

## Article

Symposium 2008:  
Data Collection: Challenges, Achievements and New Directions

### **The Effect of Non-response Follow-up in a Survey on Living Conditions among Immigrants in Norway**

by Johan Fosen, Gustav Haraldsen, and Øyvin Kleven

2009



## The Effect of Non-response Follow-up in a Survey on Living Conditions among Immigrants in Norway

Johan Fosen, Gustav Haraldsen, and Øyvin Kleven<sup>1</sup>

### Abstract

In a survey where results were estimated by simple averages, we will compare the effect on the results of a follow-up among non-respondents, and weighting based on the last ten percents of the respondents. The data used are collected from a Survey of Living Conditions among Immigrants in Norway that was carried out in 2006.

Key Words: Non-response bias, Follow-up, Weighting

### 1. Introduction

At the end of the data collection process for a survey, we sometimes face the question of whether or not to extend the data collection period. This additional data collection period is often denoted *second wave collection* or *follow-up*, an approach originally discussed by Hansen & Hurwitz (1946). The aim usually is to decrease the nonresponse bias, but sometimes just to increase the net sample. In a survey on living conditions among immigrants, that was carried out in Norway in 2006, such considerations had to be made.

The Survey on Living Conditions among immigrants had an originally scheduled end of data collection at June 1 2006. However, since the response rate was as low as 54 percent as opposed to a planned response rate of 70 percent, it was decided to prolong the data collection period, i.e. perform a follow-up or a second wave data collection. The survey will be presented in more detail in the next section. In Section 3, we will compare the results for some key variables before and after the follow-up, using the simple estimation method that Statistics Norway used when they disseminated the results. Previous studies of other surveys have revealed that follow-up may reduce the bias (e.g. Thomsen et al. 2006; Fosen & Kleven 2007), and that was what we hoped for in this survey as well. Hence, the first question is if the follow up effort paid off.

In Section 4, we will study the effect of replacing follow-up with a more advanced estimation method where we assume the latest arrived respondents to be those most similar to the non-respondents. We will denote this method *late-respondents weighting* and use the procedure described by Bartholomew (1961), and used e.g. in Thomsen & Siring (1983). At the end of Section 4, we will also consider the effect and cost efficiency of using both follow-up and late-respondents weighting.

### 2. The Survey of Living Conditions among Immigrants in Norway

Figure 2-1 summarizes the main characteristics of The Survey of Living Conditions among Immigrants in Norway (Gulløy 2008). As we can see the survey covered immigrants from 10 countries. The clients were concerned about nonresponse and the Survey division, who carried out the survey, set a goal of 70% response rate. When the original field period was ended in summer 2006, however, the response rate was only 54%.

In hope of increased contact rates and thereby higher response rates, major resources were put into an extension phase of data collection. New procedures for follow up were introduced, and the survey unit was re-assigned to staff

---

<sup>1</sup>Johan Fosen, Statistics Norway, P. O. Box 8131 Dep, Norway, N-0033 Oslo (Johan.Fosen@ssb.no);  
Gustav Haraldsen, Statistics Norway, P. O. Box 8131 Dep, Norway, N-0033 Oslo (Gustav.Haraldsen@ssb.no);  
Øyvin Kleven, Statistics Norway, P. O. Box 8131 Dep, Norway, N-0033 Oslo (Oyvin.Kleven@ssb.no)

with higher dedication possibilities. After another seven months, the response rate had risen to 64 percent, and the survey was closed.

**Figure 2-1**

**Some key information on The Survey of Living Conditions among Immigrants in Norway.**

- Gross Sample = Stratified sample of 500 from 10 most important immigrant groups = 5000
- Net Sample = 64%
- Data Collection Methods: Multilingual interviewers mainly recruited from immigrant groups and trained for this survey
  - Personal interviews = 62%
  - Telephone interviews = 38%
- Waves of data collection:
  - September 2005 to Summer 2006
  - Summer 2006 to January 2007

### 3. The effects of the follow-up

When it was decided to do a second wave collection in order to increase the response rate, the underlying assumption was that this would also lead to less *non-response bias*. Non-response bias is the systematic error in the survey results caused by non-response. Bias is only meaningful in relation to a specific variable, and we can talk of non-response bias for any of the estimated results in the survey. In this paper, the term ‘bias’ will refer to the *conditional bias* given the gross sample<sup>2</sup>. We measure it simply by the difference between the net sample result and the gross sample result, i.e.  $\text{bias } \bar{Y} = \bar{Y}_{net} - \bar{Y}_{gross}$ .

