Difficult Groups in Survey Research and the Development of Tailor-made Approach Strategies

Remco C.W. Feskens

Difficult Groups in Survey Research and the Development of Tailor-made Approach Strategies

ISBN 978-90-393-4975-5
NUR 916

Difficult Groups in Survey Research and the Development of Tailor-made Approach Strategies

Moeilijk Waarneembare Groepen in Enquêteonderzoek en de Ontwikkeling van op maat gesneden Benaderingsstrategieën

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof. dr. J.C. Stoof, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op vrijdag 9 januari 2009 des ochtends te 10.30 uur

door

Remco Cornelis Wilhelmus Feskens

geboren op 1 april 1978 te Breda

Promotoren: Prof. dr. J.J. Hox
Prof. dr. J.J.G. Schmeets

Acknowledgement

The views expressed in this publication are those of the authors and do not necessarily reflect the policy of Statistics Netherlands.

Preface

This thesis is one of the results of the collaboration between Utrecht University and Statistics Netherlands. I would like to thank Joop Hox, Gerty Lensvelt-Mulders and Hans Schmeets for supporting me in writing this thesis. Contributions made by them were not only invaluable for scientific reasons, but assistance in a more practical way was also greatly helpful. They managed to create a perfect balance between on the one side allowing much individual freedom to conduct research and on the other side guiding this process. Most important, working with them was always a pleasure.
I would like to thank people working at different divisions at Statistics Netherlands who contributed in several ways to this thesis. I am grateful to Edith de Leeuw and Maria Hage for their helpful ideas. Furthermore, I would like to thank my colleagues at the department of Methods and Statistics. I highly appreciated the collegial atmosphere. In particular I would like to mention Elly Korendijk and Roland Holdinga. I enjoyed exchanging ideas with them about all kind of things, ranging from statistical issues to personal things or the latest highly unimportant football match.
Last but not least I would like to thank Levi and Ida. Although both very interested in my work, they also contributed much to necessary leisure time by experiencing other more important things than nonresponse with them.

Contents

1 Introduction

Statistics is sometimes viewed as the scientific tool to give citizens a voice in the public space (cf. Verba 1996). Technological innovations and more output- and voter-oriented policy-making have led to an even greater urge to measure the needs and preferences of the general public (Davies, Nutley, and Smith 2002; The Economist 2007). Human behavior can be measured using many numbers of techniques, e.g. direct researcher observation or technological equipment, just as television viewing is measured by a small recording box. Opinions and evaluations can however almost only be measured by asking questions directly. This is the field of survey research, which is all about correctly representing properties of a specified population (Dalenius 1985).

As opinions become increasingly important in modern-day democracies, the tool to gather these opinions, survey research, has attained more attention from media, political decision-makers and general public as well. Not only the results of survey research are more widely discussed in the public arena, survey research itself is also questioned in this same arena. Of course the quality of survey research has been debated for much longer in the scientific field, although largely outside the traditional academic disciplines (Groves et al. 2004). One of the most important notions is that the quality of survey research is determined by four cornerstones of survey research (Hox, De Leeuw, and Dillman 2007), i.e. *measurement, coverage, sampling* and *nonresponse.*

Two other fundamental concepts are *reliability* and *validity.* Reliability refers to the random deviation in research results. A study is reliable if the same values are found in repeated measurements and the value of interest has not changed in the population. Validity refers to the concept of interest the researcher wants to measure. So reliability refers to random variations in the research results and validity to systematic variations in the research results.

Errors in all four central cornerstones of survey research may threaten a survey results' reliability and validity. The validity of survey results is threatened if they are biased. Bias in survey estimates refers to systematic deviations from the true but unknown population values. Biased estimates may lead to erroneous conclusions in survey research. This is why it is so important to optimize the four cornerstones of survey research and minimize bias in survey estimates. Survey methodology focuses on improving the quality of the data within cost constraints and on understanding why errors arise in survey estimates (Groves et al. 2004). The four cornerstones of survey research determine the quality of survey results. Survey quality refers to validity and reliability alike, in other words to systematic and variable errors. Although reliability is a necessary condition for validity, researchers are currently most concerned about the validity of the survey results. This is because reliability refers to random variations and validity refers to systematic variations in research results. There is much more of a risk of drawing erroneous conclusions if there are

systematic deviations from the true values in the survey outcomes. In order to maximize the validity of a study, survey methodologists thus are in particular trying to minimize the systematic variations in the survey results.

The first cornerstone of survey research is measurement. Quality of measurement means adequately measuring the concept in question in a consistent way. A measurement is *valid* if the concept is measured in the same way as the relevant theoretical concept. Given that the value of interest has not changed, a measurement is *reliable* if a measurement tool yields the same value over several measurements. There are many potential threats to the measurement quality in a survey, since, among others, the respondent and interviewer, question wording and answer categories as well as theoretical concept misspecification and coding errors can all influence the reliability and validity of the results (cf. De Leeuw, Hox, and Huisman 2003). Optimizing the quality of the questionnaire and how the field work is organized can decrease the risk of errors in the survey results.

The second cornerstone of survey research quality is coverage. If you want to study a certain specific target population, you need a frame that lists all the elements of that specific population. Two issues may threaten the coverage and thus the validity of the survey: (1) elements of the population may not be covered in the study, i.e. *undercoverage,* (2) elements that do not belong to the target population are represented in the study, i.e. *overcoverage.* Overcoverage may also result from double listings of the same elements in the sample frame (Bethlehem 2004). Undercoverage and overcoverage can both impact the reliability and validity of a study.

The third cornerstone of survey research is sampling. Most studies use a sample to study a population, and for good reasons. An integrated study of all the elements of a population or census can be very expensive or impossible. In addition, conducting a census requires a great deal of effort on the part of *all* the elements of the target population. Both these problems can be solved if only part of a population of interest is studied rather than all of it. So since the beginning of the twentieth century, samples have been used as an alternative to census studies. Under the key condition of *random* sampling, inferences can be made from the sample to the target population. Each random sample will almost always differ from the other random samples from the same target population and this leads to random variance in the study. Since this variance is random, it especially affects the reliability of the study. Horvitz and Thompson (1952) demonstrate that unbiased estimates can only be generated from samples created by a probability mechanism.

The fourth cornerstone of survey research is nonresponse. A distinction is usually drawn between *item-nonresponse* and *unit-nonresponse.* Item-nonresponse occurs if a respondent fails to provide all the information asked for given that he or she did provide an answer to at least one question being asked for in the survey. Unit-nonresponse occurs if a sampled unit does not provide any information at all for the survey. Item-nonresponse and unit-nonresponse both lead to missing values, which may threaten the reliability and validity of survey research.

One of the starting points of this Ph.D. project was the high and increasing unit-nonresponse rate in survey research conducted by Statistics Netherlands in particular and throughout the Western world in general (see De Heer 1999; Steeh, Kirgis, Cannon, and DeWitt 2001; De Heer and De Leeuw 2002; Snijkers 2003). Increasing nonresponse rates in survey research jeopardize the ideal of survey research, which is to correctly represent citizens in the public space.

There can be several reasons for unit-nonresponse. Firstly, the researcher may not be able to contact the sampled unit. Or the sampled unit can be contacted but may be unwilling to provide the required information. Lastly, the sampled unit may be contacted and willing to cooperate, but unable to participate in the interview process due to language difficulties or physical or mental limitations.

Nonresponse can have several consequences (cf. Biemer and Lyberg 2003). It can decrease the effective net sample size. With fewer cases to analyse, the variances will increase and the reliability of the survey will decrease. If a research agency is determined to have a certain sample size, anticipating nonresponse can lead to higher research costs and delay the completion of the survey. In addition, non-response rates are a visible indicator of survey quality and can threaten the reputation of a survey agency. Most importantly, nonresponse can bias survey estimates and thus threaten the validity of survey results. Nonresponse will bias survey estimates if respondents differ systematically from nonrespondents with respect to the variables under study (Groves and Couper 1998). This also implies that, as Groves (2006) demonstrates empirically, there is no simple relation between nonresponse rates and nonresponse bias. The occurrence of nonresponse bias depends on the degree to which the survey variables and response propensity are correlated.

In their seminal work, Little and Rubin (1987) establish a framework for determining the mechanism of missing data. They distinguish between data missing completely at random (MCAR), data missing at random (MAR), and data not missing at random (NMAR). In MCAR, the missing value is not related to variable missing or to other variables. Data is missing at random if the missing value is not related to the variable itself, but is related to other variables. Lastly, if the decision to respond is related to the variables under investigation, the nonresponse is not missing at random. Whether or not missing data will bias survey estimates depends on the missing data pattern. If the missing data is not at random, bias will be introduced in survey estimates.

Nonresponse bias is more likely to occur if nonresponse is selective (cf. Laurie, Smith, and Scott 1999; Keeter, Miller, Kohut, Groves, and Presser 2002; Te Riele 2002; Bethlehem and Schouten 2004; Stoop 2004). In this case, nonresponse is not equally distributed among societal groups, which makes it more likely that respondents will differ systematically from nonrespondents with respect to the survey variables. This is why survey methodology experts give priority to studying *difficult groups in survey research,* which are currently under-represented groups in survey research (Hox, De Leeuw, and Snijkers 2003). At Statistics Netherlands, five

of these groups have been identified by experts (Reep 2003), i.e. the homeless, people living in non-private households, ethnic minorities, asylum-seekers and illegal persons. The first priority is given to studying ethnic minorities and people living in non-private households since they are by far the largest of the five groups. Moreover, the societal and political need for information is probably most urgent for these two groups.
In this thesis, the focus is therefore on ethnic minorities and people living in non-private households. The latter group is excluded from the sample frame before the actual survey is conducted. This is common practice at most national statistical institutes (cf. Feskens, Lensvelt-Mulders, Beukenhorst, Kockelkoren, and Wetzels 2008). Ethnic minorities exhibit below-average response rates and are consequently a problematic group in survey research.
To study these issues, Statistics Netherlands launched the strategic research program "Nonresponse and Difficult Groups" in 2001 in collaboration with several universities including Utrecht University (Snijkers 2003). In this program, several Ph.D. students were able to work on various aspects of difficult groups in survey research. This thesis is one of the results of this research program.

Outline

The second chapter of this thesis, *Looking for Homogenous Groups of Respondents and Nonrespondents Using Latent Class Analysis,* is a quantitative explorative study of groups with equal response probabilities. This is done by performing a latent class analysis where we look for unobserved latent classes. In this explorative analysis, two groups with below-average response rates are found, i.e. the already in the above mentioned qualitative study group of ethnic minorities, and a second group consisting of native Dutch residents of small households living in urban areas.

Nonresponse and ethnic minorities are addressed in Chapters 3 and 4. In Chapter 3, *Nonresponse among Ethnic Minorities: A Multivariate Analysis,* we examine the effect of ethnicity on response rates controlled for various socio-economic and socio-demographic variables. It appears that the effect of ethnicity on response is almost entirely mediated by the degree of urbanization. In Chapter 4, *Nonresponse among Ethnic Minorities in an International Perspective,* strategies to reduce non-response among ethnic minorities are discussed. Nonresponse rates and data collection strategies among ethnic minorities in six European countries are reviewed in order to develop tailor-made strategies to increase survey participation among ethnic minorities.

A frequently cited method for decreasing nonresponse rates involves the use of incentives (e.g. Church 1993; Singer, Van Hoewyk, Gebler, Raghunathan, and McGonagle 1999). In Chapters 5 and 6, the results of a large-scale experiment with incentives will be discussed. General results of this experiment are described in Chapter 5, *Impact of Prepaid Incentives in Face-to-Face Surveys: A Large-Scale Experiment*

with Postage Stamps. Incentives appear to be especially effective among native Dutch urban residents, the second problematic group of respondents found in the latent class analysis. Chapter 6, *Incentives and Ethnic Minorities: Results of a Controlled Randomized Experiment in the Netherlands,* shows that using incentives only increases response rates among ethnic minorities to a limited extent.

I close with Chapter 7, *Studying People Living in Non-Private Households: Results of a Large Pilot Study in the Netherlands,* in which the possibility of conducting survey research among residents of homes for the elderly and nursing homes, the two largest groups living in non-private households, is investigated. Moreover, the potential bias of excluding these groups from standard survey research is examined.

2 *Looking for Homogeneous Groups of Respondents and Nonrespondents Using Latent Class Analysis*

Concerns about decreasing response rates have led to many nonresponse studies. Unfortunately, usually only a limited amount of data is available on nonrespondents. Linking administrative data makes information available on all the sampled units including nonrespondents. Statistics Netherlands has constructed a database in which several administrative records and several surveys are connected. In this study we use this dataset to look for homogenous groups of respondents and nonrespondents in survey research using latent class cluster analysis. Latent class techniques can provide insight into the problem of selective nonresponse and currently under-represented societal groups. We identify four different latent classes in a recent Dutch survey and replicate the findings for the same survey conducted four years later. Two of the types of sampled units have above-average response rates and the other two have below-average response rates. We also evaluate various latent class models with other response outcomes to gain insight into the contact and cooperation process.[1)]

[1)] This chapter had been submitted for publication as Feskens, R.C.W., Hox, J.J., and Schmeets, J.J.G. (2008). Looking for Homogeneous Groups of Respondents and Nonrespondents Using Latent Class Analysis. *Submitted manuscript.*

2.1 *Introduction*

Increasing nonresponse rates in survey research and with that an increasing potential to bias survey estimates is a matter of great concern in many countries (De Heer 1999; DFG 1999; De Heer and De Leeuw 2002). It is the very nature of nonresponse – information missing for a sampled unit – that makes it difficult to assess. One way to find out more about nonresponse is by using background information available for all the sampled units. Statistics Netherlands has constructed a database with links to municipal records and survey results (Houbiers 2004). It provides rich background information about respondents and nonrespondents that makes it possible to more thoroughly examine the nonresponse patterns. This database has been used in various studies about nonresponse and more specifically the selectivity of nonresponse (Schmeets and Michiels 2003; Schouten and Cobben 2007; Feskens, Hox, Lensvelt-Mulders, and Schmeets 2007). These studies grant insight into the nonresponse mechanism of specific societal groups such as ethnic minorities (Schmeets and Michiels 2003; Feskens et al. 2007) and provide information on how representative the response composition is in surveys (Schouten and Cobben 2007). In this study we search in an explorative way for homogenous groups with equal response patterns. The most important problem of nonresponse is that due to systematic differences between respondents and nonrespondents with respect to the variables under study, survey estimates are biased (Groves 1989; Groves and Couper 1998). Biased estimates are more likely to occur if specific groups exhibit below-average response rates. Since the nonresponse is not random, this makes it more likely for that the nonrespondents differ systematically from respondents as regards to the variables under study. In this case, the nonresponse is said to be selective. Finding homogeneous groups with equal response patterns can provide insight that can be used to prevent selective nonresponse by using tailor-made nonresponse reduction strategies for hitherto under-represented groups in survey research. In this study we conduct an explorative search for societal groups with equal response propensities.

A problem inherent in exploratory analysis is how to summarize and present the available frame information in a meaningful manner (Magidson 2003). A common way to do it is by summarizing the available frame data on socio-demographic and socio-economic variables as correlates of nonresponse rates. It is possible, however, that due to high correlations between the variables, summaries of this kind merely present redundant information (Magidson 2003). To overcome this problem, we look for homogeneous responding and nonresponding groups by using latent class techniques, which do not require as many assumptions as traditional techniques (McCutcheon 1987; Vermunt and Magidson 2002). Moreover, looking for societal groups that exhibit similar response patterns can be viewed as investigating a concept that cannot be observed directly. Latent class analysis (LCA) is a statistical method that enables us to find this unobservable concept by assuming the existence of underlying sub-groups or latent classes in a dataset (McCutcheon 1987). These

classes are called latent because they are not directly observed. LCA gives insight into whether commonly unobserved variables or latent classes can explain the observed relations between observed variables (Goodman 2002). A latent class model is thus a tool for classifying cases into sub-populations. Using LCA, we can generate typologies of sampled units with respect to their response behavior. In this study, we look for sub-populations of respondents and nonrespondents in order to gain more knowledge about potential selective nonresponse in survey research. We use the program Latent GOLD (Vermunt and Magidson 2003) to search systematically for latent groups in the 1998 and 2002 Dutch Continuous Survey on Living Conditions (POLS).

This results in the following research questions:

1) *Are there any related cases or types of respondents and nonrespondents in the 1998 Dutch Continuous Survey on Living Conditions (POLS)?*
2) *Were these various types of respondents and nonrespondents different when the same survey was held again in 2002?*

By addressing these questions, we hope to gain a better understanding of the response sample selectivity due to nonresponse. This information can be helpful in developing efficient tailored strategies for specific societal groups to decrease nonresponse selectivity and response bias. The following two sections describe the data used and the methodology of LCA. The fourth section presents the results and the final section states the conclusions.

2.2 Data

We have analysed the survey files of the 1998 and 2002 Continuous Survey on Living Conditions (POLS) conducted by Statistics Netherlands. About 40,000 interviews are conducted every year. POLS is an integrated survey on the living conditions of the Dutch population in private households. The observation units are individuals. The sample frame is the Population Register of all the basic Dutch municipal records. POLS is a two-stage sample. First communities are drawn and then people. Large cities are automatically included (Schouten 2003). Communities and persons are drawn in such a way that the first-order inclusion probabilities are equal across all the sampled units with the exception of age, since the target population in some modules has age restrictions. Participation is voluntary in POLS and the survey is solely administrated in Dutch. Every month a sample of about 3,500 people is drawn. In 1998, there is a two-month fieldwork period for the twelve consecutive samples. In the first month, Statistics Netherlands collects data with a CAPI mode and nonrespondents with a known telephone line are re-approached with a CATI technique. Nonrespondents without a known telephone

line and sampled units who cannot cooperate due to illness are re-approached with CAPI (Schouten 2003). In 2002 the design and fieldwork strategy are similar to these of 1998, except that *all* the nonrespondents are re-approached with a CAPI mode.

The POLS survey is supplemented by administrative data from the Population Register and information about employment and social benefits (Schmeets and Michiels 2003; Schouten 2003). Linking the municipal records makes socio-demographic and socio-economic information available on the nonrespondents at the individual and postal code level. This information has been gathered by Statistics Netherlands in the Social Statistical Database, in which several registers are linked to each other as well as to data from sample surveys (Houbiers 2004).

The additional information provided by the link to municipal records makes it possible to study the nonrespondents. However, no extra information is available on 1,143 of the sampled units (2.9% of the total sample) for the 1998 survey. Since no systematic missing data pattern is found regarding important background variables for these 1,143 cases, they are considered missing completely at random and deleted from the data file. Since these numbers are quite small, they can be dropped from the sample without a significant loss of information.

In the 1998 POLS survey, 39,431 sampled units are drawn. Since we only include sampled units aged 15–65, 28,542 sampled units are left for analysis. The number of respondents is 17,123, resulting in a response rate of 60.4%, which is not unusually low in the Netherlands (see e.g. De Heer 1999; De Heer and De Leeuw 2002).[2)] In the 2002 POLS survey, 39,572 residents of the Netherlands are sampled, 30,199 between the ages of 15 and 65. The response rate is 55%. The response outcomes presented in Table 2.1 are not totally similar. Due to fieldwork problems during the 1998 survey, interviewers had no opportunity to interview a substantial percentage of the sampled units. So these sampled units are only partially approached or not at all during the fieldwork period and are coded as *not handled.* Moreover, in 2002 changes in response coding lead to the introduction of the new response outcome *no opportunity.* This means respondents can indicate that although they are willing to participate in the survey, they have no opportunity to be interviewed during the fieldwork period.

To avoid capitalization on chance, we have randomly split both the 1998 and 2002 datasets into an exploration and validation file.

2.3 *Method*

LCA, introduced by Lazarsfeld (1950a, 1950b) and expanded upon by Lazarsfeld and Henry (1968) and Goodman (1974a, 1974b), is a statistical tool for classifying

[2)] According to AAPOR response definition two (AAPOR 2006).

Table 2.1
Response outcomes in 1998 POLS and 2002 POLS for sampled units aged 15–65

	1998 POLS		2002 POLS	
	n	%	*n*	%
Response	17,123	60.0	16,382	54.2
No contact	1,621	5.7	2,552	8.4
Refusal	6,905	24.2	7,297	24.2
Not capable	598	2.1	755	2.5
Not eligible	100	0.4	225	0.7
Not handled	1,848	6.5	*	.
No opportunity	§	.	2,179	7.2
Moved	347	1.2	809	2.7
Total	28,542	100.0	30,199	100.0

Note: Figures have been rounded off and may not add up to 100%.
§ = Not a response outcome in 1998 POLS.
* = Not a response outcome in 2002 POLS

cases into related sub-populations or latent classes. These classes are called latent because they cannot be observed directly. The idea of LCA is that the observed associations between a set of (categorical) variables are accounted for or explained away by a reduced number of latent classes. The observed variables are seen as indicators of unobserved sub-populations that may exist in one dataset. As such, LCA can be seen as an expanded version of the basic idea that the observed correlation between two variables may have resulted from a third variable or common dependence (Evans and Mills 1998). This common dependence is assumed to be the result of unobservable sub-populations existing in a dataset. In LCA, these unobserved sub-populations explain the observed correlations between the set of variables. One of the key goals of LCA is thus to determine the smallest number of latent classes needed to account for the observed associations among the observed variables. McCutcheon (1987) remarks that LCA can also grant insight into typologies in different populations. In this study we use a latent class cluster analysis. Latent class cluster analysis is used analogously to cluster analysis. The central aim of the analysis is to find out if there are a small number of discrete categories in which cases with similar characteristics can be classified. Vermunt and Magidson (2002) remark that an important difference between the two techniques is that latent class cluster analysis is a model-based cluster approach. Class membership probabilities are estimated from this model-based approach and then cases are classified into classes (Vermunt and Magidson 2002).

A central assumption in latent class analysis is the criterion of *conditional independence.* Conditional on membership in a latent category, the indicator variables are statistically independent of each other (Vermunt and Magidson 2004).The local independence criterion can grant insight into whether observed associations among a set of indicators are the result of some latent explanatory variable (McCutcheon 1987).

Parameters are estimated by first applying the EM algorithm and switching – when close enough to the final solution – to the Newton-Raphson to find the maximum likelihood (ML) estimates (Vermunt and Magidson 2002). The ML estimates are the ones most likely to account for the observed results. Several methods can be used to evaluate the model fit. Several formal statistical tests (e.g. Pearson Statistic, the likelihood-ratio test and the Cressie-Read statistic) evaluate the extent to which ML estimates for the expected cell frequencies differ from the observed frequencies (Magidson and Vermunt 2004). Lower values indicate a better model fit. However, this method is not suitable for comparing models with various numbers of latent classes. The most widely used method for assessing the model fit are, therefore, information criteria (Vermunt and Magidson 2002). This holds especially true in exploratory analyses, as in this study. These criteria can be used to determine the number of classes that fit the data most properly. These measures are based on the –2 times log likelihood of the model, adjusted for the number of parameters in the model and the sample size. These criteria seek to strike a balance between model fit and parsimony. The model with the minimal value is chosen. Usually a combination of a formal test and an information criterion is recommended to select the best fitting model (Eid, Langeheine, and Diener 2003). In this article we use the BIC values and the likelihood-ratio test to determine the number of classes.

Variables in model

Indicators are dependent variables used to define or measure the latent classes. Latent class models have recently been expanded to use variables of mixed scale types in the same analysis, thus enlarging the possibility of including indicators (Vermunt and Magidson 2002). Since we investigate different types of responding clusters in the 1998 POLS and 2002 survey, we choose to use indicators where information is available for both surveys. Some indicators have to be left out of the analysis because there is too much overlap with other indicators. Although this limits the number of indicators we can include in the models, the available variables nevertheless provide important response and nonresponse correlates. Authors including Goyder (1989), Groves and Couper (1998), Steeh et al. (2001), Van Goor, Jansma, and Veenstra (2005), Stoop (2006) and Feskens et al. (2007) show that socio-demographic and socio-economic information can affect the contact and cooperation propensity of sampled units. We tried to incorporate information on these two concepts in the models as much as possible. This has led to the following indicators:

1) Gender
2) Ethnicity
3) Indicator of having a job
4) Urbanicity
5) Size of the household
6) Response outcome.

Ethnic minorities or *ethnicity*, a somewhat ambiguous concept, are defined in the Netherlands as 'everyone residing in the Netherlands with either one or two parents born abroad' (Reep 2003; Feskens, Hox, Lensvelt-Mulders, and Schmeets 2006). A further distinction is usually drawn between people with either one or two parents born in Europe, North America, Australia, Japan or Indonesia and people with either one or two parents born in non-Western countries (mainly Turkey, Morocco, Surinam and the Netherlands Antilles).[3] Almost one in five residents of the Netherlands are members of ethnic minorities (http://statline.cbs.nl). The Western and non-Western groups of foreigners are of approximately the same size. In this analysis we merge native Dutch sampled units with Western foreigners and contrast this group with non-Western foreigners. We do so because several analyses demonstrate that unlike non-Western foreigners, Western foreigners exhibit very similar background characteristics and response behavior to native Dutch sampled units (Schmeets 2005a; Feskens et al. 2007). A description of all the variables that are used can be found in the Appendix.

We estimate several models with different numbers of latent classes using the six selected indicator variables available for both datasets. The first model, the independence model, assumes that all the indicators are independent of each other. The second assumes the existence of two latent classes to account for the associations among the indicators and so on up to six latent classes. Since this is an exploratory study, we impose no restrictions on the forms of the clusters.

The fit statistics for these models are shown in Table 2.2, where the BIC values and the Chi-square and likelihood-ratio statistics from the exploration and validation files in the 1998 and 2002 datasets are presented. In addition, the Chi-square values from the exploration file are shown.

Table 2.2
Model selection

	BIC exploration file	BIC validation file	Degrees of freedom	Chi-square	LL
1998					
Independence Model	158922.272	155237.829	466	5,478	3,812
2 Latent classes	157317.134	153860.175	459	3,043	2,140
3 Latent classes	156652.884	153141.642	452	1,414	1,409
4 Latent classes	156150.967	152630.696	445	840	850
5 Latent classes	156127.470	152565.276	438	765	749
6 Latent classes	156136.194	152576.119	431	694	704
2002					
Independence Model	161215.757	163394.448	466	3,866	2,941
2 Latent classes	160097.851	162304.407	459	2,310	1,756
3 Latent classes	159587.943	162047.386	452	1,177	1,179
4 Latent classes	159289.396	161446.834	445	843	813
5 Latent classes	159291.234	161419.250	438	752	744

[3] For reasons of simplicity we use 'Western foreigners' and 'non-Western foreigners' in this article.

Following the Bayesian Information Criterion, five latent classes are found in the 1998 and four latent classes in the 2002 exploration data file. These models are validated in the validation files. The lowest BIC values, indicating the best model fit, are found for this number of classes. Investigating this empirical result reveals that the fifth latent class in the 1998 dataset does not differ much from the second class and cannot be distinguished from it on substantive criteria. Although the likelihood-ratio test and Chi-square statistic indicate a solution with more than four classes, the decrease in both the Chi-square and likelihood-ratio test values is most substantial until the expansion to four latent classes. Fit indices provide useful empirical information about the model fit and the interpretability of the substantive results from the final model is ultimately the most important criterion. The most important goal of this study is to find groups with equal response propensities in order to develop tailor-made approaching strategies. The model with four classes can be interpreted in the most useful way. That is why we address four classes in the 1998 as well as the 2002 file. We are supported in this decision by the empirical results in the validation files. The results do not differ substantially from the ones in the exploration files. In the next section, we discuss the interpretation of the four classes found in the datasets.

2.4 Results

2.4.1 *Are there any related cases or types of respondents or nonrespondents in the 1998 Dutch Continuous Survey on Living Conditions (POLS)?*

In LCA, interpretation is based on the conditional probabilities. In Table 2.3 the conditional probabilities are presented of the four clusters or types or respondents for all the indicators in the model. The conditional probabilities indicate that a sampled unit "in a latent class will score a particular way on an observed measure" (McCutcheon 1987, p. 33). The latent class probabilities refer to the size of each cluster. For example, about 48% of all the sampled units belong to Response Type 1 in this situation. Lastly, the observed proportions found in the original dataset are shown in the last column.

