A strong case for smaller sample sizes?

Koen Beullens koen.beullens@soc.kuleuven.be and Geert Loosveldt Geert.Loosveldt@soc.kuleuven.be

Centre of Sociological Research, KU Leuven

NRWorkshop, Ottawa, 4^{th} - 6^{th} September 2012

Abstract: Assuming the omnipresence of bias due to nonresponse, increasing the sample size will not necessarily produce smaller errors. Given the same bias structure, a sample of only 500 cases might be more attractive from a cost perspective than a sample containing 5000 cases. Furthermore, the small sample will have larger standard errors, implying the reduction of potential type I errors. Even if the cost argument is not so decisive, the option of smaller sample sizes, combined with more efforts toward low propensity profiles might be more attractive in order to reduce the bias potential. This presentation seeks to discuss and illustrate the pros and cons of small(er) sample surveys and explore ways to determine an optimal quality-cost trade-off.

Introduction

Larger survey samples are not necessarily better samples. Indeed, systematic error such nonresponse bias cannot be solved by increasing the sample size. When considering $MSE = bias^2 + var$, increasing the sample size will only reduce the second term; the favourable effect of a larger sample will gradually reduce to (practically) zero, provided that the bias-component is constant and independent of the sample size (which seems to be a very reasonable assumptions). This starting point may have serious consequences for both the design of surveys as well as their analysis. This paper addresses two questions:

1. What is a reasonable amount of bias that should be accepted, when calculating the MSE?

2. How can we anticipate the unfavourable effects of nonresponse bias in the survey design (particularly with regard to the sample size)?

1 Determine MSE

As it is quite easy to find differences between respondents and nonrespondents (to the extent that available auxiliary information allows such a comparison), the risk of having to deal with biased estimates is quite considerable. Therefore, simply building on the principles of probability sampling and ignoring systematic error may be considered to be naive (Särndal, 2010). On the other hand, estimating the detrimental systematic effect of bias is very difficult as nonresponse is not observable by definition. Nevertheless, assuming that $bias = 0$ and thus $MSE = var$ is simply not an option.

A possible way to estimate the bias to which the survey parameters may be exposed is to look at auxiliary variables such as register information or interviewer observations. Such variables are observed for both respondents and nonrespondents and allow a realistic but still imperfect estimate of the damage nonresponse potentially causes. The imperfection relates to the assumption that the target variables should undergo a similar bias than the interim target variables (=auxiliary variables), which is obviously untestable.

The more (and the more diverse) auxiliary variables are available, the better the bias-distribution can be reproduced. In this respect, Peytchev and Biemer (2011), building on the work of Groves (2006) and Peytcheva and Groves (2008), have proposed the *Bias effect Size (BES_k)*, or the bias with regard to some auxiliary variable k (available for both respondents and nonrespondents), relative to its (population) standard deviations. The average BES of several auxiliary variables can be considered as an indication of the bias that applies to the target variables. Since the sampling design and the sample size are usually known, the MSE can be determined by $MSE = bias^2 + var$, where $bias$ is empirically estimated from the auxiliary variables, var is theoretically known from the sample design. Relative to the naive MSE ($MSE_{naive} = var$), the MSE can be used to estimate to what extend the confidence interval of target parameter needs to be inflated in order to protect a survey estimate against nonresponse bias. Also the effective sample size can be easily be derived ($n_{eff} = \frac{MSE_{naive}}{MSE} \times n_{naive}$, where $\frac{MSE}{MSE_{naive}}$ is the variance inflation factor and n_{naive} is the number of completed interviews/questionnaires.

