Oversampling in Relation to Differential Regional Response Rates

Jan Pickery
Research Centre of the Flemish Government, Belgium

Ann Carton
Research Centre of the Flemish Government, Belgium

Response rates of face-to-face surveys often show regional variation. In larger cities e.g., response typically will be lower than in smaller villages. Following current survey practices, substitution of survey non-respondents is no longer recommended. In order to achieve an adequate regional representation of the population in a survey, differential regional oversampling can be an option. We show how regional ineligible rates and response rates of previous surveys can be used in a multilevel analysis to obtain residuals that form the basis for the computation of an ineligible correction and a regional oversampling factor for subsequent surveys. We argue that this oversampling design is a good alternative or complement to nonresponse weighting. We illustrate our approach with the sampling procedure used for the last edition of the yearly survey on social and cultural changes in the Flemish region.
Keywords: sample design, unit nonresponse, multilevel analysis

Introduction

Survey research is inextricably bound up with unit non-response. Only in very exceptional cases all sampled persons or organisations respond to a survey. Sampled persons can appear ineligible (because of language problems, decease after selection from the sample frame ...), they can refuse to cooperate or it might be impossible to contact them. The causes and consequences of (unit) nonresponse have been studied comprehensively, as well as the different ways to deal with it (Groves, Dillman, Eltinge and Little 2002).

A consistent finding in the literature is the variation in response rates according to urbanisation (Groves and Couper 1998, pp. 140-154). Urbanicity is an indicator of response rates in all survey modes (Groves 2006), but especially in face-to-face surveys the response rates vary considerably depending on the respondents' place of residence. In larger cities response typically will be lower than in smaller villages. In a recent example Abraham, Maitland and Bianchi (2006) show that response probabilities are significantly lower for people living in a central city as compared with people who live in a nonmetropolitan area. Ignoring the regional variation in response will lead to an overrepresentation of inhabitants of smaller villages in the final sample. When analysing survey data, this can be a severe problem depending on the subject of interest. Think e.g. of research into commuting and the use of different means of transport that show large regional variation. In this case nonresponse will lead to severe bias when estimating e.g. population means, because of the differences between respondents and nonrespondents. The nonresponse bias is a function of these differences and the nonresponse rate.

Three ways of reducing the nonresponse bias that is re-

lated to differential regional response rates are weighting, oversampling and substitution. Nonresponse weighting assigns a weight to the respondents based on the probability of response. This weight can be combined with the sample selection weight. Since (non)response probabilities for the population are unknown - as opposed to sample selection probabilities - they have to be estimated from the data (Dillman, Eltinge, Groves and Little 2002). The weights will normally reduce (nonresponse) bias, but they increase the variances of survey estimates (Kish 1992), although for nonresponse weighting this does not always need to be the case (Little and Vartivarian 2005).

Survey researchers sometimes use regional oversampling as an 'alternative'. Oversampling applies various sampling fractions for different regions. More particularly urbanized regions are oversampled in the sampling design, taking the expected response rates into account. The National Travel Survey in the UK e.g. oversamples London as response rates in the capital are lower than elsewhere (Kershaw 2002). Several countries participating in the European Social Survey use oversampling as well. In Poland e.g. unequal probabilities of selection are used in different urbanicity categories. In Spain areas with a population size of more than 500,000 are oversampled. In Portugal the population is divided into 23 strata (based on region and municipality population size) and oversampling is used for the strata where the anticipated response rate is lower than the average (ESS 2004). Oversampling rates are usually based on previous survey experiences, although it is not always clear how the response results of earlier surveys are translated into the oversampling procedure.

Regional oversampling implies unequal selection probabilities by definition. Consequently it requires weighting rather than being an alternative for it. It is actually more an alternative for substitution. When applying nonresponse substitution, survey researchers or survey designers substitute a person who cannot be interviewed by another respondent. When the substitute respondent lives in the same mu-

Contact information: Jan Pickery, Research Centre of the Flemish Government, Boudeijnlaan 30, 1000 Brussels, Belgium, Email: jan.pickery@dar.vlaanderen.be

nicipality (or regional unit) as the original respondent, this procedure accounts for the differential regional response. However, old and more recent literature objects to this kind of substitution. Deming (1950) states that substitution is not a good method for reducing nonresponse because it is only equivalent to building up the size of the initial sample, leaving the bias of nonresponse undiminished. Vehovar (1999) argues that substitution introduces larger bias, extends the fieldwork period, causes fieldwork control problems and results in higher nonresponse rates since interviewers know that substitution addresses are available. In the International Social Survey Programme the use of substitution is discouraged because of two possible risks. The sample may become a convenience sample that overrepresents easy-to-contact and compliant respondents and interviewers may reduce their efforts to obtain interviews from originally selected respondents (ISSP 2003). In the European Social Survey, which has the ambition to become a kind of quality standard when it comes to survey research, substitution is not allowed because of the same arguments (ESS, 2004). None of the varieties of substitution that exist meet the requirements of probability sampling (Lynn, Häder, Gabler and Laaksonen 2007:111). However, substitution is still common practice in many surveys and the discussion on the advantages and disadvantages continues (Chapman 2003; Vehovar 2003; Lynn 2004).

