“When there is nonresponse, we need to be careful”: But what do we actually mean when we say that?

Koen Beullens*

Centre for Sociological Research, Leuven (Belgium)

Nonresponse Workshop, Leuven, 2 - 4 September 2015

Abstract: Many (short) papers that have been prepared for the nonresponse workshop in the past start their abstracts or introductions with the phrase or a variation on the tune: “When there is nonresponse, we need to be careful”. But what does it actually imply when we say we want to be careful? Be transparent about the response rate? Do a thorough nonresponse analysis? Apply weighting adjustments? Deliberately increase the standard errors of the analysis? Or maybe completely reject any conclusion drawn from a survey? This presentation will develop an ordinal scale reflecting how careful a survey researcher can be when dealing with survey nonresponse. The purpose of this presentation is to provoke discussion among the workshops’ participants.

1 Introduction

Many opening statements in papers on nonresponse are along the lines of “The lack of full participation in sample surveys threatens the inferential value of the survey method” (Groves, Cialdini, & Couper, 1992), “Declining response rates raise concerns among survey organizations and data users about the non-response bias and precision of survey estimates.” (Kreuter et al., 2010) or “The immediate consequence of nonresponse is that, without taking special measures, it is not possible to compute reliable estimates of population characteristics. Validity of inference about the population is at stake.” (Bethlehem, Cobben, & Schouten, 2011). Also, many contributions for the nonresponse workshop have introduced mini-papers with a variation on the tune “When there is nonresponse, we need to be careful”.

But what does it actually mean to be careful? I will outline a list of possible dispositions nonresponse researchers, survey researchers or the general public can take, considering the issue of nonresponse. This attitudinal scales moves in an ordinal way from *ignoring nonresponse* to *rejecting any kind of survey research that is subject to nonresponse*:

1. Ignore nonresponse
2. Transparency with regard to response rates
3. Provide a nonresponse analysis or take weak protective measures against nonresponse (weighting adjustments)
4. Take strong protective measures against nonresponse (deliberately augment the standard errors or p-values)

*koen.beullens@soc.kuleuven.be

5. Abandon any inferential claim based on survey data (use surveys only for exploratory purposes)

6. Reject any kind of survey research that is subject to nonresponse

During the presentation of this scale, I may often refer to a meta-analysis where 221 peer reviewed articles are taken to assess to which extent nonresponse is referred in survey literature. All articles were based on data from the European Social Survey and published in 2013. Inclusion conditions are that the articles should be in English and actively use ESS data. Also, 24 strictly methodological articles have been omitted, so that only substantive research articles are included. Some articles may have used ESS in combination with other sources of data, such as country-level variables (e.g., GDP per capita, crime rates, immigration data) or other survey sources (e.g., SHARE, EVS). Some of the journals that frequently publish empirical articles based on the ESS are Sociology of Health & Illness (3), European Sociological Review (17), Journal of European Social Policy (3), Journal of Cross-Cultural Psychology (3), Party Politics (6), Social Indicators Research (19), European Journal of Ageing (3), International Journal of Public Opinion Research (2), Comparative Political Studies (2), Journal of Happiness Studies (2), Personality and Social Psychology Bulletin (2), PloS one (2) and Electoral Studies (2).

Further on, I will also use data from the European Social Survey (Belgium, round 6) to illustrate the impact of nonresponse (before and after propensity weighting) on type I error. I will consider the percentage of Belgians living in an apartment in the full sample, respondents-only and respondents-only after weighting for age, gender and region.

2 An attitudinal scale towards survey nonresponse

This scale can be seen as a classification of the current treatment and attitudes of the nonresponse problem in contemporary substantial survey research.

1. *Ignore nonresponse*

   In the era when response rates were high, nonresponse was simply not an important issue. Following the decline in response rates, researchers started worrying about the validity of surveys being affected by nonresponse. It is clear that simply ignoring nonresponse or assuming it has no noteworthy impact can be considered as too naive. However, many polls that appear in newspapers or other media do not seem to take into account any survey error at all, and nonresponse in particular.

