



Improving survey fieldwork with paradata



Centraal Bureau
voor de Statistiek

Annemieke Luiten

Improving

survey fieldwork

with paradata

Publisher

Statistics Netherlands
Henri Faasdreef 312, 2492 JP The Hague
www.cbs.nl

Prepress: Textcetera, The Hague and Grafimedia, The Hague
Printed by: Tuijtel, Hardinxveld-Giessendam
Design: Edenspiekermann

Information

Telephone +31 88 570 70 70, fax +31 70 337 59 94
Via contact form: www.cbs.nl/information

Where to order

verkoop@cbs.nl
Fax +31 45 570 62 68

ISBN 978-90-357-2113-5

© Annemieke Luiten, The Hague / Heerlen, 2013.

The author is fully responsible for the text of this dissertation. Reproduction is permitted for own or internal use. The text does not necessarily correspond with the official point of view of Statistics Netherlands.

Improving survey fieldwork with paradata

Verbeteren van enquêteveldwerk met behulp van paradata

(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. G.J. van der Zwaan, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op dinsdag 1 oktober 2013 des middags te 2.30 uur

door

Johanna Maria Mathilda Luiten
geboren op 15 februari 1957 te Delft

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

Acknowledgements

Twenty years ago, my husband inadvertently threw my – not quite finished – PhD dissertation and all the supporting material in the garbage, during the move to a new home¹⁾. A busy life with a new job and small children reconciled me with the inevitable for a time. When the opportunity arose, however, to write a dissertation on the subjects that became dear to me in my work for Statistics Netherlands, I jumped to it. The subjects of survey nonresponse and its implications, survey fieldwork, interviewers and their behaviour, all the topics of this dissertation, have in the past years involved me as a researcher, a practitioner and a manager. To get the time to fundamentally reflect on these issues, to research them and write about them, has been a privilege and an immensely rewarding and enriching experience, for which I thank my employer wholeheartedly.

None of the work in this thesis would have been possible without my colleagues. First and foremost are the interviewers with whom I have had numerous – sometimes heated, but always interesting and informative – discussions. The work of interviewers lies at the heart of a large part of social statistics. Their willingness to share their insights has been indispensable for a better understanding of the influence of interviewers on data quality. Likewise, the interviewers' regional managers have always been extremely helpful in interpreting why things went as they went.

Colleagues from the department of data collection helped me with the experiments described in this thesis, on top of their already overloaded daily work. Willem Wetzels and Sabine Kockelkoren have played an important part in the design and analysis of two of the experiments. Other colleagues, like Barry Schouten, Dirkjan Beukenhorst and Ger Snijkers helped with advice, or served as sounding boards for new ideas. Numerous colleagues kept me on track with their interest in and questions about the advancement of the work.

A special word for Hans Schmeets, colleague and co-promotor. Hans, thank you for your inspiration, and for your company over many a glass of wine. Your critical reading of my work and your eye for detail has contributed immensely.

Bas and Thomas, I am already immensely proud of you, but having you as my paranymphs during the defense of this dissertation will be the highlight of the day. Rob, I kept this one carefully away from you. And that worked; it is finished.

Annemieke Luiten
Maarheeze, August 2013

¹⁾ Yes, we are still married.

Contents

Acknowledgements 5

1. Introduction 10

2. Measures to enhance response to social surveys 18

2.1 Introduction 19

2.2 Methods to reduce noncontact in CAPI/CATI surveys 21

2.3 Noncontact in web and /or mail surveys 26

2.4 Methods to increase cooperation 27

2.5 The role of the interviewer 36

2.6 Cooperation in mail and web surveys 40

2.7 Interviewer training and monitoring 42

2.8 Conclusion 43

3. Personalisation in advance letters does not always increase response rates. Demographic correlates in a large scale experiment 44

3.1 Introduction 45

3.2 Method 48

3.3 Results 51

3.4 Summary and discussion 55

4. Tailored fieldwork design to increase representative household survey response. An experiment in the Survey of Consumer Sentiment 58

4.1 Introduction 59

4.2 Method 62

4.3 Results 67

4.4 Summary and discussion 77

5. Predicting contactability and cooperation in survey research 81

5.1 Introduction 82

5.2 Method 85

5.3 Results 88

5.4 Summary and discussion 101

6. Fieldwork strategy at Statistics Netherlands 105

- 6.1 Introduction **106**
- 6.2 One of the first two visits should be after 5 pm or on Saturday **109**
- 6.3 First visits should be in the first half of the fieldwork period **112**
- 6.4 Noncontacts should be visited six times **113**
- 6.5 Visits should be spread over time, days and weeks of the field-work period **115**
- 6.6 Summary and discussion **117**

7. Understanding factors leading to interviewers' non-compliance with fieldwork rules 119

- 7.1 Introduction **120**
- 7.2 Method **124**
- 7.3 Results **129**
- 7.4 Summary and discussion **135**

8. Do interviewers introduce bias by not following fieldwork strategy? The interplay between interviewer behaviour, sample unit characteristics and non-response bias 140

- 8.1 Introduction **141**
- 8.2 Method **145**
- 8.3 Results **147**
- 8.4 Summary and discussion **157**

9. Summary and discussion 160

- 9.1 Discussion **168**

10. Samenvatting (Summary in Dutch) 169

- 10.1 Maatregelen om de respons bij enquêteonderzoek te verhogen **171**
- 10.2 Het persoonlijk maken van aanschrijfbrieven verhoogt niet altijd de respons. Samenhang met demografische kenmerken in een grootschalig onderzoek **172**
- 10.3 Representatieve respons door gedifferentieerd veldwerk. Een experiment in het Consumenten Conjunctuur Onderzoek **173**
- 10.4 Het voorspellen van contact- en coöperatiegeneigdheid in enquêtes **175**
- 10.5 Veldwerkstrategie van het Centraal Bureau voor de Statistiek **176**
- 10.6 Waarom interviewers zich niet houden aan de regels van de veldwerkstrategie **177**

10.7	Veroorzaken interviewers vertekening door zich niet te houden aan de veldwerkstrategie? De samenhang tussen interviewergedrag, kenmerken van de steekproefeenheid en vertekening	178
10.8	Conclusie	179
	References	180
	Appendix 1	200
	Appendix 2	202
	Appendix 3	203

1.

Introduction

In 1955 Isaac Asimov wrote a story, published in the Science Fiction magazine 'If', about the presidential elections of 2008. Norman Muller, the most representative American, would that year choose the new president, and in one stroke determine the outcome of thousands of other elections. He would not actually choose, but his choice would be gauged from his reactions to a set of well chosen questions, like 'what do you think about the price of eggs?'

Asimov showed himself to be a visionary, with an early grasp of concepts that we are still grappling with almost 60 years later: what is a representative respondent, and what do we need to know of such a person to determine if indeed he is representative for others. Multivac, the kilometre long computer that pinpointed Norman Muller, must have had an impressive and enviable array of paradata at its disposal. It's a pity that Asimov does not reveal more about how Multivac went about this task. He does not tell us more about Norman than that he has a wife and daughter, a job as a salesman, and a blond, though greying moustache.

Although the 2008 presidential elections in the USA were special in a way Asimov did not fathom, we did not reach the point on the horizon he sketched. We still need considerably more than one respondent to give us reliable answers to our survey questions. But we have made progress. We do have a fair understanding of how representative our several thousand respondents are of the population, and can express that understanding in a figure, the R-indicator. We calculate representativeness by using paradata that are perhaps not yet as sophisticated as Multivac's, but the start is there. We have also made progress in the art of asking questions, although when we want to know whom the respondent would choose as president, we still use the crude method of asking him just that.

This thesis discusses some of the subjects sketched above. There are chapters on survey response, on representativeness and on fieldwork. In all chapters paradata are used to aid survey design, survey management and survey evaluation. Paradata, a term coined by (Couper, 1998), are data that are collected as a by-product of the data collection process (Durrant & Kreuter, 2013). Examples are data from interviewer call records or contact attempts (e.g., Durrant, D'Arrigo & Steele, 2013), keystroke files and audit trails (e.g., Snijkers & Morren, 2010), time stamps, reflecting the length of a question-answer sequence (e.g., Couper & Kreuter, 2013), vocal properties of the interviewer and respondent (e.g., Conrad et al., 2013) and interviewer observations about a sampled household or neighbourhood (e.g., West, 2013). In this manuscript we extend the definition to include data on sample persons or households that are collected in other ways than the present survey. Examples are data that are available in registries like the Communal registries.

Paradata are increasingly used by survey organizations during data collection to guide interventions, after data collection to improve next rounds, and during non-

response adjustment (Couper & Lyberg, 2005; Biemer, Chen & Wang, 2013; Kreuter, 2013).

When I started out on the endeavour of writing this thesis in 2006, the original aim was to do research on and write about non-response. Non-response is the bane of survey statisticians, and Statistics Netherlands had a few years earlier started a large scale research program called 'non-response and difficult groups in survey research', in collaboration with several universities. The aim was doing research in three areas: 'non-response adjustment', which resulted in a PhD thesis by Fannie Cobben (2009), 'difficult groups in survey research', resulting in a PhD thesis by Remco Feskens (2009) and non-response reduction. Up until that time, the consensus was that higher response rates are better response rates. By 2006 the understanding and views on the subject had changed, however. Keeter, Miller, Kohut, Groves & Presser (2000) and Curtin, Presser & Singer (2000) had shown that the increase in response through refusal conversion had no effect on the estimate of the target parameter and increasingly survey methodologists became aware of the dangers of 'blind pursuit of high response rates' (Groves, 2006).