We argue that adequate regional oversampling can ensure the advantage of substitution (proportional representation of regional units) while maintaining the principles of probability sampling, since every sampled unit has a known probability of being selected. The main disadvantages of substitution are overcome because there is just one list of sampled units, who have to be contacted all together. In some cases it might also be preferable over nonresponse weighting after the data collection. It will increase the number of sampled units in otherwise underrepresented regions, resulting in more precise estimated for those regions. Moreover it can also simplify the fieldwork. We will return to this evaluation in the discussion.

The main focus of this paper is to demonstrate a way to set up an adequate oversampling design, based on the response rates of previous surveys. As mentioned above, oversampling is used frequently, but the oversampling design is not always substantiated. In section 2 we explain the procedure that implies a multilevel analysis of regional response rates. In the third section we illustrate our approach with the sampling procedure used for the last edition of the yearly survey on social and cultural changes in the Flemish region. We conclude with an assessment of the procedure.

A multilevel approach to compute regional oversampling factors

The approach we propose to define the adequate regional oversampling factors implies a multilevel analysis of the regional (non)response rates of preceding surveys. The regional residuals of that analysis can be incorporated in the computation of the oversampling factors for subsequent surveys.

For the multilevel analysis data are needed at the regional level that is used as primary sampling unit in the sampling design. This can be the municipality level, the postcode sector level or any other relevant regional identification, provided that the number of regional units allows for a multilevel analysis. Snijders and Bosker (1999:44) state that multilevel modelling can become attractive when the number of higher level units is larger than 10. Even though the optimal design for a multilevel analysis depends on the parameter of interest (see also Moerbeek and Wong 2002), this number can serve as a rule of thumb. We are particularly interested in higher level residuals. These will be estimated more precisely when the number of sampled persons within a regional unit is higher. A small number of persons in a regional unit does however not obstruct the analysis. The oversampling design will take the precision of the estimate into account (see below).

The researcher has to collect response rates for the regional units from previous surveys. In principle all possible surveys can be considered, although a similar survey set-up will undoubtedly produce more consistent results: same mode, same or similar sample frame The data have to discriminate between ineligibles and nonresponse.

These data allow for a multilevel logistic regression, or actually two logistic regressions. In the first one the (in)eligible rate is the dependent variable, in the second one the response rate, or actually the product of the eligible rate and the response rate. Both models include the number of sampled persons in the regional unit as an offset. That way the model is equivalent to a multilevel binomial logistic model with respondents nested within regional units.

The analysis can handle unbalanced designs: the number of sample selections of regional units does not need to be the same and the number of sampled respondents may vary across regional units as well (see Snijders and Bosker 1999:166 ff). Actually both play an important role. They will define the confidence interval around the level 2 residual, which will determine the decision whether or not to use a specific ineligible correction or a specific oversampling factor for the regional unit concerned. Apart from the parameters of the multilevel equation, the model indeed can be used to estimate residuals for the regional units. These residuals are sample estimates with a degree of uncertainty. They have (comparative) standard errors that depend on the number of respondents in the regional unit and the between and within variation (Goldstein and Thomas 1996). With the residuals and the confidence interval, regional units can be compared to the general mean. If the residual of a regional unit for the (in)eligible analysis differs significantly from 0, that regional unit has more or fewer ineligibles than on average would be expected. Consequently we will modify the probability of being selected for that regional unit. If the residual of a regional unit for the (non)response analysis is significant, the number of completed interviews in relation to the number of selected persons will be lower or higher in that regional unit. As a result we will modify the number of sampled respon-

dents within the regional unit, if that unit is selected.

We will illustrate this procedure, which is simpler than this explanation probably leads one to suspect, in the following section.

Illustration with the Flemish survey on socio-cultural changes in Flanders

Description of the survey

Since 1996 the Research Centre of the Flemish Government has been conducting an annual face-to-face survey on "Socio-cultural changes in the Flemish region and in Brussels" .[1] The sample size of this survey is about 1500 respondents, who are 18 to 85 years old and speak Dutch. Traditionally a two-stage sampling design was adopted. In the first stage regional units (municipalities) were selected (possibly several times) with chances based on their population size. In the second stage sets of 10 respondents were drawn from the National Population Register in each selected municipality. This procedure retained the equal probability of selection for elementary units (respondents/persons).

The old habit (1996 - 2003) of the survey researchers was to employ nonresponse substitution. All unit nonresponse independent from the reason e.g. refusal, non-contact had to be substituted. A person who could not be interviewed was substituted by a respondent of the same municipality and with a similar age. Of course no new sample was used for this substitution. In the first step about 6000 persons were sampled from the National Register. They were divided in a group of prime sampled units and three separate groups of substitutes, all clustered within the same municipalities. In all four groups the sampled persons were ranked according to their age. If a person appeared to be a nonrespondent he or she was replaced by a sampled person of the next group with the same rank in the same municipality.

Apart from theoretical concerns about the substitution process (see above) we also experienced problems in the cooperation with the bureau responsible for the fieldwork due to the substitution procedure we adopted. We found that interviewers received the substitution addresses before the required number of contact efforts were made (contrary to explicit agreements) (Carton et al. 2005:1/2-3). Consequently, since 2004, we have been using oversampling instead of substitution, still aiming at 1500 respondents.

To define the necessary sample size we dispose of an elaborate documentation from the previous surveys. On average we encountered 6% ineligibles in the surveys in the former years and response rose up to 72% of the eligible persons, resulting in almost 68% completed interviews out of all sampled units (0.676 being the product of the eligible rate (0.94) and the response rate).