In 1998, the units in the first cluster (about 48% of all the sampled units are in this cluster) have a high conditional probability of being a native Dutch or Western foreigner unit and having a job. Moreover, members of this cluster predominantly live in non-urban areas and average-sized households. They have a higher conditional probability of being male and have above-average response rates. Units in the second cluster (about 31% of the sample) have a higher conditional probability of being female and a very low conditional probability of having a job. Besides these two differences concerning the first type of respondents, about the same conditional probabilities are found for the other indicators as in the first

Table 2.3
Conditional probabilities response, 1998 and 2002

	1998	2002	1998	2002	1998	2002	1998	2002	1998	2002
Response 1998 & 2002	Type 1		Type 2		Type 3		Type 4			
Latent class probabilities	0.48	0.49	0.31	0.31	0.17	0.15	0.04	0.05		
	Conditional probabilities								Observed proportion	
Female										
No	0.60	0.58	0.36	0.40	0.52	0.53	0.49	0.50	0.50	0.51
Yes	0.40	0.43	0.64	0.60	0.48	0.47	0.50	0.50	0.50	0.49
Non-Western										
No	0.99	0.99	0.99	0.98	0.85	0.84	0.01	0.02	0.93	0.92
Yes	0.01	0.01	0.01	0.02	0.15	0.16	0.99	0.98	0.07	0.08
Job										
No	0.01	0.01	0.98	0.99	0.43	0.41	0.59	0.55	0.40	0.40
Yes	0.99	0.99	0.02	0.01	0.57	0.59	0.41	0.45	0.60	0.60
Urbanicity										
1 = low	0.37	0.17	0.40	0.20	0.02	0.00	0.04	0.01	0.31	0.14
2	0.22	0.26	0.23	0.28	0.04	0.02	0.07	0.06	0.19	0.22
3	0.19	0.24	0.18	0.23	0.11	0.09	0.16	0.15	0.17	0.21
4	0.16	0.25	0.14	0.22	0.33	0.38	0.35	0.41	0.19	0.27
5 = high	0.06	0.08	0.05	0.06	0.50	0.51	0.39	0.37	0.15	0.16
Size of household										
1	0.07	0.08	0.08	0.09	0.42	0.43	0.06	0.02	0.13	0.14
2	0.28	0.28	0.28	0.30	0.44	0.43	0.08	0.11	0.30	0.30
3	0.22	0.21	0.22	0.21	0.10	0.09	0.13	0.14	0.20	0.19
4	0.29	0.28	0.28	0.26	0.03	0.04	0.34	0.35	0.25	0.24
5	0.10	0.10	0.10	0.09	0.00	0.00	0.25	0.22	0.09	0.09
6 or more	0.04	0.04	0.04	0.03	0.00	0.00	0.18	0.16	0.04	0.04
Response										
No	0.33	0.40	0.35	0.41	0.64	0.63	0.59	0.50	0.40	0.44
Yes	0.67	0.60	0.65	0.59	0.36	0.37	0.41	0.50	0.60	0.56

1998, n = 14,454; 2002, n = 14,515.

cluster. This implies that sampled units in this cluster are also more inclined to take part in survey research. The sampled units in the third cluster (about 17% of the sample) mainly live in urban areas in small households. Although the conditional probability of being a non-Western foreigner is higher than in the first two clusters, it is still very small (0.15). Most of this segment thus consists of native Dutchmen or Western foreigners. This group has the lowest response probability. The first two segments are clearly distinguished on the gender and job indicators, with the other conditional probabilities on indicators being about the same, but the third and fourth segments have more or less the same conditional probabilities on the gender, job, urbanicity and response indicators. They differ, however, on the ethnicity (non-Western foreigners) and size of household indicators. The sampled units in the fourth segment are mainly non-Western foreigners (conditional probability of 0.99)

and live in above-average sized households. This is also a cluster with below-average response rates. It is striking that the size of this cluster (less than 4% of all the sampled units) is relatively small, especially bearing in mind the percentage of non-Western foreigners in the Dutch population (8.2% in 1998, the year this survey was conducted).

In sum, the four clusters can be described as follows:

1) Non-urban male sampled units with a job, above-average response rate (48% of sampled units)
2) Non-urban female sampled units without a job, above-average response rates (31% of sampled units)
3) Urban sampled units living in small households, below-average response rates (17% of sampled units)
4) Urban non-Western foreigners living in large households, below-average response rates (4% of sampled units).

Table 2.4 presents the conditional probabilities for the four latent clusters in the 1998 and 2002 contact process.

Apart from some small differences, the same four types of responding groups are found. Although the size of the first cluster one has decreased in favor to cluster size two. Again the first two clusters have below-average conditional probabilities of being contacted and the last two have below-average conditional probabilities.

To examine the process of cooperation (defined as AAPOR cooperation definition No. 2, AAPOR 2006), we include only the sampled units contacted in the first place so that 13,623 sampled units remain for the exploration file. Table 2.5 shows the results for the refusal outcome.

These findings correspond with earlier somewhat counterintuitive and unexpected results. In an experiment carried out within the Dutch Labor Force Study in 2005 with postage stamps as an incentive, the effect of the incentive on response selectivity is examined (Wetzels, Schmeets, Van den Brakel, and Feskens 2008). It was found that using incentives decreases the selectivity with respect to urbanicity but increases the selectivity with respect to ethnicity, although ethnic minorities predominantly live in urban areas. These findings suggest that two major types of sampled units live in highly urbanized areas, mainly the three largest cities in the Netherlands, Amsterdam, Rotterdam and The Hague. One cluster consisting of mainly native Dutch sampled units living in small households with above-average refusal rates is probably very susceptible to instruments like incentives introduced in an effort to decrease refusal rates. Another segment found in highly urbanized areas consists of non-Western foreigners living in large households

who exhibit a smaller conditional probability of refusing to take part in survey research. Incentives are probably less effective among the sampled units who belong to this cluster.

Table 2.4
Conditional probabilities, 1998 and 2002 contacts

	1998	2002	1998	2002	1998	2002	1998	2002	1998	2002
Contact 1998 & 2002	Type 1		Type 2		Type 3		Type 4			
Latent class probabilities	0.41	0.52	0.39	0.29	0.16	0.14	0.04	0.05		
	Conditional probabilities								Observed proportion	
Female										
No	0.04	0.57	0.98	0.38	0.54	0.55	0.48	0.50	0.50	0.51
Yes	0.96	0.43	0.02	0.60	0.46	0.45	0.52	0.50	0.50	0.49
Non-Western										
No	0.99	0.99	0.99	0.98	0.86	0.85	0.01	0.02	0.93	0.92
Yes	0.01	0.01	0.01	0.02	0.14	0.15	0.99	0.98	0.07	0.08
Job										
No	0.51	0.07	0.28	0.99	0.40	0.39	0.60	0.54	0.40	0.40
Yes	0.49	0.93	0.72	0.01	0.60	0.61	0.40	0.46	0.60	0.60
Urbanicity										
1 = low	0.37	0.17	0.39	0.19	0.02	0.01	0.05	0.02	0.31	0.14
2	0.22	0.26	0.22	0.28	0.04	0.04	0.08	0.06	0.19	0.22
3	0.19	0.23	0.18	0.23	0.12	0.11	0.16	0.15	0.17	0.21
4	0.16	0.25	0.15	0.22	0.33	0.38	0.34	0.40	0.19	0.27
5 = high	0.07	0.10	0.06	0.08	0.49	0.46	0.36	0.38	0.15	0.16
Size of household										
1	0.07	0.07	0.07	0.09	0.46	0.52	0.01	0.02	0.13	0.14
2	0.29	0.28	0.28	0.32	0.43	0.40	0.10	0.12	0.30	0.30
3	0.23	0.21	0.22	0.22	0.08	0.06	0.15	0.16	0.20	0.19
4	0.28	0.29	0.29	0.26	0.02	0.02	0.35	0.25	0.25	0.24
5	0.10	0.10	0.10	0.08	0.00	0.00	0.23	0.21	0.09	0.09
6 or more	0.04	0.04	0.04	0.03	0.00	0.00	0.16	0.15	0.04	0.04
Contact										
No	0.03	0.05	0.04	0.04	0.18	0.32	0.11	0.15	0.06	0.09
Yes	0.97	0.95	0.96	0.96	0.82	0.68	0.89	0.85	0.94	0.91

1998, n = 14,454; 2002, n = 14,515.

2.4.2 *Are there any related cases or types of respondents or nonrespondents in the 2002 Dutch Integrated Survey on Household Living Conditions (POLS)?*

To see whether the response pattern in the 1998 data files of the Survey on Living Conditions is stable over time, we also look for homogenous responding and nonresponding groups in the 2002 data files of the same survey. The conditional probabilities for this survey can be found in Table 2.3 as well.

Table 2.5
Conditional probabilities refusal, 1998 and 2002

	1998	2002	1998	2002	1998	2002	1998	2002	1998	2002
Refusal 1998 & 2002	Type 1		Type 2		Type 3		Type 4			
Latent class probabilities	0.49	0.54	0.32	0.28	0.15	0.13	0.04	0.05		
	Conditional probabilities								Observed proportion	
Female										
No	0.60	0.57	0.36	0.37	0.51	0.51	0.50	0.50	0.50	0.51
Yes	0.40	0.43	0.64	0.63	0.49	0.49	0.50	0.50	0.50	0.50
Non-Western										
No	1.00	1.00	0.99	0.99	0.87	0.88	0.01	0.01	0.93	0.93
Yes	0.00	0.00	0.01	0.01	0.13	0.12	0.99	0.99	0.07	0.07
Job										
No	0.01	0.09	0.97	0.99	0.41	0.39	0.57	0.55	0.40	0.40
Yes	0.99	0.91	0.03	0.01	0.59	0.61	0.43	0.45	0.60	0.60
Urbanicity										
1 = low	0.38	0.17	0.40	0.20	0.02	0.01	0.06	0.02	0.32	0.15
2	0.22	0.27	0.23	0.28	0.04	0.04	0.08	0.07	0.19	0.23
3	0.19	0.23	0.18	0.23	0.12	0.12	0.17	0.16	0.17	0.21
4	0.15	0.25	0.14	0.22	0.34	0.40	0.35	0.41	0.19	0.27
5 = high	0.06	0.08	0.05	0.07	0.49	0.43	0.34	0.34	0.13	0.14
Size of household										
1	0.06	0.06	0.07	0.08	0.44	0.47	0.01	0.02	0.12	0.12
2	0.28	0.28	0.29	0.32	0.45	0.44	0.09	0.13	0.30	0.30
3	0.22	0.21	0.23	0.22	0.09	0.07	0.14	0.16	0.20	0.19
4	0.29	0.29	0.28	0.27	0.03	0.02	0.35	0.35	0.25	0.25
5	0.10	0.11	0.10	0.09	0.00	0.00	0.24	0.21	0.09	0.09
6 or more	0.04	0.04	0.04	0.03	0.00	0.00	0.17	0.14	0.04	0.04
Refusal										
No	0.76	0.73	0.75	0.70	0.70	0.70	0.80	0.82	0.75	0.72
Yes	0.24	0.27	0.25	0.30	0.30	0.30	0.20	0.18	0.25	0.28

1998, n = 13,623; 2002, n = 13,228.

An inspection of the table with conditional probabilities for the 2002 latent class model essentially reveals the same four clusters as the 1998 data. Moreover, the model with four classes for the 2002 data is also the one that should be chosen solely on empirical grounds. As is clear from Table 2, a model with four latent classes is the one with the lowest BIC value. Again, the first two clusters exhibit above-average response rates and have high conditional probabilities of living in non-urban areas and being native Dutch or Western foreigners and living in medium-sized households. Sampled units in the first segment (49% of all the cases) are more likely to be male and have a job. Sampled units in the second class (31% of all the cases) are more likely to be female and not have a job. The third and fourth segments are again problematic in terms of response probabilities. In particular, the sampled units in the third segment have a small conditional proba-

bility of responding (37%). As in the 1998 dataset, the sampled units in this cluster are mostly members of small households who live in urban areas. There is a high conditional probability of their being native Dutch or Western foreigners. The fourth cluster can be classified as the non-Western foreigner cluster, and its sampled units mainly live in large households in urban areas. Members of this segment also exhibit below-average response probabilities (0.50), although not to as extreme an extent as in the third segment.

Table 2.4 presents the results of the latent class model describing the process of establishing contact with the sampled units in 2002. Establishing contact with the sampled units in the third cluster is very problematic; the conditional probability of contact being established is very low, especially compared to the sampled units in the first two clusters. As in 1998, the fourth cluster also has below-average contact propensities.

Table 2.5 shows the conditional probabilities for the model including the indicator *sampled unit refuses to participate in the survey.* Again, only the sampled units who are contacted in the first place are considered in this model. The conditional probabilities of refusing a request to participate have increased for the first two clusters and are now similar to these of the third cluster. As is noted above, a new response outcome is added in the interviewers' registration files in 2002. Interviewers can code a sampled unit as *no opportunity to be interviewed* if a person indicates he is willing to participate but has no opportunity to do the interview during the fieldwork period. This category can however also contain sampled units who would have refused otherwise, so called *soft refusals.* This is why we analyse this variable to see whether one of the four clusters exhibits deviant conditional probability on *no opportunity to be interviewed.* It appears that this is not the case. The clusters do not differ in conditional probabilities on this variable, indicating that it does not introduce response selectivity with respect to the variables in the model. All the results in the exploration files of the 1998 and 2002 datasets have been re-examined in the validation files. The results of the validation files are very similar to those of the exploration files.

2.5 *Conclusions and discussion*

In this study, we systematically look for homogenous types of responding and nonresponding groups using latent class cluster techniques. We find four latent groups in the 1998 and 2002 Dutch Survey on Living Conditions, a large-scale annual survey conducted by Statistics Netherlands. The first two clusters exhibit above-average response rates and live in non-urban areas. Sampled units in these two clusters are most likely to be native Dutchmen or Western foreigners. Sampled units in the second cluster are also more likely to be female and have no job. The first two groups are the two largest clusters, containing about 80% of all the sampled units. The other two clusters have below-average response probabilities. The third

cluster mainly contains sampled units living in small households (mostly single households) in urban areas and most of them are native Dutchmen or Western foreigners. This third cluster is the most problematic in terms of low response probabilities. The fourth cluster can be classified as a non-Western foreigners cluster. It is also a cluster whose members have low conditional response probabilities, although not as problematic as in the third cluster. The 2002 clusters are more or less similar to those in the 1998 dataset. Examining other response outcomes – contact and refusal, shows that the third type of sampled units (native Dutch, small households living in urban areas) exhibits below-average contact rates and above-average refusal rates. Sampled units in the fourth cluster (non-Western foreigners living in large households in urban areas) also exhibit below-average contact rates, but have lower conditional probabilities of refusing a request for survey participation.

These findings correspond with earlier studies, which note that using incentives to decrease refusal rates decreases response selectivity in urban areas but increases response selectivity in terms of ethnicity. This is somewhat unexpected, since ethnic minorities mainly live in urban areas. This study demonstrates that two different types of sampled units live in urban areas. Both these groups are currently underrepresented in survey research due to low response rates, and each needs its own tailor-made approach to improve response rates. The contact and participation rates of sampled units living in small households (third cluster) needs to be increased, and the sampled units in the fourth cluster need special attention to increase their contact rates. Earlier studies show that increasing the number of contact attempts from three to six in cases of prior non-contact is especially successful among ethnic minorities (Schmeets 2005b; Feskens et al. 2006) and including incentives in the advanced letter is especially successful among residents of the three largest cities in the Netherlands (Wetzels et al. 2008). For practical reasons, the two strategies – raising the number of contact attempts and offering incentives – can be combined in urban areas.

An important limitation to this study is the availability of comparable variables for the two datasets. To answer our two research questions, we use a model with only six indicators. We nonetheless think these variables cover the important concepts in nonresponse attributes. We also examine models with more indicators for each dataset separately. Including more variables does not substantially alter the general picture and results.

APPENDIX

Variables

Sex; Sex of the sampled unit

0 = male
1 = female

Ethnicity; Is the sampled unit a non-Western foreigner?

0 = no
1 = yes

Job; Does the sampled unit have a paid job?

0 = no
1 = yes

Urbanization at postal code level; Urbanization at the postal code level of the address where the sampled unit is registered:

1 < 500 addresses per square km
2 500 - < 1,000 addresses per square km
3 1,000 - < 1,500 addresses per square km
4 1,500 - < 2,500 addresses per square km
5 > 2,500 addresses per square km

Size of household;

1 1-person household
2 2-person household
3 3-person household
4 4-person household
5 5-person household
6 6-or-more-person household

Response; Did the sampled unit respond or partially respond?

0 = nonresponse
1 = response

Contact; Was contact established with the sampled unit during the fieldwork period?

0 = no, contact was not established
1 = yes, contact was established

Refusals; Did the sampled unit refuse to participate in the survey?

0 = sampled unit refused to provide requested information

1 = sampled unit cooperated

3 *Nonresponse among Ethnic Minorities: A Multivariate Analysis*

This chapter examines the nonresponse among ethnic minorities in the Netherlands. We have constructed a structural equation model with various response outcomes, controlling the effect of ethnicity on the response outcomes for various socio-economic and socio-demographic variables. The effect of ethnicity on response is almost entirely mediated by the degree of urbanization. We have also performed multiple group analyses to examine differences between ethnic groups in the response outcome predictors. Here again, we note that urbanization has a negative effect on the response probabilities in all the ethnic groups and in particular on the contact probabilities. This negative effect is somewhat larger, however, among sampled units with a non-Western background.[1)]

[1)] This chapter has been published as Feskens, R.C.W., Hox, J.J., Lensvelt-Mulders, G.J.L.M., and Schmeets, J.J.G. (2007). Nonresponse among Ethnic Minorities: A Multivariate Analysis. *Journal of Official Statistics,* 23, 3, 387-408.

3.1 *Introduction*

Nonresponse rates in survey research have increased in recent years in almost all the Western countries (De Heer and De Leeuw 2002). For several reasons, this is a problem. Firstly, nonresponse reduces the number of respondents and consequently the precision of estimates. Secondly, nonresponse can increase the costs of survey research since greater efforts are needed to reach the desired sample size. Thirdly, if nonresponse is selective, the survey estimates may be biased and not accurately reflect the true values of the target population (Groves and Couper 1998; Thornberry and Massey 1988). Nonresponse is selective when nonrespondents differ systematically from respondents in terms which matter to the survey objectives (Groves and Couper 1998). The most disturbing consequence of nonresponse is the bias in point estimators (Groves 1989). Biased estimates are more likely to occur if specific groups exhibit below-average response rates. This makes it more likely that the nonrespondents differ systematically from the respondents, since the nonresponse is not random. So in order to speculate about nonresponse bias, it is important to look at response rates among various subgroups (Thornberry and Massey 1988).

Due to their above-average nonresponse rates Statistics Netherlands has difficulties in surveying the ethnic minority or immigrant population. Ethnic minorities constitute about 20% of the Dutch population (http://statline.cbs.nl). The ethnic minority or immigrant population is defined in the Netherlands as 'everyone residing in the Netherlands with one or both parents who were born abroad' (Reep 2003). A further distinction is usually drawn between people with one or both parents born in Europe, North America, Australia, Japan or Indonesia and people from non-Western countries (mainly Turks, Moroccans, Surinamese and Antilleans).[2)] The two groups are of approximately the same size.

The response problems among ethnic minorities are not restricted to Statistics Netherlands. Ethnic minorities have lower response rates in almost all the Western countries (Eisner and Ribeaud 2007; Feskens et al. 2006). Nonresponse among ethnic minorities is becoming politically relevant. Ethnic minority interest groups in the United States have organized to avoid alleged under-counts in survey research. Comparing groups and cultures is an essential feature of survey research (Harkness, Van de Vijver, and Mohler 2003). However, differences in the response rates of various ethnic groups may bias overall survey estimates. Couper and De Leeuw (2003) indicate that the under-representation of ethnic groups may threaten studies on values and norms, e.g., with regard to Sunday observance or commercial activities. Schmeets (2005a) presents other results where below-average response rates among ethnic minorities may bias survey results. Even after correction for age, ethnic minorities are found to be less happy, less healthy and less active in club life and have a greater sense of insecurity. Above

[2)] For reasons of simplicity we use 'Western foreigners' and 'non-Western foreigners' in this article.

that, better ethnic minority response rates are needed for good estimates of sub-populations.

One way to find out more about possible selectivity would be to use background information available for all the sampled units. Recent changes in Dutch legislation have enabled Statistics Netherlands to link administrative records, resulting in a unique database (Houbiers 2004). It provides rich background information about the nonrespondents, enabling us to more thoroughly examine the nonresponse patterns.

A previous study by Schmeets and Michiels (2003) demonstrate that the high ethnic minority nonresponse rates can be attributed to socio-economic status and urbanization. In particular, non-Western foreigners tend to have lower response rates than the native population. At the same time they live predominantly in urban areas, they are more often unemployed and have lower education levels than the native population. These characteristics correlate negatively with response rates (Goyder, Lock, and McNair 1992; Lavrakas 1993; Groves and Couper 1998; Stoop 2004; Van Goor et al. 2005).

The study by Schmeets and Michiels uses a logistic regression model. In this study we examine whether their conclusion also holds true if structural equation techniques are used. Structural equation modeling is more suitable than log linear modeling in several ways. Firstly, structural equation modeling makes it possible to include indirect effects. A variable can be independent as well as dependent simultaneously. Secondly, structural equation models can incorporate latent variables. In this analysis, we examine the concepts of social economic status and urbanization in a more detailed way than would otherwise be possible. Using latent variables makes it possible to measure these constructs more precisely. Thirdly, structural equation models provide more model fit statistics than simple logistic models do. We use the method proposed by Schneekloth and Leeven (2003) to assess the nonresponse bias introduced by below-average ethnic minority response rates. They use logistic regression analysis to evaluate the degree to which the sample nonresponse can be traced back to population characteristics. Pseudo R square values are used to evaluate the explanatory power of the total model. In addition, the model is elaborated by including more variables and multiple group analyses.

Lastly, the effects on non-contacts and refusals are illustrated. These considerations have resulted in three research questions:

1) *Do ethnic minorities in the Netherlands have lower response rates, contact rates and cooperation rates?*
2) *What is the effect of ethnicity on the various response outcomes if controlled for other socio-economic and socio-demographic variables?*
3) *Do response models differ between various ethnic groups?*

By addressing these questions, we hope to gain a better understanding of the nonresponse problem among ethnic minorities in a multivariate environment. The

available data are described in the second section of this chapter the methods are described in the third and the results in the fourth, which is divided into three parts. In the first part we show the response rates among ethnic minorities, the second part describes the construction of the structural equation model used to examine the effect of ethnicity on the response controlled for other variables, and the third part presents the results of our multiple group analyses. Lastly, our conclusions are given in section five.

3.2 *Data*

We have performed our analyses on the survey files of the Continuous Survey on Living Conditions (POLS) 1998 conducted by Statistics Netherlands. About 40,000 interviews are conducted every year. POLS is an integrated survey on living conditions of the Dutch population in private households. The POLS design is based on a modular structure consisting of a joint sample frame and a joint questionnaire. The observation units are individuals. The sample frame is the Population Register from all Dutch municipal basic administrations. POLS is a two-stage sample. First communities are drawn, and then people. Large cities are automatically included (Schouten 2003). Communities and persons are drawn in such a way that the first order inclusion probabilities are equal across all sampled units with the exception of age, since the target population in some modules has age restrictions. We concentrate our analyses on the joint questionnaire, with the total Dutch population (except residents of non-private households) as the target population. Participation is voluntary in POLS and the survey is solely administrated in Dutch. Every month a sample of about 3,500 people is drawn. In 1998, there is a two-month fieldwork period for the twelve consecutive samples. In the first month, Statistics Netherlands collects data with a CAPI mode and nonrespondents with a known telephone line are re-approached with a CATI technique. The nonrespondents without a known telephone line and the sampled units who cannot cooperate due to illness are re-approached with CAPI (Schouten 2003).

The POLS survey is supplemented by administrative data from the Population Register and information about employment and social benefits (Schmeets and Michiels 2003; Schouten 2003). Linking the administrative records makes socio-demographic and socio-economic information available on the nonrespondents at the individual and postal code level. This information has been gathered at Statistics Netherlands in the Social Statistical Database, in which several registers are linked to each other as well as to data from sample surveys (Houbiers 2004). For an extensive summary of the construction of this database, see Houbiers (2004). In the POLS 1998 survey, 39,431 sampled units are drawn and the number of respondents according to AAPOR response definition two is 23,993 (60.8%) (AAPOR 2006), which is not unusually low in the Netherlands (see e.g. De Heer 1999).

The additional information provided by the link to administrative data makes it

possible to study the nonrespondents. However, no extra information is available on 1,143 of the units sampled (2.9% of the total sample). Since no systematic missing data pattern is found with respect to important background variables for these 1,143 cases, they are considered to be missing completely at random and deleted from the data file. Because these numbers are quite small, they can be dropped from the sample without a significant loss of information. Since we only analyse sampled units aged 15–65[3], 28,542 sampled units are left for analysis. The response in this sub-sample is somewhat lower: 60.4% or 17,123 sampled units respond. To avoid capitalization on chance in the analysis, we randomly split this new file into an exploration file consisting of odd case numbers (14,271 cases), and a validation file consisting of the even case numbers (14,271 cases).

As is noted above, in the Netherlands ethnic minorities are defined as everyone residing in the Netherlands with one or both parents born abroad. The following table shows the ethnicity distribution in the Netherlands in 1998, the year the POLS survey is conducted.

Table 3.1
Population in the Netherlands according to ethnicity 1998 (http://statline.cbs.nl)

Total population	15,654,192	
Native population	13,033,792	83.3
Ethnic minorities	2,620,400	16.7
Western foreigners	1,341,947	8.6
Indonesia	407,885	2.6
Germany	405,911	2.6
Belgium	111,537	0.7
Non-Western foreigners	1,278,453	8.2
Suriname	290,467	1.8
Turkey	289,777	1.8
Morocco	241,982	1.5
Netherlands Antilles & Aruba	92,105	0.6

3.3 *Methods*

First of all, we looked at the bivariate relationships between ethnic groups and several response categories, and this provides information for answering the first research question. To address the second research question, we construct a structural equation model. Structural equation modeling allows us to combine latent variables and structural relationships between them and other observed variables (Kline 1998). The advantages of path models (the so-called structural component)

[3] The absolute number of ethnic minorities (especially those of non-Western descent) above 65 is very limited. In the year that the survey was held this number was 22,675 or less than 0.15% of the total population.

and factor models (the measurement component) are combined in structural equation models. Using latent variables reduces the effect of measurement errors. Structural equation modeling also makes it possible to analyse models in which variables are both exogenous and endogenous, and hence, the use of indirect effects. Indirect effects are useful for evaluating the combined effect of ethnicity, urbanization and socio-economic status (SES) on response probabilities. We first construct a structural equation model in the exploration file. This model is validated in the validation file. As Groves and Couper (1998) note, "dissecting the nonresponse phenomenon into one of non-contacts, refusals and other causes sensitizes us to considering alternative causes of each outcome." We thus not only analyse the response outcome, we also address non-contacts and refusals. Other causes of non-response only have a minor impact on the response rate, as is shown in Table 3.2 and are therefore not further analysed.

We also want to know which variables the two groups (native/Western foreigners vs. non-Western foreigners) differ on as regards the response phenomenon. We address this point by conducting a multiple group analysis and statistically comparing the path coefficients of the groups. The multiple group analyses are conducted on the total sample aged 15–65. Here again we dissect the response outcomes, which makes it possible to examine the various response predispositions in the sub-groups. Since the sample size is large, the assessment of model fits is based on two goodness-of-fit indices that are less sensitive to sample size, namely the Bentler comparative fit index (CFI; see Bentler 1990) and the root mean square error of approximation (RMSEA) value (Browne and Cudeck 1993). The CFI value indicates the degree of improvement of the overall fit of the specified model relative to an independence model in which the variables are assumed to be uncorrelated (Kline 2005).

$$CFI = 1 - \frac{\chi_m^2 - df_m}{\chi_b^2 - df_b} \qquad \textbf{(1)}$$

The RMSEA fit index is an exact fit in which the null hypothesis states that the model corresponds to the data (RMSEA= .00). This value is calculated as follows:

$$RMSEA = \sqrt{\frac{\chi_m^2 - df_m}{ndf_m}} \qquad \textbf{(2)}$$

3.4 Results

3.4.1 Do ethnic minorities in the Netherlands have lower response rates, contact rates and cooperation rates?

The overall response rates for sampled units aged 15–65 in the POLS 1998 survey is 60.4% (AAPOR response definition No. 2). Further dissecting the nonresponse in alternative outcomes is as follows for the three ethnic groups:

Table 3.2
Response outcomes among ethnic groups in POLS 1998 in percentages

	Native population	Western foreigners	Non-Western foreigners
Response	62.0	57.2	39.7
Non-contacts	12.0	15.9	26.4
Refusals	24.7	23.0	20.1
Language problems	0.0	2.5	13.0
Other	1.3	1.8	1.0
n	24,005	2,511	2,026

Figures may not add up to 100% due to rounding.

The distribution of nonresponse outcomes is more or less similar among Western foreigners and the native population. The relative numbers of non-contacted sampled units and nonresponse due to difficulties with the survey language are higher among non-Western foreigners than among the native population, a result also found in other surveys (Feskens et al. 2006). These are the primary reasons for the lower response rates among non-Western foreigners.

We also address the bivariate relationships between ethnicity and various socio-demographic and socio-economic variables. Here again, the observed characteristics of Western foreigners are very similar to those of the native population, whereas those of non-Western foreigners differ substantially (Schmeets 2005a). We consequently decide to focus our analysis on non-Western foreigners. Ethnicity is a dichotomous variable in the following analyses with non-Western foreigners as the first and Western foreigners and the native population as the second group.

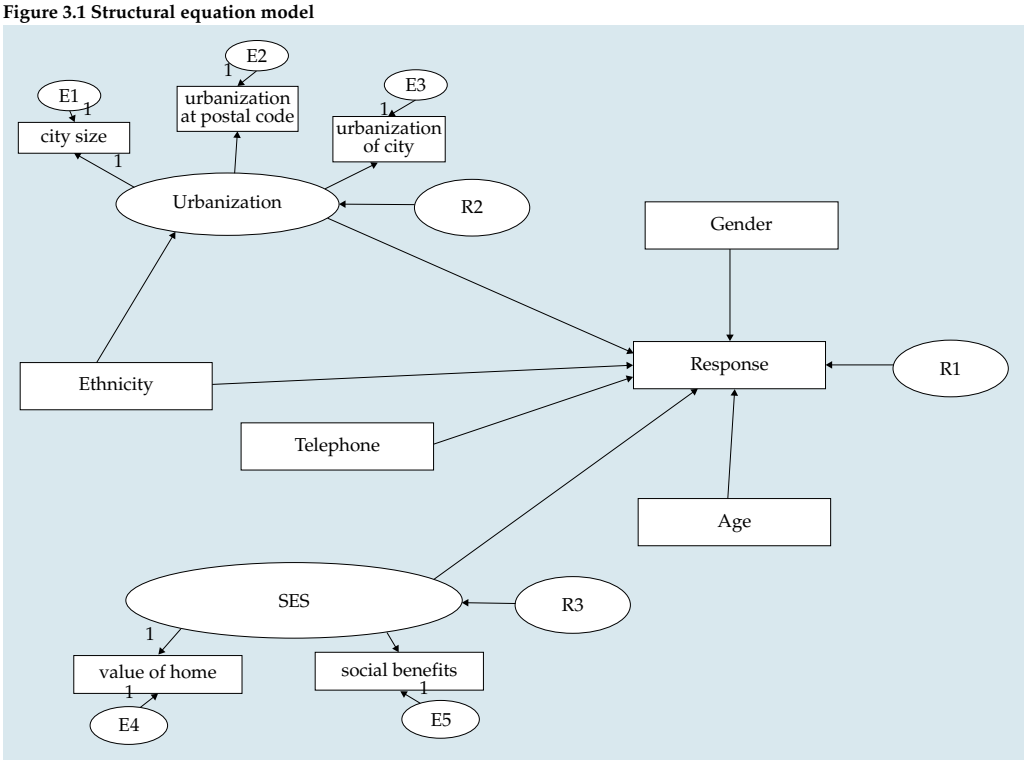
3.4.2 What is the effect of ethnicity on the various response outcomes if controlled for other socio-economic and socio-demographic variables?