Instead of concentrating on response rates, the focus of attention became the bias that is introduced, not only by nonresponse, but also by the measures to reduce nonresponse, like refusal conversion. Groves (2006) describes the mechanisms by which measures to stimulate response may introduce nonresponse bias. This may happen when persons with distinctive values on the survey variables are differentially sensitive to the measures to stimulate response. Groves speculates that advance letters may introduce bias in semiliterate populations for example. Incentives may likewise introduce bias if subgroups react differently to it. This is illustrated by Wetzels, Schmeets, Van den Brakel and Feskens (2008) in an incentive experiment that showed that some ethnic minority groups in the Netherlands did not react at all to the incentive, while other groups showed a response rate change of ten percentage points.

Nonresponse as such is not necessarily detrimental to the quality of a statistic. A meta-analysis by Groves and Peytcheva (2008) showed that some surveys have a high response rate but large bias, while others combine a low response rate with little bias. In a special issue on nonresponse bias in household surveys in the *Public Opinion Quarterly*, Singer (2006, p. 643) concluded 'We are at the beginning of a new period of research on nonresponse bias, one that will have large consequences for survey practitioners and survey methodologists'. She sketched how the new orientation would have implications for best practices in an era of low response rates.

It is hard to predict when nonresponse will lead to bias. The difference between respondents and nonrespondents is not always known, and may be different for each survey and even for each substantive variable within one survey. The amount of bias in a survey is a function of the difference between respondents

and nonrespondents in relation to the target variable, but also of the amount of nonresponse (Bethlehem, 2002). Therefore, in the absence of firm evidence of bias, it is prudent to strive for high response rates, but with a critical eye for the possibility of doing harm.

With this warning in mind, the first chapter in this thesis is on 'old school' nonresponse reduction. Chapter 2 gives a review of measures that can be employed in CAPI¹⁾, CATI and web surveys to increase the chance that sample persons participate in our survey. In a survey statistician's heaven, all persons who are invited to cooperate in a survey are located, and are able and willing to participate. Reality is otherwise. In a typical Statistics Netherlands social survey, about 65% of the sample participates. The other 35% are nonrespondents. For the largest part these are people who refuse to participate in the survey. Other people cannot be contacted, do not speak the language of the interview (Dutch) or are too ill to participate. Chapter 2 describes the measures that can be taken to increase the chance that persons are contacted, and subsequently cooperate with the survey request. Where possible, the measures are illustrated by Statistics Netherlands' experiences.

In chapter 3 an experiment is described with the advance letters that are sent to households prior to an interviewer's call or visit. The study was set up to examine whether personalising advance letters, by adding names and appropriate salutation, affects the survey cooperation of subgroups in the general population differently, in analogy to findings that subgroups react differently to advance letters. Differential reactions could be an explanation for the mixed findings in the literature on personalisation of advance or cover letters. In this study, paradata from communal registries made it possible to study if subgroups reacted differently to personalisation. Other paradata, from the 5% quality re-interview that Statistics Netherlands routinely performs for most surveys, showed whether the advance letter was read by more households when the household was addressed by name. The study focussed on groups that the literature pinpointed as possibly reacting differently to advance letters, i.e., different age, ethnic, gender, household composition and income groups, and groups with or without a listed telephone number. It was found that there was hardly any difference in the overall level of cooperation whether or not a personalised letter was used. Differential reactions were found for listed versus unlisted telephone owners, where only listed households reacted positively to personalisation. In the other subgroups studied, no firm evidence of differential reactions was found.

¹⁾ CAPI is the acronym of Computer Assisted Personal Interviewing, meaning that an interviewer visits a sample person at home or another place of his choice and interviews him with the help of a computer (or laptop or tablet). CATI is Computer Assisted Telephone Interviewing, where interviewers call the sample person on his landline or cell phone.

The positive reaction of listed telephone owners is one example where response stimulating measures can lead to bias. As a result of the intervention, the response in the group with a listed line, a group that is already over-represented in our survey results (Bethlehem & Schouten, 2004), increased, while the rates in the group without a landline decreased. It is not unimaginable that the sample composition worsened as a result of addressing households by name.

So, when stimulating response rates, the survey organisation needs to be guided by knowledge of the relation between the measures they take, which groups are sensitive to them, and what the influence is on survey estimates. That may necessitate a different approach for different groups. Survey designs where not all groups receive the same treatment are called adaptive or dynamic designs (Wagner, 2008), responsive designs (Groves & Heeringa, 2006) or tailored designs (Luiten & Schouten, 2013).

Chapter 4 describes a study in which a tailored survey design was used to obtain a more representative response in the Consumer Sentiments Survey, traditionally a CATI survey. Paradata from previous rounds of the survey and register information were used to identify groups that differed in contact and cooperation rates. In an experiment we varied which modes the sample persons received (paper questionnaire, web questionnaire or CATI), and in which combination. Contact chance was manipulated by timing of telephone calls and by prioritising calls to important groups. Chance of cooperation was manipulated by assigning the best interviewers to cases where we foresaw the most problems. We were quite accurate in predicting who would be hard to contact, but less so in who would cooperate. Nevertheless, the tailored fieldwork strategy successfully increased the representativeness of the resulting response, according to the representativeness- or R-indicator (Schouten, Cobben & Bethlehem, 2009).

The finding that we were not able to correctly predict who would cooperate with our surveys lead to the research described in chapter 5. Statistics Netherlands surveys with different sampling methods, topics, modes and lengths were linked to registry information and completed with paradata provided by interviewers. A large number of socio-demographic characteristics of persons and households thus became available. Again it became apparent that contactability is fairly predictable, but cooperation is not. The statistically significant relations between socio-demographic characteristics and survey cooperation that are often reported in survey research must be reconsidered and possibly re-interpreted as mostly spurious.

Statistics Netherlands has long been notorious for having the lowest response rates in the western world (De Leeuw & De Heer, 2002). However, the situation has improved markedly in recent years. Response rates in other countries drop as well, while ours have stabilized and have for some time even risen, with a respectable 72% for the Dutch Parliamentary Election Study in 2006 as a high point

(Schmeets, 2010). The circumstance that changed was that interviewers were employed in 2001, and from 2003 were given rules to adhere to in the field and were monitored on compliance to those rules. The fieldwork strategy has the triple purpose of securing a high contact rate as quickly as possible, with the lowest risk of bias introduced by interviewers going for 'low hanging fruit'. Chapter 6 gives an overview of the fieldwork strategy and the effects of the strategy on response rates and costs.

Chapter 6 serves as an introduction to the fascinating world of the field interviewer, which is described in the next two chapters.

Chapter 6 will show how adherence to fieldwork strategy has a positive effect on contact rates and costs. Nevertheless, in a substantial number of cases, interviewers do not adhere to this strategy. In chapter 7, a large scale research project among interviewers is described, in which the reasons for this given are investigated. Literature suggests that compliance with rules is, among others, a factor of perceived legitimacy, social norms, and contextual factors. In all, 12 dimensions were identified that could potentially be of influence in the decision (not) to follow a rule. A questionnaire was designed in which these dimensions were measured. The results show that the dimensions that underlay compliance with fieldwork rules are different for each rule. Some rules were considered as not legitimate (a polite way of saying 'nonsense'), while other rules were not followed as a result of miscommunication between the interviewers and the management.

Chapter 8 shares the subject of interviewers who do not follow fieldwork rules. In this chapter I address the important issue if interviewers introduce bias by doing so. If for example interviewers would shirk going into difficult neighbourhoods, or if they would never visit those neighbourhoods during the evening, that could introduce serious bias.

Groves (2006) summarizes five ways in which the existence of nonresponse bias can be studied. The easiest to accomplish, but the least informative, is to compare response rates across subgroups in the response. This does however not give direct estimates of the bias on key statistics. The second method is to use rich sampling frames or supplemental matched data. The additional data supply identical measures for all members, respondents and nonrespondents. By studying the relation between substantive variables and the frame variables, a sense of likely nonresponse bias can be gleamed. A third way is to make use of a comparison with similar estimates from other sources, like a census or high quality government survey. If estimates are comparable, independent of the survey, that will give confidence in their validity. Often, however, key variables are not present in other surveys, the measurement may differ and the coverage and nonresponse characteristics of the second survey may be unknown. A method that is often used is to compare early response to later response or response follow-up, a method used for example by Stoop (2005). Although it is often found that nonrespondents

are different from original respondents, the difference is seldom enough to result in different estimates. Information on nonrespondents is furthermore indirect, and depends on an assumed 'continuum of resistance'. It is however unknown to which extent converted nonrespondents resemble the final nonrespondents. The last method Groves (2006) describes is the comparison of different post-adjustments. This method is easily performed and gives the researcher confidence if results from different adjustments converge. The method only informs on convergence or differences, but says nothing about the true value.