With these average numbers different approaches are possible to compute the necessary number of sampled persons. The overall or average number of persons to be sampled can be computed and the oversampling fraction can be applied in the first or in the second stage of the sampling design. The first option draws more municipalities, but keeps the sets of 10 sampled units in each municipality. In total 222 municipalities will be drawn, which corresponds with 2218 respondents that have to be sampled.[2] Actually less than 222 municipalities will be in the sample, because some will be drawn several times, with the probabilities based on population size. In the second option 150 municipalities will be drawn, just as before. But in each municipality sets of 15 persons will be selected or 14.8 times the number of sample selections of the municipality.[3] Because of rounding error probably more respondents will be selected than 2218. The second option will be cheaper than the first, since interviewers can do more interviews in the same municipality.

However, both approaches have the disadvantage that they do not account for regional variation in response rates. Moreover they disregard the distinction between ineligibility and nonresponse. The ineligible and response rates of our survey vary indeed considerably depending on the respondents' place of residence with higher response rates in rural areas as compared with larger cities. To avoid overrepresentation of inhabitants of smaller villages in our final sample we opted for differential regional oversampling. That implies that the second option is partially followed, but that the oversampling factors differ from municipality to municipality. To calculate these oversampling factors we apply multilevel modelling. The data used for that analysis are presented in the following section.

Data structure

We restrict our clarification to the sampling procedures in the Flemish region where (approximately) 1460 interviews have to be accomplished. The remaining 40 interviews have to take place in Brussels, where additional sampling complexities arise. For reasons related to the cooperation with the fieldwork bureau – especially to limit interviewer travel time and expenses – from 2005 on we chose the postcode sector as primary sampling unit instead of the municipality. Consequently, in the following paragraphs the regional unit is always the postal sector, instead of the municipality. For the analysis and the approach we propose, this does not make any difference at all.

We collected response data (ineligible rate and overall completion rate) at the postcode level from 2000 until 2006. We could not use data from the older surveys since these surveys used another form to register the outcomes of the contacts with the respondents. In the seven surveys that we analysed 436 postal sectors were at least once part of the sample (out of a total of 516 possibly sampled ones) and 15474 respondents were sampled in these postal sectors. Our multilevel analysis starts from the dataset that is represented in table 1.

---

[1] Before 2006 the Planning and Statistics Administration of the Ministry of the Flemish Community (APS), a precursor of the Research Centre of the Flemish Government, was responsible for this survey.

[2] 1500/0.676 = 2218

[3] 10/0.676 = 14.8

Table 1: Data structure

Postal sector	Survey	Number of selected respondents	Number of ineligibles	Ineligible rate	Number of completed interviews	Overall Completion rate
1500	2000	24	2	0.08	10	0.42
1500	2001	16	3	0.19	8	0.5
1500	2002	28	2	0.07	17	0.61
1500	2005	33	1	0.03	19	0.58
1501	2000	4	0	0	3	0.75
1501	2001	1	0	0	1	1
1501	2002	9	0	0	6	0.67
1502	2000	9	1	0.11	5	0.56
1502	2001	2	0	0	0	0
1502	2002	10	0	0	7	0.7
1540	2001	13	1	0.08	8	0.62
1541	2001	5	3	0.6	2	0.4
1541	2005	15	4	0.27	3	0.2
1547	2001	14	3	0.21	10	0.71
1560	2000	10	0	0	10	1
1560	2004	9	2	0.22	3	0.33
…	…	…	…	…	…	…
9000	2000	50	6	0.12	19	0.38
9000	2001	82	4	0.05	41	0.5
9000	2002	54	5	0.09	27	0.5
9000	2003	60	7	0.12	38	0.63
9000	2004	39	7	0.18	24	0.62
9000	2005	69	8	0.12	46	0.67
9000	2006	36	6	0.17	24	0.67
…	…	…	…	…	…	…
9992	2005	15	0	0	10	0.67

Table 1 shows that postal sector 1500 was four times part of the sample: in 2000, 2001, 2002 and 2005. The number of selected persons varies between 16 and 33, the ineligible rate for that sector goes from 0.03 to 0.19 and the overall completion rate ranges from 0.42 to 0.61. This completion rate is the product of what generally will be called the eligible rate and the response rate (see e.g. The American Association for Public Opinion Research 2006). Postal sector 1501 is smaller, it was sampled only three times and the number of selected respondents was also smaller. Note that until 2004 the level of clustering was the municipality. Municipalities usually are larger than postcode sectors. Therefore it is perfectly possible to have only one sampled person in a postal sector in 2001. Postal sector 9000 is the centre of Ghent; it is the postal sector with the highest number of inhabitants in Flanders: about 86000. That is more than 30000 inhabitants more than in any other postal sector. Accordingly that sector was always in the sample and the number of sampled persons was rather high as well.

It is clear that the design is highly unbalanced. The number of sample selections of a postal sector varies and the number of sampled units is not the same in the different sectors either. But this is not a problem for the multilevel analysis.

These data allow for two multilevel logistic regressions, one with the ineligible rate as dependent variable and another with the overall completion rate as dependent variable.

With the number of sampled persons as an offset, both models are equivalent to multilevel analyses of persons nested within postal sectors. In both multilevel models we will only use one independent variable: survey or year. That way we account for varying response rates over the years. We effect coded this variable (see e.g. McClendon 1994:215 ff). The main advantage of the effect coding is a more interesting interpretation of the intercept. It can be converted into the grand mean of the expected response and ineligible rates for all surveys and does not refer to the particular rates for one arbitrarily chosen reference year. The results of the analyses and the way to incorporate these results in the calculation of the oversampling factors are presented in the following paragraphs.