As the expected bias grows, it is obvious that the distance between n_{eff} and n_{naive} will also increase. It is also obvious that MSE has a lower limit (due to the bias), even if var converges to (practically) zero. This explains the decreasing marginal effect on the effective sample size, when increasing the sample size (provided that bias remains constant). When the sample is relatively small (e.g. n=100), it may be worth the effort to enlarge the

Table 1: Overview of non-response bias with regard to some auxiliary variables of the ESS - Belgium, 3^{rd} round

	Full sample	Respondents only		Full sample	Respondents only
Age class			*Percentage of non-Belgians in municipality (in %)*		
<20	8.51%	9.52%	<2	20.53%	22.88%
21-40	29.01%	29.15%	2-5	31.02%	32.73%
41-60	35.77%	36.47%	5-15	31.94%	31.02%
>60	26.72%	24.86%	>15	16.50%	13.37%
Population density in municipality (in inh./km²)			*Average annual per capita income in municipality (in €)*		
<200	11.99%	12.43%	<12000	16.98%	14.85%
200-400	26.55%	29.04%	12000-14000	38.03%	37.90%
400-700	18.38%	19.97%	14000-16000	33.34%	35.42%
700-2500	31.94%	30.80%	>16000	11.65%	11.83%
>2500	11.14%	7.76%			
Male	47.69%	46.64%	*Multi-unit building*	16.26%	11.61%
Region			*Neighbourhood quality*		
Flanders	59.69%	62.87%	poor	20.40%	17.55%
Brussels	8.75%	5.50%	good	31.53%	33.39%
Wallonia	31.57%	31.63%	excellent	48.07%	49.06%

sample somewhat, as there is still potential for the *var*-part to decrease. When the sample size is already relatively large (e.g. n>1000), increasing the sample size might only have a small or practically no added value to the statistical power of the sample.

As an illustration, consider table 1, where information about auxiliary variables is presented for the European Social Survey (Belgium - 3^{rd} round - 2006). The gross sample (respondents and nonrespondents) was 2927 of whom 1798 (response rate = 61.43%) participated. The average absolute standardized bias for the available auxiliary variables is 0.05 (=expected standardized bias). The *var*-part is 1/1798, as the variables are first standardized. The implied $MSE = 0.05^2 + 1/1798 = \underline{0.0031}$ as compared to $MSE_{naive} = 1/1798 = \underline{0.0005}$. As a consequence, the variance inflation is 5.50 and instead of 1798 nominal cases, the effective sample size is only 327.

2 Consequences for design

Given the results of section 1, we may reconsider the choice of an adequate sample size (gross or net). From a mere cost-perspective, we might select for example only 1500 cases instead of 2927, loosing only a small fraction the statistical power of the survey (see situation (1) in figure 1). In the graph, the initial situation of ESS3-BE starts from a gross sample (1798 respondents + 1129 nonrespondents) of 2927 units, corresponding to an effective sample size of 327. When moving to a lower gross sample size of 1500, following the

