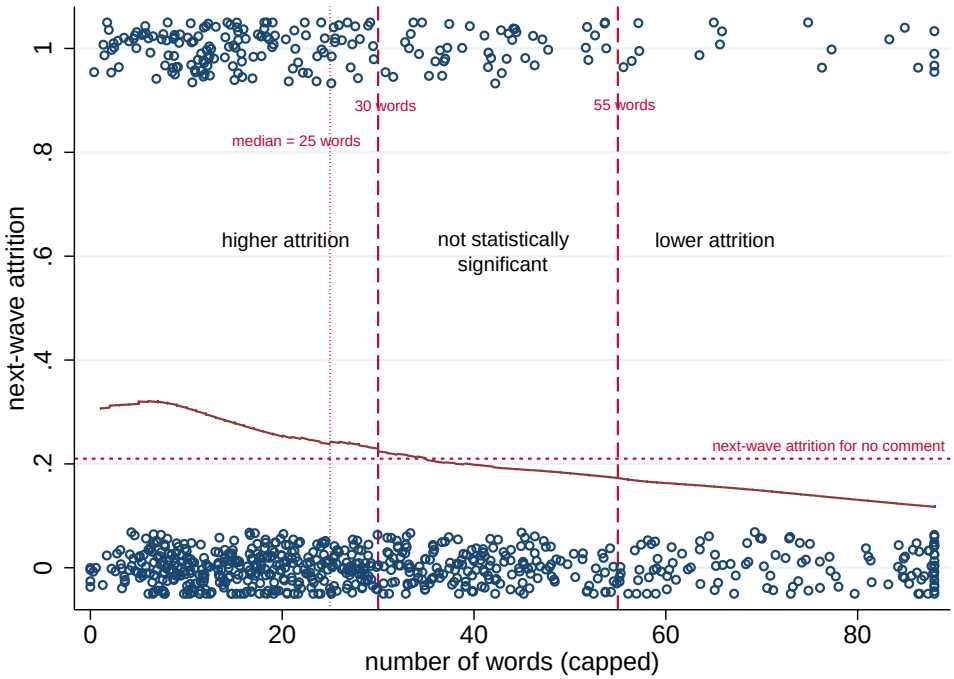

Figure 2. Lowess plot of the Immigrant panel's next-wave attrition as a function of final comment length. Reference lines are provided for the next-wave attrition of respondents who do not make a final comment, and the model-based boundaries of final comment lengths that are significantly associated with next-wave attrition relative to not making a comment. The median length of a comment is 25 words.

sions find next-wave attrition is associated with older age. Both regressions also find that marital status is not associated with attrition. In the LISS panel, gender is not significantly associated with next-wave attrition, but a higher monthly net income and level of education is associated with increased next-wave attrition. In the Immigrant panel, male gender is associated with increased next-wave attrition, while net income and education are not associated with attrition.

We next look at the second research question: is the content of a final comment associated with next-wave attrition above and beyond comment length? Instead of a single indicator variable for presence of a comment, now indicator variables for individual comment categories (relative to not making any comment) are used. We accounted for the measurement error in comment categorization using the SIMEX method. We found no significant interactions. Table 4 gives the results of this regression for the LISS panel, and Table 5 for the Immigrant panel.

In the LISS panel, comment length is again not associated with next-wave attrition. Adjusted for covariates and measurement error, the odds of next-wave attrition for those who make a neutral comment of median length (24 words) are about half the odds of next-wave attrition of those who do not make a comment ($p = 0.04$):

$$0.52 = e^{-0.88+0.0096 \cdot 24} \tag{3}$$

To put this in context, the probability of next-wave attrition for a respondent with values set to their reference levels and who makes no comment is 0.13. If that same respondent instead makes a neutral comment with 24 words, the probability of next-wave attrition decreases to 0.08.

In the Immigrant panel, longer final comments are again associated with decreased next-wave attrition, while several comment categories are found to associate with increased next-wave attrition after adjusting for comment length. Specifically, the odds of next-wave attrition of those who made a final comment about "unclear survey questions" of median-length (25 words) are 4.30 times the odds of next-wave attrition of those who do not make a comment ($p < 0.001$):

$$4.30 = e^{1.68-0.0089 \cdot 25} \tag{4}$$

In other words, the probability of next-wave attrition for a respondent with values set to their reference levels increases from 0.19 (no comment) to 0.49 (25-word comment about unclear questions).