Ineligible rate

The first analysis considers the ineligible rate. The results of that analysis are reported in table 2.

Since survey was effect coded, the intercept is an unweighed mean: the average logit that can be transformed into the average ineligible rate for all surveys:

$$\frac{exp\,(-2.702)}{1 + exp\,(-2.702)} = 0.063 \qquad (1)$$

Apparently there were no significant differences in this ineligible rate over the years. None of the survey variables

Table 2: Results of the ineligible analysis

	Parameter	s.e.
Fixed		
Intercept	-2.702	0.052
survey2001	-0.038	0.082
survey2002	0.057	0.084
survey2003	-0.109	0.088
survey2004	0.115	0.084
survey2005	0.066	0.081
survey2006	0.045	0.083
Random		
postal sector level		
σ^2_{cons}	0.441	0.068

has a significant effect. There is however substantial variation at the postcode level. A significance test based on standard errors is only indicative for the random part. Such a test lacks power (Longford 1999; Berkhof and Snijders 2001). Nevertheless the municipality variance of the intercept is considerable and rather high compared to the standard error.

Apart from this general indication of regional variation, we can also have a look at the level 2 residuals and their standard errors. The residuals can be represented graphically (with the rank on the x-axis, so that they go up from the lowest to the highest) and their confidence interval can be displayed by an error bar. With the residuals and the confidence interval, postal sectors can be compared to the general mean. The graphical representation of the residuals with the [±1.96 s.e.] confidence intervals shows the significant nonoverlap with the general mean. Graph 1 reports such a representation for the ineligible analysis.

The graph shows that there is no postal sector for which a significantly negative residual was found. The smallest residual (−0.86, at the left hand side of the graph) was registered in postal sector 2900, but its confidence interval includes zero. The conclusion of this graph is that there is no sector in which a significantly lower ineligible rate was recorded.

There are however postal sectors where significantly higher ineligible rates were encountered. Actually 20 sectors at the right side of the graph have residuals that differ significantly from zero. Their confidence intervals show no overlap with the zero axis. Most of those sectors with high residuals are suburbs of Brussels or are situated near the Walloon region. They have a large number of French speaking inhabitants and that is the explanation for the higher number of ineligible respondents. The highest residual (2.42) at the right end of the graph is e.g. for the sector with the postal code 1950. That is the municipality of Kraainem, situated in the Flemish region, but bordering on the Brussels region. It is forbidden by law in Belgium to register native languages of residents, but, based on election results, the amount of Dutch speaking residents in Kraainem can be estimated to be less than 30%. In the municipality elections of 2000 actually

22% of the voters voted for the joint Flemish list. All other votes went to French speaking parties.[4] Election results are only one indication, but it is clear that the language composition of the municipality is the explanation for the exceptional ineligible rate. Moreover, when examining the contact forms for municipalities like Kraainem, it also becomes clear that language problems are predominantly cited for the ineligibility of the respondents.

Another postal sector attracts the attention in graph 1. The residual of that sector is also highlighted. It has rank 404 (out of 436), but this residual is significant whereas the residuals of several sectors with a higher rank are not. The reason is the smaller confidence interval, which relates to the number of sample selections of the postcode sector and the number of sampled respondents in that sector. The particular sector is the centre of Ghent, postcode 9000, the largest postal sector in Flanders (see above).

We can incorporate the results of this multilevel analysis in the sampling procedure. Actually the ineligible results are used for the first stage of the sampling design: the selection of postcode sectors. That selection uses probabilities based on the population size. We refine the selection mechanism and base the probabilities on the estimated numbers of eligible respondents. For most postcode sectors we estimate that number to be 0.937 times (= 1−0.063) the number of inhabitants according to the National Register. This goes for all the postal sectors that haven't been sampled in the previous years and for sectors with residuals that were not significant in the multilevel analysis (in total 496 out of 516 postal sectors in Flanders).

For the other 20 postcode sectors we take the residual into account. Take the example of Kraainem, the sector with the highest ineligible rate. The residual amounts to 2.422. The number of eligible respondents in Kraainem is estimated to be 0.430 times the number of inhabitants of Kraainem according to the National Register. This is the result of the following formula:

$$\frac{exp\,(-2.702 + 2.422)}{1 + exp\,(-2.702 + 2.422)} = 0.430 \qquad (2)$$

After having calculated the estimated number of eligible respondents for all postal sectors, it is easy to base the probabilities of selection of the sectors (primary sampling units) on that number.