3.4.2.1 Structural equation model

We want to see whether the nonresponse is still affected by ethnicity if controlled for socio-demographic and socio-economic variables. First we consider the bivariate relationships between each of the socio-demographic and socio-economic variables and the response for all the ethnic groups separately. Then we looked at the bivariate relationships between the ethnic groups and the socio-economic and

socio-demographic variables. With this information and the theoretical consideration that urbanization and SES have been related to survey nonresponse for many years (Goyder et al. 1992; Lavrakas 1993; Groves and Couper 1998; Stoop 2004; Van Goor et al. 2005), we construct a structural equation model enabling us to control the effect of ethnicity on the various response outcomes for the other variables, see Figure 3.1.

Figure 3.1 Structural equation model



The measurement part of the model consists of the latent variables urbanization and SES. The latent variable urbanization is measured by the observed variables urbanization at postal code level, degree of urbanization of the city, and city size. These variables are sufficiently correlated with each other, but not correlated to the extent that they measure the same. The latent variable SES is measured by home values and an indicator for receiving social benefits. For identification purposes, we fixed the factor loadings of the indicators urbanization of the city and home value at one.

Probit regressions are estimated for the categorical factor indicators, and simple linear regressions are estimated for the continuous factor indicators (Muthén 1998-2004). We construct the two latent variables because using the extra information of all the indicators makes it possible to identify urbanization and SES more precisely.

Moreover, some indicators are subject to subjective classification. Using latent variables with more indicators reduces this form of measurement error. The relationships between the other observed variables and latent variables on response constitute the structural part of the model. These observed variables are ethnicity, an indicator for having a known telephone land-line (telephone), gender and age. These observed variables are regressed on the binary outcome variable response.
Since we want to see whether the relation between ethnicity and response is mediated by SES and urbanization, we also regress these latent variables on ethnicity. The relationships are assumed to be unidirectional, and the latent variables SES and urbanization and also SES and having a known telephone land-line are assumed to covary (these relationships are not drawn in the figure). We also explored including interaction terms in the model, but this did not improve the model fit significantly. The regression coefficients are estimated with the unweighted least square estimator and are interpreted as probit regression coefficients. The estimates between parentheses are standardized coefficients using the variances of the continuous latent variables as well as the variances of the background and outcome variables (Muthén and Muthén 1998–2004). The sign of the regression coefficients in Tables 3.3 – 3.8 shows the extent to which this characteristic changes the probability in the nonresponse (–) or response direction (without sign). We assume that the measurement errors are uncorrelated.

3.4.2.2 Response

The results of the structural equation model with the dependent variable response defined as the AAPOR response definition number two are presented in Table 3.3.

Table 3.3
Structural equation model with dependent variable response

	Explained variables					
Predictors	Estimates on response	S.E.	Estimates on urbanization	S.E.	Estimates on SES	S.E.
Size of city	.	.	1.000 fixed (0.874)	0.000	.	.
Urbanization of city	.	.	0.825** (0.914)	0.011	.	.
Urbanization of postal code	.	.	0.870** (0.887)	0.015	.	.
Value of home	.	.	.	.	1.000 fixed (0.648)	0.000
Social benefits	.	.	.	.	–0.370** (–0.491)	0.020
Gender	0.075** (0.037)	0.021	.	.	.	.
Age	–0.002** (–0.033)	0.001	.	.	.	.
Urbanization	–0.097** (–0.142)	0.011	.	.	.	.
SES	0.047** (0.063)	0.018	.	.	.	.
Telephone	0.341** (0.139)	0.031	.	.	.	.
Ethnicity	–0.285** (–0.072)	0.048	1.453** (0.250)	0.054	–1.805** (–0.340)	0.073

Note: dependent variable coded 1 = response, 0 = nonresponse; gender coded 1 = female, 0 = male;
ethnicity coded 1= non-Western foreigners, 0 = native population & Western foreigners.
**$p < .01$
Chi square = 125.432 (df = 4); RMSEA = .046; CFI = .972, R square = .072
Validation file: Chi square = 91.611 (df = 3); RMSEA = .045; CFI = .980, R square = .082

The relation between ethnicity and response is almost entirely mediated by urbanization and the socio-economic status of the sampled units. A large amount of the negative effect of ethnicity on response is mediated by urbanization. The standardized probit regression of urbanization on response is –.142 and the standardized effect of ethnicity on urbanization is .250. In particular, urbanization and telephone have a large impact on the response probability. The latent variable SES does not strongly affect the probability of responding. The total standardized effect of ethnicity on response is –.129, which is the sum of the direct effect and all the indirect effects. The sum of all the unstandardized indirect effects is –.227, which is the sum of the product of the unstandardized coefficients for the paths from ethnicity via SES to response and ethnicity via urbanization to response. The standard error for all the indirect effects of ethnicity on response is .025.

The sum of the direct and indirect effects of ethnicity (–.129) is still less than the direct effects of urbanization (–.142) and owning a landline telephone (.139). Ethnic minorities have lower response rates than the native population, but ethnic minorities also disproportionally live in urban areas. These results suggest that in particular, it is this urbanization effect that 'causes' lower response rates among ethnic minorities, and not ethnicity itself. If controlled for other variables, ethnicity only has a small effect on the response probability and a fairly small standardized coefficient of ethnicity on response (–.072) remains in this multivariate environment.

The socio-economic status of the sampled units barely affects the response probabilities, suggesting that the negative effect of ethnicity on response is mainly mediated by socio-demographic rather than socio-economic characteristics. This also holds true of the native population. Regardless of their ethnicity and socio-economic status, people who live in large cities have lower response probabilities than those who do not. The availability of a land-line proves to be a strong indicator for response. This not only holds true of the second part of the fieldwork period when the re-approaching strategies differ between a CATI mode for nonrespondents with a known phone line, or otherwise CAPI, it also holds true of the first part of the fieldwork period when no telephone calls are made. Age and gender only have a small influence on response; older sampled units have slightly smaller response probabilities.

We not only inspect path coefficients to examine the effect of ethnicity on response, we also use the Schneekloth and Leeven (2003) method to evaluate nonresponse bias introduced in this case by ethnicity. This is done by examining the pseudo R square values of the total model. This value can serve as an indicator of the amount of nonresponse bias introduced while including background variables. If the model does not, or only poorly, predicts whether the sampled units will or will not respond, the nonresponse pattern can be seen as random and thus as following the basic logic of probability sampling. According to Andreß, Hagenaars, and Kühnel (1998), values below 0.05 indicate low and negligible correlation, and values above 0.2 indicate a strong correlation. Of course, the explanatory power is heavily dependent on the availability of information for the respondents and nonrespondents. However, as is

noted above, the construction of the Social Statistical Databases at Statistics Netherlands provides detailed socio-demographic and socio-economic information. The effect of ethnicity decreases substantially if the model controls for other variables. Not only does the path coefficient decrease if the indicator for ethnicity is included in the analysis, the pseudo R square only increases by 1.0% (from 6.2% to 7.2%). Small path coefficients and a low pseudo R square suggest that predicting response is still fairly difficult, even if rich background information is available. This suggests that the effect of ethnicity on response and the response bias is not as high as some bivariate relationships seem to indicate. However, there may be a certain amount of selective nonresponse in urban areas. Urban residents, regardless of the ethnic group they belong to, are somewhat under-represented in this survey.

3.4.2.3 Contact

Fieldwork strategies that can successfully increase contact rates have been described in greater detail in recent years (Groves and Couper 1998; Bates 2004). The number of contact efforts and time of contacting sampled units are among the factors noted to explain contactibility. Unfortunately, this kind of fieldwork information is not available for this survey and thus cannot be included in the model examining the pattern of contacting the sampled units. For this reason, and for the comparability between the models, we use the same model to explain the contact process and in the following section the cooperation process. The results of the structural equation model with dependent variable contact defined as AAPOR contact definition No. 3 are presented in Table 3.4.

Table 3.4
Structural equation model with dependent variable contact

	Explained variables							
Predictors	Estimates on contact		S.E.	Estimates on urbanization		S.E.	Estimates on SES		S.E.
Size of city	.	.		1.000 fixed	(0.874)	0.000			
Urbanization of city	.	.		0.825**	(0.914)	0.011			
Urbanization of postal code	.	.		0.870**	(0.887)	0.015			
Value of home	.	.		.		.	1.000 fixed	(0.649)	0.000
Social benefits	.	.		.		.	–0.368**	(–0.489)	0.020
Gender	0.060**	(0.030)	0.027	.		.	.		.
Age	0.004**	(0.056)	0.001	.		.	.		.
Urbanization	–0.183**	(–0.267)	0.012	.		.	.		.
SES	–0.014 ns	(–0.019)	0.020	.		.	.		.
Telephone	0.386**	(0.156)	0.035	.		.	.		.
Ethnicity	–0.215**	(–0.054)	0.052	1.453**	(0.250)	0.054	–1.806**	(–0.340)	0.073

Note: dependent variable coded 1 = contact, 0 = no contact; gender coded 1 = female, 0 = male;
ethnicity coded 1 = non-Western foreigners, 0 = native population & Western foreigners.
**$p < .01$, ns = not significant
Chi square = 123.290 (df = 4); RMSEA= .046; CFI = .972. R square = .104
Validation file: Chi square = 90.715 (df = 3); RMSEA = .045; CFI = 0.980, R square = .113

In line with the argument formulated by Groves and Couper (1998), the role of urbanization is even more pronounced in the contact process. The standardized coefficient of urbanization on the contact probability is –.267. The effect of ethnicity on contact is less than on the response rate, indicating that in this model, the effect of ethnicity is mediated even more strongly by urbanization. The unstandardized estimate for all the indirect effects from ethnicity to contact is –.240 here, with a standard error of .029. Groves and Couper (1998) note that it is more difficult to establish contact with urban sampled units than non-urban sampled units.

These results show that the same is true for ethnic minorities. Nonresponse among ethnic minorities is heavily determined by low contact rates, which are not unique to ethnic minorities since they are largely mediated by urbanization. This also helps explain why ethnic minorities have lower response rates. Contact difficulties are mainly concentrated in urban areas. Nonrespondents with a known landline are re-approached in the second month with a CATI mode. Other nonrespondents are re-approached again with a CAPI mode. However, due to capacity problems in the interview staff, not all the nonrespondents without a known land line are re-approached, or fewer contact efforts are made than in the case of nonrespondents in the CATI mode. Not surprisingly, the regression coefficient from telephone on the contact probability is high. Nonetheless, this positive effect of having a known land phone on the contact probability is also found in the first month of the interview process when no telephone calls are made. Furthermore, the results show that women and the elderly are somewhat easier to contact, albeit with small probabilities.

Table 3.5
Structural equation model with dependent variable cooperation

	Explained variables								
Predictors	Estimates on cooperation		S.E.	Estimates on urbanization		S.E.	Estimates on SES		S.E.
Size of city	.	.		1.000 fixed	(0.860)	0.000			
Urbanization of city	.	.		0.856**	(0.911)	0.012			
Urbanization of postal code	.	.		0.896**	(0.882)	0.017			
Value of home	.	.				.	1.000 fixed	(0.623)	0.000
Social benefits	.	.				.	–0.400**	(–0.501)	0.025
Gender	0.070 ns	(0.035)	0.024			.			.
Age	–0.003**	(–0.046)	0.001			.			.
Urbanization	–0.041**	(–0.057)	0.013			.			.
SES	0.042*	(0.054)	0.021			.			.
Telephone	0.257**	(0.102)	0.034			.			.
Ethnicity	0.188**	(0.045)	0.058	1.324**	(0.226)	0.057	–1.715**	(–0.321)	0.082

Note: dependent variable coded 1 = cooperation, 0 = refusal; gender coded 1 = female, 0 = male;
ethnicity coded 1 = non-Western foreigners, 0 = native population & Western foreigners.
*$p < .05$, **$p < .01$, ns = not significant
Chi square = 177.247 (df = 6); RMSEA = .048; CFI = .949, R square = .025
Validation file: Chi square = 142.455 (df = 5); RMSEA = .047; CFI = .956, R square = .033

3.4.2.4 Cooperation

To examine the process of cooperation (defined as AAPOR cooperation definition No. 2) we include only those sampled units who are contacted in the first place so that 12,202 sampled units remain for the exploration file and 12,366 for the validation file. The results are presented in Table 3.5.

Somewhat surprisingly, ethnicity has a positive effect on the cooperation probability. Non-Western foreigners tend to refuse a request to participate in a survey less often than the other sampled units. Because of the low contact rates and the high nonresponse due to language problems among ethnic minorities, the results should be interpreted with caution. Sampled units who are not contacted do not have an opportunity to refuse a request to take part in a survey (Hox and De Leeuw 1998). Moreover, sampled units can use language problems as a friendly way to refuse to participate. The indirect effect of ethnicity on cooperation via SES and urbanization is negative, –.127 (standard error .028), which also suggests that the positive effect of ethnicity on the cooperation rate should not be over-interpreted. The effect of urbanization, which is large in the response and contact process, is also lower, suggesting that there is a contact problem and not a participation problem in urban areas. The effect of SES is very small. Again, having a known land phone proves to be a strong predictor. Older sampled units have a somewhat higher probability of refusing to participate. As indicated by a low pseudo R square value, the cooperation process is even more difficult to predict than the response and contact process.

3.4.3 Do response models differ between various ethnic groups?

In order to examine whether predictors for various response outcomes have the same influence for various ethnic groups, we perform a multiple group analysis with the same two groups. In multiple group analyses, the significance of group differences on model parameters can be tested by imposing cross-group equality constraints (Kline 1998). More general references on multiple group analysis can be found in Bollen (1989) and the Mplus technical appendix (www.statmodel.com). The native population and Western foreigners constitute the first group, and the non-Western foreigners the second. The results of the multiple group analysis with outcome variable response are presented in Table 3.6.

We perform our analysis on 26,479 native sampled units and Western foreigners (first group) and 1,893 non-Western foreigners (second group). Group membership moderates the relationship between having a known land phone and response. The negative effect of urbanization on response probability is only slightly larger for the non-Western foreigners. Age has a somewhat larger negative effect on the response rate on the second group than on the first group. This reflects the higher nonresponse due to language problems, which is almost entirely found among older non-Western foreigners. In sum, the two groups do differ in their response

Table 3.6
Multiple group analysis on response

	Explained variables					
	Native population & Western foreigners			Non-Western foreigners		
Predictors on response	Estimates		S.E.	Estimates		S.E.
Gender	0.063**	(0.032)	0.016	0.153**	(0.076)	0.058
Age**	–0.002*	(–0.022)	0.001	–0.008*	(–0.092)	0.002
Urbanization**	–0.085**	(–0.127)	0.006	–0.180**	(–0.244)	0.032
SES	0.021**	(0.036)	0.007	0.016 ns	(0.027)	0.032
Telephone*	0.318**	(0.129)	0.020	0.173*	(0.080)	0.074

Note: gender coded 1 = female, 0 = male.
*$p < .05$, **$p < .01$, ns = not significant
* after variable name reflects significant difference between groups for this variables at $p < .05$, **$p < .01$
Chi square= 14.512 (df = 5) p = .0126; RMSEA = .012, CFI = .698; R square1 = .047, R square2 = .097

process on the path coefficients age, urbanization, and having a known land-line. Table 3.7 shows the results for the two groups on the dependent variable contact, which demonstrate that the response probability of the first group is more influenced than that of the second by having a known land phone.

Table 3.7
Multiple group analysis on contact

	Explained variables					
	Native population & Western foreigners			Non-Western foreigners		
Predictors on contact	Estimates		S.E.	Estimates		S.E.
Gender	0.076**	(0.038)	0.020	0.152*	(0.075)	0.085
Age*	0.006**	(0.082)	0.001	–0.002 ns	(–0.019)	0.003
Urbanization**	–0.138**	(–0.206)	0.007	–0.247**	(–0.336)	0.036
SES	0.002 ns	(0.004)	0.007	0.002 ns	(0.004)	0.035
Telephone**	0.291**	(0.118)	0.022	0.198*	(0.092)	0.096

Note: gender coded 1 = female, 0 = male.
*$p < .05$, **$p < .01$, ns = not significant
* after variable name reflects significant difference between groups for this variables at $p < .05$, **$p < .01$
Chi square = 13.198 (df = 5) p = .0215; RMSEA = .011, CFI = .739; R square1 = .075, R square2 = .143

Again, the parameter indicator for having a known land phone varies across groups. Contact probabilities of the first group are heavily influenced by this predictor. In the second group, the negative effect of urbanization is larger than in the first group. Nonetheless, urbanization has also a relatively large negative effect on the contact rate among the native population and Western foreigners. Table 3.8 shows the results of the multiple group analysis with the outcome variable refusal.

For this multiple group analysis, where we examine the predictors for the sampled units who refuse to participate across the groups, we again only include the sampled units who are contacted. The selection of contacted units results in 23,210 remaining sampled units for the first group, and 1,359 non-Western foreigners with whom contact has been established remain for analysis. For the non-Western foreigners, urbanization has a somewhat more negative effect on the response outcome, in this case the refusal rate. Having a known land phone again proves to be a strong positive predictor for the first group but does not have much impact on the second group.

Table 3.8
Multiple group analysis on cooperation

	Explained variables					
	Native population & Western foreigners			Non-Western foreigners		
Predictors on refusals	Estimates		S.E.	Estimates		S.E.
Gender	0.044*	(0.022)	0.018	0.142*	(0.071)	0.072
Age*	–0.005**	(–0.062)	0.001	0.003 ns	(0.043)	0.003
Urbanization*	–0.004**	(–0.059)	0.007	–0.136**	(–0.179)	0.050
SES	0.015*	(0.028)	0.006	–0.022 ns	(–0.039)	0.048
Telephone	0.264**	(0.104)	0.022	0.185 ns	(0.088)	0.096

Note: gender coded 1 = female, 0 = male.
*$p < .05$, **$p < .01$, ns = not significant
* after variable name reflects significant difference between groups for this variables at $p < .05$, **$p < .01$
Chi square = 2.804 (df = 2) p = .2429; RMSEA = .006, CFI = .911; R square1 = .023, R square2 = .047

3.5 *Conclusions*

Ethnic minorities are a growing part of Western societies, and are increasingly relevant for policy-makers. According to the Statistics Netherlands definition, almost 20% of the Dutch population have a foreign background and are called *"allochtonous"*. Predictions are that the percentage of ethnic minorities in the Netherlands will increase to 35% by 2050 (De Jong and Hilderink 2004). With an increasing demand for data about ethnic minorities and decreasing response rates among them, more attention is devoted to the quality of the data about ethnic minorities. Nonresponse itself does not automatically imply bias in point estimates. However, nonresponse rates can serve as an indicator for potential bias problems. Nevertheless, reducing nonresponse should focus on reducing nonresponse error. Simply trying to increase response rates can actually increase the survey error (Merkle and Edelman 2002). Therefore it is important to know which societal groups have high nonresponse rates, so that tailored strategies can be developed to reduce nonresponse in these under-represented sub-groups.

The analyses in this article are based on the results of the survey on living conditions in the Netherlands in 1998. Bivariate tables of response and ethnicity show large differences in the response rates between various ethnic groups. One of the most interesting findings is the high non-contact rate among ethnic minorities and more specifically among non-Western foreigners. Surprisingly, the cooperation rate among ethnic minorities is higher than among the native population. If sampled units are not contacted, of course it is impossible for them to refuse a request to participate. An increase in the minimum number of contact efforts in cases of earlier non-contact at Statistics Netherlands in March 2004 nonetheless shows a substantial increase in the contact and response rate among non-Western foreigners, but not in the refusal rate (Schmeets 2005b).

The results of the structural equation models show that the negative effect of ethnicity on response partially disappears if we control for other variables. In particular, urbanization has a strong effect on nonresponse. However, ethnicity still affects the response probability. Knowing that nonresponse among ethnic minorities is especially found in highly urbanized areas nevertheless enables the researcher to focus on this very specific group. Of course there is a high correlation between urban areas and ethnic minorities, but in the Netherlands about 70% of the population of foreign descent and 70% of the non-Western foreigners do not live in one of the four major cities. Urbanization has been related to survey nonresponse, and more specifically nonresponse due to non-contact for many years (Groves and Couper 1998, Steeh et al. 2001). Indeed, one of the reasons for higher nonresponse rates in the Netherlands than most other countries might be the urban nature of the western part of the country.

Survey response remains a process greatly influenced by chance. Some groups, like ethnic minorities, do have lower response rates. But, fortunately, even with the inclusion of background information, it is extremely difficult to predict whether a specific person will respond or not. Although a great deal of administrative information is available, a low pseudo R square value (.072) suggests that the nonresponse is not very systematic. In particular, the refusal process seems to be greatly influenced by chance, as indicated by even lower pseudo R square values. Although regression coefficients and R square are small, the focus should be on enlarging the response in urban areas, where there is a serious contact problem. Tailoring data collection strategies for higher contact rates will lead to increased response rates among ethnic minorities.

To sum up, urban sampled units are more difficult to contact and more contact efforts are needed. This holds particularly true of non-Western urban sampled units. However, this extra negative effect is limited. Steeh et al. (2001) observe a trend of increasing non-contact rates in some American metropolitan areas, and urge survey methodologists to plan for this eventuality. To keep fieldwork procedures at a manageable level, a tailored fieldwork strategy might be conceivable, involving all the urban sampled units receiving more contact efforts by earlier non-contact. In the Netherlands, the minimum number of contact efforts by earlier non-

contact was changed from three to six in March 2004[4] (Snijkers and Kockelkoren, 2004). This has had a very positive effect on the contact and response rates, in particular among ethnic minorities (see Feskens et al. 2006; Schmeets 2005b). This supports our findings on the contact problem among ethnic minorities. Increasing contact efforts will decrease the nonresponse among ethnic minorities. However, the question still remains as to the kind of nonrespondents there will be after a change in the fieldwork procedure. Future research could focus on this aspect.

The multiple group analysis results show that ethnic groups are not homogeneous in their response processes. Age and urbanization have a more negative impact on the response and contact probability of non-Western foreigners than of the native population and Western foreigners. However, having a known land phone proves to be a very strong indicator for responding among the native population and Western foreigners. This positive effect is much smaller among non-Western foreigners.

This study has a number of limitations. Although the Social Statistical Database contains a rich amount of data on respondents and nonrespondents, not all the theoretical considerations can be taken into account. And since non-Western foreigners are only about 8% of the population in the 15–65 age group, they might not have much impact on the overall estimates. Although they may not have much impact on the overall estimates, better response rates among ethnic minorities may still be needed to obtain better estimates on the sub-populations. In addition, since these results are based on a Dutch survey, they can only be partially generalized to other countries. Future research could focus on these limitations and use our findings as hypotheses to study in further detail. Nevertheless, our analysis outcomes suggest that although ethnic minorities have lower response rates, the focus should be on enlarging the response in urban areas. Ethnic minorities do not respond as well as the native population, but the explanations and hence the solutions have less to do with a divergent response behavior among ethnic minorities, and more to do with living conditions. Additional efforts should be made to increase the contact rate in urban areas. A possible solution can be to approach the sampled units with specially tailored strategies, e.g., a higher minimal number of contact efforts than for non-urban sampled units.

[4] It is also the maximum number, except for appointments in the sixth attempt. In that case a seventh attempt is allowed.

APPENDIX

Variables

Sex; sex of the sampled unit

0 = male
1 = female

Age; age of the sampled unit

15–65 years

Size of city; size of the community where the sampled unit is registered

1 = small
8 = large

Urbanization of community; urbanization of community where the sampled unit is registered

1 none
2 weak
3 moderate
4 strong
5 very strong

Urbanization at postal code level; urbanization at postal code level of the address where the sampled unit is registered

1 < 500 addresses per square km
2 500 - < 1,000 addresses per square km
3 1,000 - < 1,500 addresses per square km
4 1,500 - < 2,500 addresses per square km
5 > 2,500 addresses per square km

Indicator for telephone; Does the sampled unit have a known registered land phone?

0 = no known registered land phone
1 = known registered land phone

Response; Did the sampled unit respond (partially)?

0 = nonresponse
1 = response

Value of home; value of the home in Dutch guilders where the sampled unit is registered

1 < 50 thousand
2 50 - 75 thousand
3 75 - 100 thousand
4 100 - 125 thousand
5 125 - 150 thousand
6 150 - 200 thousand
7 200 - 250 thousand
8 250 - 300 thousand
9 300 - 350 thousand
10 350 - 400 thousand
11 400 - 500 thousand
12 > 500 thousand

Indicator for social benefits; Does the sampled unit receive some form of social benefits?

0 = no
1 = yes

Ethnicity; Is the sampled unit a non-Western foreigner?

0 = no
1 = yes

Contact; Was contact established with the sampled unit during the fieldwork period?

0 = no, no contact
1 = yes, contact

Refusals; Did the sampled unit refuse to participate in the survey?

0 = sampled unit refused to provide requested information
1 = sampled unit cooperated

4 *Nonresponse among Ethnic Minorities in an International Perspective*

This chapter examines strategies to reduce nonresponse rates among ethnic minorities. The authors review nonresponse rates and data collection strategies among ethnic minorities with respect to response rates and response bias in six European countries. The national statistical institutes of these six countries use different definitions of ethnic minorities. This is why the definitions of ethnic minorities and their impact on the number of members of ethnic minorities in the six countries are compared. Nonresponse rates are usually higher among ethnic minorities than among the native population. Dissecting the nonresponse phenomenon shows that contact rates among ethnic minorities are lower, nonresponse due to an inability to produce the required information is higher, and cooperation rates are higher among ethnic minorities than among the native population. Increasing the response rates among ethnic minorities should focus on enhancing the contact rate and reducing the number of nonrespondents who are unable to produce the required information.[1)]

[1)] This chapter has been published as Feskens, R.C.W., Hox, J.J., Lensvelt-Mulders, G.J.L.M., and Schmeets, J.J.G. (2006). Collecting Data among Ethnic Minorities in an International Perspective. *Field Methods*, 18, 3, 284-304.

4.1 Introduction

In recent years, nonresponse rates in survey research have increased in almost all developed countries (De Heer and De Leeuw 2002). Nonresponse can bias estimates of the target population if nonrespondents systematically differ from respondents with respect to the studied variables. Bias is more likely if the response is not equally distributed among various societal groups. In this case, some groups are under- or overrepresented and the nonresponse is selective. To reduce nonresponse bias, it is essential to increase the response rates of societal groups with below-average response rates. Ethnic minorities tend to have below-average response rates (Snijkers 2003; Schmeets 2004). The terms used to define them are not straightforward: *foreigners, immigrants, people of foreign descent,* and *ethnic minorities* are used throughout Europe (European Commission, Employment and Social Affairs DG 2004). These terms cover various realities, as is noted in this article in greater detail. The term *ethnic minorities* is used below to describe the specific target population. Response rates among ethnic minorities were not always lower than those of the native population. In the 1980s, Bronner (1988) noted above-average response rates among ethnic minorities in the Netherlands. Nowadays, however, response rates among ethnic minorities are usually considerably lower than among the native population (Schmeets 2004). Concerns about bias in survey estimates have stimulated survey methodologists to develop measures to reduce nonresponse among special groups (see Hox et al. 2003). This study was motivated by the growing percentage of ethnic minorities, which is already considerable in Western societies (e.g., almost 20% in the Netherlands), and the reluctance to publish results about ethnic minorities because of their low response rates (Statistics Netherlands 2005). To gain insight into the issue of collecting data among ethnic minorities, we examine strategies for collecting data among ethnic minorities at several European National Statistical Institutes (NSIs) and semiofficial statistical bureaus. Until now, there has been only a limited focus on response rates and potential bias and ethnic minorities. Much of the material in this article is thus explorative and should be interpreted as such. This study has two aims. First, we are looking for successful strategies for collecting data among ethnic minorities that produce high response rates and low nonresponse bias. Second, cross-national research is becoming essential to support policies of international organizations and governments, and there is an increasing need for valid and reliable cross-national survey data (De Heer 1999). Nonresponse is an issue in cross-national research. As Couper and De Leeuw (2003) noted, "Only if we know how data quality is affected by nonresponse in each of the countries can we assess and improve the comparability of international and cross-cultural data." This is why we also compare relevant definitions of ethnic minorities as a necessary step toward evaluating nonresponse differences among countries. In the next section, we describe our methods and data. The definitions and percentages of ethnic minorities are discussed in the third section. Nonresponse rates among ethnic minorities in the six countries are examined in the fourth

section, and strategies to reduce nonresponse rates among ethnic minorities and the discussion are presented in the next two sections.