Unfortunately, the gross sample result is generally unknown for the survey variables. However, the relative difference in bias before and after the second wave is

$$\frac{\text{bias } \bar{Y}_{net,orig} - \text{bias } \bar{Y}_{net,final}}{\bar{Y}_{gross}} = \frac{\bar{Y}_{net,orig} - \bar{Y}_{net,final}}{\bar{Y}_{gross}},$$

where  $\bar{Y}_{net,orig}$  and  $\bar{Y}_{net,final}$  are the averages in the original and final net sample respectively. The relative bias is proportional to the difference between the two net sample averages, hence the difference itself tells something about the relative difference in bias due to follow-up.

For some non-survey variables, we actually know the gross sample average. In Statistics Norway, many variables from administrative registers are potentially available since in Statistics Norway we have a unique identification key that we can use to merge the survey data with the register data. We can regard each register variable as a survey variable where the response rate is 100 percent, and we can find the bias of the register variable in each of the net samples.

---

<sup>2</sup> The *unconditional* bias is the average difference from the population result, over many different samples. Since we have only one sample, we will consider the bias given this gross sample, although the unconditional bias could have been studied using resampling techniques.

**Figure 3-1**  
**The non-response bias of distribution of countries. For preliminary net sample (before follow-up) and final net sample.**