Overall completion rate

The overall completion rate is used to define the number of sampled persons within each selected postcode sector. This overall completion rate comprises the eligible rate and the response rate. The correction based on the completion rate will produce an equal number of eligible and responding units per primary sampling unit, rather than an equal number of selected units. The completion rate is the dependent variable of the second multilevel analysis. The results of that analysis are in table 3.
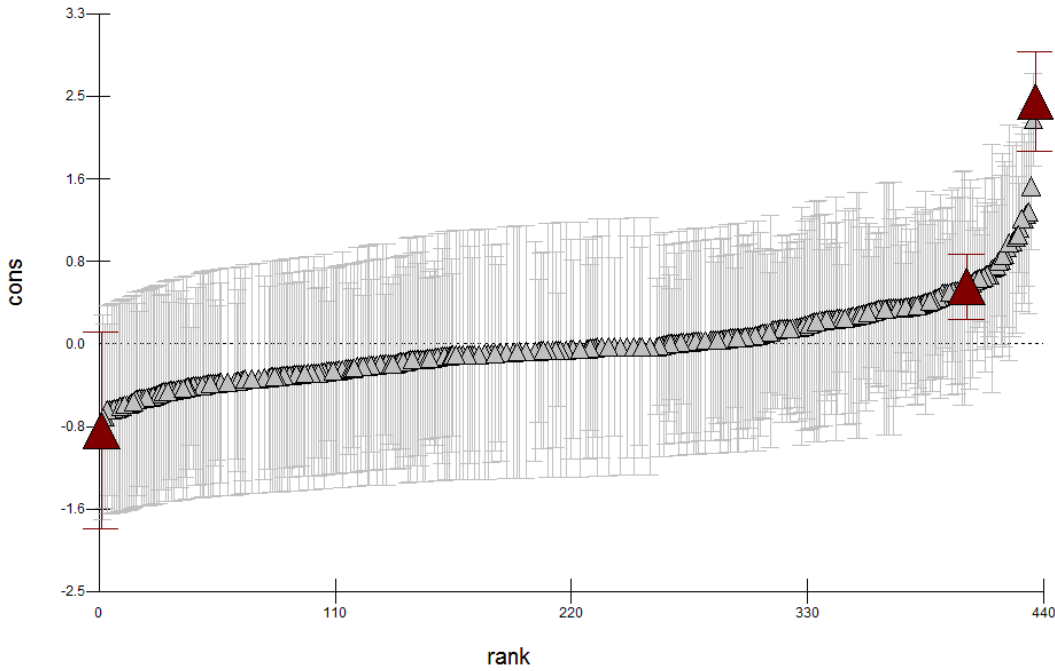
---

[4] http://www.elections.fgov.be/

Figure 1.   Sector residuals in the multilevel analysis of the ineligible rate

Table 3: Results of the analysis of the completion rate

	Parameter	s.e.
Fixed		
Intercept	0.737	0.030
survey2001	-0.152	0.044
survey2002	0.227	0.048
survey2003	-0.037	0.046
survey2004	-0.002	0.047
survey2005	0.110	0.046
survey2006	0.075	0.047
Random		
postal sector level		
σ^2_{cons}	0.177	0.023

Again survey was effect coded; the intercept is the un-weighed mean: the average logit that can be transformed into the average overall completion rate:

$$\frac{exp\,(0.737)}{1 + exp\,(0.737)} = 0.677 \qquad (3)$$

Contrary to the ineligible analysis, table 3 shows significant differences in completion rates over the years. In 2002 and in 2005 response was better than on average, in 2001 it was worse.

Again, there is evidence of regional variation, as the level 2 variance suggests (without interpreting it as an exact test of significance). But for our purpose the postal sector residuals are more interesting than this general indication of regional variation. Those sector residuals are represented in graph 2.

At the right side of the graph there are 5 postal sectors that have significantly higher completion rates. Postcode sector 2275 has the highest residual (0.656). This sector is Lille, a smaller village in the north of Flanders with about 12000 inhabitants. The estimated completion rate for Lille is 80%.

$$\frac{exp\,(0.737 + 0.656)}{1 + exp\,(0.737 + 0.656)} = 0.801 \qquad (4)$$

At the left hand of the graph there are not less than 30 sectors with a significantly negative residual, a number of completed interviews that is significantly lower than on average. The lowest residual was found in postcode sector 1970, with a residual of -1.198. This sector is the municipality of Wezembeek-Oppem. Like Kraainem, Wezembeek-Oppem also borders the Brussels Region and it faces a high ineligible rate, which is reflected in the estimated overall completion rate of 38.7%.

$$\frac{exp\,(0.737 - 1.198)}{1 + exp\,(0.737 - 1.198)} = 0.387 \qquad (5)$$

Finally in graph 2 the centre of Ghent again stands out because of the small confidence interval.

These sector residuals (or the estimated completion rates) are used to compute the necessary oversampling factor. They define the number of sampled persons within each sampled postcode sector. For 481 postal sectors (516 - 35) the oversampling factor is the same as the overall factor: 1.477 (= 1/0.677). In the other 35 sectors we take the municipality
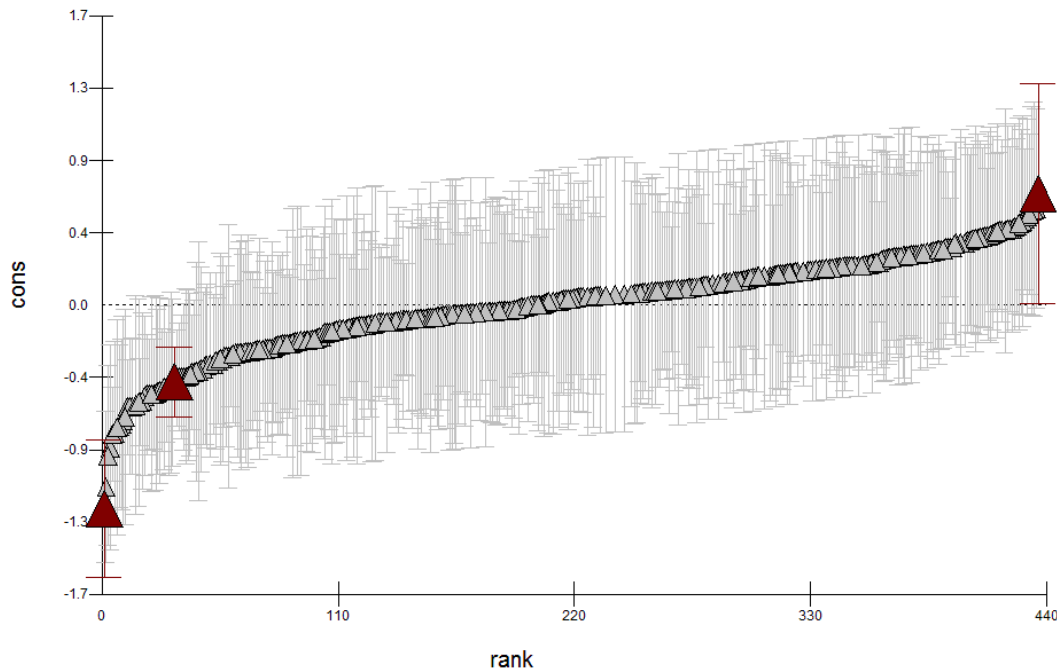
Figure 2. Sector residuals in the multilevel analysis of the completion rate

