Anticipation of unit nonresponse or not in the sampling designing

From the point of view of the European Social Survey (ESS)

Seppo Laaksonen

University of Helsinki

Seppo.Laaksonen@Helsinki.Fi

Abstract

Various sampling designs in household surveys are applied. There can be one or more stages. If two or more stages are used, cluster sampling is one part of the design in most cases. It is possible to design the sampling for the whole target population, or to stratify. This paper concentrates on the latter case, thus to sample allocation in the case of non-ignorable unit missingness. My question is: is it good to include the anticipated response rates into sample allocation or not? This question is considered in the European Social Survey (ESS) that since round 6 recommends proportional allocation of the gross sample. If another allocation could be used, explicit strata are necessarily needed. Many countries are applying the design without explicit strata, and even the gross sample weights are varying much. Hence proportional allocation might be motivated. This however is not any good reason to require the allocation without anticipated response rates. If this disproportional allocation for the gross sample would have been successfully applied, the weights of the respondents could be more equal than those of the gross sample. My conclusion is that anticipated response rates in gross sample allocation should be used. The examples of the four ESS country have been used that show concretely how recent response rates vary by strata and how the allocation can be done. Some open questions are still to discuss. For example, which response rates to use in allocation and for how many strata?

1. Introduction

The European Social Survey (ESS) has been conducted since 2002 biannually. Its sampling design is probability-based so that each country should follow the so-called sampling guidelines (e.g. round 7: https://www.europeansocialsurvey.org/docs/round7/methods/ESS7_sampling_guidelines.pdf.). This paper is focused on unit nonresponse in the sampling design. The anticipation or the prediction is a big part when designing the sample, that is, it is needed to 'guess' in the best way the certain figures, and then to hope that these will be realized later in the sampling and the fieldwork. There have been some differences during the first 8 rounds in these rules, and these might be changed in Round 9 again, if found appropriate. We here introduce you to one possible new rule, how to take unit nonresponse better into account in sampling design.

The target effective sample size (*neff*) is one requirement needed to satisfy in each country. The size is 1500 for most countries but 800 for smallest ones. It is such a net sample size that corresponds to the size achieved with simple random sampling. Naturally, this cannot be in real-life possible to get but it makes it easier to design the sample so that the same rules are used in each country. The sampling guidelines should have followed in each country but the central sampling team must have been flexible so that the design can be approved sometimes still, mainly due to

financial problems; otherwise the country had not participated in survey at all. For example, the target 70% unit response rate cannot be achieved in many countries; hence a lower rate has been accepted given that the rate is as high as in best surveys in the country.

Currently we have already anticipated both due to unit nonresponse and in-eligibility but at the whole country level. To describe the whole procedure, t*he target* effective sample size depends on the four components:

- the two design effects,
1. One due to varying sampling weights (*DEFFp*); On the other hand, the *DEFFp* can be called due to varying inclusion probabilities or more specifically due to varying sampling weights is not clear what to use in each case. This is a big issue in this presentation.
and
2. The other due to clustering (*DEFFc*) but this latter is not here considered in details but it is good to point out that *DEFFc* includes the unit nonresponse at primary sampling unit (PSU) level in its average since the anticipated *DEFFc* =1+(average anticipated net cluster size-1)*intra-class correlation includes unit nonresponse via net cluster size,

3. Anticipated unit response rate based on the best possible recent information = rr,

4. Anticipated in-eligibility rate respectively = ir.

Using these components the target gross sample size = (*neff*/(rr*(1-ir)))**DEFFp***DEFFc* in which *neff* thus should be at least either 1500 or 800.

This is not necessarily achieved even in the sampling design phase, often due to limited resources. For example in the oncoming round 8 this formula for the UK gives = (1278/0.50*(1-0.087))*1.27*1.407= 5000. The gross sample size thus is 5000 but the *neff* is below 1500.

Most countries fortunately can achieve the minimum, for example Estonia. Their anticipated gross sample is = (2000/0.65*(1-0.02))*1*1=3140, their *neff* target is even as high as 2000. We here see that both *DEFFp* and *DEFFc* are equal to one. As we later find, this *DEFFp* is not necessarily fair since Estonia's sampling design is simple random that gives this figure by definition. The UK design is three-stage where such *DEFFp* would be difficult to get even though it would be somewhat smaller than this 1.27.

2. Design effect due to varying sampling weights with examples

DEFFp thus depends on the sampling design:

- If it is simple random sampling without stratification, it is automatically = 1. This is used in many register countries such as Denmark, Finland or Estonia, and sometimes also in Norway, Switzerland, Belgium and the Netherlands (at least approximately).

- Many countries apply two- or three-stage design in which case *DEFFp* can be larger than one, sometimes 'too much' so that the sample allocation is too far from the proportionality that is currently recommended by the ESS central coordination committee.

Examples in round 7: PT=1.51, FR=1.48, LT=1.38, …, AT=1.05, PL=1.02; the first three ones are too high but a lower *DEFFp* has been difficult to get.

