Nonresponse in sample surveys: methods for analysis and adjustment

Cobben, F.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 10

Summary and Conclusions

In this thesis, we have developed methods for the treatment of unit nonresponse in sample surveys of households. Sample elements either respond to the survey request, or they do not. However, the reasons for nonresponse vary. This leads to different types of respondents. For example, persons that cannot be contacted during the fieldwork period often have different reasons not to respond than persons that refuse participation. To compare surveys, over time or across countries, it is important to distinguish between different types of respondents. When the composition of the nonresponse varies over surveys, and in the treatment of nonresponse no distinction is made between different types of respondents, the conclusions based on the surveys will reflect changes in the composition of the response instead of real changes in society. Furthermore, the causes, correlates and the effect on survey estimates are known to be different for different response types. The methods for nonresponse analysis and adjustment for nonresponse bias hence fit the data better when different (non)respondents are not combined into one group, but distinguished by cause. Throughout this thesis, we therefore distinguish between different response types.

Traditionally, treatment of nonresponse starts by reducing nonresponse in the field. Despite methods to prevent nonresponse, the response will never be 100%, nor will it be completely non-selective. Hence after the data have been collected, an estimation technique is applied to the data to adjust for nonresponse bias. Recently, nonresponse reducers have shifted focus from increasing response rates to obtaining a more balanced composition of the response. Amongst others, the research into representativeness indicators described in Chapter 5 of this thesis has contributed to this shift in paradigm. The tendency of research
on nonresponse reduction to balance the composition of the response resembles nonresponse adjustment, where the difference is in the timing since nonresponse adjustment balances the composition after the data have been collected.

On the interface of nonresponse reduction and adjustment for nonresponse bias lies analysis of response behaviour. Nonresponse analysis reveals how respondents differ from nonrespondents. This information is useful for the construction of weighting models to adjust for nonresponse bias, as well as for the construction of models to predict response behaviour. At the same time, analysis of nonresponse points at characteristics of the sample that have become unbalanced due to nonresponse. The nonresponse can be efficiently reduced by concentrating nonresponse reduction efforts on these unbalanced groups. In Chapter 5 of this thesis, we have developed an indicator for the quality of survey response that can be seen as a tool for the analysis of response behaviour that bridges the gap between adjustment and reduction of nonresponse. This so-called R-indicator is an indicator for the representativeness of survey response. It describes how balanced the composition of the response is compared to the composition of the sample, with respect to a set of predefined characteristics. The R-indicator can be used for multiple purposes. It can serve as an indicator for survey quality, supplementary to the response rate. But it can also be used during the fieldwork, to efficiently enhance response by concentrating fieldwork efforts on underrepresented groups. By using the R-indicator in an early stage of the data collection process, different strategies can be followed based on the course of the data collection. This will lead to a responsive survey design (Groves and Heeringa, 2005), or dynamic data collection strategies (Bethlehem and Schouten, 2008). At a time when survey costs and respondents’ burden have to be reduced while preserving survey quality, dynamic survey design has great potential.

The analysis of response behaviour becomes increasingly important in dynamic survey design, since response behaviour will serve as the input for constructing different strategies. The R-indicator serves as a tool for analysing response behaviour; it summarizes multivariate information on response behaviour in an easy to interpret, univariate figure. The R-indicator is based on estimated response probabilities, or response propensities. Several methods can be used to compute these. As we have outlined before, different response types should be distinguished, leading to a sequential representation of the response process. For example, sample elements have to be contacted first, before we can observe whether or not they participate. In Chapter 8 of this thesis, we have developed methods that account for the sequential nature of the response process. These methods consist of a number of equations; one for every response type. Like
this, the response process is closely followed and it becomes possible to use different variables in each equation, as well as to allow for different relations between variables and the type of respondent. Another issue that we have dealt with, is the correlation between different types of response. A correlation arises when there are unobserved characteristics that influence more than one type of response, or when sample elements are misclassified.