   Surprisingly, scientific output also seem to regularly ignore this issue. As an example from the 221 peer reviewed journal articles that appeared based on the European Social Survey, only 51 (23%) mention the notion of 'nonresponse', 'non-response', 'nonrespondent' or 'response rate'[1]. So, even though response rates in the ESS roughly range between 35% and 85%, the majority of the empirical work based on its data seems to ignore the issue of nonresponse.

2. *Transparency with regard to response rates*

   Many international survey authorities are highly committed to survey transparency. Organizations such as the American Association for Public Opinion Research (AAPOR), the Council of American Survey Research Organizations (CASRO), the European Society for Opinion and Marketing Research (ESOMAR), the International Statistical Institute (ISI), and the American Statistical Association (ASA) mention the importance of providing the necessary information about the sample design or the methods of data collection. The American Association for Public Opinion Research (2015) has even developed a set of definitions and standards for the calculation of response rates.

   Nonetheless, only 24 articles (out of 221 articles as mentioned above; 11%) report response rates. These articles even include the ones that do not literally report response rates, but

[1] Notice that some of the article only mention nonresponse in the context of item nonresponse.

refer the ESS-website for more information instead. So, the call from many survey authorities or professional institutes to provide information about response rates only seem to be very moderately followed.

3. *Provide a nonresponse analysis or take weak protective measures against nonresponse (weighting adjustments)*

   Many survey researchers have become aware that the response rate alone is a very poor indicator of survey quality, particularly when response rates are low and leave more potential for bias to occur. Comparisons between respondents and nonrespondents with regard to some auxiliary variables often suggest a substantial danger that survey statistics are biased. Other methods such as the comparison between weighted and unweighted estimates also seem to indicate that nonresponse unfavorably affects the quality of surveys.

   Of the 221 articles that were considered, exactly 2 mentioned the concept of nonresponse bias potentially affecting their results. Nevertheless, they do not provide any results of a bias analysis.

   Many survey researchers have resorted to corrective measures such as weighting adjustments or imputation. These are believed to (partially) improve the quality of survey statistics. However, their effectiveness is hard to prove because the full-sample target statistics are unknown by definition.

   Not in a single article, a reference could be found related to 'adjustment' or 'weighting' for nonresponse. This may be due to the fact that post-stratification weights are only available from round 6 of ESS, which was only disseminated in 2014. References to survey adjustments related to the sampling design, however, could frequently be observed in the articles.

   Recently, alternative quality indicators that measure the impact of nonresponse have been introduced. For example, the R-indicator (Schouten, Cobben, & Bethlehem, 2009) measures the variance of response propensities, indicative for the imbalance of an obtained sample or the capacity of that sample to represent the targeted population. One can also consider the fraction of missing information (FMI) (Wagner, 2010). Both indicator assume that nonresponse operates under the 'missing at random' mechanism, stating that nonresponse is exclusively conditional on the observed auxiliary variables.

   Not a single trace of such alternative nonresponse related quality indicators could be found during the article screening.

   At this point, it becomes clear that the nonresponse issue escapes the attention of researchers in sociology, political sciences, ... Even though more strict attitudes can be adhered to (see point 5 and further), it can be expected that survey researchers do not go beyond the point of taking weak corrective measures such as raking, propensity weighting, imputation, ...

4. *Take strong protective measures against nonresponse (deliberately augment the standard errors or p-value)*

   Only relatively recently, more attention has been given to the possibility that survey damage due to nonrepsonse is not completely taken into account after weighting for age, gender or level of education (or other common available auxiliary variables). Bias or uncertainty can still hinder researchers to draw valid conclusions from survey data. Therefore, a shift from MAR (missing at random) to MNAR (missing not at random) should be considered. Some scholar have already made their first contribution, for example Andridge and Little (2011).