Groves (2006) concludes that each of the techniques has value but each has drawbacks too. Chapter 8 addresses the issue of nonresponse bias. To measure the extent that interviewers introduce bias by not following fieldwork procedure, two methods are used. The first one uses a rich sample frame and supplemental data to study under what circumstances interviewers transgress the rules: in what neighbourhoods and with what kind of households and persons. This is an indirect measure of possible bias. The assumption is that when interviewers structurally favour or disfavour certain groups of people, that will influence results if key variables correlate with levels of these groups. If, for example, interviewers give less effort to underprivileged neighbourhoods, they may misrepresent the number of unemployed people in the Labour Force Survey. Likewise, if they hesitate to work evening hours, they may fail to reach working people, especially single working people.

The second method that is employed to study bias in chapter 8 is comparison to similar estimates. The key variable of the Labour Force Survey is compared with a registry that approximates this variable, by showing who is employed.

It was found that interviewers do indeed differentiate, although not in the direction implicitly assumed. The interviewers generally give more effort to sample units with a low response propensity, thus reducing the chance of biased results. Interviewers do as a matter of course what we painstakingly try to achieve: they use a tailored design in which they give more attention to more difficult cases. As a result the net bias, the difference between the number of working people in the workload and the number in the response, was extremely small.

Analysis of individual differences, however, showed that a minority of interviewers does tend to shirk difficulties. Those interviewers biased results by finding too few working people. These interviewers were balanced by a number of interviewers who exceeded response expectations. They too biased results, but in the other direction: they found too many working people.

Chapter 9, finally, presents a synthesis of the findings and a look into the future. The work in this thesis has a marked 'Statistics Netherlands' focus. This comes as no surprise of course. All data come from SN surveys; all experiments are performed within this context. This focus has advantages. The most important is the access to a rich reservoir of paradata and auxiliary variables that allows precise insight in

the effects of experimental measures on subgroups within the population, like in chapter 3 and 4, and even the rare and coveted possibility to study nonresponse bias with the help of registry data of substantive variables, like in chapter 8. The focus has disadvantages as well, however. The organisation of SN's fieldwork, with interviewers in permanent employment with a fixed remuneration and an infrastructure of elaborate monitoring, has implications for findings like those in chapters 7 and 8. The generalizability of some of these results to other settings is therefore not a priori obvious. Other findings, like those in chapters 5 and 6, corroborate and extend other research in very different settings, surveys and cultures and do not seem to be SN specific. In all, the availability of this wealth of paradata and registry information opens a treasure trove of possible research, of which this thesis only scratches the surface.

2.

Measures

to enhance response

to social surveys

2.1 Introduction¹⁾

In an ideal world, all persons who are invited to cooperate in a survey are located, and are able and willing to participate. Reality is otherwise. In a typical Statistics Netherlands (SN) social survey, about 65% of the sample participates. The other 35% are nonrespondents.

Nonresponse may negatively influence the quality of the statistics if nonrespondents systematically give different answers than respondents. Bethlehem (2009) gives examples of several Dutch surveys where this was the case. A follow-up study of the Dutch Victimization Study showed that people who were afraid to be home alone at night, are less inclined to participate in the survey. In the Housing Demand Survey, people who refused to participate had less housing demands than people who responded. In the Survey of Mobility, more mobile people were underrepresented among the respondents.

Nonresponse is not necessarily detrimental to the quality of a statistic. A meta-analysis by Groves and Peytcheva (2008) showed that some surveys have a high response rate but large bias, while others combine a low response rate with little bias. However, it is hard to predict when nonresponse will lead to bias. The difference between respondents and nonrespondents will generally be unknown, and may be different for each substantive variable within one survey. Although there has been research into other measures of survey quality (e.g., the Representativeness Index of Schouten, Bethlehem and Cobben, 2008), it is advisable in the absence of sound information about the probability of bias to aim for the highest possible response.

This chapter describes the measures a survey organization can take to optimize the probability that sample units respond favourably to the request to participate in a social survey. The measures are, where possible, illustrated by Statistics Netherlands' experiences.

Increasingly, survey methodologists become aware of what Groves (2006) calls the dangers of 'blind pursuit of high response rates'. This blind pursuit may actually increase bias, for example if the mechanisms causing non-response are different for different groups within one survey. When stimulating response rates, the survey organisation needs to be guided by knowledge of the relation between response-stimulating measures, the groups sensitive to them, and their influence on survey estimates. This awareness has led to the development of designs in which not all groups receive the same treatment. These designs are called adaptive and dynamic designs (Wagner, 2008), responsive designs (Groves & Heeringa, 2006) or tailored

¹⁾ This chapter is an updated version of a chapter published in the Statistics Netherlands' Methods Series (Luiten, 2009).

designs (Luiten & Schouten, 2013). The use of designs that thus differentiate the treatment that groups of sample units receive, may mean that the response-enhancing measures described in this chapter should not be applied to the entire sample. The techniques that are described in this chapter do not change, however.

The three most common reasons why a sample person does not respond are:

- Failure to contact the sample person (*noncontact*);
- *Refusal* of sample persons to cooperate in the survey;
- *Inability* of the sample person to take part, for example because of illness or absence.

These three types of nonresponse have different causes, and may have different influences on the quality of the statistics.

Bias may be anticipated if the subject of the survey is related to the chance of noncontact. People who are difficult to contact because they are almost never at home may cause serious bias in surveys of travel behaviour. Likewise, people who cannot be reached because they are at work will bias a labour force survey. Section 2.2 of this chapter describes measures that may be taken to maximize the probability of contact in surveys that involve an interviewer. Section 2.3 briefly describes the contact phase in surveys where no interviewer is present, like in web or mail surveys.

Section 2.4 describes factors that determine the willingness of a sample person to participate in a survey. Groves et al., (2004) categorize these factors in four dimensions:

1. the social environment (e.g., there are more refusals in cities than in rural areas; there are more refusals in single person households than in multi-person households);
2. the person (e.g., men are more likely to refuse than women);
3. the interviewer (e.g., experienced interviewers are better at persuading people than their less-experienced colleagues);
4. the design of the survey (e.g., incentives reduce the number of refusals).

Only the last two dimensions can be influenced by the survey organization.

Section 2.4 describes the relation between survey design and subjects' willingness to participate, and the influence of interviewers.

The third major nonresponse category comprises people who are unable to participate in the survey, either because of illness, absence throughout the entire survey period, or inability to speak or read the language in which the survey is administered. Sometimes a person's inability is related to the survey subject, e.g., where people are too ill to take part in a health survey. Groves et al., (2004) state that in surveys of the entire population the bias introduced by this nonresponse category is slight. However, inability to participate may be a significant source of

bias in surveys of specific populations (e.g., elderly people, or ethnic minorities). There are various measures available to a survey organization to reduce the probability of nonresponse because of inability to participate, including the use of foreign-language interviewers or questionnaires, or extending the fieldwork period. A longer fieldwork period may also reduce noncontact, and is described in section 2.2. The use of foreign-language interviewers also may prevent refusals, and is therefore discussed in section 2.4. The chapter concludes with a section about interviewer training and monitoring.

2.2 Methods to reduce noncontact in CAPI/CATI surveys

The success of a survey organization in contacting sample units depends on characteristics of the persons or households (like at home patterns and family composition), and on characteristics of the building they live in (e.g., impediments like gates or gate keepers). However, several aspects of the survey design have an important influence on the chance of contact: the number and timing of the contact attempts, the length of the fieldwork period and the number of addresses to be approached by an interviewer in the course of the fieldwork period (the workload). Other factors that may affect the probability of success include the option to make contact in a different mode, and the availability of background information. These aspects are described in this section.

Despite their influence on contact, the success of measures to increase contact rates is limited by practical aspects, such as the number of evening hours that interviewers are willing and able to work. The best method in terms of response sometimes runs up against practical limits of what can be demanded of interviewers.

Table 2.1.1 gives an overview of the measures to reduce noncontact that are described in the paragraphs below and in section 2.3 for web surveys. The table summarizes the gain in response as a result of each measure. The more '+'es, the higher the gain in response. The number of '+'es was determined by weighting research results. For instance, 'planning of contact attempts' was given one +, because, while spreading attempts over time is very important in making contact rapidly, it is not absolutely necessary for making contact at some point in the future.

2.1.1 Methods to reduce noncontact in CAPI, CATI, web and mail surveys

	CAPI / CATI	Web / Mail
The number of contact attempts	+++	+/-
A lengthy fieldwork period	++	++
Call scheduling	+	+
(Optimal) interviewer workload	++	N/A
Contact attempts in a different mode	+/- ¹⁾	++
Use of auxiliary information	+/-	Unknown

¹⁾ +/- means that variable results were found, with no clear trend in either direction, or no effect.

Heighten the number of contact attempts

One of the most obvious measures to reduce noncontact is to increase the number of permitted contact attempts. There is a clear relationship between the number of contact attempts and the contact rate. Fieldwork organizations generally prescribe a minimum number of contact attempts before writing off a survey unit as a noncontact. This number will be lower for CAPI than for CATI surveys, because of differences in the costs of contact attempts in the various modes. Fieldwork organizations will however not permit unlimited contact attempts, because at some point the costs of additional approaches will no longer outweigh the benefits in terms of bias reduction of marginal improvement in response.