For researchers that are interested in the quality of the obtained response, the R-indicator can be computed to supplement the response rate. If, however, the aim of analysing response behaviour is to build a weighting model for nonresponse adjustment, the described methods in Chapter 8 can be applied. Depending on the amount of auxiliary information, both socio-economic and demographic as well as information about the data collection process or paradata, different models can and should be used. For instance, when only little information is available it is not possible to closely describe the response process. However, when more detailed information is available, we have shown in Chapters 3 and 8 that by using more complex methods we are better able to understand the causes and correlates of (non)response. It is therefore important to collect good auxiliary information. Preferably, this information is obtained by linking the survey sample to one or more registers. Like this, individual data about both respondents and nonrespondents is obtained. Other ways to collect data on nonrespondents are: using the population characteristics published by the national statistical office; collecting interviewer observations and other information from the data collection process, i.e. paradata; or re-approaching nonrespondents.

A topic for future research is the integration of methods to compute response propensities with different types of respondents and the R-indicator. In addition, a different R-indicator can be constructed for each type of respondent. This information could prove useful in dynamic survey design, for example when the fieldwork organisation employs measures to reduce the non-contact rate but does not attempt a refusal conversion. The R-indicator can be employed to compute the representativeness of the subsample of respondents, if auxiliary information is available for respondents and nonrespondents. Organisations that do not have this type of information cannot compute the indicator. However, national statistical offices (NSO’s) publish population distributions for specific domains. This information is publicly available. Hence, another topic for future research is the development of an indicator for representativeness that can be computed with aggregate population information. Within the 7th Framework of the European Union, in 2008 a project has started that aims at the development of these indicators (Bethlehem and Schouten, 2008). This RISQ project (Rep-
resentativity Indicators for Survey Quality) investigates how we can measure representativeness and how we can use the R-indicator to compare surveys or to construct dynamic survey designs.

Even with a dynamic survey design, some nonresponse will always remain. Therefore, the need for nonresponse bias adjustment once the data have been collected also remains. In Chapter 6 of this thesis, we have presented an overview of nonresponse adjustment methods. Depending on the type of auxiliary information, different estimation techniques can be applied. For instance, when the researcher knows individual values for the respondents only and in addition has the population distributions published by the NSO, then calibration estimators (under which the generalised regression estimator and post-stratification) can be applied. However, when the researcher can obtain information on the nonrespondents by linking the survey sample to a register, besides the calibration estimators also the propensity score methods can be used. Another way of obtaining information about nonrespondents, is to re-approach nonrespondents. In Chapter 4 of this thesis we have described the re-approach of nonrespondents to the Dutch LFS with the call-back approach and the basic-question approach. To validate the results from both re-approaches we used linked data from several registers. It turned out that the responding households in the call-back approach were different from the regular LFS-respondents. Furthermore, these households resembled the remaining nonrespondents. Hence, the composition of the response improved by adding the call-back respondents. The results for the basic-question approach were less satisfactory, but this was caused by the design of the approach which led to a confounding effect of telephone ownership. The additional information that is obtained with the re-approach can also be used for nonresponse bias adjustment. How this should be done is described in, for example, Bethlehem et al. (2006) or Laaksonen and Chambers (2006).

The methods that we describe in Chapter 6 do not account for the different causes of nonresponse. Therefore, in Chapter 8 we have developed nonresponse adjustment methods that take into account the different types of respondents. These methods consists of separate equations for the response types, and an additional equation for the survey item. They are in fact an extension of the methods for nonresponse analysis in Chapter 8. The nonresponse bias in the survey item is modelled by means of correlations between the different response types and the survey item.

The reason to distinguish different types of response comes from the intuition that the fieldwork process encompasses information that is useful in explaining response behaviour and, ultimately, in adjustment for nonresponse bias. However, the amount of information about the data collection process, or paradata,
is different for each response type. For instance, interviewer observations are available for contacted sample elements only. The methods that we have developed in Chapter 8 facilitate the inclusion of paradata because each response type is described by its own equation. However, again the applicability of these methods depends on the available information. In the best situation, there is detailed socio-economic and demographic information for all sample elements, as well as a large amount of paradata. This is the situation at NSO’s in the The Netherlands, the Scandinavian countries and some other North-European countries. However, persons working in commercial market research and academics are usually less fortunate when it comes to auxiliary information. The methods described in Chapter 8 can only be applied when the information is available for all elements in the sample. When there is not much available information it is not possible to closely follow the response process and the methods described in Chapter 6 will be more appropriate. However, when there is detailed information about the sample elements, but no paradata, it is still possible to apply the methods described in Chapter 8.