Also, the odds of next-wave attrition of those who make an "other negative comment" are 1.75 times the odds of

Table 4
The association between final comment category and next-wave attrition in the LISS panel with sample size of 378,505 from 11,959 panel members. Results reported are based on the SIMEX method with logistic regression and robust standard errors.

	Coef.	S.E.
Comment Length		
Number of words	0.01	0.01
Comment Category		
No comment made [+]	-	-
Difficult question	−0.07	0.66
Technical error	−0.42	6.03
Too long	−0.34	9.13
Not applicable	1.49	1.33
Question unclear	−1.21	1.10
Other negative	−0.40	0.62
Positive	0.10	2.09
Neutral	−0.88*	0.42
Gender		
Female[+]	-	-
Male	−0.03**	0.01
Age		
Aged 15-24	0.43***	0.02
Aged 25-34	0.42***	0.02
Aged 35-44	0.25***	0.02
Aged 45-54[+]	-	-
Aged 55-64	−0.41***	0.02
Aged 65 or older	−0.61***	0.02
Marital Status		
Married[+]	-	-
Divorced/Separated/ Widowed	0.05**	0.02
Single	−0.03	0.01
Monthly Net Income		
No income	−0.07***	0.02
EUR 500 or less	−0.13***	0.02
EUR 501-1000	−0.03	0.02
EUR 1001-1500	−0.04**	0.01
EUR 1501-2000[+]	-	-
EUR 2001-2500	0.05**	0.02
EUR 2501-3000	0.03	0.03
EUR 3001-3500	0.18***	0.03
EUR 3001 or more	0.25***	0.03
Highest Level of Education		
Secondary Education or Less	0.01	0.02
Selective Secondary Education[+]	-	-
Intermediate Vocational	0.03	0.02
University	0.04*	0.02
Constant	−1.86***	0.02

[+] reference level * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Table 5
The association between final comment category and next-wave attrition in the Dutch Immigrant panel with sample size of 55,841 from 2,867 panel members. Results reported are based on the SIMEX method with logistic regression and robust standard errors.

	Coef.	S.E.
Comment Length		
Number of words	−0.01	0.01
Comment Category		
No comment made^{+}	-	-
Difficult question	0.40	0.59
Technical error	−0.72	7.86
Too long	1.91	18.81
Not applicable	0.20	1.36
Question unclear	1.68***	0.42
Other negative	0.78*	0.37
Positive	1.82	1.30
Neutral	0.08	0.38
Gender		
Female^{+}	-	-
Male	0.06*	0.02
Age		
Aged 15-24	0.38***	0.04
Aged 25-34	0.34***	0.04
Aged 35-44	0.14***	0.03
Aged 45-54^{+}	-	-
Aged 55-64	−0.30***	0.04
Aged 65 or older	−0.47***	0.04
Marital Status		
Married^{+}	-	-
Divorced/Separated/ Widowed	0.06	0.03
Single	−0.04	0.03
Monthly Net Income		
No income	0.01	0.04
EUR 500 or less	0.02	0.05
EUR 501-1000	0.04	0.03
EUR 1001-1500	0.09**	0.03
EUR 1501-2000^{+}	-	-
EUR 2001-2500	0.05	0.04
EUR 2501-3000	0.00	0.05
EUR 3001-3500	0.02	0.07
EUR 3001 or more	0.14*	0.06
Highest Level of Education		
Secondary Education or Less	0.08*	0.04
Selective Secondary Education^{+}	-	-
Intermediate Vocational	0.03	0.04
University	0.03	0.04
Constant	−1.45***	0.05

$^{+}$ reference level * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

next-wave attrition as those who do not make a comment ($p = 0.04$):

$$1.75 = e^{0.78-0.0089\cdot 25} \quad (5)$$

Likewise, the probability of next-wave attrition for a respondent with values set to their reference levels increases from 0.19 (no comment) to 0.28 (25-word "other negative comment").