   A possible way to take strong protective measures against nonresponse is to deliberately increase the margins of uncertainty. Consider Figure 1 in this respect. In the case of full response, the percentage of Belgians living in an apartment is about 20%. However, if only respondents are taken into account (response rate = 59%), it seem that only 17.3% are believed to live in an apartment[2]. Because of this bias, relying solely on respondents implies a coverage rate of 13%

[2] The densities about these estimates represent the uncertainties of the considered parameters due to sampling error.

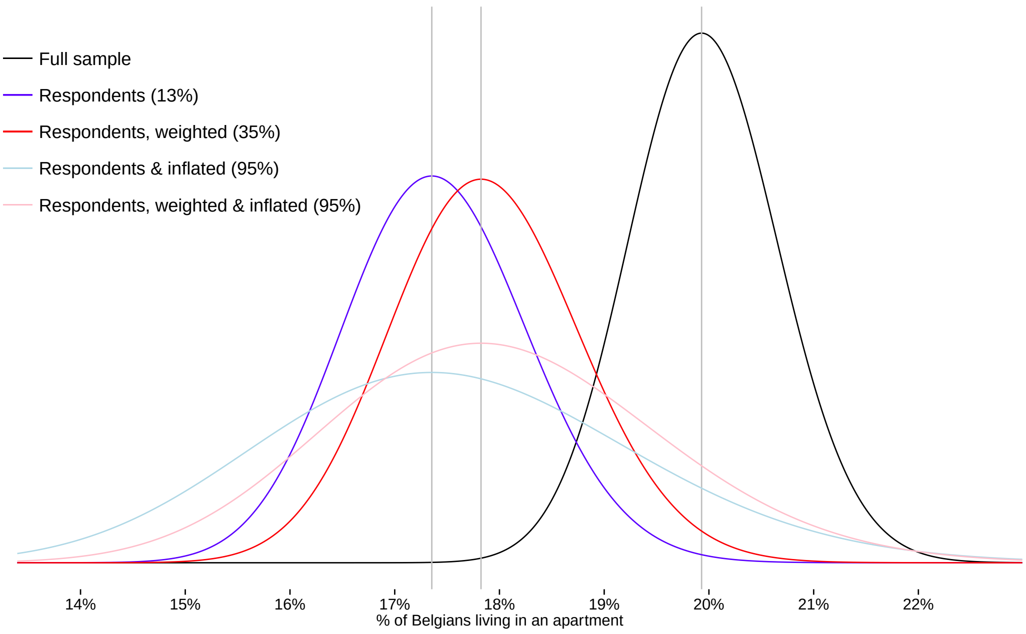

Figure 1: Percentage of Belgians living in an apartment - European Social Survey, round 6

(the respondents-only confidence interval only covers the density of the estimate for the full sample by 13%), or a type I error[3] of 87%. After weighting for age (four age classes), gender and region (10 Belgian provinces + Brussels), the estimate shifts a bit toward the full sample estimate, but still only covers the full sample density by 35%. This corrective measure may not be so satisfactory since it still leaves a type I error of 65%.

If one wants to be completely safeguarded against type I error, variance inflation may be a solution. In Figure 1 the variances of the respondent-only and weighted respondent-only estimates have been inflated so that the coverage rate is 95% in both cases. As a result, a variance inflation factor needed to be accepted of respectively 4.13 and 3.05. Given that ESS6 in Belgium counts 1867 eligible respondents, the effective sample sizes are only 1867/4.13=452 (respondents-only) and 1867/3.05=612 (weighted, respondents-only).

The term (inflated) "standard error" may be misleading here, since that notion usually refers to random variation in estimates such as sampling error. On the contrary, nonresponse can be seen as a systematic error (imagine the experiment where multiple surveys would be carried out in Belgium that continually underestimate the percentage of apartment dwellers). In this respect, one might introduce the terms "uncertainty bound" or "region of ignorance" (see, for example, Vansteelandt, Goetghebeur, Kenward, & Molenberghs, 2006) to indicate the combined uncertainty of random (e.g. sampling error) and systematic error such as nonresponse.