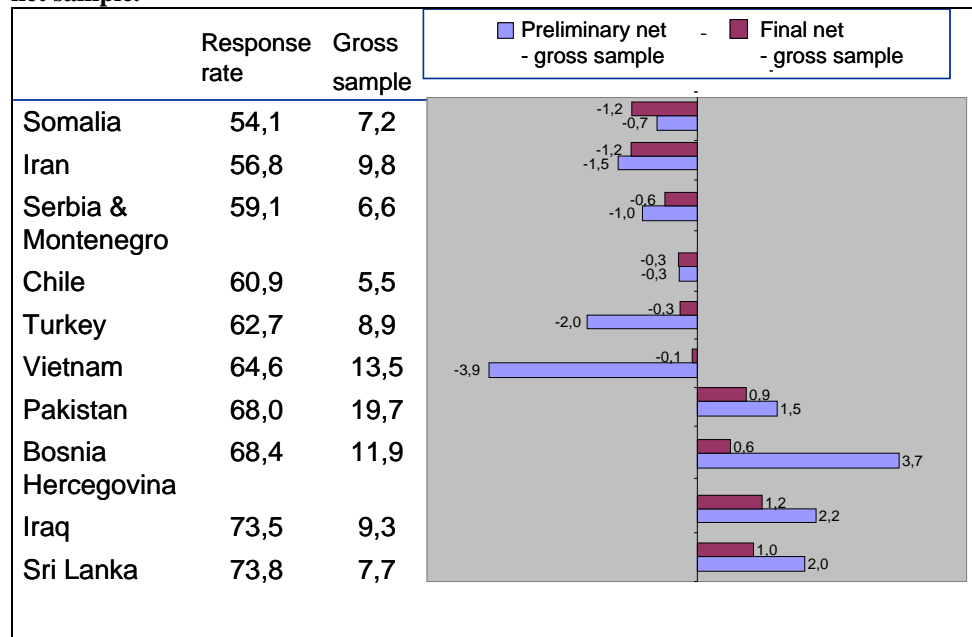


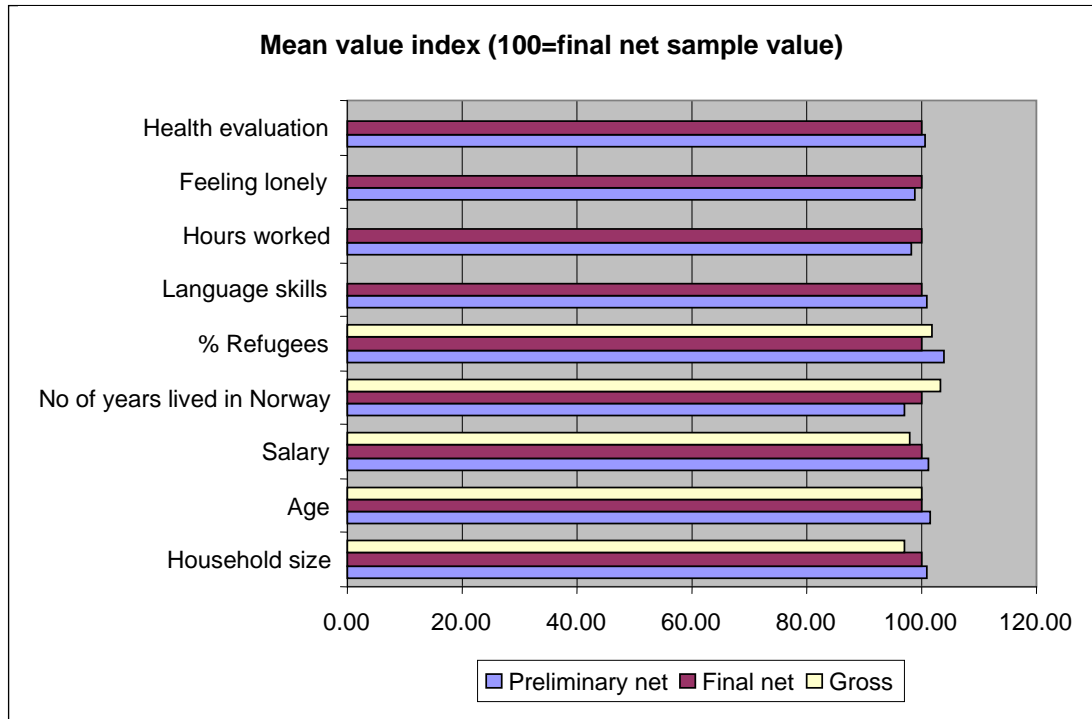
Figure 3-1 shows how follow-up affected the distribution of country of origin. The countries are sorted by ascending response rate, ranging from just above 50 percent to almost 75 percent. For each country, the table considers the variable ‘proportion of respondents having this country as country of origin’, and the difference between the bars is the change of bias due to the follow-up. As we can see, the bias decreased for most countries, and especially for Turkey, Vietnam and Bosnia Herzegovina. The only country with an increased bias was Somalia. Consequently, it seems fair to conclude that the follow up led to a better distribution between countries than in the preliminary net sample.

For several survey and register variables, we wanted to compare the average values before and after follow-up. Where possible, we also wanted to compare with the true value of the gross sample. This was done for two kinds of register variables. Firstly, we looked at traditional auxiliary variables like household size, age and salary. Secondly, we also picked some variables that would have been survey variables unless already available from the registers. These were the number of years lived in Norway and the percentage that came to Norway as refugees (contrasted to other immigration causes). In addition to this, we look at some surveys variables, and these variables were chosen in order to cover a wide range of question types:

- perceived health condition was chosen because health was the main issue in the survey
- perceived loneliness was chosen in order to include a subjective evaluation variable
- hours worked was chosen in order to include a variable that has to do with behavioural patterns
- language skill was chosen as an indicator of social integration

Using the variables above, Figure 3-2 shows the relative size of the preliminary and gross sample results, both compared to the final net sample. Looking at those variables that can be compared with gross sample information, the results came closer during the second wave, but even before the follow up the results were surprisingly close to the true values. We can e.g. see that the preliminary household size average was about one percent higher than that of the final net sample. The survey variables also changed during the follow up, but did not change very much.

**Figure 3-2**  
**Mean value as percent of final net sample value, for selected variables. For preliminary net, final net, and gross sample.**



In the following, we will look closer at how the results for some of these variables changed during the data collection period. The graphs in Figure 3-3 and Figure 3-4 shows how the survey estimates changed as the response rate increased. Where available, the gross sample average is indicated by a straight line (Figure 3-3).