4.2 *Methods and data*

We developed a questionnaire on response rate issues (final response rates, noncontact rates, refusal rates, and language difficulties), response definitions, definitions of ethnic minorities, approach strategies, and results of experiments. The questionnaire was sent to informants at the national statistical agencies. For practical reasons, we had to confine our analysis to six countries. The countries are included in the analysis based on a purposive sampling rationale (Creswell 1998). The selected countries all have decreasing response rates (De Heer 1999). As developed Western European countries, they constitute a homogeneous cluster. The countries in the final sample are Belgium, France, Germany, the Netherlands, Sweden, and the United Kingdom. In this article, we describe successful strategies survey organizations used to reduce nonresponse rates among ethnic minorities. Germany and Belgium have mandatory participation in major surveys. This makes it difficult to compare the response rates since the number of refusals under these survey conditions is systematically lower than in other EU countries. In addition, the decentralized nature of the Federal Republic of Germany has resulted in sixteen state-level data-collecting institutes with their own far-reaching competences (Allum 1998). This federal data collection in Germany makes it very difficult to obtain comparable nationallevel data. Belgian law (Article 24) prohibits the Belgian national statistical institute from giving third parties data that are classified as to ethnicity (D. Luminet, personal communication, 2004). This makes it impossible to include response rates classified into ethnic groups. Because of these difficulties in obtaining comparable data, we decided to include information provided by ZUMA in Germany and APS Belgium, both semipublic organizations. National statistical institutes use different designs, fieldwork strategies, and other fixed factors related to survey organization (see De Heer 1999). With respect to these differences, the procedures at ZUMA and APS are comparable to those at national statistical institutes. We collected data from various surveys. Ideally, to compare trends internationally, the data for analysis should contain a wide range of survey types over years. However, it is difficult to obtain comparable data, and sometimes it is simply not available. Nonetheless, informants at the statistical institutes provided extensive information. We mainly focus here on the Survey on Living Conditions and the Labor Force Survey. Labor Force Survey information is available for all six countries, although France, the Netherlands, and the United Kingdom could not provide statistics by ethnic minorities. Belgium, France, the Netherlands, and Sweden have provided Survey on Living Conditions or Survey on Health information, although not all of them could provide response rates classified according to ethnic groups. The Labor Force Survey is mandatory in Germany and Belgium,

so of course it gives less information on response rates and more specifically on cooperation rates. The German institute ZUMA conducts the Allgemeine Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS; German General Social Survey), a general social survey. Results of the ALLBUS are available for the period from 1994 to 2002. We have also collected information on relevant definitions, sampling methods, and survey organizational information and can consult the rich results of several experiments on ethnic minorities conducted by the selected institutes.

4.3 *Definitions and number of ethnic minorities*

One of the most striking findings is that different definitions of ethnic minorities are used in the various countries. The size of the ethnic minorities and their countries of origin are among others determined by historical developments such as colonization or by specific legislation. The official definition of ethnic minorities in each country also affects official estimates of the size of its ethnic minorities, as Table 4.1 shows.

Table 4.1
Percentages of ethnic minorities in total national populations

	Official national definition	Percentage ethnic minorities according to official definition	Percentage ethnic minorities according to citizenship
Belgium, 2004	Citizenship	8.2	8.2
France, 2004	Citizenship	12.2	12.2
Germany, 2003	Citizenship	8.9	8.9
The Netherlands, 2004	At least one parent born abroad	18.8	4.1
Sweden, 2003	Born abroad	12.0	5.3
United Kingdom, 2001	Self-identify	7.9	*

* no data available.
** Sources Table 1 and Table 2: all the information is available on the NSI websites.
For France the percentages are based on extrapolated estimations of 20% of the total population.

The official national definition in the six countries is given in the second column. Three countries (Belgium, France, and Germany) classify residents as members of ethnic minorities if they have a foreign nationality. Statistics Netherlands uses the following definition: 'every person residing in the Netherlands of whom one or both parents were born abroad' (Reep 2003). Statistics Sweden usually defines members of ethnic minorities as people who were born abroad. Residents of the United Kingdom selfidentify, stating which ethnic group they belong to. Examples of questions on the ethnicity of a respondent can be found in a guide recently published by the Office of National Statistics (ONS 2003). In the other countries, registration in the municipal base administration or census information are used to

determine how many people are members of ethnic minorities. The percentages of ethnic minorities in the total population according to this official national definition are shown in the third column of Table 4.1. According to the official definition, almost one in five residents of the Netherlands are members of an ethnic minority, which is the largest percentage in the six countries. The differing definitions make it difficult to compare ethnic minorities internationally. A possible solution might be to use the citizenship criterion for all the countries. Then the percentage of ethnic minorities in the Netherlands and Sweden decreases considerably, as is shown in the fourth column. If residents of the Netherlands with a Dutch and at least one other nationality are counted as members of ethnic minorities, the percentage of ethnic minorities in the total population increases to 10%. Unfortunately, there is no information available on double citizenships for Swedish citizens. If nationality is used as indicator, France has the largest percentage of ethnic minorities (12.2%). Comparing the relative percentages of ethnic minorities is difficult, however, even if the same definition is applied to all the countries. Applying the citizenship criterion yields different results depending on the laws for acquiring citizenship. Another promising option might be the country of birth criterion. Unfortunately, the required data are largely unavailable. Table 4.2 shows where the ethnic minorities in the six countries are from. Relative percentages of the various ethnic groups are presented, and the most important countries of origin are given in percentages of the total population.

Unlike the definitions of ethnic minorities, the definitions of response rates are similar in the six countries. All their national statistical institutes use the American Association for Public Opinion Research response definition 2, which includes partial interviews in the numerator and excludes noneligible sampled units from the denominator given equal inclusion probabilities.[2)]

4.4 Nonresponse among ethnic minorities and current strategies for collecting data among ethnic minorities

The variety in the surveys, the different definitions of ethnic minorities, and the limited data available make it difficult to compare nonresponse rates directly in a meta-analysis. Nonetheless, the data corroborate the anecdotal evidence that in all six countries, the response rates for ethnic minorities are mostly lower than for the native population. The countries vary, however, in the degree to which the ethnic minority nonresponse rates differ from the native ones, as is shown in Table 4.3 in greater detail.

[2)] In the appendix, different response definitions according to American Association for Public Opinion Research definitions are given in a more formal way.

Table 4.2
Ethnic minorities in six countries according to geographical background

		Population						
			Native population	Europe	Africa	Asia	Other	Total
		m	%					
Belgium			Belgian nationality					
2004		10,396	91.7	5.9	1.2	0.8	0.2	100.0
	Italy			1.8				
	Morocco				0.8			
	Turkey[c]					0.4		
France			French nationality					
2004		61,684	87.8	6.1	5.0	0.9	0.2	100.0
	Maghreb[b]				4.0			
Germany			German nationality					
2003		82,537	91.9	4.8	0.4	3.4	0.4	100.0
	Italy			0.7				
	Yugoslavia[a]			0.7				
	Turkey[c]					2.3		
The Netherlands			Allochtonous					
2004		16,258	81.0	5.9	3.1	6.5	3.6	100.0
	Germany			2.4				
	Morocco				1.9			
	Turkey[c]					2.2		
	Indonesia					2.5		
	Suriname & Antilles						2.8	
Sweden			Born in Sweden					
2003		8,975	88.0	7.0	0.7	3.3	1.0	100.0
	Scandinavia			0.3				
	Yugoslavia[a]			0.1				
United Kingdom			'White'	'Mixed'	'Black' (British)	'Asian' (British)		
2001		58,789	92.1	1.2	2.0	4.4	0.4	100.0
	India					1.8		
	Pakistan					1.3		

[a] Serbia and Montenegro
[b] Tunisia, Algeria, Morocco, Mauritania and Libya.
[c] Turkey is classified as an Asian country.

Table 4.3
Response rates among ethnic minorities based on the official definition of ethnic minorities in each country in the six studied countries

Response rates Belgium

Labor Force Survey	**Mandatory**		

Response rates France

Health 2003 Mode: Face-to-Face	**French**	**Non-French**	**Difference scores**
Response	67.0	68.0	1.0
Non-contact	9.3	12.0	2.7
Refusals	7.9	8.4	0.5
Inability	1.4	2.8	1.4
Other	14.3	8.8	–5.5
n	22,083	3,003	

Labor Force Survey	**not available**		

Response rates Germany

ALLBUS 1994 Mode: Face-to-Face	**Germans**	**Non-Germans**	**Difference scores**
Response	54.6	50.4	–4.2
Non-contact	2.7	5.0	2.3
Refusals	38.1	20.5	–17.6
Inability	3.7	22.9	19.2
Other	0.9	1.2	0.3
n	5,788	341	

ALLBUS 1996 Mode: Face-to-Face	**Germans**	**Non-Germans**	**Difference scores**
Response	53.9	60.4	6.5
Non-contact	4.2	5.8	1.6
Refusals	37.7	16.8	–21.3
Inability	3.7	15.8	12.1
Other	0.5	1.6	1.1
n	6,109	379	

ALLBUS 2000 Mode: Face-to-Face	**Germans**	**Non-Germans**	**Difference scores**
Response	47.6	48.1	0.5
Non-contact	5.8	10.5	4.7
Refusals	41.0	20.7	–20.3
Inability	1.9	15.7	13.8
Other	3.7	4.9	1.2
n	5,010	324	

Table 4.3 (cont'd)

ALLBUS 2002 Mode: Face-to-Face	**Germans**	**Non-Germans**	**Difference scores**
Response	48.0	38.2	–9.8
Non-contact	4.6	6.9	2.3
Refusals	31.8	21.3	–10.5
Inability	4.6	15.8	11.2
Other	11.0	17.7	6.7
n	5,052	361	

Labor Force Survey	**Mandatory**

Response rates the Netherlands

Survey on Living Conditions 1998 Mode: CAPI/CATI	**Native population**	**Ethnic minorities**	**Difference scores**
Response	63.2	51.6	–11.6
Non-contact	4.5	9.5	5.0
Refusals	23.3	20.5	–2.8
Inability	1.9	8.0	6.1
No opportunity to interview	5.8	7.4	1.6
n	32,467	5,850	

Survey on Living Conditions 2004 Mode : CAPI	**Native population**	**Ethnic minorities**	**Difference scores**
Response	65.6	56.8	–8.8
Non-contact	4.5	9.5	5.0
Refusals	23.2	18.4	–4.8
Inability	0.1	7.7	7.6
No opportunity to interview	6.6	7.6	1.0
n	25,735	4,857	

Labor Force Survey	**not available**

Response rates Sweden

Survey on Living Conditions 2001 Mode: CATI/Face-to-Face	**Native population**	**Ethnic minorities**	**Difference scores**
Response	79.3	66.8	–12.5
Non-contact	4.6	14.7	10.1
Refusals	14.6	15.2	0.6
Inability	1.3	3.1	–1.8
n	6,488	986	

Labor Force Survey 2003 Mode: 99.8% CATI	**Native population**	**Ethnic minorities**	**Difference scores**
Response	85.0	74.8	–10.2
Non-contact	7.4	17.3	9.9
Refusals	6.7	6.7	0.0
Inability	0.7	1.2	0.5
n	79,506	12,301	

Response rates United Kingdom

Labor Force Survey	**No classification table available**

Figures have been rounded off and may not add up to 100%.

Dissecting nonresponse into noncontact, refusals, and other outcomes shows that in all six countries, ethnic minorities have lower contact rates (defined here as the ratio between the contacted sampled units and all the eligible sampled units; see the appendix) and relatively more nonrespondents due to inability to participate in survey research than the native population. This might help explain why France and Sweden, countries with a high number of minimum contact attempts, have relatively high response rates among ethnic minorities. Unlike the lower contact rates among the ethnic minorities, almost all the ethnic minority cooperation rates are higher than the native ones. Cooperation is defined here as the ratio between completed interviews and completed interviews + partial interviews + refused eligible units (Groves 1989). However, participation in some of the surveys is not voluntary, so no statements can be made about the cooperation rate. Moreover, in a group in which the noncontact rate is higher, the cooperation rate may decline if more sampled units are contacted. Establishing contact gives the people in the sample a first chance to refuse a request to take part in the survey. In fact, the ethnic minority refusal rates are increasing. Nonresponse due to inability to provide the required information, the third main reason for nonresponse, is always higher among ethnic minorities. Difficulty with the survey language is the main reason why this form of nonresponse is higher among ethnic minorities. It is, however, important to note that the three reasons for nonresponse—noncontact, refusal, and inability—are not so clearly distinguished. Sampled units who are not contacted have no opportunity to refuse to take part in a survey (Hox and De Leeuw 1998). Furthermore, sampled units can use their problems with the survey language as a friendly way to refuse a request to take part in a survey. Members of ethnic minorities might also poorly communicate their reluctance in such a way that the interviewer codes it as a nonresponse due to language problems instead of as a refusal. Various ethnic minority response outcomes are presented in Table 4.3. Response outcomes are divided into response, noncontact, refusals, inability to provide the required information, and other reasons for nonresponse. As is clear from Table 4.3, detailed information on the various ethnic minority response rates is not always available. Except for some minor adjustments, no special strategies are currently in use at the NSIs for collecting data among ethnic minorities. In the United Kingdom, however, the ONS uses language cards that are distributed to interviewers for use at ethnic minority households if prospective respondents cannot speak English (M. McConaghy, personal communication). ONS may also have relatives act as translators or use paid translators. This method is described in greater detail in the Reducing Inability section. In Germany, the Federal Office of Statistics somewhat oversamples ethnic minorities (Gruber 1997).

4.5 *Reducing nonresponse among ethnic minorities*

Obtaining response is a process influenced by several factors. The first step is to locate the sampled unit. The sample frame might not provide the necessary information for locating sampled units. Second, contact has to be established. Once this is done, the sampled unit has to agree to take part in the survey and has to be able to provide the required information. The failure of any of these steps will lead to nonresponse. To reduce nonresponse, it is important to distinguish its potential causes. In this section, alternative causes of nonresponse and ways to reduce nonresponse outcomes are discussed, so that tailored strategies for collecting data among ethnic minorities can be developed. Recommendations are mainly based on the personal experiences and judgments of practitioners from the selected survey organizations. Experimental evidence of ethnic minority nonresponse reduction is rare. Some of the recommendations have already been tested in experimental settings. Others need to be tested in future experiments.

4.5.1 Reducing noneligible sampled units

Because of high ethnic minority mobility and complex household structures (Statistics Netherlands 2004), the ethnic minority sample frame is usually not as good as the native one. This can result in more noneligible sampled units among ethnic minorities. Of course, researchers and survey organizations are often unable to control the quality of the sample frame, and this is consequently mainly of theoretical interest. The fact remains, though, that more sampled units are needed to reach a certain desired sample size.

4.5.2 Reducing noncontacts

One efficient way to reduce the number of noncontacted sampled units is by increasing the number of contact attempts after earlier noncontact. In the Netherlands, the minimum number of contact attempts after earlier noncontact was increased from three to six in March 2004 (Snijkers and Kockelkoren 2004). This had a very positive effect on the contact and response rate, in particular among ethnic minorities. Response rates of first generation ethnic minority members or residents of the Netherlands born abroad with at least one parent born abroad (Reep 2003) increased in the Dutch Survey on Living Conditions from 47.1% to 53%[3] since this rise in the number of contact attempts. For the native population, this increase was only 1.5% from 64.4% to 65.9%. The increase was most striking in the response rates among first-generation non-Western foreigners, which rose from 43.0% to 51.5%. The general

[3] A comparison is made between the response rates of the first 2 months of 2004 and in the period from March to December 2004.

native and ethnic minority response outcomes in this survey before and after this fieldwork procedure adjustment are presented in Table 4.4.

Table 4.4
Response outcomes before fieldwork adjustment (three contact attempts, January and February 2004) and after fieldwork adjustment (six contact attempts, March-December 2004) in the Dutch Survey on Living Conditions.

SLC 2004	Native population Jan.-Feb	Native population Mar.-Dec.
Response	64.2	65.9
Non-contact	5.4	4.3
Refusals	23.6	23.1
Inability	0.1	0.1
No opportunity to interview	6.5	6.6
n	4,553	21,182
SLC 2004	**Ethnic minorities Jan.-Feb**	**Ethnic minorities Mar.-Dec.**
Response	50.8	58.1
Non-contact	13.0	8.8
Refusals	20.4	18.0
Inability	7.2	7.8
No opportunity to interview	8.6	7.4
n	852	4,005

A similar result was noted in Germany, where the number of minimum contact attempts in the ALLBUS 2002 was ten instead of four. As a result, the contact rate among sampled units with the German nationality increased from 94.2% in 2000 to 95.4% in 2002. For non-Germans, this increase was from 89.5% in 2000 to 93.1% in 2002. Because many features of the design changed simultaneously in the two surveys, these results should be interpreted with care. Countries with a high minimum number of contacts, such as Sweden, where Statistics Sweden uses twelve minimum contact attempts in the Labor Force Survey, also have relatively high ethnic minority contact and response rates. The ethnic minority response rate in Sweden was 74.8% in the Labor Force Survey 2003 and 66.8% in the Survey on Living Conditions 2001. Another way to reduce the number of ethnic minority noncontacted sampled units is by extending the data collection period in hours and days. Ethnic minorities relatively often do shift work (Seifert 1992) and are often in their country of origin for lengthy periods of time although officially registered in the country of the survey organization (Blohm and Diehl 2001). A longer fieldwork period increases the probability of finding a sampled unit at home. The mode of data collection can also have a different and greater impact on the contact rate among ethnic minorities than among the native population. Telephone coverage among ethnic minorities is usually lower; therefore, a computer-assisted personal interview contact mode is preferred to a computer-assisted telephone interview (CATI) mode. CATI should be at least held in a mixed-mode design.

4.5.3 *Reducing refusals*

One common way to increase survey participation is by offering monetary incentives to sampled units. In the literature, references are often made to experiments demonstrating the positive effect of incentives on response rates (see, e.g., Groves and Couper 1998; Singer 2002). With regard to ethnic minority groups, however, the evidence is mixed. To persuade respondents in the 2004 Statistics Netherlands experiment "A Tailored Approach Strategy for Young Moroccans and Turks for the Dutch Family and Fertility Survey," a gift voucher of €10.00 was promised in an advanced letter and by the interviewers. However, the incentive did not produce any major effect on the response rates (Van den Brakel, Vis-Visschers, and Schmeets 2006). Unlike earlier ALLBUS surveys, the ALLBUS 2002 survey included a €10.00 coin as incentive. The German cooperation rate (see the appendix) increased from 53.7% in 2000 to 60.1% in 2002, but the cooperation rate among non-Germans decreased in the same period from 70.0% to 64.2%.[4)] These figures should be interpreted with care. The incentives were not given in an experimental setting, and the ALLBUS 2002 was conducted by a different survey agency than the ALLBUS 2000. So the changes in nonresponse may also be due to differences between the survey agencies. This might, nonetheless, suggest that ethnic minorities are more indifferent to opportunity costs and social exchange hypotheses. Anti Athiainen (personal communication, 2004) from Statistics Sweden noted that incentives for ethnic minorities should be used with caution, especially lottery tickets, since Muslims are forbidden by their religion to gamble or bet. Refusing to take part in a survey can also have to do with not being familiar with the survey organization. This problem can be tackled by announcing the upcoming survey in the popular ethnic minority media (A. Athiainen, personal communication, 2004). This can inform people about the survey organization and the upcoming survey and decrease anxiety about providing personal information (Dumas and Théroux 2004). Publicly announcing the upcoming survey also increases the perceived legitimacy of the survey (see Cialdini 1993; Groves and Couper 1998).

4.5.4 *Reducing inability*

This category of nonresponse is always higher among ethnic minorities, mainly because of difficulties with the survey language. There are three main ways to reduce the number of sampled units who cannot participate in a survey due to language problems: (1) the survey organization can use questionnaires in other

4) In 2002, all the interviews conducted in the past 4 weeks were rejected due to doubts about whether all the rules had been followed (Blohm, Harkness, Klein, and Scholz 2003). This resulted in far more nonanalyze interviews than in Allgemeine Bevölkerungsumfrage der Sozialwissenschaften (German General Social Survey) 2000. If the nonanalyzed interviews were interpreted as refusals, there would be an increase in the cooperation rate from 51.5% to 52.8% among the German sample units and a reduction from 65.3% to 49.5% among the non-German sampled units.

languages, (2) respondents can be interviewed by interviewers with the same ethnic background, and (3) nonresponse due to language problems can be reduced by allowing relatives to act as translators. None of the NSIs in the study have questionnaires translated into ethnic minority languages. Statistics Sweden has had experience using multilingual surveys. There used to be survey translations in twelve languages. However, this was reported to be not worth the effort (O. Wessberg, personal communication, 2004). Now only an English translation is available. ONS United Kingdom is obliged by law to have questionnaires in Welsh as well as English. Translating questionnaires can be a very costly operation. Ethnic minorities in the six countries are extremely heterogeneous, so except perhaps in Germany, with a large Turkish minority, translations into more languages are needed to cover the sampled units who cannot participate because of language problems. Even if a whole ethnic minority is from one country, there can be various regional languages. Moreover, cultural differences and meanings can complicate translations (Schoua-Glusberg and Miller 2004). Nevertheless, Blohm and Diehl (2001) noted that, at least in Germany, using bilingual questionnaires can reduce nonresponse due to language problems. The same problems apply to interviewers with the same ethnic background as the respondents. Moreover, unforeseeable problems can arise. Sensitive questions posed by interviewers with the same ethnic background can result in more traditional answers (Dotinga et al. 2005). Sometimes the interviewer and the sampled person come from different political resistance or guerrilla groups in the home country. Or there can be a fear of refugee spying by the translator on behalf of the home country (A. Ahtiainen, personal communication, 2004). In practice, it is not usually easy to recruit qualified interviewers with the same ethnic background as potentially difficult respondents (Blohm and Diehl 2001). Nevertheless, interviewers with the same ethnic background can play a useful role in the tracing and contacting before the actual interview (A. Ahtiainen, personal communication, 2004). But at least for general surveys with the target population consisting of all the residents of a country, using interviewers with the same ethnic background as the respondents would probably not be cost efficient. Another way to reduce nonresponse due to language problems is to have a relative older than the age of fifteen translate, perhaps a nonresident. This method is sometimes used by ONS United Kingdom to avoid nonresponse, provided the quality of the study does not suffer as a result. Posing questions about sensitive topics with a young relative as a translator could produce socially desirable answers. Another problem with relatives acting as translators is the introduction of the measurement error involved in simultaneous translations.

4.6 Recommendations and discussion

Collecting data from ethnic minorities is not easy. Getting a sufficient response is particularly difficult among ethnic minorities. Nonresponse can have serious

consequences for researchers since the survey estimates may be biased. Overall response rates have declined virtually all across the globe in recent years (De Heer 1999). In the Netherlands as well as elsewhere, this trend of decreasing response rates is even more troublesome among ethnic minorities. Weighting techniques can partially make up for nonresponse bias, but correcting for selectivity cannot be unlimited. This is especially the case if specific societal groups have very low response rates. Then the assumption that the responding segment is more or less representative of the nonresponding segment is questionable. The only good solution to missing data is thus not to have any (Allison 2001). Of course, this is an impossible scenario, but it is still essential to put great effort into designing and executing research projects for minimizing missing data. To do so, we have studied strategies for reducing nonresponse internationally. Most of the recommendations in this article are best practices, as experiments in this field are rare. In the perspective of a growing European Union, more experiments and better contact between the NSIs are called for. Based on the response rates and experiences of some other European statistical offices, we have considered various ethnic minority nonresponse reduction strategies. Comparing the ethnic minorities in different countries is quite a daunting task in itself since each country has its own history with its own ethnic minorities, and matters are complicated even more by the different definitions of ethnic minorities. Ethnic minorities are not the same thing in every country. Different definitions of ethnic minorities are used by different governments. A resident of Germany with German nationality but born in Turkey is classified in Germany as a German, but a Dutch resident born in the Netherlands whose Dutch father was born in Belgium is classified as an immigrant in the Netherlands. These different definitions greatly affect the size and structure of ethnic minorities. A recently published internal document of the European DG Employment and Social Affairs gives some clear examples of the effect of using different definitions of ethnic minorities on some socio-economic statistics (European Commission, Employment and Social Affairs DG 2004). Low ethnic minority response rates can bias survey estimates. Depending on the definition, the ethnic minority percentage of the total population is already about 10% and increasing in all six countries. Other countries (e.g., the United States) now already have a much higher percentage of ethnic minorities in their population. Better ethnic minority response rates are also needed to give good estimates of subpopulations. As Couper and De Leeuw (2003) noted, differences in response rates may threaten the validity of comparative studies. Differences between samples may not reflect differences between populations but might result from response and definition differences. Detailed information on ethnic minority response outcomes is not always available, as Table 4.3 shows. More information on specific fieldwork procedures (e.g., time of contact attempts) is needed to more precisely evaluate response processes among various societal subgroups. This information is often not centrally available or not accessible to outsiders. Increasing ethnic minority response rates should involve tailoring the survey design in such a way that the response probability is maximized

under time and budget constraints (Snijkers 2003). Of course, this also holds true of strategies for encouraging ethnic minorities to take part in survey research. However, this study show that, except for some small modifications, none of the six countries had special strategies for collecting data among ethnic minorities. Dissecting the nonresponse phenomenon into contact, refusals, and other causes means considering alternative reasons for each outcome (Groves and Couper 1998). Ethnic minorities seem to have lower contact rates and higher nonresponse rates than the native population due to language problems. The low contact rates are closely connected to socio-demographic and socio-economic characteristics of ethnic minorities. For example, ethnic minorities are more likely to be urban residents, who are known to be difficult to contact (Groves and Couper 1998; Feskens, Hox, Lensvelt-Mulders, and Schmeets 2004). This suggests that researchers should concentrate on enhancing ethnic minority contact rates to enhance ethnic minority response rates. An efficient way to increase the ethnic minority contact rate is by increasing the minimal number of contact attempts after earlier noncontact and using a longer data collection period. In particular, raising the minimal number of contact attempts can have a positive effect on the response rates and can be tailored for use with ethnic minorities. Increasing the number of contact attempts in Germany and the Netherlands resulted in a relatively higher rise in contact and response rates among ethnic minorities than among the native population. This might also explain why the largest difference between ethnic minority and native response rates is observed in the Netherlands. Until March 2004, the Netherlands and Belgium were the only ones of the six countries with a low number of contact attempts (De Heer 1999). Unlike Belgium, contact rates cannot be increased in the Netherlands by substitution. Reducing the number of noncontacted sampled units can also have a positive effect on bias reduction, as Lynn, Clarke, Martin and Sturgis (2002) noted, "It is the difficult to contact who are most different from the easy to get." However, establishing contact with more ethnic minority sampled units by increasing the number of contact attempts can also lead to an increase in the measurement error by bringing in more respondents with language difficulties. This could be a possible hypothesis for a later experimental test. Nonresponse rates due to refusals are not as high as one might initially think. In fact, they are usually lower among ethnic minorities than among the native population. This form of nonresponse is nonetheless also growing among ethnic minorities. Incentives are usually used to increase the cooperation rate among sampled units. The results of an experiment in the Netherlands and a fieldwork adjustment in Germany show, however, that this had only a very limited effect if any among ethnic minorities. Future research could focus on why incentives seem to be, at least in European countries, less effective in raising the cooperation and response rates among ethnic minorities. Ethnic minority unfamiliarity with the survey or survey organization can be overcome by announcing the upcoming survey in a more tailored way. Nonresponse due to inability to provide the required information is higher among ethnic minorities, specifically the older members, mainly due to language problems.

Depending on the sensitivity of the survey topic, language problems can be overcome by having younger relatives act as translators. Translating questionnaires or using interviewers with the same ethnic background are probably not cost effective. Obtaining responses from ethnic minorities is not as easy as from native populations, but it is not impossible. Response rates and experiments show that special attention should be devoted to lower contact rates among ethnic minorities. Countries with high numbers of contact attempts already have relatively high response rates among ethnic minorities, and recent developments at Statistics Netherlands also reveal striking increases of contact and response rates among ethnic minorities.

5 *Impact of Prepaid Incentives in Face-to-Face Surveys: A Large-Scale Experiment with Postage Stamps*

In this chapter, the authors discuss the results of an experiment with various prepaid incentives in the Dutch Labor Force Survey among 13,000 households. The response rate increases from 65 to 73 percent when booklets of postage stamps valued four euro are included in the advance letter in a face-to-face survey. This is caused by a lower refusal rate, which decreases from 23 to 16 percent. For almost all examined characteristics such as income, gender, age, and household composition, the increase in response rates did not differ substantially. The results show a decrease of the response selectivity in the degree of urbanization, as response rates improved in particular in the three largest cities – Amsterdam, Rotterdam and The Hague, areas with high nonresponse rates. On the other hand, an increase of selectivity was found in ethnicity, because incentives did not improve response rates for ethnic minorities with a non-Western background. As a result there is no overall improvement in the selectivity of the response. Furthermore, incentives did not affect the key indicators of the survey – the labor force and unemployment rates.[1)]

[1)] This chapter has been accepted for publication as Wetzels, W., Schmeets, J.J.G., van den Brakel, J., and Feskens, R.C.W. (in press). Impact of Prepaid Incentives in Face-to-Face Surveys: A Large-Scale Experiment with Postage Stamps. *International Journal of Public Opinion Research.*

5.1 Introduction

Nonresponse is a general problem in survey research. Although response rates and the nonresponse trends differ between countries, the general trend is that response rates have been declining over the years (De Leeuw and De Heer 2002). A high nonresponse rate is a problem as it reduces the number of respondents and consequently the precision of estimates. In addition, nonresponse can be selective. This occurs when nonrespondents differ systematically from respondents as to the survey objectives. As a result, the survey estimates of the key indicators may be biased. Nonresponse is not only a statistical problem, it is also a financial problem as declining response rates cause inclining survey costs.

In the nineties, a response rate of 55 percent in face-to-face surveys conducted by Statistics Netherlands was not unusual (De Heer 1999). Due to a substantial reorganization of the fieldwork department and the raise of the minimum number of contact efforts from three to six, the response in Statistics Netherlands' face-to-face surveys gradually increased to 65 percent in 2005 (De Bie and Luiten 2005). In order to realize a further increase in the response rate, we decided to focus on the refusals. As Statistics Netherlands does not use refusal conversion policy – which is a common practice in many other fieldwork organizations, including National Statistical Institutes – we had to consider other measures. One option was the use of incentives. Prepaid cash incentives are known to be most effective, but sending cash by mail is not allowed in the Netherlands. Therefore, postage stamps are chosen as a prepaid incentive, which are expected to be an opportunity to bridge the gap between cash and material incentives. In this chapter, the effect of stamps as a prepaid incentive in face-to-face surveys on response is tested in a large-scale field experiment embedded in the Dutch Labor Force Survey.