5 Discussion

We found that (1) shorter final comments are associated with increased next-wave attrition relative to making no final comment while longer final comments have the opposite association in the Immigrant panel only, and (2) different comment categories are differentially associated with next-wave attrition after accounting for comment length in both panels.

Our first research question regarding the association between comment length and next-wave attrition contains a significant finding in the Immigrant panel. Longer final comments (> 55 words) are associated with reduced next-wave attrition, and shorter final comments (< 30 words) are associated with increased next-wave attrition relative to not making a comment. In the LISS panel, comment length is not significantly associated with next-wave attrition. Longer comments may indicate a greater level of engagement than those who make no final comment, and engaged respondents may be less likely to attrite.

Our second research question regarding the association between comment category and next-wave attrition contains significant findings in both panels. In the Immigrant panel, respondents making a final comment regarding survey questions being unclear have over four times the odds of next-wave attrition as those not making a comment ($p < 0.001$), and those making an "other negative" comment (other than difficult questions, technical error, survey too long, questions do not apply to respondent, or unclear questions) have nearly twice the odds of next-wave attrition as those making no final comment ($p = 0.04$). These estimates assume the final comment is of the median length of 25 words and incorporate the measurement error in categorising comments. It makes intuitive sense that respondents who find questions unclear and have negative things to comment on are more likely to attrite.

In the LISS panel, respondents making a neutral comment have half the odds of next-wave attrition as those making no comment ($p = 0.04$). Neutral comments include thoughts about the survey topic or personal information which all suggest some level of engagement with the survey process. Further, respondents are taking the time to make a non-negative comment. Because there were very few positive comments, it is not surprising that "positive comment" was not significantly associated with next-wave attrition. Our finding may therefore suggest that engaged respondents are less likely to attrite.

For the second research question the findings, while different in each panel, are qualitatively similar: Because there are few positive comments, the comment categorization is effectively a contrast between neutral comments and different categories of negative comments. In the LISS panel we find decreased odds of next-wave attrition for neutral comments; in the Immigrant panel we find increased odds of next-wave attrition for the two largest negative comment categories (see Figure 1). Both these findings suggest that (most) negative comments lead to increased next-wave attrition relative to neutral comments.

Like all research, our research has some limitations. First, the observed Kappa values in the categorization for research question 2 are moderate and would ideally be higher. While we were unable to achieve higher reliability, we accounted for measurement error using the SIMEX method. This increases the variance, making results potentially less significant, but is preferable to reporting false significances. Most analyses with a higher reliability fail to take measurement error into account which may lead to false significances.

Finally, results reported here are from the LISS panel and Dutch Immigrant panel. While it is always a question how results will generalize to other panels, these two probability-based panels are as good a source of data as one might hope for. The Immigrant population is a special type of population and it is not too surprising that results for the two panels differ. Note that respondents in the Immigrant panel leave 50% more final comments than those in the LISS panel (5.5% versus 3.7%) and have a higher rate of next-wave attrition (21.4% versus 13.3%). However, the median comment length for the Immigrant panel and the LISS panel are similar (25 versus 24 words). The distribution of response categories is also very similar for both panels (Figure 1). The two panels also differ in that the older LISS panel has more waves than the Immigrant panel. However, analogous analyses based on the first 40 waves of the LISS panel did not yield results similar to those seen in the Immigrant panel. Respondents in the Immigrant panel may take the final comment question more seriously than respondents in the LISS panel, as suggested by a higher rate of making final comments. For recent immigrants, language skills may also play a role in that any frustration about questions perceived to be unclear might have been amplified. Unfortunately, we had no direct measure of language ability.

What are the implications for the operation of survey panels? For immigrant populations, panels should pay increased attention to question wording to avoid respondents perceiving questions as unclear. More generally, panels should strive to avoid respondents' negative experiences while they interact with the survey. For example, additional pretesting of surveys might be desirable for immigrant populations.

Panels might also want to identify respondents at high risk of attrition. Presence of a final comment and comment length are clearly more easily measured than categorizing a final comment into categories. For immigrants, short final comments are a strong indicator of next-wave attrition. For immigrants, final comments could also be scanned for synonyms of the word "unclear" in Dutch.