The number of visits determines not only the contact rate, but also the quality of the data: there are differences between the people you reach immediately and those you find only after several visits. For instance, the response rate for some ethnic minority groups went up by ten percentage points after SN augmented the number of mandatory visits in CAPI surveys from three to six in 2003, thereby reducing the response gap between this group and the native Dutch (Schmeets, 2005). The extent to which differences in contact rate lead to differences in substantive variables may depend on the type of survey. For instance, the people who are hard to contact for the Labour Force Survey and the Dutch Mobility Survey are workers and travellers (see also Lynn, 2002 for comparable results). Abandoning contact attempts prematurely, or allowing a relatively high noncontact rate, will introduce bias in these kinds of surveys. Careful nonresponse analysis needs to determine if and how bias is related to contact rates.

Extend the fieldwork period

Obviously, the number of possible contact attempts is related to the length of the fieldwork period. The length of the fieldwork period also has an independent influence: the longer the fieldwork period, the greater the probability that all sample persons will be aware of the request to participate. Even people who spend their winter in Spain will return eventually.

Optimize the probability of contact by call scheduling

Survey organisations strive to obtain the highest chance of contact with the lowest number of visits or calls. They do this by carefully planning the timing of visits or calls. This planned sequencing is called call scheduling. There has been a considerable amount of research into optimal call schedules for both CATI and CAPI. The areas investigated include the best times for the initial contact, and what should be done if the first contact fails: spreading the subsequent contact attempts over times, days and weeks of the fieldwork period.

Call or visit in the evening

The most likely time to reach people is in the evening, and then in particular an evening before a working day (i.e. Sunday to Thursday evenings). This has been repeatedly found in research in the (Western) world (e.g., Campanelli, Sturgiss & Purdon, 1997; Durrant, d'Arrigo & Steele, 2011; Groves & Couper, 1998; Luiten, 2006). Only a small proportion of people cannot be reached in the evening (Groves & Couper, 1998). However, the first visit is frequently not in the evening, while most interviewers prefer to find out about the nature and location of an address during day time.

Spread visits over times, days and weeks of the fieldwork period

If the first visit was in vain, a second visit must be planned. The most promising time for a second attempt after an unsuccessful daytime visit is the evening of the same day or else a visit in the evening of another working day (Luiten, 2006b; Purdon, Campanelli & Sturgiss, 1999; Stoop, 2005). Opinions differ about what to do after an unsuccessful evening visit. Purdon et al., (1999) suggest that not making the subsequent visit in the evening holds the most promise. Groves and Couper (1998), and Luiten (2008, see also chapter 6) find conversely that an evening visit is more likely to succeed regardless of the time of the previous visit. Contact attempts must be spread well over the available time, in terms of the time of day, day of the week and the weeks of the fieldwork period. Spreading well over time contributes most to a high contact ratio at the minimum expense (Luiten,

2006b). In call centre-based CATI surveys, as in SN, the CATI management system takes care of an optimum spread of contact attempts. In CAPI surveys, interviewers must currently perform this task themselves. However, the evidence from research (Luiten, 2008) is that interviewers are not good at arranging an optimal spread over time for all their sample persons. Further research is needed into tools to support interviewers in this regard.

Visit weekends?

Much research recommends weekend visits: both Saturday and Sunday are promising days in many countries for making contact with a survey unit (Campanelli, Sturgis & Purdon, 1997; Durrant, D'Arrigo & Steele, 2011; Groves & Couper, 1998). However, this has not been confirmed in SN surveys (Luiten, 2006a). Saturday visits are no more successful than daytime visits on a working day. SN interviewers do not as a rule visit or call on Sundays. Groves and Couper (1998) duly observe that there are cultural differences in the acceptance of weekend visits.

Adjust workload to available interviewer capacity

Interviewers need sufficient time to approach their allotted addresses with due care. If too many cases are allotted for a given fieldwork period, both the contact and cooperation rates will deteriorate.

The exact calculation of the workload that interviewers can cope with is therefore a vital part of achieving acceptable response rates. For CATI surveys the factors that are involved in this calculation are the predicted response rate, the rate and kind of nonresponse to be expected, the predicted number of calls and the time needed for responding and nonresponding cases. For CAPI surveys, the calculation also needs the travel distance to be expected, the expected speed at which this distance can be travelled, the expected number of visits and the time needed to gain entry. See Van Berkel and Vosmer (2006) for the method of calculating the workload of SN interviewers.

CAPI-interviewers should have as much productive working time and as little travel time as possible. To secure the optimal allocation of interviewers to sample addresses, SN uses software ('Axis') that minimises travel distance between addresses and available interviewers.

Allow contact attempts in a different mode

In CAPI surveys it may be expedient to attempt contact in a different mode after a number of unsuccessful visits. For instance, after three unsuccessful face-to-face

visits, a SN interviewer will leave a card in the letterbox with a proposal for an appointment and her cell phone number. If the respondent does not respond to the request, the interviewer may attempt to contact the respondent by phone. An advantage of using the telephone is that more frequent contact attempts can be made at limited expense, and that it is also easier to make the contact attempts in the evening. However, a disadvantage of the telephone is that people find it easier to refuse than in a face-to-face situation. On balance the higher number of refusals counteracts the lower number of noncontacts, so that the net response is hardly any different (Blom & Blohm, 2007).

Written or telephone contact attempts therefore already occur in CAPI surveys in practice, but other combinations are also conceivable, such as a telephone reminder with a web questionnaire, written approaches to obtain a telephone number, or a written request to phone the survey institute for an inbound CATI interview. Fouwels, Wetzels and Jansen (2006b) show that with a web questionnaire, a telephonic (third) reminder does not generate a higher response than a written third reminder.

Use background information of an address or person

Providing interviewers with information about the sample person may be relevant in planning the contact phase as efficiently as possible and raising the probability of contact. For instance, retired people can be successfully approached in the daytime, while the same would not be a good idea for a young double-income household. Groves and Couper (1998) identify relevant background information:

- the number of adults in a household (the probability of contact is greater where there is more than one);
- people over the age of 70 in the household (greater probability of contact);
- children younger than 6 in the household (greater probability of contact);
- single person household (lower probability of contact);
- all members of the household younger than 30 (lower probability of contact).

Schouten (2007), Cobben and Schouten (2007) and Luiten, Schouten, Gevers and Cobben (2007) in addition identify ethnic background (lower probability if non-Dutch), urban density (lower probability the higher the urban density), property value and / or income (lower probability the lower the value or income), and the employment status of members of the household core (lower probability if all members work, but also if none work).

If possible, the survey organisation may provide information about relevant characteristics to interviewers, prior to the start of fieldwork. SN could relay available information from registry data. Interviewers may also pick up relevant information during the course of the fieldwork.

Increasingly, knowledge about sample units, whether derived from registry information, or by interviewers during fieldwork, may be used to differentiate the effort directed at groups of sample persons (Durrant, D'Arrigo, & Steele, 2011). If groups are known to be typically underrepresented in the response, they may be approached differently or more intensively, with the aim of obtaining a more balanced, or representative, response. Experiments in which knowledge of sample persons was thus used to obtain more representative response were performed within the framework of the European RISQ project (Schouten, Cobben & Bethlehem, 2009). Fosen et al., (2010) and Luiten and Schouten (2013) describe examples of how this can be accomplished.

2.3 Noncontact in web and /or mail surveys

Mail surveys are paper questionnaires that are sent through the mail. Sample units will receive the questionnaire, accompanied by a letter that explains the purpose of the survey. In a Web survey, the data are collected via the internet. There are several ways in which a sample unit can be approached in a web survey (Bethlehem & Biffignandi, 2012). Most common is to send an email, e.g., to respondents in a panel, or to invite people in a letter to log on to a website. Several of the measures to stimulate contact rates that are described in the sections above also apply to mail and web surveys. Both are helped by reminders, a long fieldwork period, etc. Nonetheless, web and mail surveys have several specific aspects.

Noncontact in CAPI and CATI surveys is caused by people who are not at home, or by people not answering the phone. The cause of noncontact in mail surveys is that people do not receive the letter. The cause for web surveys may be more diverse. An important consideration is how sample persons are selected (Bethlehem & Biffignandi, 2012). If the sample frame is a list of email addresses, noncontact occurs if the email with the invitation does not reach the sample person, for example because the email address is wrong, or the mail is blocked by a spam filter. If the sample frame is a list of addresses, and an advance letter is sent with a login code, noncontact may be caused if people do not receive the letter. Finally, if sample persons are recruited by CAPI or CATI interviewers to participate in a web survey, noncontact may be caused by people not at home, or not answering their phone.

General population surveys will not be based on sample frames of email addresses, but will use the same sample frames as CAPI or CATI surveys, i.e., a population register or a list of addresses. Even if an email address is available, however, the invitation to participate in a web survey should preferably be in the form of a letter, to prevent the blocking of the invitation by spam filters.

If the survey request actually reached the sample person is generally not known to the survey organisation. The fine distinction that we make in CATI and CAPI surveys between nonresponse categories is not possible in web and mail surveys. With the few exceptions when sample units explicitly let you know that they will not participate, you only know that sample persons did not respond. Assuming that the advance letter with the survey request reached its destination, we will treat nonresponse on web and mail surveys as soft refusals. Measures to overcome these soft refusals, like sending reminders, will be treated in the next session, on cooperation.