5.2 Past research on the use of incentives

The literature shows that incentives can be effective to increase the response rate of surveys. Simmons and Wilmot conclude (2004): 'The general finding from the literature is that the use of incentives, however small in monetary terms, is effective in increasing response rates in postal, telephone and face-to-face surveys. This seems to be the case for all types of surveys, not just those where there is a high burden for the respondent, and it appears to be true for panel survey.' In an extensive meta-study Church (1993) shows the significant and positive effect of prepaid incentives on response rates in mail surveys. Church also found that monetary incentives offered with the initial mailing yield higher response rates than material incentives. In the meta-analysis, a strong correlation is also found between the cash value and the effect on the response rate: response rates increase with increasing amounts of money. For mailings in which the incentive was made contingent to returned responses no meaningful increase in response was found.

James and Bolstein (1992) examined the effect of monetary incentives in a mail survey. The results indicate that an incentive of $1 significantly increased the response rate, regardless of the number of mailings. The response rate increased also as the incentive was raised from $1 to $5 and from $5 to $20. A promise of $50 however did not result in a higher response rate over a no-incentive control group. In an earlier article (1990) James and Bolstein have demonstrated that even very small amounts of money like $0.25 and $0.50 do improve the response rate in mail surveys. Dillman (2000) demonstrates that an incentive of $1 yields a substantially increased response rate, while an additional increase diminishes as higher amounts ($2, $3, etc.) are used.

In a meta-analysis with interviewer-mediated surveys, Singer et al. (1999) come to comparable conclusions as Church (1993). Paying an incentive is effective in increasing response rates in telephone and face-to-face surveys. In the five studies in the analysis in which a comparison between prepaid and promised incentives could be made, prepayment yielded significant higher response rates than promised incentives. Another important conclusion is that also in interviewer-mediated surveys gifts are also less effective than cash in increasing the response rate, even controlling for the value of the incentive. Singer, Van Hoewyk and Maher (2000) found that a 5-dollar bill enclosed in an advance letter in a telephone survey had a significant and large effect on the response and cooperation rates.

Groves and Couper (1998, p. 281) reach the same conclusion: 'The literature shows that incentives appear to increase overall response rates. In both modes of data collection (telephone and personal visit), prepaid incentives lead to increased response rates over no incentives and promised incentives.' For a face-to-face survey with the use of diaries, Nicolas and Stratford (2005) show a rise in the response rate from 51 to 58 percent by using an incentive of 5 pounds, and to 61 percent for a 10 pound incentive. Other studies also demonstrate the positive effect of incentives on response rates (see e.g. Singer 2002; Teisl, Roe, and Vayda 2006).

The effect of prepaid incentives on the response is often explained with the social exchange theory. The potential respondents receive an incentive from the research institution without having to give something back. An incentive is seen as an act of kindness and a token of trust, and this kindness evokes the norm of reciprocity. In accepting the gift, the potential respondent feels obliged by social norms to respond in kind.

Given the literature, a logical choice for face-to-face surveys would be the use of a relatively small cash prepaid incentives. However, sending cash by mail is not allowed in the Netherlands. Therefore, we had to use a material incentive and made the choice to use postage stamps. We think incentives like postage stamps provide an opportunity to bridge the gap between cash and material incentives. Postage stamps can be regarded as close to cash, the value is exactly known and almost everyone uses them more or less frequently.

5.3 *Experimental design*

To obtain insight in the effect of prepaid incentives on response rates, their selectivity and costs, we decided to conduct a large-scale experiment embedded in the Labor Force Survey (LFS). The LFS is designed as a rotating panel with five waves. Each month a sample of about 7,500 addresses are drawn by means of stratified two-stage sampling. All households on selected addresses, up to a maximum of three households per address, and all household members aged 14 years or over are included in the survey. In the first wave data are collected by computer assisted personal interviewing (CAPI). The households are re-interviewed four times at quarterly intervals, with data being collected with computer-assisted telephone interviewing.
The monthly samples of the first wave of November and December 2005 were randomized over four treatment groups with no stamps, 5 stamps, 10 stamps and 20 stamps. In total, 6,960 addresses were approached in November and 6,029 in December. The addresses were allocated over the four treatment groups as:

1. No incentive, 6,195 addresses;
2. Incentive valued 2 euro (5 stamps), 3,146 addresses;
3. Incentive valued 4 euro (10 stamps), 3,148 addresses;
4. Incentive valued 8 euro (20 stamps), 500 addresses.

The purpose of this experiment is to investigate the following research questions:

1) What is the effect on the response rates of 2-, 4-, and 8-euro incentives?
2) Do incentives improve the selectivity, i.e. decrease the variation in response rates between different subpopulations?
3) Do incentives have any impact on the core indicators in the survey?

Besides the gain in overall response, for the 5- and 10-stamps versions we were interested in changes in selectivity. A reduction of selectivity would indicate improved quality of the data. This would be of value as the variation of the weights decreases and probably the bias of the estimates of the core indicators would be reduced.
Cost effectiveness is another important factor. Introduction of incentives in face-to-face surveys has to be done preferably without or with very limited extra costs. Berlin et al. (1992) found that offering an incentive can reduce the total survey costs. Simmons and Wilmot (2004) argued that prepaid incentives reduced interviewer effort and consequently at least partially paid for themselves. Whether incentives are cost effective depends on the obtained response increase, the value of the incentive and – in case of face-to-face surveys – the average costs of visiting an address. In the case of a fixed number of responses a 2 euro incentive (5 stamps of 39 eurocent each) would be cost neutral if the response rate increased by 4 percent; whereas for an introduction of a 4 euro incentive (10 stamps) the response

rate would have to go up by 6 percent. In addition to the 2 and 4 euro test groups, a smaller group with 8 euro (20 stamps) was created to determine the additional overall response effect of a larger incentive.

5.4 Result

5.4.1 Results: response rates

A response account of the fieldwork under the four treatment groups of the experiment is given in Table 5.1. The response figures are based on households living alone at an address. Addresses with multiple households are excluded from this analysis because in such cases it is not known which household received the incentive.[2)] The response increases substantially if stamps are included in the advance letter. Compared to the group without incentives, the gain in response rate equals 5.3, 7.8 and 7.4 percent for the group with 5, 10 and 20 stamps respectively (according to AAPOR definition No. 2; AAPOR 2006). Note that this is caused by lower refusal percentages. The percentage refusals dropped with 5.1, 7.1 and 10.1 percent respectively.

A logistic regression analysis reveals that incentives contribute significantly to the response rate. This is true for 5, 10 and 20 stamps ($p < .01$). Also the difference in response rates between the 5-stamps and 10-stamps group is significant ($p = .04$). However, due to the small number of households (434), for 20 stamps we can not prove a surplus value in comparison to 5 stamps ($p = .36$) and 10 stamps ($p = .88$).

Table 5.1
Response and nonresponse in households by different incentives

	No stamps		5 stamps		10 stamps		20 stamps	
	Count	*%*	*Count*	*%*	*Count*	*%*	*Count*	*%*
Refusal	1,211	23.3	476	18.1	432	16.1	57	13.1
No opportunity	240	4.6	125	4.8	123	4.6	26	6.0
Language problem	78	1.5	37	1.4	51	1.9	5	1.2
No contact	262	5.0	127	4.8	106	4.0	29	6.7
Response	3,419	65.6	1,865	70.9	1,964	73.4	317	73.0
Total	5,210	100.0	2,630	100.0	2,676	100.0	434	100.0

So far, we discussed the impact of incentives on response rates in the first wave, where data are collected with face-to-face interviews. However, the LFS is designed

[2)] If we include such cases in the analyses, the figures are very similar. The response rates are slightly lower, except for 20 stamps. No incentive: 64.9%, 5 stamps: 70.1%, 10 stamps: 72.8%, 20 stamps: 73.2%. The increase for 20 stamps is probably a consequence of the small number of cases. (Wetzels and Schmeets 2006, p. 7).

as a rotating panel with five waves. This raises the question whether an unconditional small incentive also affects participation in later waves. There are no differences observed in panel attrition between the groups with and without incentives. This indicates that the positive effect of incentives on response rates remains stable in a multi-wave panel design.

5.4.2 Results: selectivity in response

The second research question deals with the selectivity in the response, i.e. will incentives decrease or increase the variation in response rates between different subpopulations?

The Labor Force Survey is supplemented by administrative data from the *Social Statistical Database.* In this database several registers are linked to each other as well as to data from sample surveys. The core of the database is the Population Register. Linking the administrative and survey records to the Population Register, which is also the sample frame of the LFS, makes socio-demographic and socio-economic information available on the respondents and nonrespondents of the LFS at the individual and household level. On the level of individual sampled units, crucial information as age, gender and ethnicity is available. On the household level, we could use information on the household size and household income. We used the standardized household income which is the disposable income corrected for differences in household size and composition. The standardized income is a measure for the prosperity of households.

Response behavior on the personal level was analyzed first since the main parameters of the LFS are defined on the personal level. On the personal level the response rate is slightly higher than on level of households (68.1, 73.7, 75.1 and 76.2 percent for the group with no incentive, 5, 10 and 20 stamps respectively, compare with Table 5.1). This is caused by a higher response rate among multiple-person households.

The response distributions of the explanatory variables for the four treatments show for most variables only minor effects of the incentives on the selectivity. There are, however, effects on the response distributions over regions and country of origin. The influence on response distributions is particularly strong for the three largest cities – Amsterdam, Rotterdam and The Hague – with the highest nonresponse rates. A 10-stamp incentive results in an increase from 53 to 65 percent, whereas outside of those three cities a moderate increase, from 70 to 76 percent, was obtained.

On the other hand, the response composition is probably getting worse for the respondent's country of origin. Ethnic minorities constitute about 20 percent of the Dutch population (http://statline.cbs.nl). The ethnic minority or immigrant population is defined in the Netherlands as 'everyone residing in the Netherlands with one or both parents who were born abroad.' A further distinction is usually drawn between people with one or both parents born in Europe, North America, Australia, Japan or Indonesia and people from non-Western countries (mainly

Turks, Moroccans, Surinamese and Antilleans).[3] The two groups are of approximately the same size. The lower response rate among ethnic minorities is a problem that is not restricted to Statistics Netherlands. Ethnic minorities have lower response rates in almost all the Western countries (Feskens et al. 2006). We find that incentives do not have an impact on the response rate of the non-Western ethnic minorities. As a result, incentives increase the response gap between non-Western foreigners and the native Dutch population. This is also true for the gap between the non-Western and Western foreigners.

The effect of incentives on response behavior is also tested with a logistic regression analysis. Interaction effects between the incentives and the other social-demographic variables on response rates indicate an increase or decrease of selectivity. Therefore response behavior (1 = response; 0 = other response categories) is described in a logistic regression model using the following explanatory variables:

- Treatment (no incentive, 5 stamps and 10 stamps)
- Age (15–34 / 35–44 / 45–54 / 55–64 / 65 and older);
- Gender;
- Household size (1 / 2 / 3 / 4 / 5 or more members);
- Household income (standardized);
- Region (three largest cities / other areas);
- Country of origin (native Dutch population /Western foreigners /non-Western foreigners).

As the 10-stamp incentive was the most promising one in terms of an increased response, in particular for the three largest cities, we started with a model in which the 10-stamp incentive was compared with the no-incentive group. Apart from main effects, we included second order interactions with incentive in the regression model. Two significant interactions are found (a) incentive * region and (b) incentive * country of origin and are therefore included in the model (see Table 5.2). Table 5.2 shows that there is a positive effect from the incentive on response, which indicates that a 10-stamp incentive increases the response. In addition, we find higher response rates of persons belonging to households consisting of more members. A higher income also results in a higher response rate. Furthermore, the 45–54 age group responds less often than the 15–34 age group. The difference between response rates of non-Western foreigners and both Western foreigners and native Dutch population increases due to the incentive. The discrepancy in response between the three largest cities and the other areas is getting smaller. In other words, if we focus on the region, there is less selectivity if we would use a 10-stamp incentive. But, looking at the country of origin indicates an increase in selectivity.

For the 5 stamp incentive it follows that the interaction between region and incentive is no longer significant, whereas the interaction with country of origin remains. Therefore we conclude that a 5-stamp incentive does not improve the

[3] For reasons of simplicity we use 'Western foreigners' and 'non-Western foreigners' in this article.

selectivity of region. For the 20-stamp incentive we find that all interaction effects disappear, very likely due to the small number of cases (results are not presented here).

Table 5.2
Logistic regression: incentive treatment (5 versus 0 Stamps and 10 versus 0) and socio-demographic characteristics on response (no/yes)

	5 versus 0 stamps					10 versus 0 stamps				
	B	*S.E.*	*Wald*	*df*	*Sig.*	*B*	*S.E.*	*Wald*	*df*	*Sig.*
Treatment	0.433	0.157	8.0	1	0.005	0.775	0.158	23.9	1	0.000
Household size			93.6	4	0.000			61.1	4	0.000
1 person (ref. cat.)										
2 persons	0.385	0.066	33.6	1	0.000	0.327	0.067	23.9	1	0.000
3 persons	0.513	0.080	41.5	1	0.000	0.397	0.079	25.0	1	0.000
4 persons	0.596	0.080	55.0	1	0.000	0.522	0.081	41.8	1	0.000
5 or more persons	0.827	0.106	61.0	1	0.000	0.618	0.104	35.3	1	0.000
Income household	0.127	0.027	21.8	1	0.000	0.121	0.027	20.2	1	0.000
Age			16.5	4	0.002			9.8	4	0.045
15–34 years (ref. cat.)										
35–44 years	–0.135	0.078	3.0	1	0.084	–0.101	0.079	1.7	1	0.197
45–54 years	–0.187	0.079	5.6	1	0.018	–0.181	0.079	5.3	1	0.021
55–64 years	0.097	0.080	1.5	1	0.228	0.039	0.081	0.2	1	0.628
65 years and older	0.037	0.095	0.1	1	0.700	–0.076	0.096	0.6	1	0.426
Country of origin			7.2	2	0.028			7.3	2	0.027
Native Dutch population (ref. cat.)										
Western foreigners	–0.233	0.103	5.1	1	0.023	–0.235	0.102	5.3	1	0.021
Non-Western foreigners	–0.187	0.113	2.7	1	0.980	–0.184	0.113	2.7	1	0.102
Region	0.641	0.094	46.3	1	0.000	0.661	0.094	49.4	1	0.000
Region by Treatment	–0.164	0.162	1.0	1	0.311	–0.364	0.162	5.0	1	0.025
Country of origin * Treatment			9.0	2	0.011			11.0	2	0.004
Western foreigners by Treatment	0.117	0.188	0.4	1	0.532	–0.084	0.178	0.2	1	0.635
Non-Western foreigners by Treatment	–0.541	0.190	8.1	1	0.004	–0.621	0.187	11.0	1	0.001
Constant	–0.214	0.108	3.9	1	0.048	–0.148	0.108	1.9	1	0.171

5.4.3 Results: response bias

Two key variables of the LFS are the unemployment rate and the labor force rate. A cross-table shows that there is no correlation between the treatment (0, 5, 10 and 20 stamps) and the unemployment rate ($p = .71$). The unemployment is 5.6 percent for the group without an incentive and 5.4 percent for 5 stamps and 10 stamps (see Table 5.3). For the 20 stamps group we found a higher unemployment rate of 6.7 percent. However, this estimate is based on a very small sample (29 cases out of 432).

The labor force rate varies between 67.7 percent (no incentives) and 69.4 percent (10 stamps). There is obviously no correlation between the labor force rate and the treatment ($p = .29$ based on a Chi-squared test for independency). In addition, the effect of the incentive treatments on the unemployment rate and the labor force rate is tested in a logistic regression model with age, gender, household composition,

standardized income, region and country of origin as covariates. No effects were found of the incentive treatment on the unemployment rate, neither main effects nor interaction effects. The same holds for the labor force.

Table 5.3
Unemployment rate and labor force by incentive treatments

	No stamps		5 stamps		10 stamps		20 stamps	
	Count	*%*	*Count*	*%*	*Count*	*%*	*Count*	*%*
Unemployment	257	5.6	135	5.4	145	5.4	29	6.7
Labor Force	4,598	67.7	2,518	68.7	2,682	69.4	432	68.7

5.5 *Conclusions and discussion*

The experiment clearly demonstrates that including postal stamps with the advance letter results in a substantial increase of response. Compared with the response rate of 65 percent in the control group, a 5-stamp incentive (2 euro) increased the response by 5 percent, a 10-stamp incentive (4 euro) resulted in an 8 percent higher response rate and a 20-stamp incentive (8 euro) in a 7 percent response raise. As expected this was caused by a decrease of the refusals. The refusal rate of 23 percent in the control group declined by 5, 7 and 10 percent, respectively. These figures show that, in a face-to-face survey, postal stamps included with the advance letter persuade potential refusers to participate and that there is a strong correlation between the value of the incentive and the decrease of the refusals. This result supports our assumption that stamps are perceived similar to money. In situations where sending cash by mail is not allowed or giving cash is seen as not appropriate, stamps can be a good alternative to cash incentives. It appears that stamps provide an opportunity to bridge the gap between cash and material incentives.

On the other hand, incentives do not seem to improve the selectivity. Although the sample size is quite large, there is hardly any impact on the response distributions over the different subpopulations. This implies that the additional group of respondents participating in the survey does not differ in their socio-demographical composition compared to the group of respondents that is reached without offering an incentive. There are two exceptions. First, the response rate of non-Western foreigners is not increased with a 5 or 10 stamp incentive. Second, the differences in response rates between regions decreases, since the response rate in the three largest cities of the country increases stronger than in other areas when a 10-stamp incentive is given.

Finally, the incentives did hardly change the key characteristics in the LFS. If small unconditional incentives, in our case stamps, do not have an impact on the key characteristics, why should Statistics Netherlands change its policy into submitting

incentives? We think this study is not sufficient to provide a clear answer. It is our belief that more research is necessary to determine the effects of incentives on selectivity as well as on response bias. We nevertheless think that the increases in response rates are substantial and encouraging. This is very promising, all the more, as our experiment shows, the use of small unconditional incentives can be done cost-neutral or even may reduce the total survey costs.

6 *Incentives and Ethnic Minorities: Results of a Controlled Randomized Experiment in the Netherlands*

In this chapter we examine the effect of prepaid incentives on ethnic minority cooperation rates in the Netherlands. We find that the incentives do have a substantial positive effect on the cooperation rates of native Dutch sampled units and Western foreigners. This effect is only modest among non-Western foreigners. We also match ethnic minorities with native Dutch sampled units using propensity score matching to compare the effect of incentives on the cooperation rates of ethnic minorities and comparable native Dutch sampled units. We find that the increase in cooperation rates is larger on the part of the native Dutch than ethnic minorities.[1)]

[1)] This chapter has been accepted for publication as Feskens, R.C.W., Hox, J.J., Schmeets, J.J.G., and Wetzels, W. (in press). Incentives and Ethnic Minorities: Results of a Controlled Randomized Experiment in the Netherlands. *Survey Research Methods.*

6.1 *Introduction*

Nonresponse rates in survey research threaten the validity of survey research and have increased in recent years in almost all Western countries (De Heer and De Leeuw 2002). Biased estimates are more likely to occur if specific groups exhibit below-average response rates. This makes it more likely that nonrespondents differ systematically from respondents with respect to the survey objectives. Ethnic minorities are one group internationally known for below-average response rates (Eisner and Ribeaud 2007; Feskens et al. 2006). Almost one in five residents of the Netherlands are members of ethnic minorities (http://statline.cbs.nl). The ethnic minority or immigrant population is defined in the Netherlands as 'everyone residing in the Netherlands with either one or two parents born abroad' (Reep 2003). A further distinction is usually drawn between people with either one or two parents born in Europe, North America, Australia, Japan or Indonesia and people with either one or two parents born in non-Western countries (mainly Turkey, Morocco, Surinam and the Netherlands Antilles).[2)] The Western and non-Western groups are of approximately the same size. Ethnic minorities also exhibit lower response rates than the native Dutch (Schmeets 2005b; Feskens et al. 2006). Consequently, the response is selective and survey estimates may be biased. One way to reduce nonresponse rates and more specifically refusal rates in survey research is to use incentives. However, the effect of incentives on ethnic minority response rates is still unclear.

In the literature, references are often made to experiments demonstrating the positive effect of incentives on response rates (see e.g. Berk, Mathiowetz, Ward, and White 1987; Dodd 1998; Groves and Couper 1998; Singer 2002; Simmons and Wilmot 2004; Berger 2006; Teisl et al. 2006). Several studies point out that in particular, prepaid incentives have a significant and positive effect on response rates, whereas the effect of promised incentives is less clear or even non-existent (Church 1993). As regards ethnic minorities, however, there is no such clear evidence. There is some evidence that incentives, particularly monetary ones, can be especially effective in increasing ethnic minority response rates in survey research. Mack, Huggins, Keathley and Sundukchi (1998) note that offering a $20.00 incentive in the first wave of a SIPP panel is much more effective in increasing the response rates of African-American and poor households than of other households (see also Singer 2002). Beebe, Davern, McAlpine, Call, and Todd (2005) also find a positive effect of incentives among most ethnic groups in a survey of Medicaid enrolees in the United States. They find higher response rates among all ethnic groups to a $2.00 incentive, although the difference on the part of Latino enrolees is not significant. Studies of the effectiveness of prepaid incentives on the ethnic minority response rates in Europe are however limited. Some studies note that the

[2)] For reasons of simplicity we use 'Western foreigners' and 'non-Western foreigners' in this article.

effects of promised incentives on ethnic minority response rates are less clear or altogether non-existent. To persuade respondents in the 2004 Statistics Netherlands experiment *A Tailored Approach Strategy for Young Moroccans and Turks for the Dutch Family and Fertility Survey,* a gift voucher of €10.00 is promised in an advance letter and by the interviewers. However, the incentive does not produce any major effect on response rates (Van den Brakel et al. 2006). Unlike earlier ALLBUS surveys, the ALLBUS 2002 survey includes a €10.00 coin as incentive. The cooperation rate of German nationals increases from 53.7% in 2000 to 60.1% in 2002, but the cooperation rate of non-Germans decreases in the same period from 70.0% to 64.2%[3)] (Feskens et al. 2006).

The effects of incentives can be studied from various theoretical perspectives. Groves et al. (2000) introduce the *leverage-salience theory.* According to this theory, the ultimate effect of incentives on survey participation not only depends on the incentives themselves, but is also related to many other survey features such as the topic and sponsor (see also Groves, Singer, and Corning 2004; Stoop 2005). Another approach is the *social exchange theory,* which views human behavior as an exchange of rewards between actors (Zafirovski 2005). From this perspective, incentives can be viewed as a reward for survey participation. This theory also implies that the marginal utility of an incentive should be larger among sampled units with a lower socio-economic status. Ethnic minorities have a below-average socio-economic status, so from this theoretical perspective they should be more positively disposed to receiving an incentive, reflecting the higher marginal utility gained by an incentive. It has however been implied that ethnic minorities have higher cooperation rates than native sampled units (Schnell 1997; Feskens et al. 2006). The above-average ethnic minority cooperation rates make it less feasible to increase response rates with incentives, although there is sometimes the impression that ethnic minorities use language problems as a friendly way to refuse to participate in a survey. The so-called *soft refusals* are following this argument reflected in the higher ethnic minority *unable to participate* rates.

In sum, various theoretical perspectives generate different expectations about the effect of incentives on ethnic minority response rates. Economic exchange theory predicts that a higher marginal utility of incentives in groups with a lower socio-economic status like ethnic minorities can lead to a larger positive effect on ethnic minority response rates. Sociological theory on nonresponse recognizes the already above-average ethnic minority cooperation rates (Schnell 1997), possibly reducing the potential effect of strategies to increase survey participation. Several experiments show the positive effect on response rates in survey research of incentives. However,

[3)] In 2002, all the interviews conducted in the previous four weeks are rejected due to doubts about whether all the rules have been followed (Blohm et al. 2003). This results in far more non-analyzed interviews than in *Allgemeine Bevölkerungsumfrage der Sozialwissenschaften* (German General Social Survey) 2000. If the non-analyzed interviews are interpreted as refusals, there is an increase in the cooperation rate from 51.5% to 52.8% among the German sampled units and a reduction from 65.3% to 49.5% among the non-German sampled units.

the effect of incentives on the ethnic minority response rates in Europe is less clear. We know though that simply increasing response rates can actually increase the bias introduced by nonresponse rates (Merkle and Edelman 2002; Stoop 2005). To end with a balanced sample that can be used to draw inferences about the target population with as little nonresponse bias as possible, it is important to study the effect of incentives on currently under-represented groups in survey research (see also Singer et al. 2000, p. 187).
For this purpose, a controlled randomized experiment among 13,000 households at Statistics Netherlands is conducted in the autumn of 2005. In this experiment, standard stamps enclosed in the introduction letter are offered to sampled units. In addition to a control group where no incentive was given, three other variants are introduced: booklets of five, ten or twenty stamps are sent with the introduction letter, representing a monetary value of approximately two, four or eight euros. We explicitly focus on the effect of the incentives on the ethnic minority response and cooperation rates. We evaluate the absolute effect on the ethnic minority survey attrition as well as the relative effect of the incentive on ethnic minorities compared with native Dutch sampled units selected by propensity score matching.

This results in two research questions addressed in this study:

1) *What is the effect of incentives on ethnic minority cooperation rates?*
2) *Are there differences between the cooperation rates of ethnic minorities and comparable native Dutch sampled units?*

For a more general study of the results of this controlled experiment, we refer to the study by Wetzels et al. (in press). The following two sections describe the data used and the research design. The fourth section presents the results and the final section gives the conclusions.

6.2 Data

The experiment is conducted within the Dutch Labor Force Survey (EBB) in November and December 2005. We briefly describe this survey and the obtained data for the experiment below.
The Labor Force Survey is a rotating panel study conducted by Statistics Netherlands. After receiving an introduction letter, sampled units are visited at their homes by interviewers for a CAPI (Computer-Assisted Personal Interviewing) administrated interview. Respondents are re-approached for a CATI (Computer-Assisted Telephone Interviewing) for four more short interviews. The Labor Force Survey is conducted among private households in the Netherlands. The Labor Force Survey is a stratified two-stage sample. The sampling frame is a list of all the addresses constructed from the Population Register. These registered addresses are

the sampling units. Addresses with multiple households registered are excluded, because in such cases it is not known which household receives the incentive.[4] Response figures are based on households. Communities are drawn in the first stage and households are selected in the second stage using systematic sampling. The cases are allocated to an experimental condition in a random way. The size of the experimental conditions is however determined beforehand; 6,000 addresses do not receive an incentive, 3,000 addresses receive an incentive of five stamps (value two euros), 3,000 addresses receive an incentive of ten stamps (value four euros) and 500 addresses receive an incentive of twenty stamps (value eight euros).

To answer our first research question, we include all the ethnic minority households (1,861) and native Dutch households (9,089), see also Table 6.1. To answer the second research question, we include all the ethnic minority households with complete background information (see also next section); 1,777 ethnic minority households remain for this second analysis. We match a native Dutch sampled unit with comparable background characteristics and the same experimental condition to each ethnic minority household. So to answer the second research question, we also include 1,777 native Dutch households. The propensity score matching method as proposed by Rosenbaum and Rubin (1983) is explained in greater detail in the following section.

Table 6.1
Experimental conditions and ethnicity

	Native Dutch population	%	Western foreigners	%	Non-Western foreigners	%
No stamps	4,334	47.7	462	47.3	412	46.6
5 stamps	2,204	24.2	221	22.6	206	23.3
10 stamps	2,196	24.2	255	26.1	225	25.5
20 stamps	355	3.9	39	4.0	41	4.6
Total	9,089	100.0	977	100.0	884	100.0

As is noted in the Introduction, ethnic minorities are defined in the Netherlands as everyone residing in the Netherlands with either one or two parents born abroad. The following table shows the ethnicity distribution in the Netherlands in 2005, the year the experiment is conducted.

[4] If we include such cases in the analyses, the figures are very similar.

Table 6.2
Ethnic distribution in the Netherlands in 2005 (http://statline.cbs.nl)

	Population	Percentage of total population
Total population	16,305,526	
Native population	13,182,809	80.8
Ethnic minorities	3,122,717	19.2
Western foreigners	1,423,675	8.7
Non-Western foreigners	1,699,042	10.4
Suriname	329,430	2.0
Turkey	358,846	2.2
Morocco	315,821	1.9
Netherlands Antilles & Aruba	130,538	0.8
Other non-Western foreigners	564,407	3.5

6.3 *Methods*

As is noted above, this article concentrates on the effect of prepaid incentives on ethnic minority response and more specifically cooperation rates. To evaluate the effect of incentives on ethnic minority cooperation rates, we include all the sampled units (10,950). To study possible differences between ethnic minority and native Dutch households as regards the effect of incentives, we select for each ethnic minority unit a comparable native Dutch sampled unit. We do so in such a way that we are able to evaluate the relative effect of incentives on the ethnic minorities as compared to the native Dutch as well as the absolute effect of incentives on ethnic minority cooperation rates.

We know from previous studies (Schmeets and Michiels 2003; Eisner and Ribeaud 2007; Feskens et al. 2007) that ethnic minorities differ from the native population as regards their living conditions and socio-economic status. Ethnic minorities mainly live in urban areas, are more often unemployed and have lower education levels than the native population. These characteristics correlate negatively with response rates (Goyder et al. 1992; Lavrakas 1993; Groves and Couper 1998; Stoop 2004; Van Goor et al. 2005). To compare the relative effect of incentives on the ethnic minority survey attrition with that of the native Dutch, we select native Dutch sampled units with background characteristics comparable to those of ethnic minorities. In other words, to attain a more fair comparison we select more or less similar sampled units. We do so by utilizing the available background information and using propensity score matching (Rosenbaum and Rubin 1983).