The left panel of Figure 3-3 reveals that there was under-coverage of smaller households. This is as expected. Respondents in smaller households are generally more difficult to get into contact with than respondents in bigger households. There was a small but steady decrease in this bias during the follow up period.

Perhaps more surprisingly, we see in the right panel that those who had stayed in Norway for a long time were more difficult to recruit than those who have lived in Norway for a shorter time. This result questions the notion that low integrated people tend to have a higher nonresponse rate than well-integrated people. We notice that the results changed very little during the second wave of data collection. Yet it ended up slightly closer to the gross sample value after the follow up.

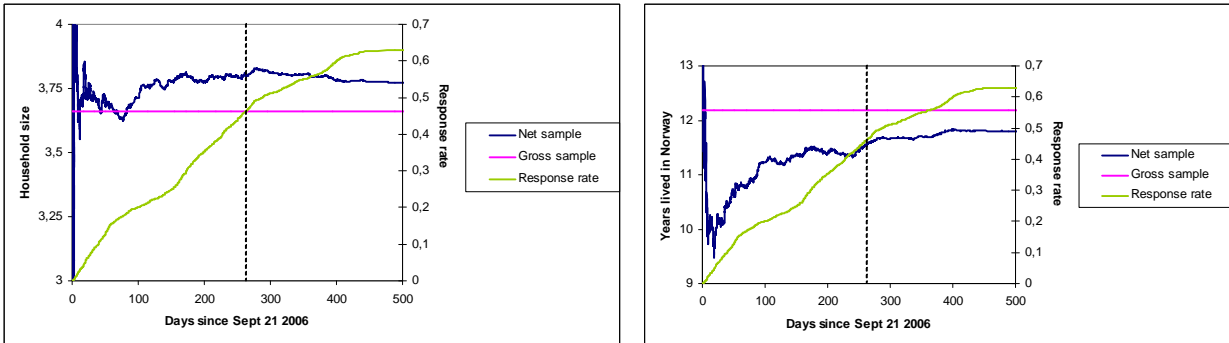
Figure 3-4 shows the average of some survey variables as time goes by. The average of loneliness (upper left panel) is hardly changing during the second wave. Especially when we take into consideration that the average value is based on the simplifying assumption that loneliness is a numeric variable, we cannot determine in which direction the loneliness variable moves during the second wave.

Before the follow up period, only four percent characterised their health to be poor or very poor. Looking at the trend as the response rate increase in the right panel of Figure 3-4 we may suspect that the health situation is not this positive. But again, the second wave of data collection does not seem to add much information on this.

As already stated, those who worked long hours were more difficult to recruit than other were, and the lower panel of Figure 3-4 confirms this. In this case, the estimates seem to improve a little bit during the follow up period.