2.4 Methods to increase cooperation

Survey design

Survey design has an important influence on the chance that respondents will cooperate with the interviewer's survey request. The probability of response can be optimized by writing an effective advance letter, enclosing an information folder, and including an incentive. The use, or not, of interviewers, the way of selecting respondents in a household, the length of the questionnaire, possible attempts to persuade refusers to change their minds, and the kind of training given to interviewers, are all design aspects with an influence on the ultimate response. The actual outcome is further determined by characteristics of the respondent, and the interaction with the interviewer.

Table 2.4.1 gives an overview of the design aspects that relate to cooperation. The table summarizes the gain in response as a result of each measure. The more '+'es, the higher the gain in response according to research results. The following sections subsequently describe each measure.

2.4.1 Methods to improve cooperation in CAPI, CATI, web and mail surveys

	CAPI / CATI	Web / Mail
Advance letters	+	+++
Folders	+/-	+/-
Corporate image	+	+
Incentives	++	+
Reminders	+ ¹⁾	+++
Use a different mode	+/-	+++
Good survey material	+	++
A short questionnaire	+/-	+++
The use of interviewers	+++	N/A
Respondent rules	++	++
Allowing the respondent to choose the mode	-	+/-
Foreign-language interviewers and/or questionnaires	+/-	+/-
Refusal conversion	++	+++

¹⁾ Probably. No empirical research at SN, no other known empirical research.

Advance letters

A letter to announce a forthcoming survey is called an advance letter. In general, but in particular for statistics bureaus, an advance letter leads to a higher response (De Leeuw, Callegaro, Hox, Korendijk, & Lensvelt-Mulders, 2007). The letter is an opportunity to explain what the survey is about, why it is important for the respondent to participate, and how and why the respondent was selected. The letter should explain and underline the survey organisation's authority (Cialdini, 2001). The letter gives interviewers a point of reference, and is a source of confidence for them. Referring to the letter is an important element of the initial contact (Becks, 2008). SN sends advance letters for all surveys.

Both the sending of the letter as such and its contents are relevant for the response process. Experiments with advance letters (e.g., Luiten, Campanelli, Klaassen & Beukenhorst, 2008; White, Martin, Benneth & Freeth, 1997) have shown that minor adjustments to the contents of the letter can have an effect on the response rate. Advance letters are a vehicle for applying socio-psychological theories in order to be as persuasive as possible towards respondents (Cialdini, 2001; Dillman, 2009). The timing of the advance letter is a point of consideration: it should precede the first visit or call by no more than a few days, so that respondents do not forget receiving it.

Folders

An advance letter must be brief, no longer than one A4 page (Dillman, 2009), which implies very concise statements of much relevant information about the survey objective, the sampling method, the interview procedure, and what will be

done with the data. Many survey bureaus solve this problem by enclosing a folder with a more detailed explanation. It is unclear how effective sending a folder is as a response-enhancing measure. SN encloses folders with the advance letter for social surveys, but no experimental research has been performed to test the effect. The folder briefly touches on questions that people may have (who uses this information, what is the survey about, how was I selected), but is mainly oriented to how data figure in press releases.

Corporate Image

People are generally more inclined to take part in surveys run by a National Statistical Institutes than by an academic or commercial organization. Groves and Couper (1998) speculate that this predisposition stems from a general awareness that a democratic administration cannot function without information from the public. For this mechanism to work, the public needs to be aware who you are. Despite daily mention of SN figures in newspapers, magazines, radio and TV, far from everyone knows who we are. The yearly satisfaction survey among respondents and nonrespondents of the Labour Force Survey found that 12% of respondents and 14% of nonrespondents had never heard of SN before receiving the advance letter (Arends-Tóth, Meertens, Engelen & Kroeze, 2008). SN has embarked on a charm offensive to improve its image with respect to business statistics. A similar approach, aimed at improving name awareness, has often been considered for social statistics. It has never been implemented, though. In view of the almost equal proportion of respondents and nonrespondents who didn't know SN before being approached, name awareness does not appear to be an overriding response mechanism. A clear statement of SN' authority is therefore a primary aim of the advance letter and folder. The statement involves sending a serious, attractive, and well drafted letter, using professional folder material, giving a free number for the respondent to call with any questions, and explaining clearly what will happen with the results of the survey (Dillman, 1978, 2000; Dillman, Smyth & Christian, 2009).

Incentives

An incentive is a token of the survey organization's gratitude to the respondent. Incentives lead to higher response rates, and higher incentives lead to even higher response rates, albeit that the relationship is not linear, but declines as the incentive increases, and can even become negative if respondents feel that the incentive is excessive for what is asked (Singer, 2002). The increase in response rate depends on various factors: cash incentives are more effective than gifts. Incentives given in advance to all sample persons (unconditional incentives) are more effective than incentives that are promised, conditional on response.

If what is requested from the respondent is so demanding that the incentive may be viewed as payment, it can be paid after the survey. An example is the incentive given to respondents of the SN Household Expenditure Survey, where respondents are asked to keep a diary of expenses during a number of weeks.

Incentives are not equally effective with everyone. Groves and Couper (1998) state that incentives work best with people who have a relatively low educational level and a low income. However, Wetzels, Schmeets, Van den Brakel and Feskens (2008) show that unconditional incentives in the form of a book of ten postage stamps may attain a substantial rise in response rates of ten percentage points in some groups, but do not work with members of non-western ethnic minorities, despite this group's relatively low educational and income level.

It is also possible to give incentives to specific groups, e.g., those who are likely to refuse, or people who initially refused. Ethical objections may be raised to giving incentives only to people who might not otherwise participate, or giving larger incentives to people who initially refused. However, it is unclear to what extent these objections are also shared by the public. Groves and Couper (1998) see no problem whatsoever with different incentives, since people also differ in their aversion to surveys. For some people it is enough incentive to be able to express their opinion, or chat with an interviewer for half an hour. Other people need considerably more resources to be brought into play to make the survey attractive enough for them to take part.

Survey costs may decrease with a sufficiently powerful incentive, because interviewer and follow-up expenses will be lower, and the sample can be smaller (Berlin et al., 1992).

It has not been SN policy to date to use incentives as a matter of course, except in cases of high burden on the respondent, such as in the Household Expenditure Survey. In exceptional cases incentives are used however. An example was the Dutch Parliamentary Election Study (Schmeets, 2010). The incentive that we use, depending on the increase in response envisaged, is a booklet of five or ten postage stamps or a €5 gift coupon.

Reminders

With mail surveys and web surveys it is customary to send reminders after some time has elapsed. Without these reminders the response may be 20 to 40 percentage points lower (Dillman et al., 2009). For CAPI en CATI surveys reminders can also be valuable. SN uses calling cards to alert sample persons that interviewers are trying to contact them. These cards are intended to convey to sample persons their importance, in that interviewers are prepared to go to great length to contact them. Social exchange theory predicts that respondents will as a result be more likely to take part in the survey. CATI interviewers leave a voicemail message in the event of noncontact with the same intent.

A calling card is left behind after each of the first three unsuccessful visits. There are three different cards, with successively lower barriers for the survey target person to contact SN. The first card states only that an interviewer called, and will try again later. The second card adds the help desk telephone number. The third card adds the interviewer's mobile number and suggests a date for an appointment. In the case of mixed mode surveys, where sample persons are first asked to respond by web, and then are reminded two times, interviewers will leave their telephone number at the first visit, so as not to aggravate respondents with too many reminders.

Introduce a different survey mode

Mode switching can be an effective way of enhancing response, although its actual effect depends on the initial single mode design. For instance, SN has been unsuccessful to date in enhancing the response of previously single mode CAPI or CATI surveys by adding the web as the first mode in a consecutive mixed mode design (Wetzels, Janssen & Schooten, 2007). Lynn (2012) indicates that higher response rates in mixed mode designs can be attained if the first mode in a consecutive design is the most expensive one. On top of that, each consecutive mode must be run as if it were the only mode, i.e., with the highest possible effort.

Survey material

Interviewers must have confidence in the material. If they feel the material makes them look foolish, because of questions that do not flow, illogical sequences, unnecessary repetition in the questionnaire, and so on, they will be unable to convey the response message with conviction. Interviewer confidence has a substantial influence on the ultimate response.

SN questionnaires are produced by a specialist design department, which aims to make their products respondent-friendly, relevant and unambiguous (Vosmer & Engelen, 2008). After the design and construction phase, the questionnaires are tested for internal logic (does the routing flow properly, are there any questions that make no sense after a previous answer, have any routes been missed, etc.). See Cuppen and Te Riele (2008) for an example of this test phase.

The length of the questionnaire

The confidence that interviewers have when they ring the bell, or call on the phone, is also related to the length of the questionnaire they have to administer. It is easier to persuade someone to answer a mere five-minute questionnaire than one that will take 45 minutes. However, in a summary of the role of interview length, Bogen (1996) concludes that the relationship between interview length and nonresponse is more frequently positive than otherwise, but is surprisingly low and variable across research.