The Labor Force Survey is supplemented by administrative data from the Population Register and information about employment and social benefits (Schmeets and Michiels 2003; Schouten 2003). Linking the administrative records makes socio-demographic and socio-economic information available on the nonrespondents at the individual and postal code level. This information has been gathered at Statistics Netherlands in the Social Statistical Database, in which several registers are linked to each other as well as to data from sample surveys (Houbiers

2004). The additional information provided by the link to administrative data makes it possible to study the nonrespondents and match all the ethnic minority units to comparable native Dutch sampled units in terms of important background characteristics. However, no extra information is available on 84 of the ethnic minority sampled units (4.5% of the ethnic minority sampled units). Since no systematic missing data pattern is observed as regards relevant background variables for these 84 cases, they are viewed as missing completely at random and deleted from the data file. These numbers are quite small and can be dropped from the sample without a significant loss of information.

We calculate the propensity score with the variables *household income* and *postal code urbanization.* We select these variables because they are felt to effectively measure the relevant concepts of socio-economic status and urbanization. We also want to include the education level of the sampled units in the calculation of the propensity scores, but this information is unfortunately not available on the nonrespondents. The propensity scores are calculated using logistic regression with the variables *household income* and *urbanization* as independent variables and ethnicity as dependent variable. The predicted values are saved and used as propensity scores. We then match native Dutch sampled units with the same propensity score and in the same experimental condition as the ethnic minorities and select these sampled units.

6.4 Results

6.4.1 What is the effect of incentives on ethnic minority cooperation rates?

To examine the effect of the prepaid incentive we use the cooperation rate. By using cooperation instead of response rates, we assure that sampled units who did not have a chance to participate, for example because they are not contacted, are excluded from the analysis. We use cooperation number four as defined by AAPOR in all the further analyses. This ratio is defined as completed interviews and partial interviews divided by completed interviews, partial interviews and refusals (AAPOR 2006).

Tables 6.3, 6.4 and 6.5 show the response outcomes for the three main groups in the Netherlands, the native Dutch, Western foreigners and non-Western foreigners, for all four experimental conditions. In each table the cooperation rate for each specific condition is given, enabling us to examine the effect of the incentives. We calculate exact p values to test for statistically significant differences using Fisher's exact test. First, we make a 2 x 2 matrix representing all possible outcomes (i.e. incentive, no incentive, cooperation and non-cooperation) and weight this matrix with the absolute number of sampled units in each cell. We test the null hypothesis that the cooperation rate in the null condition is equal to the cooperation rate in the experimental condition ($p < .05$ denoted by *). Then we test the hypothesis that the cooperation rate in an

experimental condition is equal to the cooperation rate in the previous experimental condition, e.g. H0: cooperation rate with five stamps = cooperation rate with ten stamps ($p < .05$ denoted by #).

Table 6.3
Response outcomes for the native Dutch in four experimental conditions

	No stamps		**5 stamps**		**10 stamps**		**20 stamps**	
	Count	*%*	*Count*	*%*	*Count*	*%*	*Count*	*%*
Refusals	1,032	23.8	404	18.3	366	16.7	52	14.6
No opportunity	198	4.6	100	4.5	97	4.4	19	5.4
Language problems	10	0.2	3	0.1	3	0.1	0	0.0
No contact	190	4.4	93	4.2	65	3.0	15	4.2
Response	2,879	66.4	1,593	72.3	1,656	75.4	268	75.5
Partial response	20	0.5	9	0.4	7	0.3	1	0.3
Broken off interview	5	0.1	2	0.1	2	0.1	0	0.0
Total	4,334	100.0	2,204	99.9	2,196	100.0	355	100.0
Cooperation rate 4	73.8%		79.9%**		82.0%**,#		83.8%**,ns	

**p value < .01, H0: cooperation rate experimental condition = cooperation rate null condition.
# p value < .05, H0: cooperation rate experimental condition = cooperation rate previous experimental condition.
ns = not significant.

As is clear from Table 6.3, the native Dutch cooperation rate is 73.8% in the control condition where no stamps are given and increases to 79.9% (five stamps), 82.0% (ten stamps) and 83.8% (twenty stamps). All these increases are statistically significant. Offering ten instead of five stamps also results in a statistically significant increase of the cooperation rate, but offering twenty instead of ten stamps does not. The sample size in the latter condition is however much smaller than in the first three experimental conditions.

Table 6.4
Response outcomes for Western foreigners in four experimental conditions

	No stamps		**5 stamps**		**10 stamps**		**20 stamps**	
	Count	*%*	*Count*	*%*	*Count*	*%*	*Count*	*%*
Refusals	116	25.1	44	19.9	38	14.9	4	10.3
No opportunity	26	5.6	10	4.5	14	5.5	4	10.3
Language problems	13	2.8	4	1.8	12	4.7	2	5.1
No contact	31	6.7	12	5.4	18	7.1	6	15.4
Response	273	59.1	147	66.5	169	66.3	23	59.0
Partial response	2	0.4	3	1.4	3	1.2	0	0.0
Broken off interview	1	0.2	1	0.5	1	0.4	0	0.0
Total	462	99.9	221	100.0	255	100.1	39	100.1
Cooperation rate 4	70.4%		77.4%*		82.0%**, ns		85.2%ns, ns	

*$p < .05$, H0: cooperation rate experimental condition = cooperation rate null condition.
**$p < .01$, H0: cooperation rate experimental condition = cooperation rate null condition.
ns = not significant.

Table 6.4 shows that the cooperation rate among Western foreigners increases from 70.4% (0 stamps), to 77.4% (5 stamps), 82.0% (10 stamps) and 85.2% (20 stamps). The effect of incentives on Western foreigners is thus very similar to the effect on the native Dutch population.

Table 6.5
Response outcomes for non-Western foreigners in four experimental conditions

	No stamps		5 stamps		10 stamps		20 stamps	
	Count	*%*	*Count*	*%*	*Count*	*%*	*Count*	*%*
Refusals	61	15.2	28	13.6	27	12.3	2	4.9
No opportunity	16	4.0	15	7.3	10	4.6	3	7.3
Language problems	55	13.7	30	14.6	36	16.4	3	7.3
No contact	40	10.0	22	10.7	23	10.5	8	19.5
Response	216	53.7	108	52.4	121	55.3	24	58.5
Partial response	11	2.7	2	1.0	0	0.0	1	2.4
Broken off interview	3	0.7	1	0.5	2	0.9	0	0.0
Total	402	100.0	206	100.1	219	100.0	41	99.9
Cooperation rate 4	79.0%		79.9%^{ns}		82.0%^{ns,ns}		92.6%^{ns,ns}	

ns = not significant.

Table 6.5 shows the response outcomes of non-Western foreigners in the four experimental conditions. The cooperation rate of non-Western foreigners increases from 79.0% in the control condition to 79.9% (five stamps), 82.0% (ten stamps) and 92.6% (twenty stamps). This last cooperation rate for non-Western foreigners who receive an incentive of twenty stamps should however be interpreted with care. Since there are only 41 sampled units in this specific experimental condition, small changes in the refusal rate can cause large changes in the cooperation rate. This is also reflected in the results of Fisher's exact test: none of the results are statistically significant. Additional doubt arises about this increase in the cooperation rate caused by the incentive when the response rate is examined. This rate only increases modestly in the last condition compared with the first three conditions. Furthermore, the cooperation rate of non-Western foreigners hardly increases with a five or ten-stamp incentive. It does however start at a higher level in the null condition. This supports the findings of earlier studies where nonresponse problems on the part of non-Western foreigners are attributed to in particular contact and language problems (Feskens et al. 2006). Both are response outcomes that do not impact the cooperation rate. In summarizing, incentives increase the cooperation rates of the native Dutch and Western foreigners, but not of non-Western foreigners.

6.4.2 *Are there differences between the cooperation rates of ethnic minorities and comparable native Dutch sampled units?*

To examine the differences between non-Western foreigners and comparable native Dutch sampled units, we match ethnic minorities with selected native Dutch

sampled units using propensity score matching. To do so, we select all 1,777 ethnic minority units and match one native Dutch sampled unit to each of them, as is described above. We also want to study the effect of incentives on various groups of non-Western foreigners in greater detail. The non-Western foreigners in the Netherlands are mainly from Surinam, Turkey, Morocco, the Netherlands Antilles and Aruba. The following table shows the cooperation rates of these ethnic groups and the comparable native Dutch sampled units. For completeness, the cooperation rates of all the ethnic minority and native Dutch sampled units are also presented in Table 6.6. Table 6.6 compares the cooperation rates of the null condition to those of an incentive (five, ten or twenty stamps). To remain with a sufficient sample size for all the ethnic groups, we combine all the incentive conditions.

Table 6.6
Cooperation rates of various ethnic groups without and with incentives

	Without incentive		**With incentive**		
	Cooperation rate (%)	*N*	*Cooperation rate (%)*	*N*	*Difference*
All sampled units	73.9	5,208	81.1	5,742	7.3%**
Native Dutch population	73.8	4,334	81.1	4,755	7.4%**
Selected native Dutch population	70.8	835	80.8	942	10.1%**
Western foreigners	70.4	462	80.1	515	9.7%**
Non-Western foreigners	79.0	402	82.0	466	2.9%ns
Surinamese	73.7	98	74.4	105	0.7%ns
Netherlands Antilles & Aruba	92.6	32	89.3	38	–3.3%ns
Turkey	78.0	80	83.3	102	5.3%ns
Morocco	72.7	66	78.8	89	6.1%ns
Other non-Western foreigners	83.0	126	87.5	132	4.5%ns

**$p < .01$, H0: cooperation rate without incentive = cooperation rate with incentive.
ns = not significant.

The first important result presented in Table 6.6 is that native Dutch sampled units selected with propensity score matching are no more similar to ethnic minorities than the whole native Dutch population sample. The selected native Dutch sampled units have a cooperation rate of 70.8% in the control condition, which increases to 80.8% when an incentive is given. So because of the lower cooperation rate in the control condition, the effect of the incentive is somewhat larger in this group with a lower average household income and a higher urbanization level than in the whole native Dutch sample. None of the effects of the incentives are significant in the groups of non-Western foreigners.
In summarizing, the cooperation rates of non-Western foreigners are still different than those of selected native Dutch sampled units.

6.5 *Conclusions*

In this study we examine the effects of prepaid incentives on ethnic minority cooperation rates in the Netherlands. To do so, a controlled randomized experiment is conducted at Statistics Netherlands, where sampled units receive either no incentive or a prepaid incentive of five, ten or twenty stamps representing a monetary value of approximately two, four or eight euros.

In general, the ethnic minority results are based on relatively small sample sizes and should thus be interpreted with great care. Future research can focus on this aspect.

We note that the native Dutch cooperation rate increases with the increasing value of the incentive. The same holds true for Western foreigners. The cooperation rate of non-Western foreigners does not however increase with a prepaid incentive of five, ten or twenty stamps. The sample size of this latter condition is very small (41 non-Western foreigners receive this twenty-stamp incentive), so this result should be interpreted with great care.

We also examine the effect of incentives on ethnic minority cooperation rates compared with those of comparable native Dutch sampled units. To do so, we match each ethnic minority sampled unit to a native Dutch sampled unit with similar background characteristics. After first creating propensity scores with available background information on household income and urbanization at the postal code level, we use propensity score matching. This information is available for respondents as well as for nonrespondents. The second step is to match a native Dutch sampled unit to an ethnic minority sampled unit with the same propensity score and in the same experimental condition. It should be noted however that with only limited information for respondents and nonrespondents alike, this propensity model might not account for all the differences between the two matched groups.

To maintain a sufficient sample size, we only compare the control condition with the incentive condition (five, ten and twenty stamps together) for this purpose. Compared to these matched native Dutch sampled units, the cooperation rates of non-Western foreigners differ substantially. The difference in the cooperation rates is even greater between the matched native Dutch sampled units and non-Western foreigners (10.1% vs. 2.9%) on the one hand and all the native Dutch sampled units and non-Western foreigners (7.4% vs. 2.9%) on the other. A further examination of the non-Western foreigners reveals that none of the cooperation rates of non-Western groups increase statistically significantly with the prepaid incentives.

Noticeable is that although the cooperation rate of non-Western foreigners does not increase with prepaid incentives, it is already high (79%) without any incentive. It is almost as high as the cooperation rate of the native Dutch sampled units who receive an incentive (81.1%). Implementing small incentives to increase the response rates of non-Western foreigners does not have an effect that is statistically significant. It does however seem to be effective with the native Dutch population and Western

foreigners. Perhaps more importantly for nonresponse bias reduction, it also has a large effect on the cooperation rate of the selected native Dutch sampled units. These sampled units with a lower average household income and mostly living in more urbanized areas, characteristics often attributed to below-average response rates, exhibit a substantial increase in the cooperation rate (70.8% vs. 80.8%).

Linking these results back to the theoretical considerations reveals that social exchange theory rightly hypothesizes a larger marginal utility of incentives for native Dutch sampled units with a below-average socio-economic position. This does not however hold true of ethnic minorities. Higher initial cooperation by ethnic minorities partially eliminates the potentially positive effect of incentives on this group. Incentives do not affect the number of sampled units that cannot participate in survey research due to language problems, sometimes viewed as a category also containing soft refusals.

The higher cooperation rate in the null condition and the non-existent effect with incentives suggest that the response problem of non-Western foreigners is largely caused by lower contact rates and nonresponse due to survey language problems. Increasing the response rates of non-Western foreigners should focus on these two aspects rather than on using incentives.

7 *Studying People Living in Non-Private Households: Results of a Large Pilot Study in the Netherlands*

Residents of non-private households are currently excluded from most general social surveys. Their inability to answer survey questions and difficulties in approaching them are viewed as obstacles to survey this population. However, excluding them can potentially bias survey estimates. These two claims are investigated in this study. Firstly, we evaluate the opportunities to study residents of elderly and nursing homes, two major groups belonging to the non-private population. Secondly, we assess possible differences between the key characteristics of these two groups and those of comparable residents of private households to explore the potential bias of excluding them from survey research. To do so, a pilot study is conducted at Statistics Netherlands. In this pilot study, 537 responses are collected from residents of elderly and nursing homes. We find that with an adjusted approaching strategy and questionnaire, conducting survey research among residents of elderly and nursing homes is certainly feasible in terms of response rates and data quality. An examination of the potential bias caused by excluding non-private households from general nation-wide social surveys reveals substantial differences between non-private and private households on almost all key survey variables.[1)]

[1)] This chapter has been submitted for publication as Feskens, R.C.W., Lensvelt-Mulders, G.J.L.M., Beukenhorst, D., Kockelkoren, S., and Wetzels, W. (2008). Studying the Non-Private Population: Results of a Large Pilot Study in the Netherlands. *Submitted manuscript*.

7.1 *Introduction*

Almost all nation-wide surveys conducted by national statistical institutes define the target population as "all residents of a country *excluding* people living in non-private households" (cf. Gatward, Lound, and Bowman 2002).[2)] In the main general surveys conducted by Statistics Netherlands, residents of non-private households are also excluded from the sample frame before the actual survey is conducted. Therefore they are not covered in the general population statistics. They are excluded by definition and as such, this is not a form of undercoverage, one of the cornerstones of data quality (Hox et al. 2007). Eurostat (2000) provides the survey quality indicator of completeness. This means that the domains for which statistics are available should reflect the needs and priorities expressed by the users of the statistics. Excluding residents of non-private households means failing to meet this quality criterion.

There are two main reasons to exclude residents of non-private households: 1) they cannot be approached in the same way as residents of private households and 2) their ability to take part in survey research is questioned, especially if they are very old.

Excluding them from survey research can, however, bias survey characteristics for the general population, which is a cause for concern. Moreover, information is then missing for a relevant part of the ageing population.

At Statistics Netherlands, residents of non-private households are defined as 'the population living (for at least six months) in an establishment providing managed residential accommodations in a professional manner' (Reep 2003). Residents of non-private households are a heterogeneous collection of several groups. Non-private households constitute about 1.3% of the Dutch population (http://statline.cbs.nl), although this percentage is much higher among older age categories. Table 7.1 shows the sizes of groups of residents of non-private households in the Netherlands in 2007.

As is clear from Table 7.1, the residents of elderly and nursing homes are by far the largest groups of residents of non-private households. A total of 14.5% of the Dutch population is above the age of 65 and this percentage is increasing every year (http://statline.cbs.nl). To anticipate the effects of changes in the population's age composition, reliable information about these two groups is becoming increa-

---

2) Other examples of surveys that exclude residents of non-private households from the target population are the General Social Survey conducted by Statistics New Zealand (http://www.stats.govt.nz/developments/gen-social-survey.htm), the Household Expenditure Survey conducted by the Australian Bureau of Statistics (http://www.abs.gov.au/Ausstats/abs@.nsf/30b7945513904945ca2570c000174361/5f1422f1af472d80ca256bd00026aee6!OpenDocument), the General Social Survey conducted by Statistics Canada (http://www.statcan.ca/cgibin/imdb/p2SV.pl?Function=getSurvey&SDDS=4501&lang=en&db=IMDB&dbg=f&adm=8&dis=2#b1) and the European Social Survey (http://www.europeansocialsurvey.org/index.php?option=com_content&task=view&id=80&Itemid=125).
All websites consulted in March 2008.

Table 7.1
Residents of various types of non-private households in 2007

Type of institution	Number of residents
Homes for the elderly	92,732
Nursing homes	28,416*
Homes for the mentally handicapped	22,015
Psychiatric hospitals	11,524
Family-replacing homes	40,819
Boarding schools	3,028
Monasteries/ convents	4,704
Penitentiaries	3,615
Total	206,732

* These figures are likely to be an undercount of the actual population. The residents of nursing homes are especially apt to remain registered at their former address. Sampling frame information indicates that the actual number of accommodations at nursing homes is about 69,000, twice the figure from the Population Register.
Source: Statistics Netherlands, http://statline.cbs.nl

singly important. This is why we focus this pilot study on the residents of elderly and nursing homes. Moreover, we expect to encounter more or less the same survey research adjustments for residents of nursing homes as for residents of elderly homes, at least in terms of approaching strategies. To gain insight into the possibility of conducting survey research among residents of elderly and nursing homes and assess the potential differences on key characteristics between the residents of non-private and private households, a large pilot study of elderly and nursing homes is conducted at Statistics Netherlands.

In surveying non-private households, problems are anticipated as regards 1) approaching sampled units and 2) their ability to take part in survey research. The problems regarding their ability to take part can be divided in two subcategories, (i) health and functioning-related problems and (ii) personal and motivational factors. Subcategory (i) consists of visual and hearing problems (Corso 1977; Havlik 1986), severe cognitive limitations (Gurland, Cross, and Golden 1980; Pfeffer, Afifi, and Chance 1987) and limited recall capacities (Poon 1985, Perlmutter 1986). The older the respondents, the more likely these problems are to occur (Havlik 1986; Havlik, Liu, and Kovar 1987). This may result in more missing data (Colsher and Wallace 1989). Subcategory (ii) consists of issues like nonresponse due to a fear of crime among the elderly (Clemente and Kleiman 1976). The issue of increasing introspection with increasing age (cf. Botwinick 1978) can decrease a general interest in survey research (Yu and Cooper 1983; Goyder 1985) or lead to more "don't know" answers (Herzog and Rodgers 1982). All these characteristics may affect the quality of answers provided by residents of elderly and nursing homes and are part of the subject of this study. We assess the possibility of studying residents of elderly and nursing homes with respect to data quality. In other words, we evaluate whether data collected from this population can be interpreted in a meaningful way.

Concerns about possible biased estimates in survey research are noted in the literature on studying residents of non-private households (cf. Rosenwaike 1985; Havlik et al. 1987; Kohout 1992; De Klerk 2001; Gatward et al. 2002), particularly if statements about categories of elderly people are made (Schmeets and Geurts 1990). Our second research objective addresses potential effects of excluding non-private households from general national surveys. If relevant differences are found on key characteristics between residents of non-private and private households, this would indicate that a relevant group is now being overlooked in general national surveys. So as a second goal of this study, we examine the potential bias on survey estimates due to not including residents of elderly and nursing homes. This results in our two research questions in this study:

1) *Is it possible to study residents of elderly and nursing homes?*
   a. *What are the institution and resident response rates?*
   b. *What is the quality of the answers of residents of elderly and nursing homes?*
   c. *What are the costs of conducting survey research among residents of elderly and nursing homes?*
2) *Do residents of elderly and nursing homes differ systematically on key characteristics from residents of private households?*

The following sections describe the design of the pilot study and the data collected. The results are presented in the fourth section. We close with our conclusions and a discussion.

7.2 *Design of the pilot study*

A guideline for studying residents of elderly and nursing homes in the Netherlands is described by Wetzels, Kockelkoren, Beukenhorst and Feskens (2008) and referred to here. The aim of this pilot study is to collect approximately 500 responses from residents of elderly and nursing homes. Unlike private households, residents of non-private households cannot be directly contacted. For practical reasons, the cooperation is required of the homes or institutions where they live. A two-stage sample needs to be drawn (Gatward et al. 2002). In the first stage of the sampling, the institutions are selected and in the second stage the residents. We draw a sample of 210 institutions from a Health Care Insurance Board (CVZ) registration file listing 1,482 registered nursing and elderly homes. The sample is representative as regards the type of institution, number of residents and location in the Netherlands. After controlling, the information on 205 institutions appears to be correct and they are approached by telephone by especially trained interviewers in April - May 2007. A two-month recruitment period gives them time to decide whether or not to take part in the survey (sometimes the approval is needed of resident councils that do not meet frequently).

In the second stage, a sample of the residents at the institutions is drawn. Temporary residents (i.e. if a specific date of leaving the house is known) at elderly and nursing homes are not approached for interviews since they are not considered eligible for this pilot study. There are different approaching strategies for elderly homes and nursing homes. The sampling procedure can be performed at Statistics Netherlands for elderly homes but not for nursing homes. For elderly as well as nursing homes, our aim is to collect five responses at each institution by residents themselves or by proxy.

For residents of elderly homes, sampling can be done using the municipal base as a sample frame, which is also the sample frame for private household surveys at Statistics Netherlands. For each elderly home, a random sample of fifteen residents is drawn at Statistics Netherlands. Following the order of the list, interviewers approach the sampled units until they have done five face-to-face interviews. Each sampled unit's ability to take part in survey research is determined in conjunction with the contact person at the institution before the interviews are conducted there. If a resident is considered unable to take part, the interviewer asks the reason for this. The interviewer then approaches five sampled units who are deemed able to take part until there are five responses at each institution.

Most nursing home residents are still listed at their previous address in the municipal records and not at the address of the nursing home. Since there is unfortunately no good alternative to be used as a sample frame for nursing home residents, interviewers have to draw a random sample of residents the first time they come to a nursing home. Interviewers randomly select thirteen nursing home residents using the institution's list. The sampling process is a simple random sampling procedure. To prevent systematic ordering affecting survey estimates (e.g. ordering by time of admission to the home), the sample is randomized again. We do not expect many nursing home residents to be able to take part in the survey, especially not psycho-geriatric cases (cf. De Klerk 2001). The interviewers decide in conjunction with the contact person at the institution which sampled units are able to take part and which are not. Interviewers are instructed to first collect data on the sampled units considered unable to take part. A proxy mail questionnaire is sent to a relative of theirs or some other close contact person. For privacy reasons, this is done by the institution. If the questionnaire is not sent back, the interviewers remind the institution two weeks later. We expect a return rate of 50% from the proxy questionnaires. Interviewers collect data from the residents if they do not expect to be able to acquire five responses at each institution by only sending proxy questionnaires (i.e. if less than ten proxy questionnaire are sent). In the telephone recruitment phase, about a quarter of the institutions appear to be combinations of elderly and nursing homes. The institutions are approached using the sample frame information. A flow diagram of this design is shown in the Appendix.

Before the field work period starts, the approaching strategy and questionnaire are pre-tested. The drawing of a sample by the interviewer and the questionnaire are successfully tested at three institutions, which are not re-approached during

the pilot study. This is done with nine residents of elderly and nursing homes and twelve proxy respondents, although the questionnaire is somewhat adjusted. The data collection period for nursing and elderly homes is July-August 2007 and the interviews are conducted by 36 interviewers. Shortly before the data collection period starts, all the interviewers, except the ones already taking part in the pretest, are given a special interview training course explaining the approaching strategies.

The questionnaire covers various topics such as household and living conditions, helping devices, psychological limitations, aspects of medical care, mental capacities, psychological and mental health, social contacts and leisure activities and ends with a section on education. If possible, the questions are adopted from the Dutch Survey on Living Conditions (POLS), though the answer categories and other aspects are adjusted to this special population. Conducting the face-to-face survey takes an average of an hour. Ideally, no one but the respondent and the interviewer are in the room during the interview. The interviewer makes a note of it if this is not possible. After the interview, the interviewer can give the respondent a small present consisting of five cards to express his thanks.

Lastly, a small mail questionnaire is filled in by the contact person at the institution with questions on background characteristics of the residents.

7.3 *Data*

7.3.1 *Institutions*

For this pilot study 210 institutions are selected of which three institutions have been approached in the pretest and are not re-approached again. 118 elderly homes and 89 nursing homes remain in the sample. Five addresses are not institutions (sample frame errors), five institutions cannot be handled during the field work period and information is missing for two institutions, so respondents are recruited from 195 institutions. The results of the recruitment are shown in Table 7.2.

Table 7.2
Recruitment results at institutions

	Elderly home	Nursing home	Total
Cooperation	83	57	140
Refusal	21	27	48
Refusal before management can be contacted	4	1	5
Appointment outside recruitment phase	1	1	2
Institutions not handled	5	0	5
Sample frame errors	3	2	5
No information available	1	1	2
Total	118	89	207

Of the 202 eligible institutions, 140 agree to cooperate in the study (69.3% AAPOR response definition 3, AAPOR 2006). The institutions each provide the name of a contact person who can be addressed by the interviewer in the actual data collection period. Of the 140 institutions that agree to cooperate, 100 institutions are selected for the study and 40 are put in a fall back pool in case any of the 100 do not take part in the pilot study after all. The 100 institutions are selected in such a way that the results can still be generalized to all the elderly and nursing homes in the Netherlands in terms of type, number of residents and location. The institution response rates are discussed in greater detail in section 7.4.1.1.
According to the sample frame, 61 of the selected institutions are elderly homes, 39 are registered as nursing homes. Controlling this information at the institutions reveals that quite a few of them are combination institutions, i.e they provide elderly and nursing care. This results in a sample containing 48 elderly homes, 30 nursing homes and 22 combination institutions.

7.3.2 Residents

In the second stage of the sampling, residents are selected. Again, the response rates are discussed in section 7.4.1.1. A total of 358 residents of elderly and nursing homes fill in the questionnaire. Their mean age is 84.5. Almost 73% of these respondents are women. Moreover, 179 proxy responses are collected via the family questionnaire. The mean age of the residents collected by proxy responses is 81.4; here as well, most of the residents are women (79%). A total of 75.8% of the respondents at elderly homes are female (direct and proxy responses). The mean age of the respondents at elderly homes is 86. A total of 73.5% of the residents at nursing homes are female and their mean age is 80.6. In total 537 responses are collected in this pilot study.

7.4 Results

7.4.1 Is it possible to study residents of elderly and nursing homes?

As is noted in the Introduction, there are two main reasons to exclude non-private households from general surveys. Firstly, it is feared that it can be hard to obtain reasonable response rates from residents of non-private households. Secondly, the quality of the answers given by residents of non-private households can be questionable. In the following two sections, we investigate these two claims. We also evaluate the costs of conducting surveys among this population. This grants insight into whether or not it is possible to study residents of elderly and nursing homes.

7.4.1.1 What are the response rates among institutions and residents of non-private households?

An initial quality indicator is the response rate among institutions and residents of non-private households. Their willingness and ability to take part can give an initial indication of whether or not it is possible to conduct survey research among residents of elderly and nursing homes.

Institutions

As is noted in section 7.3, the response rates at the first stage of the sampling procedure are quite reasonable; 69.3% of the institutions agree to take part in this study. The response rate at elderly homes is somewhat better; 72.2% agree to take part in the pilot study vs. 65.5% of the nursing homes. The reasons not to take part in the pilot study include reorganizations, mergers and the number of other studies. Despite the small sample size at the institution level, we also examine the response composition of institutions with respect to the size of the institution, the level of urbanicity and at nursing homes, and the quality of the institution. Although no statistical differences are found, the percentage of institutions that do not cooperate is higher among small elderly homes than medium-sized and large elderly homes (28% vs. 13%). Small response rate differences between institutions are found with respect to urbanicity. Although not statistically significant, the four largest cities (30% refusals) and rural communities (26% refusals) appear to cooperate less than the other urbanicity categories (24, 17 and 17% refusals). A quality indicator presented by the Ministry of Health is available for the nursing homes. A comparison of response information shows that the percentage of institutions with a high quality indicator score that do not take part in this study is somewhat higher (33%) than of institutions with a low score (28%). Again these differences are not statistically significant. We can conclude that there are indications of a somewhat lower response in the four largest cities and rural communities, although the differences are not statistically significant. The most important indicator with respect to potential nonresponse bias at the institution level, an institution quality indicator, does not reveal a selective nonresponse.

Residents

Before an interviewer speaks to residents, their ability to be interviewed is assessed in close cooperation with the contact person at the institution. Most residents at elderly homes are considered able to be interviewed. This percentage (59.5%) is somewhat lower than expected. At nursing homes, 30.4% of the residents are considered able to be interviewed, confirming our expectations. Dementia is given as the reason why residents at elderly homes cannot take part in 65.5% of the cases. This percentage is even higher at nursing homes, i.e. 74.9%. Table 7.3 shows response information for elderly homes.