If DEFFp is small or even = 1 it means that it is easier to achieve the required gross sample size, since these are linearly related to each other:

- Portugal's gross sample should thus be 51% higher if compared to those countries with *DEFFp*=1.
- This is not completely fair, since the correct sampling weights vary by response rates. One logical solution would be use stratification and create the respective weights, thus at stratum level. This respectively would lead to a higher gross sample size but not necessarily much. On the other hand, the interval estimates by strata would be smaller, i.e., the estimates more precise.

My question to you thus is now in other words: Why not to anticipate at a lower level? And if the answer is 'YES', then how to do it? My lower level here is (Explicit) Stratum. It is well known that response rates vary much by strata that are often regions or more complex such as gender*age group or region*gender or region*age group. Our examples below are concerned regional strata.

Next, in Section 3, I present some examples about the countries where stratification has already been used, and consider possible consequences if the anticipated response rates were used in the sampling design.

3. Unit non-response by stratification in some ESS countries and possible consequences

There are two types of stratification in the ESS countries. *Implicit stratification* is not any conventional stratification since it is a tool for drawing individuals by equidistance selection that factually corresponds to the simple random sampling since it would be hard to well use this design in interval estimation. It is however expected that this technique improves estimates even though it is not known how and how much.

The second stratification is conventional and can be called *explicit stratification*. This is applied in more than half of the countries but the number is increasing slightly, since the sampling team has been tried to encourage for its use. It is not still a general strategy and not used in the UK or Finland. However, Estonia and Sweden are removing to this direction in round 8. There are different strategies in stratum countries concerning the amount of strata for example. We present next four examples without mentioning the name of the country. There are in these examples some terms and symbols that are explained first:

- The response rate rr = r/(r+nr) (r=the number of the respondents; nr=the number of the non-respondents)
- *DEFFp(all)* = design effect based on the design weights = N_h/n_h

(in which *h*=stratum, *N*=target population size, *n*= gross sample size)

- *DEFFp(resp)* = design effect based on the basic sampling weights = N_h/r_h

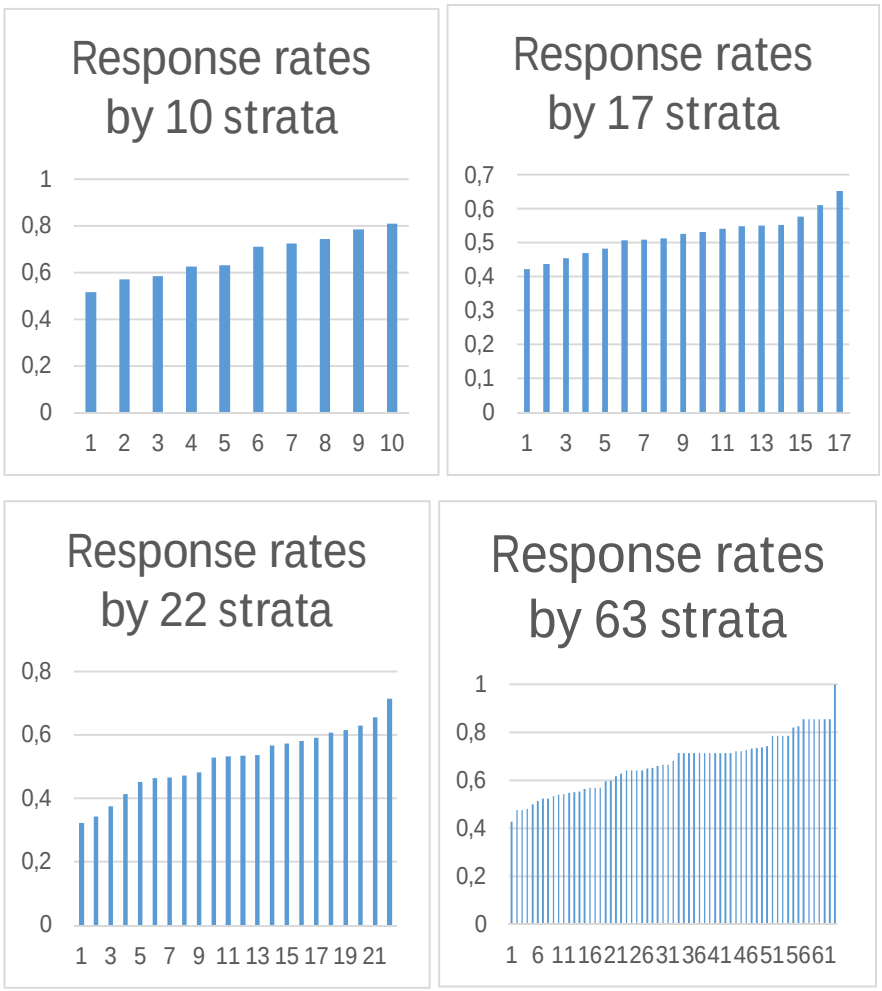

Figure 1. The response rates of the four countries with different numbers of strata

These examples do not include countries with a very low average response rate that are even as low as 40% or below. The variation by strata is clear any way. Currently the proportional sample allocation for a gross sample is required but this is not completely followed in all countries. It is difficult even definitely to know how well it has been followed unless any variable such as stratum does not exist. It is however easy to check this for these four cases using *DEFFp(all)*, Table 1. We see that these *DEFFp*'s are quite close to one, especially in the two middle cases. The design of these countries thus does not differ much from the proportionality. The one exactly is not possible due to rounding that leads to more far from the proportionality if the number of strata is big as in the last country. The *DEFFp* of the first country is surprisingly big.