Interview length is related to burden, of both respondents and interviewers. Interviewers in focus groups express more reluctance to ask for cooperation for long interviews than short ones. This resistance stems partly from a sense that it is not good etiquette, and from an expectation that a request for a long interview is more likely to meet with refusal. Interviewers feel the same way if they do not state in advance how long an interview will take. Some of the high refusal rates in long surveys is therefore caused by, possibly unconscious, interviewer behaviour. Surprisingly enough, there is no strong evidence that knowledge of the length of an interview plays any role for *respondents*. The output of research gives a mixed picture, in particular for surveys that involve an interviewer. Experimental research has revealed that relatively large differences in length like 25 minutes versus 75 minutes (Frankel & Sharp, 1981) or 20 minutes versus 40 minutes (Collins, Sykes, Wilson & Blackshaw, 1988) produce a small and insignificant difference in cooperation. Relatively few respondents ask how long the interview is (27% in a study of Groves & Couper, 1998; 19% for Becks, 2008). Many decisions about participation are thus taken without knowing the length.

In *written* surveys the length of the questionnaire has a modest effect, but more important than that is the page layout. Champion and Sear (1969) find that questionnaires with more pages, but more 'white' on the page, have a higher response than questionnaires with fewer pages and less white. A meta-analysis of the effect of various aspects of advance letters (De Leeuw, Callegaro, Hox, Korendijk & Lensvelt-Mulders, 2007) likewise showed no relationship between stating the length of the survey in the advance letter and the response rate. To what extent these results can be generalized to online surveys is unclear, because of a lack of research. Galešić (2005) shows significant differences in both login and completion rates for opt-in web interviews of 10, 20 and 30 minutes. 76% opened a questionnaire that was stated to take 10 minutes, against 64% for one of 30 minutes. 44% finished the screening questions for the 10-minute questionnaire, compared with 39% for the 30-minute counterpart. Another 28% dropped out before the end of the 10-minute questionnaire, while the corresponding figure for the 30-minute questionnaire was 50%.

The use of interviewers

The use of interviewers is the most effective measure in terms of response optimization. The more personal influence can be asserted, the higher the response. CATI surveys have a higher response than web or e-mail surveys, and CAPI surveys have a higher response than CATI surveys.²⁾ Collins, Sykes, Wilson and

²⁾ If CAPI is the only mode. CAPI surveys in a mixed-mode design, in which people with a listed telephone number are approached through CATI, will have a lower response than CATI surveys. Survey target persons with unknown telephone numbers form a selective subpopulation with a lower response rate.

Blackshaw (1988) found response rates in telephone surveys in Great Britain to be 15 percentage points lower than for CAPI surveys. Nicolaas and Lynn (2002) even found a difference of 30 per cent, also in Great Britain. A 30% response is usual in SN web surveys, albeit that both lower (around 15%) and higher rates (around 40%) are found as well, depending on the survey and the target population (Wetzels, Janssen & Schouten, 2007).

The higher response for CAPI surveys has to do both with social norms, which make it harder for respondents to give a rejection face-to-face than on the phone. It has also to do with the opportunity that face-to-face interviewers have to use features of the house and neighbourhood to gain contact. This may involve observation (e.g., toys in the garden mean that children live in the house), but also information obtained from neighbours. Together with the tailoring skills that a well trained interviewer possesses (see section 2.5), these are persuasive instruments for obtaining a response.

Rules for selecting respondents

Rules that determine who in the household is to answer the questions can have a major influence on the response rate. For instance, surveys with an otherwise similar design that allow proxy answering (i.e. where respondents also answer for other people) have a higher response than those where this is not allowed. The same applies to surveys that allow any household member to answer, rather than a randomly selected specific survey target person (such as whoever's birthday is next).

Allowing the respondent to choose the mode

It would seem obvious that response would be higher if the respondent were allowed to choose the mode in which to perform a survey. However, Griffin, Fisher and Morgan (2001) found that a choice of modes *lowered* the response. They suspect that this effect has to do with the disturbance to the response process. Dillman, Clark and West (1995) showed that the effect of allowing a choice between telephone and e-mail was that five per cent of the respondents opted for a more expensive mode (telephone), but that this had no response-enhancing effect. The five per cent concerned would have simply answered by e-mail anyway. Janssen, Wetzels and Cuppen (2008) show that a previously stated preference for a mode has no influence on the subsequent inclination to respond in a non-preferred mode. De Leeuw (2005) found no firm empirical evidence that allowing a choice improves response, but more evidence that it makes no difference.

The use of foreign-language interviewers and questionnaires

Between two and three per cent of sample persons approached by SN are written off as nonresponse because of language problems. This means that interviewers

are unable to administer the questionnaire in Dutch. This proportion is of relatively little concern in a general population survey, but can be a serious problem when surveying specific population groups. Beukenhorst (2008) reports that language is only a small part of the response problem with ethnic minorities. The response rate of members of ethnic minorities is lower than for native Dutch in initial polls, and moreover the rate of attrition of ethnic minority panel members (such as for the Labour Force Survey) is considerably higher than for native Dutch. Language problems are not the cause of that phenomenon, since the panel members have participated in Dutch in the first wave.

Korte and Dagevos (2011) show that deploying interviewers from the same ethnic minority group had a positive influence on response rates of Surinamese and Antillean sample persons, compared to using native Dutch interviewers, and had an extremely positive influence on the response rates of young Turkish and Moroccan persons.

The use of foreign-language interviewers can also affect response and data quality adversely. For instance, there may be concerns within a local community that any information disclosed could become public knowledge. Questions about sensitive subjects such as church or mosque attendance and sexual behaviour may yield considerably more socially desirable answers if the interviewer is of the same origin as the respondent (e.g., Reese, Danielson, Schoemaker, Chang & Hsu, 1986).

Refusal conversion

If a survey target person has refused, the fieldwork organization may decide to approach this person again, either in the same or a different mode. Various studies have shown that even if the repeated approach is in the same mode, between 25% and 40% of the initial refusers change their minds in a subsequent approach. This process of persuading refusers is known as refusal conversion. Depending on the initial refusal rate, refusal conversion can lead to a substantial increase in response. Refusal conversion is applied in the European Social Survey (ESS), which is conducted concurrently in European countries. Stoop and Koch (2005) report that the initial response in the Netherlands in the first ESS round was one of the lowest in Europe, but following successful refusal conversion, the eventual response rate in the Netherlands was among the highest.

Refusal conversion is generally performed by a different, often more experienced, interviewer. If possible, attempts may be made to improve the match between the interviewer's and respondent's characteristics (e.g., if an older woman appears to be afraid of a male interviewer, a female interviewer could be sent instead). Often, but not always, a new letter is sent before the new contact, specifically mentioning the objections raised by the respondent in the first round. However, De Groot and Luiten (2001) found no difference in converted response with or without a new letter.

In mixed-mode surveys, the transition from web to CATI or CAPI is not referred to as refusal conversion, because the non-respondent will not have explicitly stated an unwillingness to participate. Approaches in a different mode following a web approach also leads to a large increase in response. SN web mode surveys achieve 25% to 30% response. In the subsequent rounds with CATI or CAPI, the rate rises to the usual 65%. Although the second approach is not referred to as refusal conversion, the demands on the interviewer will often be higher than if CATI or CAPI had been used in the first place. Some respondents are annoyed at receiving two reminders following the first approach for the web questionnaire, only to be approached yet again by an interviewer. The fact that the SN advance letters do mention the follow-up approach by an interviewer does little to mitigate the irritation for some respondents. Interviewers must be trained to anticipate a more difficult task because of the mixed-mode approach.

A special form of refusal conversion involves approaching a non-respondent with an extremely short questionnaire containing the key questions of the survey. Kersten and Bethlehem (1984) call this the central question method. Lynn (2003) describes a similar approach: PEDAksi, or Pre-emptive Doorstep Administration of Key Survey Items. However, the aim of this repeated approach is not to raise the response rate, but to estimate the differences between respondents and nonrespondents, and therefore the extent of bias attributable to nonresponse. See Schouten and Cobben (2007) and Cobben and Schouten (2007) for an application of this method in the repeated approach of Labour Force Survey nonrespondents. Despite the extremely widespread application of refusal conversion, there has been little research if the additional data lead to less biased results. Keeter et al., (2000) and Curtin, Presser and Singer (2000) showed that the increase in response through refusal conversion had no effect on the estimate of the target parameter. This finding is in line with that of Groves and Peytcheva (2008), that there is a very weak relation between the response rate and the degree of bias. Schouten (2007) found demographic differences between nonrespondents and respondents in the labour force survey, but no difference in employment status. However, Stoop (2005) did find relevant improvements in the estimates as an effect of refusal conversion. Lynn, Clarke, Martin and Sturgiss (2002) also observe less bias as an effect of refusal conversion. Refusal conversion is costly. Interviewers may have to travel further to find an address, and the best interviewers are less or not at all available to do 'mainstream' work. With a finite budget, the response rate may be higher, but the number of responses lower, because of the greater fieldwork effort needed. Therefore, the use of refusal conversions must be considered carefully. At any rate, it makes no sense to apply refusal conversion to every refusing sample person. The costs involved make it advisable to reserve this technique for special groups that should be well represented in the response.

2.5 The role of the interviewer

The sections above explained how the response rate is influenced by characteristics of the survey design. Another important source of influence is the interviewer. There are considerable differences in the responses achieved by individual interviewers, in both face-to-face and telephone situations. It is therefore not surprising that much research has focused on the correlates between interviewer effectiveness and interviewer characteristics. The research has examined demographic characteristics of interviewers, as well as their experience, training, attitudes, personality and behaviour in the interview situation.