residual into account. In Lille (sector 2275) e.g. the oversampling factor equals 1.248 (= 1/0.801). If we maintain the cluster size of 10 and Lille as a sector is sampled once, we will sample 12 respondents in that sector. If Wezembeek-Oppem is drawn once as a primary sampling unit, we will sample 26 respondents in that sector (1/0.387 = 2.584). This procedure is followed for all postal sectors to complete the sampling design.

There is a clear link between the ineligible rate and the overall completion rate. In Wezembeek-Oppem the number of completed interviews is low because of the high number of ineligibles (mainly non-Dutch speaking inhabitants). But the correlation is not perfect. Postal sector 3890 e.g. has rank 7 for the completion residuals - only 6 postcode sectors have lower response rates, but its ineligible residual is not significant. Moreover this correlation is not a problem either. We will only oversample Wezembeek-Oppem when it is selected as a primary sampling unit. The probabilities for that selection are based on the estimated number of eligible respondents, which is a correction of the number of inhabitants based on ineligible results of previous surveys. So we decrease the chance that Wezembeek-Oppem will be in the sample. If it will however be sampled, we increase the number of sampled persons to obtain the desired number of completed interviews. The ineligible correction in the first stage of the sampling design avoids overrepresentation of eligible respondents in sectors with a lot of ineligible persons.

Discussion

In this paper we showed how the residuals of a multilevel analysis of response rates can be used to compute various oversampling factors. Our multilevel model was very simple. Apart from the year of the survey[5], we don't include any other variables. As a matter of fact we are not particularly interested in explaining nonresponse (at the individual level) or nonresponse rates (at the regional level). We want to identify exceptional regions (postal sectors) to determine the necessary oversampling factors. For that purpose the reasons for being exceptional are not important. One might argue however that a model with independent level 2 variables could be used to calculate oversampling factors for postal sectors that were not in the sample during the previous years. But, on the other hand, the inclusion of additional variables would change the number of postal sectors with residuals that differ significantly from zero and, since it will never be possible to explain all sector variation, it will complicate the decision process when to use separate oversampling factors: on the basis of the values of the independent variable(s) or because of a residual?

Apart from the independent variables we can also discuss the model itself. Is it better to use random or fixed effects? If we want to identify exceptional postcode sectors, the research question presumes a fixed effects model (see also Snijders and Bosker 1999, pp. 43-44). Actually we apply multilevel analysis and use posterior means for the postal sector specific intercepts - this is a random effects model. In large postal sectors with a lot of respondents, these posterior means will be practically equal to the intercept of a separate regression equation for that sector, which would result from a fixed effects model. For most sectors in this survey these

---

[5] Actually it is only a categorical variable that indicates in which year the response rates were recorded.

posterior means are however pushed a bit towards the general mean (shrinkage to the mean). They can be a rather conservative appraisal of sector differences (Snijders and Bosker 1999:59). But on the other hand the shrinkage expresses the lack of information in small sectors and takes the overall population value into account (Goldstein 1995:24). Even though the number of postal sectors is not infinite and the determination of sector specific oversampling factors conceptually calls for a fixed effects model, that is an important argument to choose the multilevel model (random effects). Moreover the flexibility of the model (when dealing with more than 400 postcode sectors) is also an advantage.

The decision to apply a specific ineligible correction or a specific oversampling factor for a postal sector is based on its residuals being significantly different from the general mean. It should be noted that the significance test we apply, does not allow for a comparison of postal sectors. A comparison of postal sectors would involve different confidence intervals (see Goldstein and Healy 1995). The consequence is that some sectors with a specific oversampling factor may not differ significantly from other sectors without such a sector specific oversampling factor. The only criterion is the significance of the deviation from the mean.

The use of this criterion and the number of postal sectors in the analysis raise the question of the stability of the estimates, especially given the small size of some geographical units. We assessed this stability by examining the effect of gradually adding data, starting with an analysis of only the 2000 data, proceeding with an analysis of the 2000 and 2001 data, and so on until all data (2000 up to 2006) are included. The overall picture of this test is that the number of postal sectors with significant residuals logically increases as more data are included. More importantly, in general there are only a few sectors with significant residuals in earlier analyses that become not significant as a consequence of adding another year to the data. It is no surprise that results tend to stabilise as more data become available. Nevertheless it is a reassurance that the application of the confidence intervals prevents us from sorting out sectors too soon. The results of our test for the analysis of the ineligible rate are reported in the appendix.