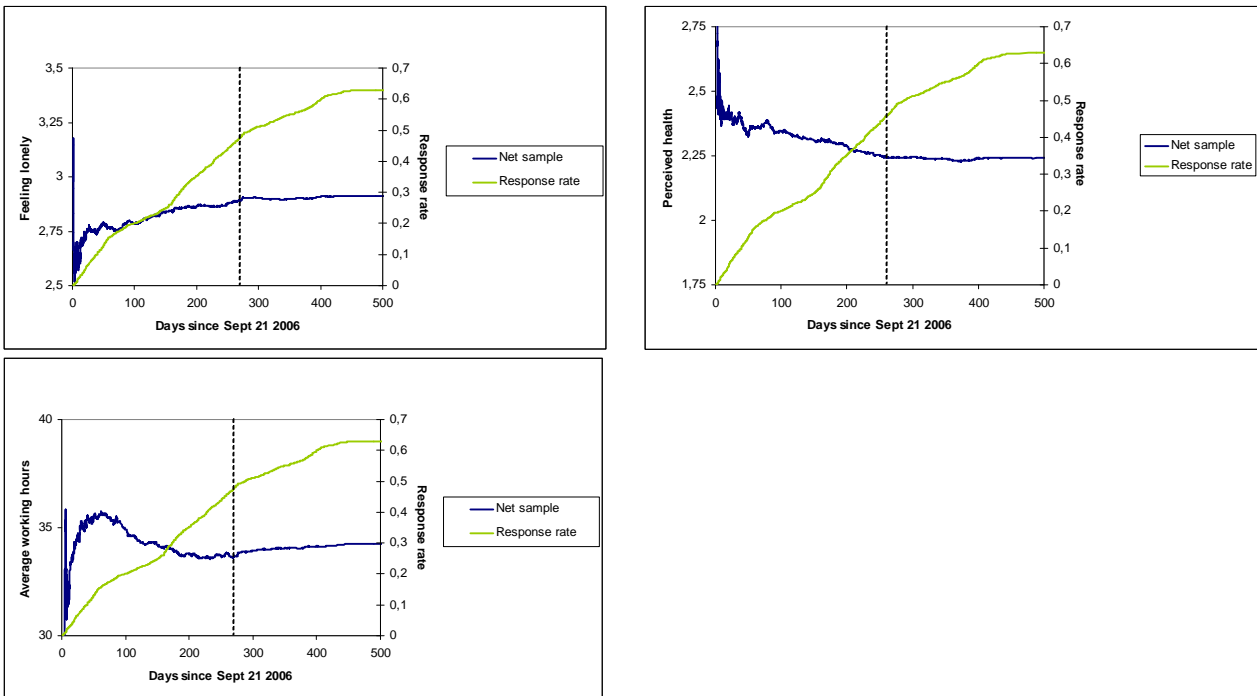
**Figure 3-3**

The average value of household size (left panel) and years lived in Norway (right panel). For each net sample of completed interviews up to each date (blue), and for gross sample (pink). Response rate (green) showed on the second y-axis.



**Figure 3-4**

The average value of feeling lonely index (upper left panel), perceived health index (right panel), and average hours worked (lower left panel). For each net sample of completed interviews up to each date (blue). Response rate (green) showed on the second y-axis.



To wrap up this far:

- The distribution of countries was better in the final net sample than in the preliminary one.
- Results known for the gross sample were surprisingly close to the net sample results even before the follow-up.
- Most results stabilised before the follow up period and very little happened during the second wave of data collection.
- Even so, non response bias probably affected some of the results.

#### 4. Late-respondent weighting

To justify putting many resources into recruiting more respondents, we implicitly assume that the extra respondents recruited during the follow-up will be more similar to the non-respondents than the original net sample is. Only then can a marginal increase in the response rate lead to a better estimate. However, for the follow-up itself, we only apply this *late respondent assumption* as a motivation for doing the follow-up. As long as we estimate based on simple averages, we assume the non-respondents to have the same average as the final net-sample (Bartholomew 1961).

The idea of late respondent weighting is to estimate the results under the late respondent assumption. In the form suggested by Bartholomew (1961), we assume that the non-respondents are equal to the second wave respondents, i.e. we predict the average value among the non-respondents to be identical to the average among the second wave respondents. This weighting differs from simple averages only by the late-respondent assumption and hence it will perform poorer than simple averages whenever the assumption fails.

We looked at the weighting method above for three of the survey variables: perceived health, feeling of loneliness and years lived in Norway.

**Table 4-1**

**The average value of perceived health index, feeling lonely index, and years lived in Norway. Using simple averages (column 1 and 2) and late-respondent weighting (column 3 and 4), and using the preliminary net sample (column 1 and 3) and final net sample (column 2 and 4).**

	Unweighted preliminary result	Unweighted final survey result	Weighting based on the last 10 % before the preliminary net sample	Weighting based on the last 10 % before the final net sample
Perceived health condition	2.25	2.24	2.21	2.27
Feeling of loneliness	2.88	2.91	2.92	2.95
Years lived in Norway	11.4	11.8	11.8	12.0
			(True value = 12.2)	

We implemented the late-respondent weighting by defining the 10 percent last arrived respondents as “the late respondents”. Then we can use the late-respondent weighting both with and without follow-up.<sup>3</sup> This leads to four alternative results shown in the four columns of Table 4-1.