Table 1. Design effect based on the design weights

Number of strata	*DEFFp(all)*
10	1.00846
17	1.00013
22	1.00013
63	1.00770

Now we can go to look what could happen if the anticipated response rates were used in gross sample allocation. It is not of course clear how to do it, since the observed response rates will not be realized necessarily in the consecutive round given that the reference survey is the previous survey. Naturally, it is good to check several rounds if available and look forward how regular response rates are. However, it seems to be clear that the response rate in big cities or urban areas often is lower than in rural areas.

We do not know the country conditions of these examples but it is easy to see that the anticipation over 63 strata exactly does not look smart. Hence we do not use this country example next when we allocate the initial gross sample size into strata again, thus so that the gross sample size will be linearly increased if the response rate is below the average, and decreased respectively.

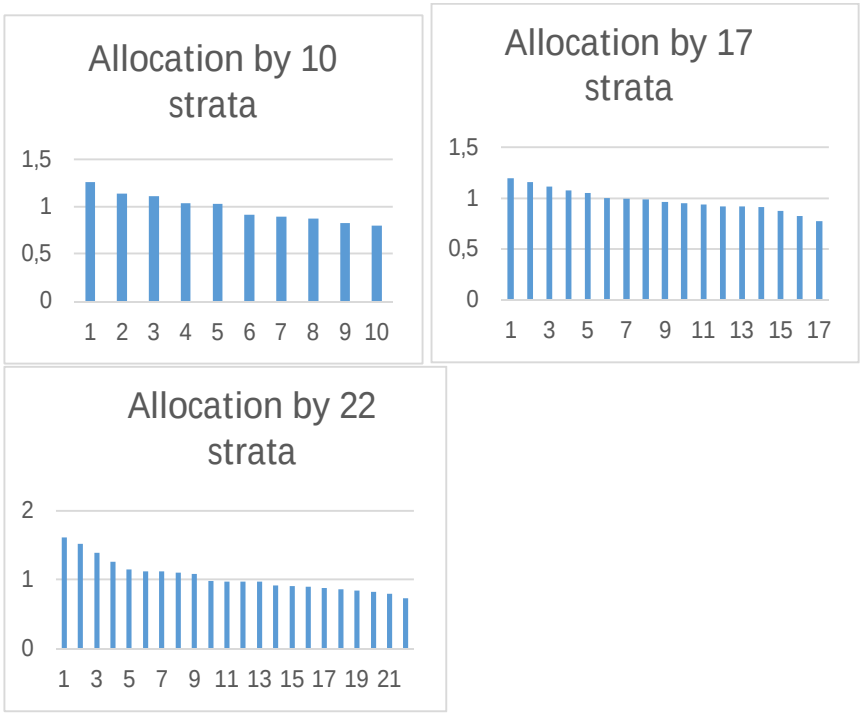

Figure 2. The ratio the re-allocated size per the initial size for three countries by strata. The re-allocation takes into account the response rates of Figure 1.

The graphs of Figure 2 thus are opposite to those in Figure 1. When looking for the ratios themselves, they do not vary very dramatically. The two first ones are fairly similar but the variation of the third is somewhat higher (the minimum= 0.73, the maximum=1.62). Table 2 illustrates the impact in the *DEFFp*-indicator.

Table 2. Design effect based on the basic sampling weights

Number of strata	*DEFFp(resp)*
10	1.0050
17	1.0136
22	1.0142

We thus observe that these design effects are not big, the smallest being in the country with the smallest number of strata. If these designs had been used in practice, the gross sample sizes for these three countries would 0.5%, 1.36% and 1.42% higher than currently. It is not however any drawback, if the anticipation is successful reasonably.

The most important point however is, that the current *DEFFp* criterion is not good since it does not matter response rates by strata. I thus suggest to use this criterion. It first requires to use explicit stratification in order to anticipate at lower level than the country. If explicit stratification with an appropriate number of strata (between 10 and 20) would be used, this technique is also good to introduce to the use.

4. Concluding points and questions

For me, it has been difficult to understand why the proportional allocation to gross sample should be used. It is understandable only for such countries whose design is far from the proportionality, as Portugal, France and Lithuania in round 7, and Slovakia or Kosovo earlier (they did not participate in round 8). This disproportionality is not due to response rates but some for unknown reasons.

One strange argument presented is also that if a relatively higher gross sample size for an urban area would have been used due to its higher nonresponse rate, it would be ‘too easy’ to get enough respondents there. I have to ask, how it is possible if the same situation has been met over all rounds and other surveys? I thus do not see any reason to resist this suggestion but it is good to start conservatively, thus not using completely the most recent response rates, for example. What else should be considered?