Our analysis suffers from the restricted possibility to distinguish interviewer and area effects (O'Muircheartaigh and Campanelli 1998). However, on the whole, the postal sectors with specific oversampling fractions are often those that were more frequently sampled for the survey with on average more sampled units (because of larger population size). Moreover the fieldwork bureau was not the same for all surveys. Consequently there were generally several interviewers doing the interviewing work in the larger sectors and in the sectors that were sampled more often. That is a reassurance that we are not modelling interviewer effects.

It is clear that our oversampling design only affects a limited number of postcode sectors. In most sectors the result will be identical. We changed the probabilities of selection of a sector only for 20 sectors and modified the number of sampled respondents in selected sectors only 35 times. Given the total number of 516 postcode sectors in Flanders, it is plain

that our oversampling design results in a limited correction of the principle of equal probabilities of selection of elementary units rather than mixing it up completely. However, although our sampling design maintains the principles of probability sampling, we cannot pretend that there is an equal probability of being selected in the sample. This impact of the oversampling can be assessed with the design effect. The design effect due to unequal inclusion probabilities for sample surveys is a function of the variance of the weights (Gabler, Häder and Lynn 2006). When combining the sample selection weight of our design with a nonresponse weight based on the regional response rates, the resulting weight is bound to have a smaller variance than the nonresponse weight itself, as the first weight is based on an estimate of the second. In our sample we can combine the sample selection weight with a nonresponse weight that is the reciprocal of the postal sector response rate. As a result of the expected smaller variance of the weight the survey will have a smaller design effect and accordingly a larger effective sample size than a survey that only weights for varying regional response rates. This is the correlate of an equal or similar probability of having a completed interview, instead of an equal probability of selection. The design with the ineligible correction and the oversampling fraction is set as to produce an equal (estimated) number of eligible and responding units per primary sampling unit, not an equal number of selected units.

Since the 2007 survey data are available we can test our hypothesis of a smaller design effect. In 2007 we expected 1460 interviews in Flanders. These interviews had to take place in 140 postal sectors - a few sectors were sampled more than once. We had interviewer problems in two sectors. In both sectors only one interview was accomplished. We do not take these sectors into account in our test, because we don't want to apply weights that amount to approximately 10. That leaves us with 1440 expected interviews in 138 sectors. Response was a bit lower than expected and finally we ended up with 1403 interviews. We calculated two different weights for these respondents: a nonresponse weight that accounts for regional variation in nonresponse and a weight that comprises the design effect and the nonresponse. In order to make a correct comparison, we have rescaled both weights so that they sum up to 1403. The variance of the nonresponse weight equals 0.058. The variance of the weight that combines the reciprocal of the selection probabilities and the nonresponse is equal to 0.053. It is a small difference but it is in favour of our approach. Note that since we have used oversampling, normally one wouldn't calculate nonresponse weights the way we have done it for the first weight. The comparison is not perfect, because we don't have a split half design. Nevertheless we have some (small) evidence that our oversampling approach results in a smaller design effect due to unequal inclusion probabilities than an approach that only weighs for nonresponse.

Another argument to use the oversampling is that nonresponse weighting will reduce bias of estimates for the total population, but it will not improve the estimates for the otherwise underrepresented regions. In the example of the National Travel Survey in the UK, nonresponse weighting might

compensate the lower response rates in London when estimating parameters for England of the UK, but it will never improve the estimates for London itself. If the underrepresentation of respondents in regional units results in too small numbers, oversampling can be an option.

A final argument is more practical, though not trivial either. Our oversampling design results in similar interviewer workloads. For all selected clusters (or for each time a sector is selected) we expect 10 completed interviews. As the arrangement into interviewer workloads usually follows the clustering, there will be much less variation in the amount of interviewing to do. Although more similarity in the number of interviews goes together with more variation in the number of selected units, which also has practical implications, this is an advantage for the organisation of the fieldwork.

Acknowledgements

References

Abraham, K. G., Maitland, A., & Bianchi, S. M. (2006). Nonresponse in the American Time Use Survey. *Public Opinion Quarterly*, *70*, 676-703.

Berkhof, J., & Snijders, T. A. B. (2001). Variance Component Testing in Multilevel Models. *Journal of Educational and Behavioral Statistics*, *26*, 133-152.

Carton, A., Van Geel, H., & De Pelsemaeker, S. (2005). *Basisdocumentatie: Sociaal-culturele verschuivingen in Vlaanderen 2004.* (Brussel: Ministerie van de Vlaamse Gemeenschap, adminstratie Planning en Statistiek)

Chapman, D. W. (2003). To substitute or not to substitute that is the question. *The Survey Statistician*, *48*, 32-34.

Deming, W. E. (1950). *Some theory of sampling*. New York: Wiley.

Dillman, D. A., Eltinge, J. L., Groves, R. M., & Little, R. J. (2002). Survey Nonresponse in Design, Data Collection, and Analysis. In R. M. Groves, D. A. Dillman, J. E. Eltinge, & R. J. Little (Eds.), *Survey nonresponse* (p. 3-26). New York: Wiley.