Statistics Norway used the alternative of column 2 when they published the results. An interesting question that Table 4-1 can cast some light over is how much worse/better the results (bias) would have been if

- follow-up just had been dropped and the results would just have been simple averages based on the original net sample (alternative in column 1)
- follow-up had been dropped and replaced by late-respondent-weighting (alternative 3)
- follow-up had been performed, but late-respondent weighting had been used instead of simple averages (alternative 4)

<sup>3</sup> On the contrary, if we were only to use late-respondent weighting for the case with follow-up, we would have defined all respondents arriving during follow-up as the late-respondents.

Table 4-1 tells us that for the variables “years lived in Norway” and “loneliness”, dropping follow-up (alternative 1) would only lead to a small change in the estimates, and more importantly that the same result could have been achieved by weighting instead of follow-up (alternative 3). Since the latter alternative is the less expensive, this alternative seems to be preferable if we should conclude on these two variables only. An even better result would probably be accomplished by weighting in addition to follow-up (alternative 4). However, the improvement is rather small, and consequently not much quality gained when compared to the cost of following up.

For “perceived health”, the situation is different from the other two variables. The most interesting feature is perhaps that whereas just dropping follow-up made almost no difference to the results (2.25 versus 2.24), the late-respondent weighting would decrease the estimates when done without follow-up, but *increase* the estimates when done *with* follow-up. This means that the last respondents before follow-up are different from the last respondents that we get in the follow-up process. In one of these two situations, the assumption behind the weighting obviously does not hold.

We get a better impression of this phenomenon when returning to the right panel of Figure 3-3. There we can see that the average value is not monotonously increasing or decreasing towards the true value, but going up in some intervals and going down in others. Before we choose to use late-respondent weighting, we should ensure that the average is monotonous except for the first part of the graph. The reason for omitting this first part is that here we would expect the change of the average to be caused by randomness.

## 5. Conclusions

Our results seem to indicate that in this survey there was only a marginal improvement achieved by doing a second wave data collection. However, we have only investigated a small part of all the variables that were collected in the survey. Also, the results were to a large extent published for sub-groups. Even if the follow up did not affect the average in the total net sample, the effects may be different in subgroups. Hence, we have insufficient empirical basis for suggesting that the second wave should have been dropped. Our point in this paper is more to point at a method that could be expanded and used to decide whether a second wave data collection should be carried out.

Late respondent weighting can sometimes be an inexpensive way of improving the results, and will sometimes accomplish just as much as the second wave. However, the late respondent assumption will sometimes not be true and this weighting method should therefore be used with great caution. Some indication of the validity of the assumption can be found in graphs such as those in Figure 3-3. Except for the first part of the graph, the trend should be monotonous when we choose to use this weighting.

## References

- Bartholomew, D.J. (1961). A Method of Allowing for ‘not-at-home’ Bias in Sample Surveys, *Applied Statistics*, 10, 52-59.
- Fosen, J. and Kleven, Ø. (2007). Non Response, Bias and Costs in the Norwegian Consumer Expenditure Survey, *18<sup>th</sup> Annual International Workshop on Household Survey Nonresponse*, Southampton, UK.
- Gulløy, E. (2008). Living conditions among immigrants 2005/2006. Documentation report (in Norwegian), Notater 2008/5, Statistics Norway.
- Hansen, M.H. and Hurwitz, W.N. (1946). The Problem of Nonresponse in Sample Surveys, *Journal of the American Statistical Association*, 41, 517-529.
- Thomsen, I.O., Kleven, Ø., Wang, J.E. and Zhang, L.C. (2006). Coping with Decreasing Nonresponse Rates in Statistics Norway. Recommended Practice for Reducing the Effect of Nonresponse, Reports 2006/29, Statistics Norway.



Thomsen, I. and Siring, E. (1983). On the Causes and Effects of Nonresponse: Norwegian Experiences, *Incomplete Data in Sample Surveys*, (Eds. W.G. Madow and Olkin) vol. 3, New York: Academic Press.