Constant differences are seldom found in research of this kind. Interviewer performance is influenced by characteristics of households, the neighbourhood, the survey, and so on. An interviewer may be good at some particular survey, but not another, or may perform excellent work in rural areas, but be lost in a city. Groves and Couper (1998) conclude that it is pointless to investigate interviewer characteristics such as race, gender and age, firstly because they yield few interpretable main effects, but mainly because of the interactions with household characteristics and survey subjects. They also consider that these kinds of interviewer characteristics are largely mediated by interviewer experience, attitudes and behaviour. Nevertheless, a summary is given below of the findings of research into correlates of interviewer effectiveness.

This subsection presents the output of research into the influence on response rates of socio-demographic interviewer characteristics, interviewer personality, voice and use of language, experience and behaviour in the interaction with respondents. Table 2.5.1 gives an overview of the influence of interviewer characteristics on response rates.

2.5.1 Influence of interviewer characteristics on CAPI and CATI response

	CAPI / CATI
Socio-demographic interviewer characteristics	+/-
Interviewer personality	+
Interviewer voice and speech characteristics	+
Interviewer experience	+/-
Interviewer behaviour (tailoring and maintaining interaction)	++

Socio-demographic characteristics

In her study on the relationship between socio-demographic interviewer characteristics and interviewer performance Carton (1999) focussed on age,

gender, education (lower than senior general secondary [HAVO], HAVO, higher vocational, or university), and work situation (home-based, on a career break, freelance, employed full time, employed part time, student, or unemployed). The analysis showed no relationship whatsoever between these characteristics and the degree of interviewer success in making contact with and persuading respondents. The only exception was the finding that men were significantly overrepresented in a (relatively small) group of less well performing interviewers. Campanelli, Sturgis and Purdon (1997) found no effect whatsoever of socio-demographic interviewer characteristics after controlling for other sources of influence.

Interviewer personality

Various studies have demonstrated relationships between certain interviewer personality traits and their response rates. However, the relationships are weak, possibly because the social skills that a good interviewer needs can be acquired (Morton Williams, 1993). The weak relationships found were emotional stability and *introversion*³⁾ (McFarlane Smith, 1972), social skills (defined as being good at negotiating) and personal organization skills (Johnson & Price, 1988) Personal organization skills were defined as making effective use of time and not being easily distracted. Snyder (1980, in Groves & Couper, 1998) mentions self-monitoring. People with a high level of self-monitoring are flexible and adaptable individuals who are adept at modifying their social behaviour to suit situational and interpersonal changes. Self-monitoring is related to what Groves and Couper call 'tailoring'. De Leeuw and Hox (1996) found that interviewers with high scores on the personality items 'conscientious' and 'formal' could more readily persuade respondents: their cooperation rate was higher than interviewers with lower scores. However, no difference was found in overall response rates, indicating that these interviewers may have been less successful in finding people at home.

Voice and paralinguistic characteristics

In telephone surveys in particular, the first few seconds often determine whether or not the interaction will lead to a response. Oksenberg and Cannell (1988) showed substantial differences in interviewers' refusal rates, even if they used a standard introduction. Based on this finding they investigated paralinguistic characteristics,

³⁾ This term is in italics because it appears to be counterintuitive. However, the relationship has been confirmed in at least three studies.

such as pitch, tempo, volume, variation and accent of interviewers. Some, but not all, of their studies showed that distinctive speech characteristics such as pitch, volume, intonation, tempo, and degree of adherence to standard pronunciation were related to cooperation rates. Successful (female) interviewers spoke quickly, loudly and with a standard (American) accent. They were also assessed as sounding competent and self-assured. Interviewers who exhibited a falling intonation on key words in the introduction also had higher cooperation ratios than those with a rising intonation.

Becks (2008) demonstrated large differences between interviewers in the degree to which respondents were immediately willing to participate, as opposed to needing persuasion (between 11% and 68% immediate response). Van der Vaart, Ongena, Hoogendoorn and Dijkstra (2006) found comparable rates when replicating Oksenberg and Cannell's (1988) study in a Dutch situation (between 14% and 51%). Van der Vaart et al., found that fluent speech and volume correlated positively with cooperation. Pitch and tempo, contrary to the studies of Oksenberg and Cannell, were found to make no difference⁴⁾. Where interviewer voices were assessed as 'pleasant', 'authoritative' or 'reliable', only the authoritative voices appeared to correlate with (immediate) cooperation.

While these findings may provide input for interviewer recruitment and training, some qualification is called for. The analyses were based on immediate consent. However, Becks' (2008) study showed no relationship between the immediate consent rate achieved by an interviewer and the ultimate result. Interviewers whose voice characteristics bring about rapid cooperation will have an easier time than colleagues who lack this natural interview voice. However, these colleagues are able to achieve equally satisfactory results through effective persuasion. An investigation of the voice characteristics that lead to immediate and resolute refusal may be more helpful in identifying people who should *not* be recruited as interviewers.

Interviewer experience

Various studies have addressed the role of interviewer experience, generally indicating that more experienced interviewers have higher response rates. Couper and Groves (1992) point out that it is hard to separate the effect of experience from the possible process of interviewer selection and self-selection. Interviewers who perform relatively poorly may leave of their own accord, or be passed over for later surveys. Carton and Pickery (2010) conclude after carefully decomposing the

⁴⁾ Van der Vaart et al. state that the variance between interviewers in this respect was slight.

different effects, that more experienced interviewers do not have higher response rates because they are getting better (there is even a slight tendency towards them getting worse), but because the less successful interviewers quit their job.

Interviewer behaviour during the introduction: tailoring and maintaining interaction

The original objective of the study presented by Groves and Couper (1998) in their book 'Nonresponse in Household Interview Surveys' was to identify 'fixed' interviewer characteristics that are linked with successful achievement of cooperation. However, as the study progressed they became convinced that what was of overriding importance was not interviewer characteristics but the interviewer's behaviour in the contact with the respondent. Groves and Couper base their findings on the work of Morton-Williams (1993). Analyses of audio recordings of interactions between interviewers and respondents in the 'doorstep' phase revealed that interviewers who did not adhere strictly to prescribed scripts achieved a higher response (59% versus 76%). These interviewers appeared better able to adapt to the specific situation, and took advantage of certain characteristics that presented themselves for maintaining an optimum interaction. Morton-Williams used these insights in developing a training course based around acquiring social skills. These skills were later dubbed 'tailoring' and 'maintaining interaction' by Groves and Couper. Tailoring is defined as using clues from the neighbourhood, the home, or the respondent's appearance to adapt the introduction. Maintaining interaction is defined as attempting to sustain the interaction, sometimes across multiple contacts.

Although several studies have demonstrated the favourable influence of tailoring and maintaining interaction on the respondent's decision to participate in a survey, it is hard to pin down exactly which actions are responsible for a favourable outcome of the contact phase. Research shows fairly accurately what an interviewer should *not* do (Becks, 2008; Dijkstra & Smit, 1999; Heerwegh, 2000; Pondman, 1998). An interviewer should not ask 'why' following a refusal, and avoid repeating the respondent's refusal. On the other hand, positive actions are harder to define. Identifying actions that the interviewer *should* carry out was the subject of studies by Heerwegh (2000) and Becks (2008). Heerwegh concludes that interviewers should recognize that there is 'strength in numbers'. A pitch with more arguments will win from one with fewer. The respondent will be more likely to cooperate if the interviewer comes up with several arguments in response to an objection. Here too, however, it is important to avoid certain behaviours, such as commenting on refusal. Heerwegh hereby replicates Pondman's (1998) results.

Heerwegh reports that giving unsolicited information can be used in the battle for the most arguments.

Campanelli and Sturgis (1997) also attempted to find empirical evidence for the role of tailoring and maintaining interaction in interviewer behaviour. However, their work revealed that most interactions at the door led either to immediate acceptance of cooperation, or outright refusal. Respondents hesitate in only relatively few interactions, thereby giving interviewers scarce opportunity to practice their arts of persuasion. Becks (2008) has replicated this observation in research in a different mode (telephone interactions). Luiten, Campanelli, Klaassen and Beukenhorst (2008) compared the rates of respondents and nonrespondents who either immediately consented or refused, and found them to be surprisingly similar to the work of Campanelli and Sturgis (1997).

Heerwegh (2000) also concludes that many interviewers never get around to tailoring or maintaining interaction. He suggests that this aspect should be addressed in interview training courses. Interviewers need to be motivated to be persuasive in these kinds of situations too. It is important to provide positive working strategies, and not to suffice with vague advice ('turn the situation to your advantage') or admonitions of what not to do ('don't repeat the objection').

2.6 Cooperation in mail and web surveys

Even more than in CATI or CAPI surveys, the way the potential respondents are approached in web or mail surveys is vital to obtain cooperation. Failure to do so may result in refusal to respond. In the absence of interviewers who can persuade reluctant persons to cooperate, respondents in web or mail surveys have to derive their motivation from the advance letter, the folder and possibly the included or promised incentive. The quality of this approach will be determined if people go to the website and open the questionnaire. The quality and attractiveness of the questionnaire will then determine whether they actually fill it in.

A well designed mail or web questionnaire can help in reducing nonresponse (Bethlehem & Biffignandi, 2012; Dillman, 2009). Both should be respondent-friendly questionnaires, with carefully organized questions in easy-to-answer formats. When constructing a web questionnaire, many design principles for paper questionnaires can be copied. However, the design of a web questionnaire is more complex than the design of a paper questionnaire. Web allows for a wide range of formats, graphics, format, route instructions, and response control mechanisms.