5. *Abandon any inferential claim based on survey data (use surveys only for exploratory purposes)*

   However, it is unclear how far measures such as deliberate variance inflation should go in order to completely safeguard survey estimates from type I error. In this regard, one might say that survey research is 'uncertain about its uncertainties'.

   Survey research seems to be forced to leave the paradigm that starts from the assets of probability sampling. In this regard, Groves (2006) points out that 'Because of falling response rates, legitimate questions are arising anew about the relative advantages of probability sample

[3] As a reminder, type I error refers to the incorrect rejection of a true null hypothesis (a "false positive"). In this particular case, based the respondent-only estimate, the 95% confidence interval runs from 15.70% to 19.14%, and only covers 13% of the full sample parameter density.

surveys. Probability sampling offers measurable sampling errors and unbiased estimates when 100% response rates are obtained. There is no such guarantee with low response rates.'

6. *Reject any kind of survey research that is subject to nonresponse*

   This last option on the scale of attitudes toward nonresponse is probably a step too far. Surveys still deliver data that is supposed to be more informative than an uninformed guess about the poverty rate, the average health, political trust, etc. Survey data at least provide a reflection of an underlying, but hard to measure reality.

   Nevertheless, in the general public, some suspicion regarding survey results can be observed. For example, in the Flemish survey on surveys (Loosveldt & Storms, 2008), 25% of the respondents (mail survey) said to disagree with the statement that 'the results of a survey are always reliable'.

3 Discussion

The workshops' participants are cordially invited to take a personal position on the scale and exchange their ideas. Some specific attention can be devoted to:

- How could we communicate, interpret and elucidate nonresponse error to the general public?
- How could data producers communicate, interpret and elucidate nonresponse error to the end-users of survey data?
- How should the fallibility of surveys, nonresponse error in particular, be addresses to students during the statistics or survey methodology courses?

References

Andridge, R. A., & Little, R. J. (2011). Proxy Pattern-Mixture Analysis for Survey Nonresponse. *Journal of Official Statistics*, *27*(2), 153–180.

Bethlehem, J., Cobben, F., & Schouten, B. (2011). *Handbook of Nonresponse in Household Surveys.* Hoboken (N.J.): John Wiley & sons.

Groves, R. (2006). Nonresponse Rates and Nonresponse Bias in Household Surveys. *Public Opinion Quarterly*, *70*(5), 646–675.

Groves, R., Cialdini, R., & Couper, M. (1992). Understanding the Decision to Participate in a Survey. *Public Opinion Quarterly*, *56*(4), 475–495.

Kreuter, F., Olson, K., Wagner, J., Yan, T., Ezzati-Rice, T., Casas-Cordero, C., ... Raghunathan, T. (2010). Using Proxy Measures and Other Correlates of Survey Outcomes to Adjust for Nonresponse: Examples from Multiple Surveys. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *173*(2), 389–410.

Loosveldt, G., & Storms, V. (2008). Measuring Public Opinions about Surveys. *International Journal of Public Opinion Research*, *20*(1), 74–89. doi: 10.1093/ijpor/edn006

Schouten, B., Cobben, F., & Bethlehem, J. (2009). Indicators for the Representativeness of Survey Response. *Survey Methodology*, *35*, 101–113.

The American Association for Public Opinion Research. (2015). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys, 8th edition* (8th ed.). `http://www.aapor.org/AAPORKentico/AAPOR_Main/media/publications/Standard-Definitions2015_8theditionwithchanges_April2015_logo.pdf`.

Vansteelandt, S., Goetghebeur, E., Kenward, M., & Molenberghs, G. (2006). Ignorance and Uncertainty Regions as Inferential Tools in a Sensitivity Analysis. *Statistica Sinica*, *16*, 953–979.

Wagner, J. (2010). The Fraction of Missing Information as a Tool for Monitoring the Quality of Survey Data. *Public Opinion Quarterly*, *74*(2), 223–243.