Table 7.3
Response outcomes for residents of elderly homes

	n	Percentage
Response	278	35.6
Partial response	4	0.5
Nonresponse	32	4.1
Not capable	316	40.5
Capable but not in sample	145	18.5
Proxy response	5	0.6
Total	780	100.0

A total of 283 completed responses are collected from residents of elderly homes, five of which have been answered by proxy respondents. This results in a response rate of 88.7%. However, residents the contact person considers unable to take part are not included because they do not belong to the target population of this study. Response information for residents of nursing homes is shown in Table 7.4.

Table 7.4
Response outcomes for residents of nursing homes

	n	Percentage
Proxy response	174	45.2
Proxy nonresponse	75	19.5
Response	80	20.7
Partial response	4	1.0
Nonresponse	11	2.9
No questionnaire	41	10.6
Total	385	100.0

A total of 80 residents of nursing homes and 174 proxy respondents fill in the questionnaire, resulting in 254 completed responses. Consequently, the response rate at nursing homes residents is 73.8%. No proxy respondent can be found for 41 residents who cannot take part in the survey themselves. If they are included in the response calculation, the response rate is 66.0%.

Combining the response figures at elderly and nursing homes, the response rates for residents and proxy respondents can be calculated. A total of 358 (278 + 80) residents of elderly and nursing homes take part in the survey themselves, resulting in a response rate of 87.5%. The proxy response rate is 70.5% or 60.7% including the sampled units no proxy respondents can be found for with a total of 179 proxy respondents.

In sum, the approaching strategy performs well in terms of response rates at the institution and resident level, which is a first indication of sound data quality.

7.4.1.2 *What is the quality of the answers given by residents at elderly and nursing homes?*

There is evidence that increasing age and reduced cognitive abilities have a negative effect on data quality (e.g. Andrews and Herzorg 1986; Bradburn, Rips, and Shevell 1987; Rodgers, Herzorg, and Andrews 1988; Colsher and Wallace 1989; Knäuper, Belli, Hill, and Herzog 1997). This is why the questionnaire is adjusted to this very specific population and extra care is devoted to making participating and answering as easy as possible. Nevertheless, we expect difficulties in the question-answering process for many of the respondents. We consequently also want to assess the quality of answers. Three methods are used to evaluate the data quality, 1) the incomplete or "no information" responses are counted, 2) the responses are compared with available background information and 3) the response patterns are analysed.

Number of incomplete or "no information" responses

Krosnick (1991) suggests that increasing the difficulty of the questions may lead to more incomplete, biased or no information at all. The frequency and severity of the potential consequences depends on the difficulty of the questions and the respondents' cognitive ability and willingness to answer. We expect respondents with less cognitive abilities to give more incomplete, biased or no information at all (cf. Kaldenberg, Koenig, and Becker 1994; Knäuper et al. 1997). Table 7.5 shows the percentage of complete answers (absence of refusals and "don't know" answers) to 81 of the relevant survey questions.

Table 7.5
Percentage of completed answers given by various respondents

	Elderly Homes	Nursing homes	Residents	Proxy
Percentage of completed answers	98.0	79.6	98.5	69.3

The percentages of completed answers is excellent among residents who respond themselves. Not surprisingly, the number of missing items is substantially higher for the proxy respondents. More than 30% of all the items are missing. Questions about aspects of caring and social contacts are the most problematic, and questions about health only yield a small amount of incomplete information.

Test-retest

The psychometric concept of reliability derived from classical test theory refers to correlational consistency between two efforts to measure the same thing using maximally similar measurements and independent of any true change in the quantity being measured (Alwin 1989). A general design strategy for measuring

reliability uses similar measures in one questionnaire. As an initial tool to assess the reliability of answers questions about the number of own children are implemented in the questionnaire. The question about the number of children is followed by one about the number of sons and daughters, which should equal the total number of children.

Since it is very difficult to ask identical questions twice in one survey, we also use a question on the respondents' age, information also available in the registration files. This is only feasible for residents of elderly homes, where frame data is available. Assuming the information in the registration files is correct, comparing it with the responses can grant insight into the data quality. Although test-retest measurements are not difficult, we think the results are a good indication of data quality. Tables 7.6a and 7.6b show the percentage of consistent answers on the respondents' number of children and their own correct age.

Table 7.6a
Percentage of consistent responses on number of children by various respondents

	Elderly Homes	Nursing homes	Residents	Proxy
Percentage of correct answers on # children	97.8	99.1	97.5	100.0

Table7.6b
Percentage of correct responses on their own age by various respondents

	Elderly Homes	Nursing homes	Residents	Proxy
Percentage age correct	96.2	NA	91.1	100.0

NA = Not available

Not surprisingly, proxy respondents have no problem consistently reporting the number of children the respondents have and their correct age. The reliability of other figures is also found to be over 90%.

Response patterns

To investigate response patterns, the questionnaire includes a scale consisting of a number of questions about the same topic. The scale should have good psychometric properties and the reliability and scalability should be high (Van den Wittenboer, Hox, and De Leeuw 1997). The "De Jong-Gierveld Loneliness Scale" measures social and emotional loneliness among different populations in studies with different designs (De Jong-Gierveld and Kamphuis 1985). Van Tilburg and De Leeuw (1991) show that this scale generally meets the psychometric requirements of items non-response, scale homogeneity and person scalability. We thus use these scales to evaluate response patterns in our survey. The social loneliness scale consists of five items and the emotional loneliness scale consists of six items. The

internal consistency of answering is evaluated using Cronbach's alpha. Following the instructions of De Jong-Gierveld and Kamphuis (1985), respondents with two or more items missing are excluded from the further analysis. The results are shown in Table 7.7.

Table 7.7
Response consistency of various respondents

	Elderly Homes		Nursing homes		Residents		Proxy	
	α	*n*	α	*n*	α	*n*	α	*n*
Social loneliness	0.727	270	0.708	139	0.722	341	0.728	68
Emotional loneliness	0.797	277	0.793	140	0.783	347	0.851	70

The internal consistency of answering goes well with Cronbach's alpha ranging from .708 (nursing home responses) to .851 (proxy respondents). The first item of the scale ("Can talk about daily problems") seems the most problematic among all the respondent groups in terms of internal consistency.
The data quality results are summarized in figure 7.1. The quality indicators are listed ranging from 0 (very poor) to 100 (excellent). Almost all the indicators indicate a reasonable level of data quality. The medium level of items completed by proxy respondents (69.3%) is probably the most problematic. The results answer the first research question: It is possible to study residents of elderly and nursing homes?

7.4.1.3 Costs of conducting survey research among residents of elderly and nursing homes

Conducting the pilot study at elderly and nursing homes is expected to require seventeen hours of interviewer capacity for each elderly home and thirteen hours for each nursing home. The hours include all the activities in the homes such as drawing a sample of residents, conducting face-to-face interviews, distributing proxy mail questionnaires and travelling to the homes.

The interviewers' accounts of their time spent at the institutions indicate that conducting the pilot study requires less time than is expected. Conducting the pilot study at elderly homes requires an average of 11.4 hours and at the nursing homes an average of nine 9 hours. Table 7.8 shows the average times for various activities.

Figure 7.1 Quality of data

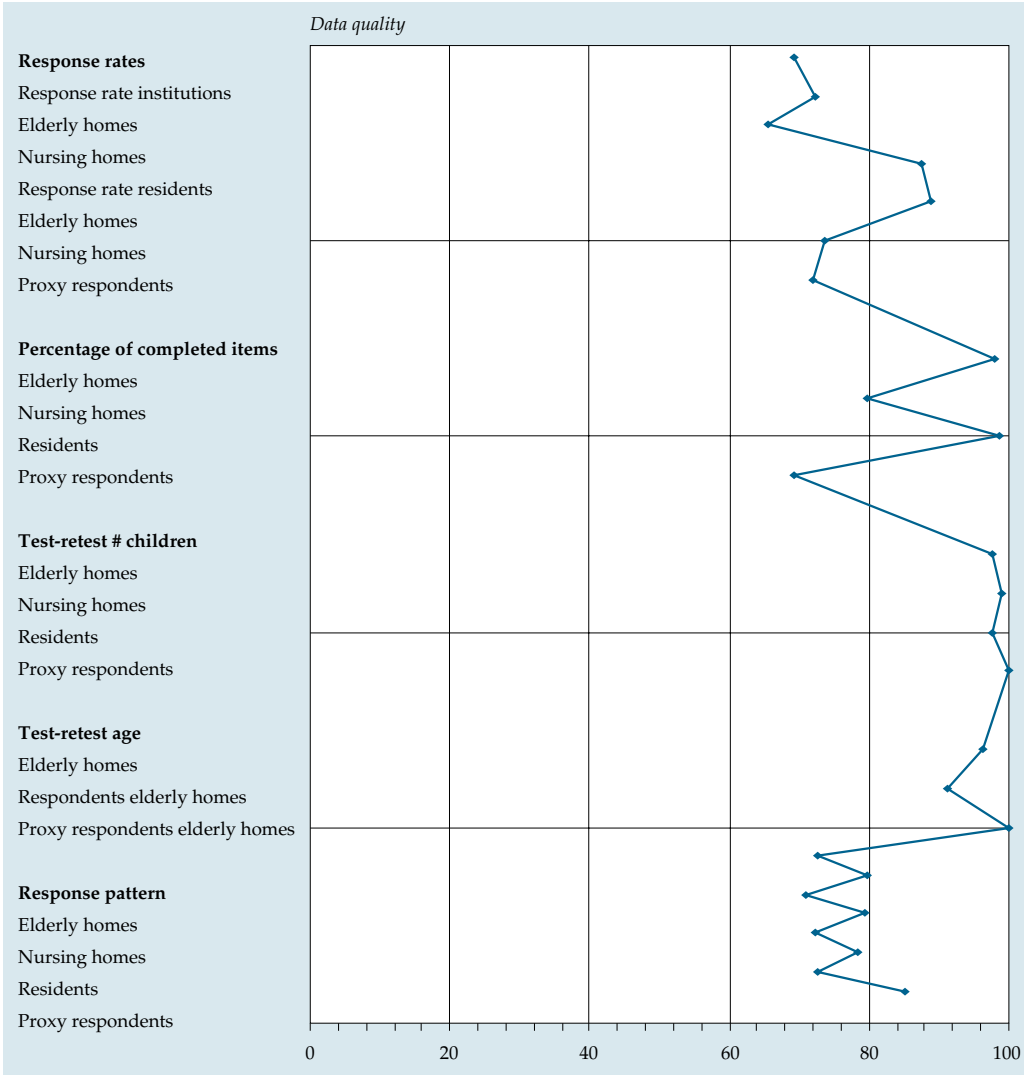


Table 7.8
Average times for various activities

Activity	Elderly homes	Nursing homes
	minutes	
Drawing samples		48
Determining ability of residents to take part	39	28
Conducting face-to-face interviews	289	117
Distributing and recalling proxy questionnaires		37
Travelling time	211	145
Clerical and other activities at home	90	120
Other activities	66	64

7.4.2 *Do residents of elderly and nursing homes differ systematically on key characteristics from residents of private households?*

Survey methodologists are concerned that excluding the residents of non-private households from general social surveys might lead to a biased observation of the general population (cf. Rosenwaike 1985; Havlik et al. 1987; Kohout 1992; De Klerk 2001; Gatward et al. 2002), especially if statements are made about certain sub-populations, e.g. older age categories (Schmeets and Geurts 1990). Their concern is based on the assumption that members of non-private households differ substantially from members of private households on key survey characteristics, e.g. health issues. If this is true, it legitimizes including the residents of non-private households in general social surveys. To evaluate this claim, we examine potential differences on key survey estimates between private and non-private households. Most questions need to be adjusted to our target population. In particular, the number of answer categories is reduced. However, to maintain comparability, a number of survey questions on key issues from the General Survey on Living Conditions (POLS) and the Survey on Health (GeZo) are included in this pilot study. Since Lee, Mathiowetz and Tourangeau (2004) note that comparing other topics may lead to underreporting by proxy respondents, we compare items that are almost all observable to the proxy respondent. Survey estimates of questions about physical limitations, helping devices and health are compared between comparable residents of private households above the age of 75 (mean age 80.18, n=1,807) and respondents and proxy respondents living in elderly and nursing homes. The estimates are shown in the Appendix.

In general, residents of elderly and nursing homes report more physical limitations, have more helping devices and are in somewhat poorer health than their peers in private households. In the health category, most of the differences between the two groups are not statistically significant. Not surprisingly, elderly home residents have less physical limitations and less helping devices and score better on health issues than proxy respondents. These results suggest that excluding residents of

elderly and nursing homes may bias survey estimates about the elderly on important population characteristics.

7.5 Conclusions and discussion

Concerns about excluding a relevant societal group are the main rationale for this study. Most nation-wide studies currently exclude residents of non-private households. They are usually excluded from the sample frame before the actual sampling procedure starts. One reason for excluding them is their different living situation. The common approaching strategies cannot be used because for practical reasons, the cooperation is required of the institutions providing managed care and housing. A second reason is the assumed difficulty of interviewing residents of non-private households. To investigate the possibility of conducting survey research among residents of elderly and nursing homes, two main groups belonging to the non-private population, a pilot study has been conducted at Statistics Netherlands. In this study we also investigate the potential for bias in survey characteristics if this population is excluded. We have collected 537 direct or proxy responses at elderly and nursing homes.

Evaluating the data reveals that the data quality is generally quite good. There are however still certain difficulties as regards surveying non-private households and several limitations to this study. Firstly, sampling non-private households remains problematic. Of course the availability of registration files that can be used as sampling frames differs from country to country, but the Netherlands is probably not the only country where there is no workable sampling frame. We have tried to solve this problem by having the interviewers partially sample the individuals at institutions. In practice, even after intensive interviewer instructions, interviewers find this difficult to do. A new data registration system, at least upcoming in the Netherlands, and more extensive interviewer instructions could solve this problem. A second problem, also related to the sampling procedure, is the rise of combination institutions that provide managed care for elderly as well as nursing home residents, thus making a separate approaching strategy more complicated. A third problem is the lower percentage of residents of elderly homes able to take part in the pilot study themselves. Together with the second problem, this leads us to believe that apart from the sampling procedure, the approaching strategies for elderly and nursing homes should be identical. We feel there is no way around using proxy respondents when surveying residents of elderly and nursing homes. However, allowing proxy respondents creates the problem of more "don't know" answers, in particular to questions about aspects of caring and social contacts.

We nevertheless believe that most of the quality indicators that are used are positive. We consequently conclude that it is certainly possible to conduct survey research among residents of elderly and nursing homes. This is the answer to our first research question. To do so, however, adjustments need to be made in the

approaching strategy (e.g., two-stage sampling plan) and questionnaire (e.g., fewer answer categories). It takes our interviewers an average of 11.4 hours to collect five responses at elderly homes and nine hours at nursing homes, which is less than expected.

We find significant differences on key survey estimates between the residents of elderly and nursing homes and their peers living in private households. We find substantial differences on survey variables about physical limitations, helping devices and general health. This indicates that including residents of elderly and nursing homes would affect survey characteristics. This is our answer to the second research question.

In sum, this pilot study shows that with certain adjustments, it is possible to conduct survey research among residents of elderly and nursing homes. We show that this population is systematically different and needs to be observed, in particular if statements are to be made about the elderly.

APPENDIX

Table 7.9
Comparison of Survey Estimates on Private and Non-private Household Residents

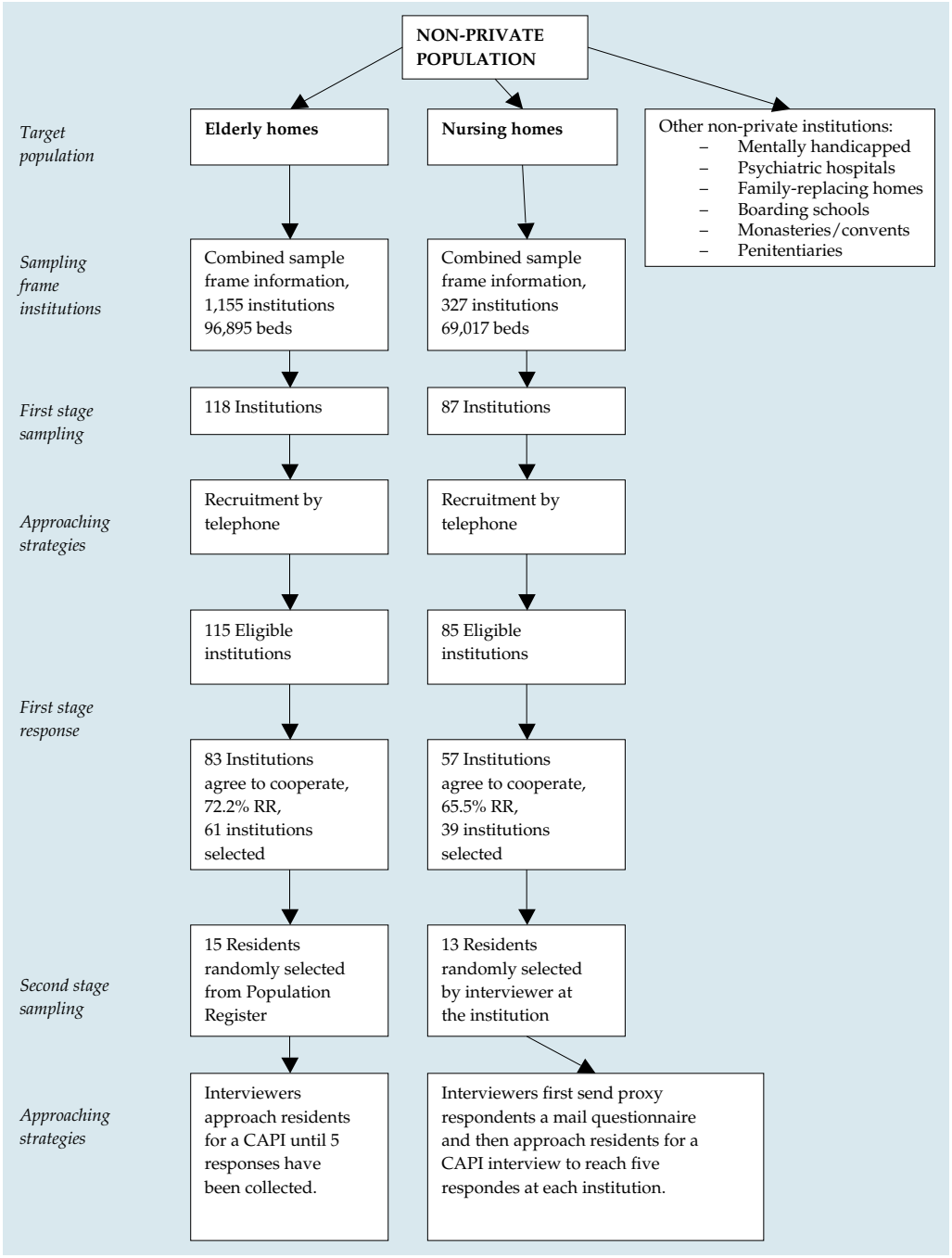
Physical limitations	Private pop. > 75	Elderly homes	Nursing homes	Residents	Proxy
Can you follow a conversation in a group of three people or more (if necessary with a hearing aid)?	1.55	1.73**	2.73**	1.74**	3.14**
Is your eyesight good enough to read the small letters in the newspaper (if necessary with glasses or contacts)?	1.45	2.02**	2.75**	2.09**	2.93**
Can you recognize someone's face from four metres away (if necessary with glasses or contacts)?	1.19	1.56**	1.87**	1.54**	2.05**
Can you bite and chew hard food, e.g. an apple?	1.77	2.04**	2.49**	2.04**	2.69**
Can you carry an object of five kilogrammes, e.g. a full shopping bag, for 10 metres?	2.00	2.99**	3.58**	3.05**	3.68**
If you are standing, can you bend down and pick up something from the ground?	1.62	2.52**	3.41**	2.68**	3.45**
Can you walk 400 metres without standing still (if necessary with a cane)?	1.82	2.70**	3.53**	2.84**	3.59**
Coding, 1 = without difficulty, 2 = with some difficulty, 3 = with great difficulty, 4 = cannot do this. * = $p < .05$, ** = $p < .01$, ns = not significant.					
Helping devices					
Do you have:					
A cane, crutch, walking frame or walker?	1.67	1.12**	1.58**	1.21**	1.59*
How often do you use it?	1.45	1.23**	1.56 ns	1.26**	1.57 ns
A wheelchair (electronically or manually run) or scootmobile?	1.95	1.66**	1.33**	1.58**	1.36**
How often do you use it?	1.58	1.49 ns	1.21**	1.35**	1.26**
Orthopedic footwear?	1.92	1.85**	1.80**	1.83**	1.83**
How often do you use it?	1.18	1.07*	1.72**	1.08*	1.98**
An arm or leg prosthesis?	1.98	1.98 ns	1.99 ns	1.98 ns	1.99 ns
How often do you use it?	1.41	1.29 ns	2.75**	1.38 ns	2.78**
An orthesis (brace or splint, not a tooth brace)?	1.98	1.96 ns	1.95*	1.96 ns	1.95 ns
How often do you use it?	1.32	1.45 ns	2.42**	1.23 ns	2.58**
Incontinence material?	1.86	1.47**	1.20**	1.46**	1.11**
How often do you use it?	1.28	1.17**	1.06**	1.17**	1.03**
A urinal or catheter?	1.99	1.94**	1.90**	1.92**	1.92**
How often do you use it?	1.47	1.12**	2.05**	1.11**	2.32**
Stoma or stoma material for urine or faeces?	1.99	1.99 ns	1.97*	1.99 ns	1.97 ns
Coding, 1 = yes, 2= no and 1= always, 2= sometimes, 3 = never. * = $p < .05$, ** = $p < .01$, ns = not significant.					

Table 7.9 (cont'd)

Physical limitations	Private pop. > 75	Elderly homes	Nursing homes	Residents	Proxy
Health					
To what extent has pain kept you from performing daily activities in the last four weeks?	2.00	1.58**	1.58**	1.60**	1.52**
Have you ever had a stroke, brain haemorrhage or brain infarct?	1.90	1.77**	1.63**	1.74**	1.64**
If so, was it in the last 12 months?	1.79	1.86 ns	1.76 ns	1.83 ns	1.76 ns
Have you ever had a heart infarct?	1.87	1.87 ns	1.87 ns	1.86 ns	1.88 ns
If so, was it in the last 12 months?	1.87	1.95*	1.93 ns	1.92 ns	1.95*
Have you ever had cancer (malignant disorder)?	1.83	1.85 ns	1.84 ns	1.85 ns	1.84 ns
If so, was it in the last 12 months?	1.71	1.73 ns	1.84**	1.73 ns	1.87**

Coding, 1 = yes, 2 = no. * = $p < .05$, ** = $p < .01$, ns = not significant.

Figure 7.2 Overview of study design



References

Allison, P. A. (2001). *Missing Data.* Sage University Papers on Quantitative Applications in the Social Sciences, 07-136. Thousand Oaks, CA: Sage.

Allum, P. A. (1998). *State and Society in Western Europe.* Oxford, UK: Blackwell.

Alwin, D.F. (1989). Problems in the Estimation and Interpretation of the Reliability of Survey Data. *Quality and Quantity,* 23, 277–331.

American Association for Public Opinion Research (2000). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys.* Ann Arbor, MI: American Association for Public Opinion Research.

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

Andreß, H.J., Hagenaars, J.A., and Kühnel, S. (1997). *Analyse von Tabellen und kategorialen Daten.* Berlin: Springer.

Andrews, F.M. and Herzog, A.R. (1986). The Quality of Survey Data as Related to Age of Respondent. *Journal of the American Statistical Association,* 81, 403–410.

Bates, N.A. (2004, May). *Contact Histories: A Tool for Understanding Attrition in Panel Surveys.* Paper presented at the 59th Annual AAPOR Conference, Phoenix, AZ.

Beebe, T.M., Davern, M.E., McAlpine, D.D., Call, K.T., and Todd, H. (2005). Increasing Response Rates in a Survey of Medicaid Enrollees; The Effect of a Prepaid Monetary Incentive and Mixed Modes (Mail and Telephone). *Medical Care,* 43, 4, 411–414.

Bentler, P. M. (1990). Comparative Fit Indexes in Structural Models. *Psychological Bulletin,* 107, 2, 238–246.

Berger, F. (2006). Effects of Incentives on Mail Survey Responses: A Review of Research Literature. *ZUMA-Nachrichten,* 58, 81–100.

Berk, M.L., Mathiowetz, N.A., Ward, E.P., and White, A.W. (1987). The Effect of Prepaid and Promised Incentives: Results of a Controlled Experiment. *Journal of Official Statistics,* 3, 4, 449–457.

Berlin, M., Mohadjer, L., Waksberg, J., Kolstad, A., Kirsch, I., Rock, D., and Yamamoto, K. (1992) An Experiment in Monetary Incentives. *Proceedings of the Survey Research Section of the American Statistical Association,* 393–398.

Bethlehem, J. (2004). *Technieken voor Survey-onderzoek.* Third Edition. Voorburg/ Heerlen: Statistics Netherlands.

Bethlehem, J. and Schouten, B. (2004). Nonresponse Analysis of the Integrated Survey on Living Conditions. Discussion Paper 04004. Voorburg/ Heerlen: Statistics Netherlands.

Bie, S. de, and Luiten, A. (2005, August). *Lifting Response Rates; Part Two.* Paper presented at the 16th International Workshop on Household Survey Nonresponse, Tällberg, Sweden.

Biemer, P.B. and Lyberg, L.E. (2003). *Introduction to Survey Quality.* Hoboken: John Wiley and Sons.

Blohm, M., and Diehl, C. (2001). Wenn Migranten Migranten Befragen. Zum Teilnahmeverhalten von Einwanderern bei Bevölkerungsbefragungen. *Zeitschrift für Soziologie,* 30, 3, 223–42.

Blohm, M., Harkness, J., Klein, S., and Scholz, E. (2003). Konzeption und Durchführung der "Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften" (ALLBUS) 2002. *ZUMA Methodenbericht,* December 2003.

Bollen, K.A. (1989). *Structural Equations with Latent Variables.* New York: Wiley.

Botwinick, J. (1978). *Aging and Behavior* (2nd ed.). New York: Springer.

Bradburn, N.M., Rips, L.J., and Shevell, S.K. (1987). Answering autobiographical Questions: The Impact of Memory and Inference on Surveys. *Science,* 236, 157–161.

Brakel, J. van den, Vis-Visschers, R., and Schmeets, J.J.G. (2006). Effects of Data Collection Modes and Incentives on the Outcomes of the Dutch Family and Fertility Survey for Young Moroccans and Turks. *Field Methods,* 18, 3, 321–334.

Bronner, A.E. (1988). Surveying Ethnic Minorities. In Saris, W.E. and Gallhofer, I.N. (Eds.), *Sociometric Research* (pp. 36–47). London: MacMillan.

Browne, M.W. and Cudeck, R. (1993). Alternative Ways of Assessing Model Fit. In Long, J.S. (Ed.), *Testing Structural Equation Models* (pp. 136–62). Newbury Park, CA: Sage.

Centraal Bureau voor de Statistiek (2004). *Allochtonen in Nederland 2003.* Voorburg/Heerlen, the Netherlands: Centraal Bureau voor de Statistiek.

Centraal Bureau voor de Statistiek (2005). *Enquêteonderzoek onder Allochtonen. Problemen en Oplossingen.* Voorburg/Heerlen, the Netherlands: Centraal Bureau voor de Statistiek.

Church, A.H. (1993). Estimating the Effect of Incentives on Mail Survey Response Rates: A Meta-analysis. *Public Opinion Quarterly,* 57, 1, 62–79.

Cialdini, R.B. (1993). *Influence: Science and Practice.* New York: HarperCollins.

Clemente, F. and Kleiman, M.B. (1976). Fear of Crime among the Aged. *Gerontologist,* 16, 207–210.

Colsher, P.L. and Wallace, R.B. (1989). Data Quality and Age, Health and Psychobehavioral Correlates of Item Nonresponse and Inconsistent Responses. *Journal of Gerontology, Psychological Sciences,* 44, 45–22.

Corso, J.F. (1977). Auditory Perception and Communication. In Birren, J.E. and Schaie, K.W. (Eds.), *Handbook of the Psychology of Aging* (pp. 384–420). New York: Van Nostrand Reinhold.

Couper, M.P. and de Leeuw, E.D. (2003). Nonresponse in Cross-Cultural and Cross-National Surveys. In Harkness, J.A., Vijver, F.J.R. v.d., and Mohler, P.P. (Eds.), *Cross-Cultural Survey Methods* (pp. 157–178). New York: Wiley.

Creswell, J.W. (1998). *Qualitative Inquiry and Research Design: Choosing among five Traditions.* Thousand Oaks, CA: Sage.

Dalenius, T. (1985). *Elements of Survey Sampling.* Stockholm, Sweden: Swedish Agency for Research Cooperation with Developing Countries.

Davies, H.T.O., Nutley, S.M., and Smith, P.C. (2000). *What Works? Evidence-based Policy and Practice in Public Surveys.* Bristol: The Policy Press.

Deutsche Forschungsgemeinschaft (1999). *Qualitätskriterien der Umfrageforschung.* Berlin.

Dillman, D.A. (2000). *Mail and Internet Surveys.* New York: Wiley.

Dodd, T. (1998). Incentive Payments on Social Surveys: A Summary of Recent Research. *Survey Methodology Bulletin,* 43, 23–27.

Dotinga, A., van den Eijnden, R.J.J.M., Bosveld, W., and Garretsen, H.F.L. (2005). The Effect of Data Collection Mode and Ethnicity of Interviewer on Response Rates and Alcohol Reports: An experimental Study. *Alcohol & Alcoholism* 40, 3, 242–48.

Dumas, L. and Théroux, J. (2004, November). *A Successful Collection sometimes Requires Convincing People by Informing them Properly.* Paper presented at Statistics Canada Symposium 2004, Innovative Methods for Surveying Difficult-to-Reach Populations, Ottawa, Canada.