In the LISS panel, making a neutral comment is associated with reduced next-wave attrition. Neutral comments include questions for the survey panel, which the LISS panel responds to already. Interpreting this as a form of listening to and engaging panel members, panels should continue to engage respondents in many ways. In that sense, it lends indirect support to existing panel efforts to engage respondents by responding to questions asked in the final comments, individualized communications (well-wishes to respondents who are sick for a prolonged period, congratulations to marriage or the birth of a child, Christmas cards for respondents that send Christmas cards to LISS), sending newsletters, and small gifts or photos of the panel management team.

Acknowledgement

This research was supported by the Social Sciences and Humanities Research Council (SSHRC, grant number 430-2013-0301). This paper uses data from the LISS (Longitudinal Internet Studies for the Social sciences) and Dutch Immigrant panels administered by CentERdata (Tilburg University, The Netherlands). We wish to thank Marcel Das and Joris Mulder in particular for their support with the data. We also thank the anonymous referees for their comments.

References

Aldridge, S. & Rowley, J. (1998). Measuring customer satisfaction in higher education. *Quality Assurance in Education*, *6*(4), 197–204.

Beekes, W. (2006). The 'millionaire' method for encouraging participation. *Active Learning in Higher Education*, *7*(1), 25–36.

Behr, A., Bellgardt, E., & Rendtel, U. (2005). Extent and determinants of panel attrition in the European Community Household Panel. *European Sociological Review*, *21*(5), 489–512.

Belia, S., Fidler, F., Williams, J., & Cumming, G. (2005). Researchers misunderstand confidence intervals and standard error bars. *Psychological Methods*, *10*(4), 389–396.

Bell, H. & Tang, N. (1998). The effectiveness of commercial internet web sites: a user's perspective. *Internet Research*, *8*(3), 219–228.

Brown, A., O'Connor, P., Roberts, T., Wakefield, R., Karim, Z., & Emery, P. (2006). Ultrasonography for rheumatologists: the development of specific competency based educational outcomes. *Annals of the Rheumatic Diseases*, *65*(5), 629–636.

Carroll, R. J., Küchenhoff, H., Lombard, F., & Stefanski, L. A. (1996). Asymptotics for the SIMEX estimator in nonlinear measurement error models. *Journal of the American Statistical Association*, *91*(433), 242–250.

CentERdata. (n. d.). LISS Data: General. Retrieved from http://www.lissdata.nl/lissdata/

Cleveland, W. (1981). Lowess: a program for smoothing scatterplots by robust locally weighted regression. *American Statistician*, *35*(1), 54.

Cohen, J. (1968). Weighted kappa: nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulleti*, *70*(4), 213.

Cook, J. & Stefanski, L. (1994). Simulation-extrapolation estimation in parametric measurement error models. *Journal of the American Statistical Association*, *89*(428), 1314–1328.

Cook, T. & Alexander, R. (2008). Major complications during anaesthesia for elective laryngeal surgery in the UK: a national survey of the use of high-pressure source ventilation. *British Journal of Anaesthesia*, *101*(2), 266–272.

Daniels, E., King, M., Smith, I., & Shneerson, J. (2001). Health related quality of life in narcolepsy. *Journal of Sleep Research*, *10*(1), 75–81.

De Vos, K. (2009). Panel attrition in LISS. Working Paper. Tilburg University, The Netherlands: CentERdata. Retrieved from http://www.lissdata.nl/assets/uploaded/Attrition%20in%20the%20LISS%20panel.pdf

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, *7*(1), 1–26.

Fitzgerald, J., Gottschalk, P., & Moffitt, R. A. (1998). *An analysis of sample attrition in panel data: the Michigan Panel Study of Income Dynamics. NBER Technical Working Paper no. 220*. Cambridge, Mass, USA: National Bureau of Economic Research.

Hart, A. M. & Macnee, C. L. (2007). How well are nurse practitioners prepared for practice: Results of a 2004 questionnaire study. *Journal of the American Academy of Nurse Practitioners*, *19*(1), 35–42.