Methods for nonresponse adjustment have improved by the use of auxiliary information, and hence profit from the availability of registers. However, nonresponse adjustment is also challenged by developments in survey design. We already briefly mentioned dynamic survey design. The first steps towards such designs have already been taken in the development and implementation of mixed mode surveys. Mixed mode data collection has become common practice to conduct a survey. By using a mixture of modes, data collection can be done cheaper and faster at the same quality, if performed in a clever way. From the perspective of nonresponse adjustment, the question arises how to combine data collected in different modes into one survey estimate? In Chapter 9 of this thesis, we have developed mixed mode-methods that combine data from different modes, while describing the fieldwork process in each mode separately. The methods closely follow the data collection process, by distinguishing different types of response as well as the sequence and combination of modes. However, if detailed information is not available to closely follow the fieldwork process and the response process, we recommend survey researchers that work with mixed mode surveys to at least include an additional variable for the mode into their nonresponse adjustment models. This will capture the general mode effect.

By explicitly modelling response behaviour (Chapter 8) and by distinguishing the different modes in the data collection process (Chapter 9), it becomes possible to include paradata besides the usual demographic and socio-economic information. However, this can also result in a large number of equations and, consequently, a large number of underlying assumptions. Estimation of both
methods, the nonresponse adjustment methods with different response types and the mixed mode-methods, requires evaluations of multiple integrals of the multivariate normal distribution. We have identified some important directions for future research. First, the estimation of these methods can be simplified by using Markov Chain Monte Carlo methods or Bayesian estimation. Secondly, since the methods depend on the distributional assumptions, future research should be directed at assessing the robustness to the assumptions and the development of alternative models. Finally, the described methods can only be applied to one survey item at the time. If the survey consists of a large number of items, it is impractical to perform a different adjustment for each survey item, or for each group of survey items. To improve the practical value of these methods, future research should focus on how to combine these methods for different survey items, both for the nonresponse adjustment methods described in Chapter 8 and the mixed mode-methods in Chapter 9.

Besides mixed mode data collection, another challenge is the increased use of registers. Especially for business surveys, register-based surveys are common use nowadays. In register-based statistics, there is no sampling error. Furthermore, nonresponse occurs less frequently or is even completely absent. The main cause of error is related to the framework and results in a coverage error. Like nonresponse, undercoverage is caused by non-observation. The R-indicator can be used as a quality indicator for the undercoverage of a register. Furthermore, in Chapter 7 of this thesis we have shown that adjustment methods for nonresponse can be applied to adjust for errors caused by undercoverage of telephone households. It is likely that in register-based surveys similar problems exist and, hence, the same methods can be applied.

The methods that we have developed in this thesis can be applied to other situations, for example nonresponse in business surveys. The available auxiliary information for business surveys is less rich than for household surveys. However, information about the size of the business, the number of employees, or the branch to which a business belongs is generally available. The representativeness of the response to a business survey can also be computed with the R-indicator. For example to follow the representativeness of the response over time, to see whether at some point in time the subsample of responding businesses is representative enough for the production of so-called flash statistics. In addition, the use of registers is more prevailing in business surveys than it is in households surveys, since personal information can less frequently be found in a register as opposed to the economic information from businesses. Therefore, the application of adjustment methods to errors caused by undercoverage in business surveys seems promising.
In this thesis we have focussed on unit nonresponse. Unit nonresponse causes missing data on survey items, but in addition some observations of survey items may be missing due to item nonresponse. Treatment of nonresponse therefore also requires a decision on how to treat item nonresponse. Sensitive survey items are more susceptible to item nonresponse than others. For instance, the income question is known to suffer severely from item nonresponse (see, for example, Frick and Grabka 2005). The R-indicator can be used to compute the representativeness of individual items. Item nonresponse is commonly dealt with by imputation, see for instance Little and Rubin (2002). This approach replaces the missing values for the survey items by proxy values. These proxy values are estimated by some imputation procedure. A combined approach of imputation and weighting can be used to handle both item- and unit nonresponse, see for example Särndal and Lundström (2005). The models described in Chapter 8 can be applied to adjust for item nonresponse. The auxiliary information that is used in these models for a description of the response process can be extended with answers to other, related survey questions.
Bibliography