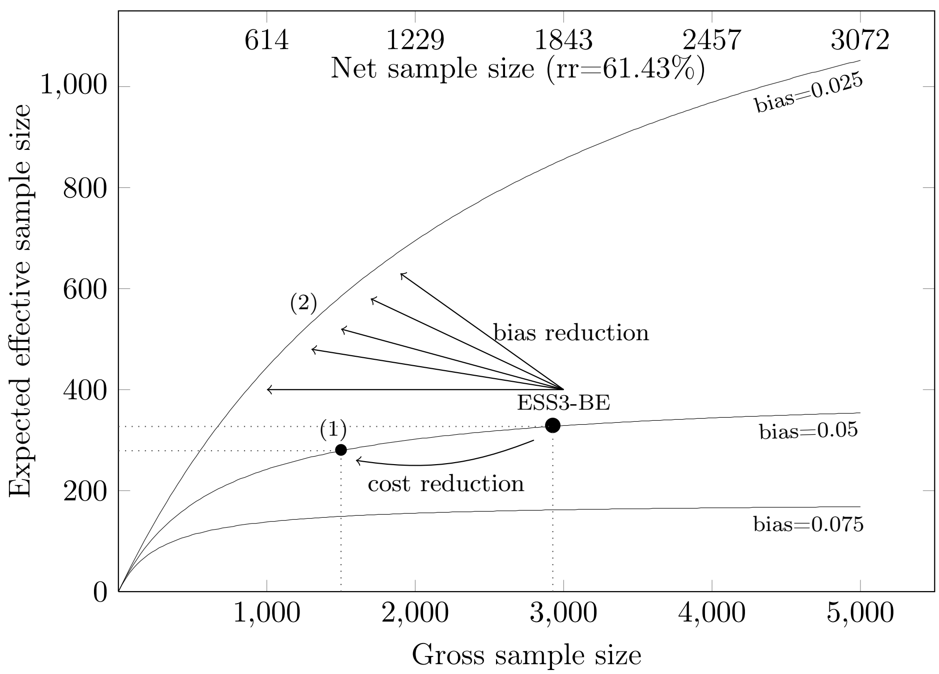

Figure 1: Expected effective sample size, conditional on the expected bias and the gross sample

same bias curve, the fieldwork cost can be considerably reduces (e.g. -50%), while the effective sample size only reduces mildly.

The second option is to reduce the gross sample as well, but to make more efforts to convert low propensity cases by incentivising hard refusals, facilitate the questionnaire access for disabled or language barriers, and so forth. The costs will probably not be reduced, but because of the increased attention for low propensity cases, we may end up in a more advantageous bias-curve, resulting is a greater effective sample size. The eventual effective sample size also depends on other (unknown) factors such as ease with which reluctant cases can be converted. This is why we are unable to exactly locate this scenario on figure 1. However, there is some evidence that survey fieldwork operations follow the 'line of least resistance', where easy cases (or so-called 'low-hanging fruit') are prioritized over difficult cases (or 'high-hanging fruit'). Prioritising the easy cases may result in higher response rates and larger samples, but also deepen the difference between respondents and nonrespondents, resulting in higher risk of bias. Alternatively, there may still be potential to minimize the risk of nonresponse bias by inverting the line of least resistance.

Discussion & Questions

This note tried to advance the idea that smaller sample sizes may be worth considering. First, a smaller (gross) sample needs less efforts and is therefore cheaper, probably outweighing the loss of statistical power. Second, instead of pursuing the highest possible sample size, it seems worthwhile to focus rather on the representativeness of the sample; from figure 1 it seems that the level of risk of bias strongly affects the effective power of the survey. In fact, increasing the sample size in the presence of (risk of) bias has a strong restrictive impact on the sample power, removing the (risk of) bias clearly offers strong potential for improvement. Third, large samples usually provide small (naive) confidence intervals. In the presence of nonresponse bias, the probability of making a type I error is therefore much higher than in a situation of a small sample with a similar burden of nonresponse bias.

Still, arguments can be found to advance large samples. First, in a multi-purpose survey such as the European Social Survey, often only a subset of the data may be used (e.g. only the 65+ or only those who have a job). Those subsets may become too small if the total set is already small. Second, as the distributions of some particular variables such as monthly expenses are very unstable, large sample sizes may be better to obtain more robust estimates.

Questions for NRWorkshop:

- Would you agree with the line reasoning as presented in this note? Why (not)?
- Are there other pro & con argument for small/large sample sizes?

References

Groves, R. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly*, *70*(5), 646–675.

Peytchev, A., & Biemer, P. (2011). A standardized indicator on survey nonresponse bias based on effect size. In *Paper presented at 22th international workshop on household survey nonresponse, bilbao, spain, 5-7 september 2011.*

Peytcheva, E., & Groves, R. (2008). The impact of non-response rates on non-response bias: a meta-analysis. *Public Opinion Quarterly*, *72*(2), 167–189.

Särndal, C.-E. (2010). The probability sampling tradition in a period of crisis. In *Keynote speech at the q2010 european conference on quality in official statistics, helsinki, may 4-6, 2010.*