The Economist (2007). The Importance of Statistics. March 1st 2007, from http://www.economist.com

Eid, M., Langeheine, R., and Diener, E. (2003). Comparing Typological Structures across Cultures by Multigroup Latent Class Analysis. *Journal of Cross-Cultural Psychology,* 34, 2, 195–210.

Eisner, M. and Ribeaud, D. (2007). Conducting a Criminological Research in Culturally Diverse Contexts. *European Journal of Criminology,* 4, 3, 271–298.

European Commission, Employment and Social Affairs DG. (2004). *Overview of third Level Indicators Used in the NAP Incl. relating to the social Inclusion of "Foreigners, Immigrants and Ethnic Minorities.* Brussels: European Commission.

Eurostat (2000). *Assessment of the Quality in Statistics.* Eurostat/A4/Quality/00/General/Standard, April 4–5, Luxembourg.

Evans, G. and Mills, C. (1998). Identifying Class Structure. A Latent Class Analysis of the Criterion-Related and Construct Validity of the Goldthorpe Class Schema. *European Sociological Review,* 14, 87–106.

Feskens, R.C.W., Hox, J.J., Lensvelt-Mulders, G.J.L.M., and Schmeets, J.J.G. (2004). *A Multivariate Analysis of Nonresponse among Ethnic Minorities.* Statistics Canada International Symposium Series—Proceedings, 2004. Innovative Methods for Surveying Difficult-to-Reach Population. Ottawa: Statistics Canada.

Feskens, R.C.W., Hox, J.J., Lensvelt-Mulders, G.J.L.M., and Schmeets, J.J.G. (2006). Collecting Data among Ethnic Minorities in an International Perspective. *Field Methods,* 18, 3, 284–304.

Feskens, R.C.W., Hox, J.J., Lensvelt-Mulders, G.J.L.M., and Schmeets, J.J.G. (2007). Nonresponse among Ethnic Minorities: A Multivariate Analysis. *Journal of Official Statistics,* 23, 3, 387–408.

Feskens, R.C.W., Lensvelt-Mulders, G.J.L.M., Beukenhorst, D., Kockelkoren, S., and Wetzels, W. (2008). *Studying the Non-Private Population: Results of a Large Pilot Study in the Netherlands.* Manuscript submitted for publication.

Gatward, R., Lound, C., and Bowman, J. (2002). *A Pilot Survey of People Living in Communal Establishments.* Technical Report, Office of National Statistics, Labor Market Trends, 2002, 141–149.

Goodman, L.A. (1974a). Exploratory Latent Structure Analysis Using Both Identifiable and Unidentifiable Models. *Biometrika,* 61, 215–231.

Goodman, L.A. (1974b). The Analysis of Systems of Qualitative Variables When Some of the Variables are Unobservable. Part 1: A Modified Latent Structure Approach. *American Journal of Sociology,* 79, 1179–1259.

Goodman, L.A. (2002). Latent Class Analysis: The Empirical Study of Latent Types, Latent Variables and Latent Structures. In Hagenaars, J.A. and McCutcheon, A.L. (Eds.), *Applied Latent Class Analysis* (pp. 3–56). Cambridge: Cambridge University Press.

Goor, H. van, Jansma, F., and Veenstra, R. (2005). Differences in Undercoverage and Nonresponse between City Neighbourhoods in a Telephone Survey. *Psychological Reports,* 96, 867–878.

Goyder, J. (1985). Face-to-Face Interviews and Mailed Questionnaires: The Net Difference in Response Rate. *Public Opinion Quarterly,* 49, 234–252.

Goyder, J., Lock J., and McNair, T. (1992). Urbanization Effects on Survey Nonresponse: A Test Within and Across Cities. *Quality and Quantity,* 26, 39–48.

Groves, R.M. (1989). *Survey Errors and Survey Costs.* New York: Wiley.

Groves, R.M. (2006). Nonresponse Rates and Nonresponse Bias in Household Surveys. *Public Opinion Quarterly,* 70, 5, 646–675.

Groves, R.M. and Couper, M.P. (1998). *Nonresponse in Household Interview Survey.* New York: Wiley.

Groves, R.M., Singer, E., and Corning, A. (2000). Leverage-Salience Theory of Survey Participation: Description and an Illustration. *Public Opinion Quarterly,* 64, 299–308.

Groves, R.M., Fowler, Jr., F.J., Couper, M.P., Lepkowski, J.M., Singer, E., and Tourangeau, R. (2004). *Survey Methodology.* Hoboken: John Wiley and Sons.

Gruber, S. (1997). *The German Mikrozensus.* Wiesbaden, Germany: Statistisches Bundesamt.

Gurland, B. D. L., Cross, P., and Golden, R. (1980). The Epidemiology of Depression and Dementia in the Elderly: The Use of Multiple Indicators of these Conditions. In Barret J.E. and Cole, J.O. (Eds.), *Psychopathology in the Aged.* New York: Raven.

Harkness, J.A., v.d., Vijver, F.J.R., and Mohler, P.P. (Eds.). (2003). *Cross-Cultural Survey Methods.* New York: John Wiley.

Havlik, R.J. (1986). *Aging in the Eighties: Impaired Senses for Sound and Light in Persons Aged 65 Years and Over.* Advance Data from Vital and Health Statistics, No. 125 DHHS Pub. No. (PHS) 86–1250. Hyattsville, MD, Public Health Service.

Havlik, R.J., Liu, B.M., and Kovar, M.G. (1987). *Health Statistics on Older Persons, United States, 1986.* Vital and Health Statistics, Ser. 3, No. 25. DHHS Pub. No. (PHS) 86–1250. Hyattsville, MD, Public Health Service.

Heer, W. de, (1999). International Response Trends: Results of an International Survey. *Journal of Official Statistics,* 15, 129–142.

Heer, W. de, and de Leeuw, E.D. (2002). Trends in Household Survey Nonresponse: A Longitudinal and International Comparison, In Groves, R.M., Dillman, D.A., Eltinge, J.L., and Little, R.J.A. (Eds.), *Survey Nonresponse* (pp. 41–54). New York: John Wiley.

Herzog, A.R. and Rodgers, W.L. (1982). *Survey of Older Americans: Some Methodological Investigations.* Final Report to the National Institute of Aging, MI. Institute for Social Research.

Horvitz, D.G. and Thompson, D.J. (1952). A Generalization of Sampling without Replacement from a finite Universe. *Journal of American Statistical Association,* 47, 663–685.

Houbiers, M. (2004). Towards a Social Statistical Database and Unified Estimates at Statistics Netherlands. *Journal of Official Statistics,* 20, 55–75.

Hox, J.J. and de Leeuw, E.D. (1998). Nonrespons in Surveys: Een Overzicht. *Kwantitatieve Methoden,* 19, 31–53.

Hox, J.J., de Leeuw, E.D., and Snijkers, G. (2003). Nonresponse in Surveys: Determining the Research Agenda for the Future. *Bulletin de Méthodologie Sociologique,* 79, 47–60.

Hox, J.J., de Leeuw, E.D., and Dillman, D.A. (2007). The Cornerstones of Survey Research. In de Leeuw, E.D., Hox, J.J., and Dillman, D.A. (Eds.), *International Handbook of Survey Methodology* (pp. 1–17). Mahwah: Lawrence Erlbaum Associates.

James, J.M. and Bolstein, R. (1990). The Effect of Monetary Incentives and Follow-up Mailings on the Response Rate and Response Quality in Mail Surveys. *Public Opinion Quarterly,* 54, 346–361.

James, J.M. and Bolstein, R. (1992). Large Monetary Incentives and their Effect on Mail Survey Response Rates. *Public Opinion Quarterly,* 56, 442–453.

Jong, A. de, and Hilderink, H. (2004). Lange Termijn Allochtonenscenario's voor Nederland. *Bevolkingstrends,* 52, 77–82.

Jong-Gierveld, J. de, and Kamphuis, F. (1985). The Development of a Rasch-type Loneliness Scale. *Applied Psychological Measurement,* 9, 3, 289–299.

Kaldenberg, D.O., Koenig, H.F., and Becker, B.W. (1994). Mail Survey Response Rate Patterns in a Population of the Elderly: Does Response Deteriorate with Age? *Public Opinion Quarterly,* 58, 68–76.

Keeter, S., Miller, C., Kohut, A., Groves, R.M., and Presser, S. (2002). Consequences of Reducing Nonresponse in a National Telephone Survey. *Public Opinion Quarterly,* 64, 125–148.

Klerk, M. de, (2001). *Survey on the Elderly in Institutions: Issues, Survey Structure and Findings.* The Hague: Social and Cultural Planning Office of the Netherlands.

Kline, R.B. (1998). *Principles and Practice of Structural Equation Modeling.* New York: Guilford Press.

Kline, R.B. (2005). *Principles and Practice of Structural Equation Modeling.* New York: Guilford Press.

Knäuper, B., Belli, R.F., Hill, D.H., and Herzog, A.R. (1997). Question Difficulty and Respondents' Cognitive Ability: The Effect on Data Quality. *Journal of Official Statistics,* 13, 181–199.

Kohout, F.J. (1992). The Pragmatics of Survey Field Work among the Elderly. In Wallace, R.B. and Woolson, R.F. (Eds.), *The Epidemiologic Study of the Elderly* (pp. 91–119). New York: Oxford University Press.

Krosnick, J.A. (1991). Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys. *Applied Cognitive Psychology,* 5, 213–236.

Laurie, H., Smith, R., and Scott, L. (1999). Strategies for Reducing Nonresponse in a Longitudinal Panel Survey. *Journal of Official Statistics,* 15, 2, 269–282.

Lavrakas, P.J. (1993). *Telephone Survey Methods: Sampling, Selection and Supervision* (2nd ed.). Newbury Park, CA: Sage.

Lazarsfeld, P.F. (1950a). The Logical and Mathematical Foundation of Latent Structure Analysis. In Stouffer, S. (Ed.), *Measurement and Prediction.* Princeton: Princeton University Press.

Lazarsfeld, P.F. (1950b). The Interpretation and Mathematical Foundation of Latent Structure Analysis. In Stouffer, S. (Ed.), *Measurement and Prediction,* Princeton: Princeton University Press.

Lazarsfeld, P.F. and Henry, N.W. (1968). *Latent Structure Analysis.* Boston: Houghton Mifflin.

Lee, S., Mathiowetz, N.A., and Tourangeau, R. (2004). Perceptions of Disability: The Effect of Self- and Proxy Response. *Journal of Official Statistics,* 20, 4, 671–686.

Leeuw, E.D. de, Hox, J.J., and Huisman, M. (2003). Prevention and Treatment of Item Nonresponse. *Journal of Official Statistics,* 19, 2, 153–176.

Little, R.J.A. and Rubin, D. B. (1987). *Statistical Analysis with Missing Data.* New York: Wiley.

Lynn, P., Clarke, P., Martin, J., and Sturgis, P. (2002). The Effects of Extended Interviewer Attempts on Non-response Bias. In Groves, R.M., Dillman, D.A., Eltinge, J.L., and Little, R.J.A. (Eds.), *Survey Nonresponse* (pp. 135–48). New York: John Wiley.

Magidson, J. and Vermunt, J.K. (2004). Latent Class Models. In Kaplan, D. (Ed.), *The Sage Handbook of Quantitative Methodology for the Social Sciences* (pp. 175–198). Thousand Oaks: Sage Publications.

Magidson, J. (2003, August). *Using Latent Class and CHAID Models to Develop Meaningful Segments from Survey Data.* Paper presented at the annual meeting of the American Association for Public Opinion Research, Sheraton Music City. Nashville.

Mack, S., Huggins, V., Keathley, D., and Sundukchi, M. (1998). Do Monetary Incentives Improve Response Rates in the Survey of Income and Program Participation? *Proceedings of the Section on Survey Methodology, American Statistical Association,* 529–534.

McCutcheon, A.L. (1987). *Latent Class Analysis.* Newbury Park: Sage.

Merkle, D.M. and Edelman, M. (2002). Nonresponse in Exit Poll: A Comprehensive Analysis. In Groves, R.M., Dillman, D.A., Eltinge, J.L., and Little, R.J.A. (Eds.), *Survey Nonresponse* (pp. 243–258). New York: John Wiley.

Muthén, B.O. (1998–2004). *Mplus Technical Appendices,* Los Angeles, CA: Muthén and Muthén.

Muthén, L.K. and Muthén, B.O. (1998–2004). *Mplus Users's Guide* (3rd ed.). Los Angeles, CA: Muthén & Muthén.

Nicolaas, G., and Stratford, N. (2005). A Plea for the Tailored Use of Respondent Incentives. In: C. van Dijkum, J. Blasius, and C. Durand (Eds.), *Proceedings of Sixth International Conference on Social Science Methodology: Recent Developments and Applications in Social Research Methodology* (pp. 1–15). Amsterdam: Budrich Verlag.

Office for National Statistics (2003). *Ethnic Group Statistics: A Guide for the Collection and Classification of Ethnicity Data.* Newport, VA: Office for National Statistics.

Perlmutter, M. (1986). A Life Span View of Memory. In Batles, P., Featherman, D., and Brim, P. (Eds.), *Advances in Life Span Development and Behavior,* Vol. 6. Hillsdale, NJ: Erlbaum.

Pfeffer, R.I., Afifi, A.A., and Chance, J. M. (1987). Prevalence of Alzheimer's Disease in a Retirement Community. *American Journal of Epidemiology,* 125, 420–436.

Poon, L.W. (1985). Differences in Human Memory with Aging: Nature, Causes and Clinical Implications. In Birren, J.E. and Schaie, K.W. (Eds.), *Handbook of the Psychology of Aging* (2nd ed.) (pp. 427–462). New York: Van Nostrand Reinhold.

Reep, C. (2003). *Moeilijk Waarneembare Groepen. Een Inventarisatie.* Statistics Netherlands report H1568-03-SOO, Heerlen: Statistics Netherlands.

Riele, S. te, (2002). Vertekening door Non-Respons. Hoe Nauwkeurig Zijn de Uitkomsten van Persoonsenquêtes? *Sociaal-economische Maandstatistiek,* 19, 4, 20–25.

Rodgers, W.L., Herzog, A.R., and Andrews, F.M. (1988). Interviewing Older Adults, Validity of Self-Reports of Satisfaction. *Psychology and Aging,* 3, 264–272.

Rosenbaum, P.R. and Rubin, D.B. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika,* 70, 41–55.

Rosenwaike, I. (1985). *The Extreme Aged in America: A Portrait of an Expanding Population.* Westport, CT: Greenwood Press.

Schmeets, J.J.G. (2004). Respons van Eerste Generatie Allochtonen in het POLS. *Bevolkingstrends,* 52, 4, 92–97.

Schmeets, J.J.G. (2005a). De Leefsituatie van Allochtonen. In Schmeets, J.J.G. and Bie, R. van der, (Eds.), *Enquêteonderzoek onder Alloctonen. Problemen en Oplossingen.* (pp. 169–176). Heerlen: Statistics Netherlands.

Schmeets, J.J.G. (2005b). Slecht bereikbare Allochtonen, autochtone Weigeraars. In Schmeets, J.J.G. and Bie, R. van der, (Eds.), *Enquêteonderzoek onder Alloctonen. Problemen en Oplossingen.* (pp. 81–88). Heerlen: Statistics Netherlands.

Schmeets, J.J.G. and Geurts, J.J.M. (1990). Deelname aan Maatschappelijke en Huishoudelijke Activiteiten door Ouderen: Een sociaal-economisch Verklaringsmodel. *Tijdschrift voor Gerontologie en Geriatrie,* 21, 249–257.

Schmeets, J.J.G. and Michiels, J. (2003). Het Effect van Non-respons onder Allochtonen. *Bevolkingstrends,* 51, 52–57.

Schmeets, J.J.G., Reep, C., and Snijkers, G. (2003). Afbakenen van Moeilijk Waarneembare Bevolkingsgroepen. *Bevolkingstrends,* 51, 62–68.

Schneekloth, U. and Leeven, I. (2003). Woran Bemisst sich eine „Gute" Allgemeine Bevölkerungsumfrage? Analysen zu Ausmaß, Bedeutung und zu den Hintergründen von Nonresponse in zufallsbasierten Stichprobenerhebungen am Beispiel des ALLBUS. *ZUMA-Nachrichten,* 53, 27, 16–57.

Schnell, R. (1997). *Nonresponse in Bevölkerungsumfragen. Ausmaß, Entwicklung und Ursachen.* Opladen: Leske und Budrich.

Schoua-Glusberg, A. and Miller, K. (2004, November). *Assessing Comprehension of Translated Questionnaires with qualitative Methods.* Paper presented at Statistics Canada Symposium 2004, Innovative Methods for Surveying Difficult-to-Reach Populations, Ottawa, Canada.

Schouten, B. (2003). *De verschillende Typen Non-respons in POLS 1998.* CBS-report 2319-03-TMO, Heerlen: Statistics Netherlands.

Schouten, B. and Cobben, F. (2007). *R-indexes for the Comparison of Different Fieldwork Strategies and Data Collection Modes.* Discussion Paper 07002, Voorburg/ Heerlen: Statistics Netherlands

Seifert, W. (1992). Die zweite Ausländergeneration in der Bundesrepublik: Längsschnittbeochbatungen in der Berufseinstiegsphase. *Kölner Zeitschrift für Soziologie und Sozialpsychologie,* 44, 4, 677–696.

Simmons, E. and Wilmot, A. (2004). Incentive Payments on Social Surveys: A Literature Review. *Survey Methodology Bulletin,* 53, 1–11.

Singer, E. (2002). The Use of Incentives to Reduce Nonresponse in Household Surveys. In Groves R.M., Dillman D.A., Eltinge J.L., and Little R.J.A. (Eds.), *Survey Nonresponse* (pp. 163–178). New York: Wiley.

Singer, E., Van Hoewyk, J., Gebler, N., Raghunathan, T., and McGonagle, K. (1999). The Effect of Incentives on Response Rates in Interviewer-Mediated Surveys. *Journal of Official Statistics,* 15, 217–230.

Singer, E., van Hoewyk, J., and Maher, M.P. (2000). Experiments with Incentives in Telephone Surveys. *Public Opinion Quarterly,* 64, 171–188.

Snijkers, G. (2003). *Programmaplan Strategisch Onderzoek 'Non-respons en Moeilijk Waarneembare Groepen': Naar een optimale Meetstrategie.* CBS-report H3318-03-SOO, Heerlen: Statistics Netherlands.

Snijkers, G. and Kockelkoren. S. (2004). *De Gestandaardiseerde Benaderingsstrategie 2003: Evaluatie en Aanbevelingen.* CBS-report H415-04-SOO, Heerlen: Statistics Netherlands.

Steeh, C., Kirgis, N., Cannon, B., and DeWitt, J. (2001). Are they really as Bad as they Seem? Nonresponse Rates at the End of the Twentieth Century. *Journal of Official Statistics,* 17, 227–247.

Stoop, I.A.L. (2004). Surveying Nonrespondents. *Field Methods,* 16, 1, 23–54.

Stoop, I.A.L. (2005). *The Hunt for the Last Respondent: Nonresponse in Sample Surveys.* The Hague: Social and Cultural Planning Office of the Netherlands.

Teisl, M.F., Roe, B., and Vayda, M.E. (2006). Incentive Effects on Response Rates, Data Quality and Survey Administration Costs. *International Journal of Public Opinion Research,* 18, 3, 364–373.

Thornberry, O.T. Jr. and Massey, J.T. (1988). Trends in United States Telephone Coverage across Time and Subgroups. In Groves R.M., Biemer P.P., Lyberg L.E., et al. (Eds.), *Telephone Survey Methodology* (pp. 25–49). New York: John Wiley and Sons.

Tilburg, T.G. van, and de Leeuw, E.D. (1991). Stability of Scale Quality under Various Data Collection Procedures: A Mode Comparison on the "De Jong-Gierveld Loneliness Scale". *International Journal of Public Opinion Research,* 3, 69–85.

Verba. S. (1996). The Citizen as Respondent: Sample Surveys and American Democracy. Presidential Address, *American Political Science Association,* 1995, American Political Science Review, 90, 1, 1–7.

Vermunt, J.K., and Magidson, J. (2002). Latent Class Cluster Analysis. In Hagenaars, J. A. and McCutcheon, A.L. (Eds.). *Applied Latent Class Analysis* (pp. 89–106). Cambridge University Press.

Vermunt, J.K. and Magidson, J. (2003). *Latent GOLD Choice User's Guide.* Boston: Statistical Innovations Inc.

Vermunt, J.K. and Magidson, J. (2004). Local Independence. In Lewis-Beck, M., Bryman, A. and. Liao, T.F. (Eds.), *The Sage Encyclopedia of Social Sciences Research Methods* (pp. 580-581). Thousand Oaks, CA: Sage Publications.

Wetzels, W. and Schmeets, J.J.G. (2006). *Tweede Adviesnota 'Experiment Postzegel Beloning'.* CBS-rapport SOO-2006-H104/intern, Heerlen: Statistics Netherlands.

Wetzels, W., Schmeets, J.J.G., Brakel van den, J., and Feskens, R.C.W. (in press). Impact of Prepaid Incentives in Face-to-Face Surveys: A Large-scale Experiment with Postage Stamps. *International Journal for Public Opinion Research.*

Wetzels, W., Kockelkoren, S., Beukenhorst, D., and Feskens, R.C.W. (2008). Onderzoek onder Bewoners van Verzorgings- en Verpleeghuizen, Pilot IIT 2007, CBS report. Voorburg/Heerlen: Statistics Netherlands.

Wittenboer, G. van den, Hox, J.J., and de Leeuw, E.D. (1997). Aberrant Response Patterns in Elderly Respondents: Latent Class Analysis of Respondent Scalability. In Rost, J. and Langeheine, R. (Eds.), *Applications of Latent Class Models in the Social Sciences* (pp. 155–162). Münster/New York: Waxmann.

Yu, J. and Cooper, H. (1983). A Quantitative Review of Research Design Effects on Response Rates to Questionnaires. *Journal of Marketing Research,* Vol. XX, 36–44.

Zafirovski, M. (2005). Social Exchange Theory under Scrutiny: A Positive Critique of its Economic-Behaviorist Formulations. *Electronic Journal of Sociology,* 2, 1–40.

Appendix

The standard definition for response rate according to the American Association for Public Opinion Research (2006) is given by:

$$\textbf{Response rate 2} = \frac{\textbf{(I + P)}}{\textbf{(I + P) + (R + NC + O) + (UH + UO)}}$$

$$\textbf{Contact rate 3} = \frac{\textbf{(I + P) + R + O}}{\textbf{(I + P) + R + O + NC}}$$

$$\textbf{Cooperation rate 2} = \frac{\textbf{(I + P)}}{\textbf{(I + P) + R + O}}$$

where I denotes the number of complete interviews; P the number of partial interviews; R the number of refusals and break offs; NC the number of noncontacts; O the number of other nonrespondents; UH number of sampling units with household or occupation unknown; UO unknown others.

Samenvatting

"Moeilijk waarneembare groepen" in enquêteonderzoek is het onderwerp van deze dissertatie. Het Centraal Bureau voor de Statistiek heeft vijf groepen als zodanig aangemerkt, waarbij allochtonen en mensen woonachtig in instituties, inrichtingen en tehuizen als belangrijkste en grootste groepen gelden (Reep 2003). Een van de redenen waarom een groep moeilijk in de waarneming kan zijn, zijn bovengemiddelde non-respons cijfers voor deze bepaalde groep. Non-respons doet zich voor wanneer een onderzoeksorganisatie er niet in slaagt informatie voor een onderzoekseenheid te verzamelen. Dit kan verschillende redenen hebben. Een onderzoeksorganisatie kan er bijvoorbeeld niet in slagen de persoon of een andere onderzoekseenheid te bereiken. Ook kan een verzoek tot deelname aan enquête-onderzoek geweigerd worden. Tenslotte kan het voorkomen dat een onderzoeks-eenheid niet in staat is tot deelname aan de enquête. Non-respons kan tot een aantal problemen leiden. Ten eerste leidt non-respons tot een kleiner aantal respondenten. Het verminderd aantal respondenten kan tot een verbreding van het betrouwbaarheidsinterval van schattingen leiden. Een onderzoeksorganisatie kan natuurlijk op deze situatie inspringen en bijvoorbeeld een grotere steekproef trekken of grotere inspanningen in het veldwerk ondernemen om tot het gewenste aantal respondenten te komen. Dit zal echter tot additionele kosten leiden, een tweede probleem van non-respons. Tevens kunnen hoge non-respons cijfers nadelig werken voor de reputatie van een onderzoeksbureau. Het belangrijkste probleem van non-respons is echter een eventuele vertekening in de enquête-resultaten. Hiermee wordt de validiteit van het onderzoek bedreigd, wat tot foute conclusies kan leiden. De vertekening veroorzaakt door non-respons is een functie van het non-respons cijfer en de mate waarin respondenten en non-respondenten in hun antwoorden van elkaar verschillen op een bepaalde enquêtevraag (Groves, 1989, Couper en De Leeuw 2003). Deze vertekening laat zich echter, in tegenstelling tot het non-respons cijfer, niet eenvoudig achterhalen. Immers, men weet over het algemeen niet wat non-respondenten zouden hebben geantwoord op enquêtevragen als ze zouden hebben deelgenomen aan het onderzoek. De relatie tussen non-respons cijfer en non-respons vertekening is niet eenduidig. Groves (2006) laat in een uitgebreide studie zien dat ook enquêtes met lage responscijfers zuivere schatters kunnen opleveren, en tegelijkertijd hoge responscijfers niet altijd een onvertekend beeld opleveren. Ondanks dit, blijft het van belang een hoog responscijfer na te streven. Het non-respons cijfer geeft immers wel informatie over de non-respons vertekening in zoverre dat het de bandbreedte van de non-respons vertekening weergeeft (DFG, 1999). Een van de groepen met bovengemiddelde non-responscijfers, zijn allochtonen. Deze situatie doet zich zowel voor in Nederland als in andere westerse landen (zie bijvoorbeeld Feskens et al. 2006).
Er kunnen ook nog andere redenen zijn dat een groep moeilijk in de enquête-waarneming is. Voor de bevolking woonachtig in instituties, inrichtingen en

tehuizen (de zogenaamde IIT bevolking of de bevolking woonachtig in niet-zelfstandige huishoudens) doen zich meerdere problemen voor. Ten eerste is er in Nederland (nog) niet de beschikbaarheid over een steekproefkader van dezelfde kwaliteit zoals voor de zelfstandige wonende bevolking. Belangrijker is echter dat deze groep mensen niet zoals de zelfstandige wonende bevolking direct benaderbaar is. Het dagelijkse leven van mensen woonachtig in instituties, inrichtingen en tehuizen wordt bedrijfsmatig door derden verzorgd. Dit betekent dat er een extra stap moet worden gemaakt om de bewoners zelf te kunnen enquêteren. Bovendien is het niet duidelijk of deze groep in staat zal zijn tot deelname aan enquêteonderzoek. Om deze redenen wordt de IIT bevolking in veel enquêteonderzoek uitgesloten uit de doelpopulatie en daarmee niet waargenomen in de enquêteresultaten.

Het tweede hoofdstuk van deze dissertatie is een verkennende studie naar groepen met vergelijkbare respons geneigdheden. Uit deze studie komen twee groepen naar voren met bovengemiddelde non-respons cijfers. Enerzijds zijn dat allochtonen en anderzijds personen woonachtig in kleine huishoudens in stedelijke gebieden.

Het derde en vierde hoofdstuk behandelt non-respons onder allochtonen. Allochtonen hebben bovengemiddelde non-respons cijfers. Deze hoge non-respons cijfers zijn echter meer het resultaat van andere factoren dan etniciteit zelf, zo komt naar voren in structurele vergelijkingsanalyses. In het bijzonder de stedelijke woonsituatie van allochtonen verklaart de hoge non-respons onder deze groep. Hoofdstuk vier is een internationale vergelijking van non-respons onder allochtonen. Hieruit blijkt dat vele landen andere afbakeningen gebruiken voor het vaststellen van iemands etniciteit. Non-respons onder allochtonen blijkt in alle landen bovengemiddeld te zijn. Het verhogen van het aantal contactpogingen lijkt een succesvolle strategie om non-respons onder allochtonen te verminderen.

Het vijfde hoofdstuk van deze dissertatie presenteert de resultaten van een experiment waarin bij de aanschrijfbrief aan de respondenten postzegels zijn bijgesloten. Dit blijkt een succesvolle maatregel te zijn om de respons te verhogen, vooral in stedelijke gebieden. Onder allochtonen blijkt deze beloning of incentive niet of nauwelijks effect op het respons cijfer te hebben. Deze conclusies worden verder ondersteund door de analyses over de effectiviteit van het inzetten van incentives op de respons onder allochtonen in hoofdstuk zes. Respons onder autochtonen die op andere kenmerken, zoals stedelijkheid en inkomen, lijken op allochtonen neemt fors toe bij het inzetten van postzegels, terwijl de respons onder niet-westerse allochtonen in dezelfde situatie vrijwel gelijk blijft.

In hoofdstuk zeven, tenslotte, worden de mogelijkheden om enquêteonderzoek uit te voeren onder bewoners van verzorgings- en verpleeghuizen besproken. Een pilot studie onder ruim 500 bewoners van verzorgings- en verpleeghuizen laat zien dat zowel hoge respons cijfers onder deze groep mogelijk zijn, als wel dat de kwaliteit van de antwoorden voldoende is. De studie laat bovendien zien dat de bewoners van verzorgings- en verpleeghuizen veelal andere kenmerken hebben dan leeftijdgenoten woonachtig in zelfstandige huishoudens.

Curriculum Vitae

Remco Feskens was born on April the first, 1978, in Breda, the Netherlands. After completing secondary school at the Onze Lieve Vrouwe Lyceem in 1996, he studied Political Sciences at the Radboud University Nijmegen (formerly known as Katholieke Universiteit Nijmegen). After graduation in 2002, he studied Statistics at the Humboldt and Freie Universität in Berlin, Germany. In 2003 he started to work at the Department of Methods and Statistics at Utrecht University to write his dissertation.