ESS (European Social Survey Round 1). (2004, June). *2002/2003 Technical Report Edition 2.* Available from `http://naticent02.uuhost.uk.uu.net/archive/tech_report.htm` (Retrieved February 9, 2007)

Gabler, S., Häder, S., & Lynn, P. (2006). Design Effects for Multiple Design Samples. *Survey Methodology*, *32*, 115-120.

Goldstein, H. (1995). *Multilevel Statistical Models*. London: Edward Arnold.

Goldstein, H., & Healy, M. J. R. (1995). The Graphical Presentation of a Collection of Means. *Journal of the Royal Statistical Society. Series A. Statistics in Society*, *158*, 175-177.

Goldstein, H., & Thomas, S. (1996). Using Examination Results as Indicators of School and College Performance. *Journal of the Royal Statistical Society. Series A. Statistics in Society*, *159*, 149-163.

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

Groves, R. M., & Couper, M. P. (1998). *Nonresponse in Household Surveys*. New York: Wiley.

Groves, R. M., Dillman, D. A., Eltinge, J. E., & Little, R. A. (Eds.). (2002). *Survey Nonresponse*. New York: Wiley.

ISSP. (2003, February). *Report of the Standing and Methodology Committees to the General ISSP Meeting.* (International Social Survey Programme)

Kershaw, A. (2002). *National Travel Survey. Technical Report 2001*. London: Office for National Statistics.

Kish, L. (1992). Weighting for Unequal p_i. *Journal of Official Statistics*, *8*, 183-200.

Little, R. J., & Vartivarian, S. (2005). Does Weighting for Nonresponse Increase the Variance of Survey Means? *Survey Methodology*, *31*, 161-168.

Longford, N. T. (1999). Standard Errors in Multilevel Analysis. *Multilevel Modelling Newsletter*, *11*, 10-13.

Lynn, P. (2004). The use of substitution in surveys. *The Survey Statistician*, *49*, 14-16.

Lynn, P., Häder, S., Gables, S., & Laaksonen, S. (2007). Methods for Achieving Equivalence of Samples in Cross-National Surveys: The European Social Survey Experience. *Journal of Official Statistics*, *23*, 107-124.

McClendon, M. J. (1994). *Multiple Regression and Causal Analysis*. Illinois: Waveland.

Moerbeek, M., & Wong, W. K. (2002). Multiple-Objective Optimal Designs for the Hierarchical Linear Model. *Journal of Official Statistics*, *18*, 291-303.

O'Muircheartaigh, C., & Campanelli, P. (1998). The Relative Impact of Interviewer Effects and Sample Design Effects on Survey Precision. *Journal of the Royal Statistical Society. Series A. Statistics in Society*, *161*, 63-77.

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel Analysis. An introduction to basic and advanced multilevel modeling*. Newbury Park, London: Sage.

The American Association for Public Opinion Research. (2006). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys* (4th ed.). Lenexa Kansas: AAPOR.

Vehovar, V. (1999). Field substitution and unit nonresponse. *Journal of Official Statistics*, *15*, 335-350.

Vehovar, V. (2003). Field substitutions redefined. *The Survey Statistician*, *48*, 35-37.

Appendix: Assessment of the stability of the estimates

Table 4 in this appendix reports the results of seven separate analyses of the ineligible rate. The analyses gradually include more data. The postal sectors that have significant residuals in the analyses are marked in the table.

It is clear and also evident that the number of sectors with significant residuals increases as more data are included. From the point of view of the stability of the estimates it is however more important that only a few sectors with significant residuals in earlier analyses, get non-significant residuals as a consequence of adding another year to the data. The most notable exception to this rule is sector 2850. We obtain significant residuals in all analyses except from the last one. The results for sector 1700 are not stable either. It has significant residuals in 4 out of the 7 analyses. Those sectors do not follow the overall picture of the table.

The results of sector 1541 are interesting, because they can be compared with the data in table 1. As table 1 showed this sector was sampled in 2001 and in 2005. In 2001 3 out of 5

Table 4: Results of seven separate analyses of the ineligible rate

	significant residuals for the analysis of the ineligible rate						
postal sector	analysis of 2000 data	analysis of 2000-2001 data	analysis of 2000-2002 data	analysis of 2000-2003 data	analysis of 2000-2004 data	analysis of 2000-2005 data	analysis of 2000-2006 data
1541						x	x
1600						x	
1630							x
1640			x	x	x	x	x
1650						x	x
1700	x			x	x	x	
1780		x	x	x	x	x	x
1800	x	x	x	x	x	x	x
1932	x	x	x	x	x	x	x
1950		x	x	x	x	x	x
1970			x	x	x	x	x
2050						x	x
2140						x	x
2220	x	x					
2800			x	x	x	x	x
2850	x	x	x	x	x	x	
3040	x	x	x	x	x	x	x
3080						x	x
3202	x	x					
3320							x
3790		x	x	x	x	x	x
3791		x					
3798		x	x	x	x	x	x
8500					x		
9000					x	x	x
9050					x	x	x
9600	x	x	x	x	x	x	x

sampled units appeared ineligible, which corresponds to an ineligible rate of 0.6. But despite this exceptional ineligible rate this sector does not get a significant residual as a consequence of the small number of sampled units and the accordingly large confidence interval. It is only after the 2005 survey, when 4 out of the 15 sampled units appeared ineligible (ineligible rate of 0.27) that the sector becomes exceptional - according to the multilevel analysis. The application of the confidence intervals prevents us from sorting out this sector too soon.

A similar table for the analysis of the overall completion rate, which contains more postal sectors, is available upon request.