Hoekstra, L., Kuijper, C., Bakx, R., Heij, H., Aronson, D., & Benninga, M. (2011). The malone antegrade continence enema procedure: the Amsterdam experience. *Journal of Pediatric Surgery*, *46*(8), 1603–1608.

Huber, P. (1976). *The behavior of maximum likelihood estimates under non-standard conditions*. Paper presented at the Proceedings of the fifth Berkeley symposium on mathematical statistics and probability.

Kendell, R. & Pearce, A. (1997). Consultant psychiatrists who retired prematurely in 1995 and 1996. *Psychiatric Bulletin*, *21*(12), 741–745.

Kingston, R., Carver, S., Evans, A., & Turton, I. (2000). Web-based public participation geographical information systems: An aid to local environmental decision-making. *Computers, Environment and Urban Systems*, *24*(2), 109–125.

Kohler, U. & Kreuter, F. (2005). *Data Analysis Using Stata*. College Station, Texas: Stata Corp.

Landis, J. & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174.

Laurie, H., Smith, R., & Scott, L. (1999). ‘‘strategies for reducing nonresponse in a Longitudinal Panel Survey.” *Journal of Official Statistics*, *15*(2), 269–282.

Lepkowski, J. M. & Couper, M. P. (2002). Nonresponse in the second wave of longitudinal household surveys. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey nonresponse* (pp. 259–272). New York: Wiley.

Liang, K. & Zeger, S. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73*(1), 13–22.

Lipps, O. (2009). Attrition of households and individuals in panel surveys. SOEP papers 164. Berlin: German Institute for Economic Analyses (DIW). Retrieved from http://www.diw.de/documents/publikationen/73/diw_01.c.96125.de/diw_sp0164.pdf

Lugtig, P. (2014). Panel attrition separating stayers, fast attriters, gradual attriters, and lurkers. *Sociological Methods & Research*, *43*(4), 699–723.

Marshall, T., Mohammed, M., & Rouse, A. (2004). A randomized controlled trial of league tables and control charts as aids to health service decision-making. *International Journal for Quality in Health Care: Journal of the International Society for Quality in Health Care / ISQua*, *16*(4), 309–315.

Martin, M. & Edwards, L. (1998). Peer learning on fieldwork placements. *The British Journal of Occupational Therapy*, *61*(6), 249–252.

Nicoletti, C. & Peracchi, F. (2005). Survey response and survey characteristics: microlevel evidence from the European Community Household Panel. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *168*(4), 763–781.

Schonlau, M. (2015). What do web survey panel respondents answer when asked “do you have any other comment?” *Survey Methods: Insights from the Field*. Retrieved from http://surveyinsights.org/?p=6899

Slonje, R. & Smith, P. (2008). Cyberbullying: another main type of bullying? *Scandinavian Journal of Psychology*, *49*(2), 147–154.

Stefanski, L. & Cook, J. (1995). Simulation-extrapolation: the measurement error jackknife. *Journal of the American Statistical Association*, *90*(432), 1247–1256.

Stein, R., Chong, S., Everaert, D., Rolf, R., Thompson, A., Whittaker, M., … Ihashi, K. (2006). A multicenter trial of a footdrop stimulator controlled by a tilt sensor. *Neurorehabilitation and Neural Repair*, *20*(3), 371–379.

Uhrig, S. N. (2008). The nature and causes of attrition in the British Household Panel Study (2008-05). University of Essex: ISER Working Paper Series. Retrieved from http://hdl.handle.net/10419/92025

Watson, N. & Wooden, M. (2004). Sample attrition in the HILDA survey. *Australian Journal of Labour Economics*, *7*(2), 293–308.

Watson, N. & Wooden, M. (2009). Identifying factors affecting longitudinal survey response. In P. Lynn (Ed.), *Methodology of longitudinal surveys* (pp. 157–182). New York: Wiley.

Webster, C., Merry, A., Gander, P., & Mann, N. (2004). A prospective, randomised clinical evaluation of a new safety orientated injectable drug administration system in comparison with conventional methods. *Anaesthesia*, *59*(1), 80–87.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, *50*(1), 1–25.

Zeger, S. & Liang, K. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, *42*(1), 121–130.