Special attention must be paid to technical aspects. When the survey is (too) sophisticated, it may be possible that the respondent's computer's software does not run the survey properly. If the survey is not accessible, or is difficult to read or to fill in, response rates will suffer.

Access to the questionnaire must be gained by one or more unique identification codes for each sample person or household. This ensures that only selected persons can obtain access. Heerwegh (2005) studied how the login procedure influences response rates. Either one or two access codes had to be entered. The results showed that the number of people who started the survey was higher with a semi-automatic login procedure, but the number of people who completed it was higher if respondents were required to enter more access codes. There was also evidence that sensitive questions were answered more accurately if more access codes had to be entered.

Once the respondent starts, it is important to keep him motivated to continue. According to Dillman (2009) no single question is more important than the first one, especially in web surveys. The quality of this question is most likely to determine whether people will cooperate. It should be salient to the respondent. Sensitive questions are better placed at the end of the questionnaire. Once people have invested time, they are less likely to stop.

One of the features that is used to accomplish that is the progress indicator. This is a device that gives feedback about the progress of the respondent in the questionnaire. The prevailing view is that this information is appreciated by respondents. Bethlehem and Biffignandi (2012) indicate that when progress surpasses the expectation of the respondent, feedback can significantly improve completion rates. When progress seems to lag behind what is expected, feedback reduces completion rates. Heerwegh (2005) found that displaying a progress indicator had no influence on finishing the questionnaire. Couper, Traugott and Lamia (2001) and Crawford, Couper and Lamia (2001) found that displaying a progress indicator had an adverse effect. SN web questionnaires do not show progress indicators either; because of the complicated routing in most questionnaires the progress indicator would behave erratically.

Send reminders

Both in mail as in web surveys, it is vital to send reminders in due course. The response without these reminders will be 20 to 40 percentage points lower than otherwise (Dillman, 2009). A reminder gives the researcher a new opportunity to emphasize the importance of completing the questionnaire, and therefore has to convey more than simply 'we have not yet received your questionnaire'. According to Dillman, each successive reminder must make a stronger appeal to the

respondent than its predecessor. Dillman proposes the following sequence: A card to all sample persons after two weeks that serves as either a reminder or 'thank you'. This card is meant as a friendly reminder to nonrespondents. It intentionally takes the form of a card, since it is more likely to receive attention than a letter, which might end up unopened on the 'to be read' pile.

The next reminder is again a letter, enclosing a new questionnaire or again a login code for web surveys. The letter resembles the original advance letter, but with stronger arguments. If no new questionnaire is added to the reminder, or no reminder of the login code, only half the increase in response is realised, compared with a reminder accompanied by a questionnaire (Dillman, Smyth & Christian, 2009).

Dillman et al., (2009) propose a third reminder after eight weeks, carefully emphasizing the particular importance, through an approach in a different mode, or by delivering the letter in some special way, such as by courier.

Use incentives

As with CAPI and CATI surveys, using an incentive can enhance the response rate. Most studies into incentives with web surveys concern *promised* incentives. As was mentioned in section 2.4, promised incentives in CATI and CAPI surveys have far less impact than unconditional incentives. A meta-analysis by Göritz (2006) showed this to be the case in web surveys as well. Response rates increased by 2.8 percentage points on average after a promised incentive. Incentives had a small effect both on starting a web survey and on finishing once the first page had been loaded. Dillman et al., (2009) however mention that including unconditional incentives in an advance letter for web results in a far greater increase in response rates.

2.7 Interviewer training and monitoring

The interviewer plays a very important role, both in contacting sample persons and in subsequently persuading them to cooperate. It is thus important to train them to perform the work as effectively as possible. Thorough interviewer training has a response-enhancing effect, and is crucial for survey quality. Thorough interviewer training gives attention to how best to approach sample persons, i.e., how often and at what times. The interviewer needs to be trained in getting cooperation:

what to answer to questions or objections, how to tailor and how to maintain interaction. Subsequently, she needs to learn how to administer the questionnaire: reading questions as they are formulated, recording answers, probing incomplete or unclear answers.

Even the best training would be tantamount to casting pearls before swine without monitoring interviewer behaviour for compliance after the course (Fowler & Mangione, 1991). There is no mention in the literature of the effect of monitoring and feedback on response. These instruments are described only in the context of preventing and detecting interviewer falsification, which is when interviewers invent data. SN are nonetheless convinced that these instruments are also vital for achieving a target response. De Vree, De Bie & Luiten (2006) demonstrate that for several fieldwork performance indicators, the mere fact that they were introduced in a monitoring program with targets imposed on both regions and interviewers, was sufficient to secure substantial improvement. Examples include the reduction of partial nonresponse in the labour force survey, and the reduction of panel attrition in the labour force survey. These examples illustrate how attention for interviewer performance and subsequent feedback are significant conditions for maintaining interviewer motivation.

2.8 Conclusion

This chapter describes the measures available to a survey organization to minimize the nonresponse that is within its control. The measures described are 'state-of-the-art', but the art concerned reached maturity in the final decades of the 20th century. Currently, two significant trends are visible in nonresponse research. The first is an understanding that the bias attributable to nonresponse is a far more important research target than nonresponse as such.

The achievement of a high response and a low bias is related directly to the survey costs. Throughout the western world, the effort needed to maintain response rates at an acceptable level is increasing. Not only is more effort needed to make contact, but people are also less willing to cooperate. In order to curb the rising costs, an increasing number of organizations are resorting to the use of a mixed-mode design, where some of the data is collected in a less expensive mode.

Much of the research described in this chapter for CAPI and CATI is currently being repeated with web surveys in the main role. Examples are the role of advance letters in web surveys, the effect of personalization of invitations, and the effect of incentives. The best way to optimize response in mixed-mode surveys remains far from clear. This area of research is still in its infancy, and will reach maturity in the next few years.

3.

Personalisation

in advance letters

does not always

increase

response rates

Demographic correlates in a large

scale experiment

3.1 Introduction¹⁾

Advance letters have long been established as a means to heighten response rates in survey research (e.g., De Leeuw, Callegaro, Hox, Korendijk & Lensvelt-Mulders, 2007; Dillman, 2000; Goldstein & Jennings, 2002; Yammarino, Skinner & Childers, 1991). Numerous studies have tried to determine what makes a successful advance letter. Attention has been given to the wording of the letter, its length, colour, paper quality, mailing quality, signature, salutation and the usage of a survey unit's name (see Dillman [2000] for an overview). The last three elements, naming, salutation and signature are used to personalize advance letters. By using a sample unit's name in addressing the letter and in the salutation (*Dear Mrs. Johnson*), and by hand signing the letter, the survey organisation gives a message of personal attention to the respondent, who will react with heightened attention to the needs of the survey organisation, resulting in higher response rates. Cialdini (2001) shows that people are inclined to reply in a similar return action to certain ways of acting. In the case of a respondent who receives a personalized letter, she/he may feel that since the researcher took the time to personalize the letter she/he should take the time to participate in his/her survey.

Personalisation of letters has been the subject of hundreds of studies and of numerous reviews and meta-analyses (e.g., De Leeuw et al., 2007; Dillman et al., 2007; Fox, Crask & Kim, 1988; Goyder, 1982; Harvey, 1987; Heberlein & Baumgarten, 1978; Scott & Edwards, 2006; Yammarino et al., 1991; Yu & Cooper, 1983). Results have been mixed, however. Worthen and Valcarce (1985) analyzed 26 personalization experiments and found that the results favoured personalization in 18 of the studies, but overall, the effect in favour of personalization was slight. In a subsequent study they failed to find a significant increase in response rates for personally addressed, individually typed, and hand signed letters versus mimeographed form letters with general salutations and facsimile signatures. Reviews by Harvey (1987) and Yu and Cooper (1983) found positive response effects of personalization, while quantitative meta-analyses by Fox et al., (1988), Goyder (1982), Heberlein and Baumgartner (1979), and Yammarino et al., (1991), suggest that personalization has no effect. Dillman et al., (2007) comment that personalization has been operationalized in so many different ways, and with so many different combinations of individual elements (e.g., envelope labels vs. typing of address on envelope, handwritten salutations, postscripts or entire letters, group salutations vs. inserted names, stationery vs. mimeographed letters,

¹⁾ This chapter was published as Luiten, A. (2011). Personalisation of advance letters does not always improve response rates. Demographic correlates in a large scale experiment. *Survey Research Methods*, 5, 11–20.

real signatures vs. stamped vs. printed, black vs. blue contrasting signatures) that it is extremely difficult to develop categories with enough studies included so that results would be meaningful. A recent meta-analysis by Scott and Edwards (2006) focussed on two aspects: including participants' names on letters, and a combination of names and handwritten signatures. In fourteen randomised controlled trials the odds of response when including participants' names on letters were increased by one-fifth. When participants' names and hand-written signatures were used in combination, the effect was more substantial, with an almost 50% increased odds of response, corresponding to an absolute increase in the proportion of questionnaires returned of between 4% and 10%.

The largest body of research entails cover letters, accompanying a mail questionnaire. Far less attention is given to personalization as a means to influence response rates in interviewer mediated survey research (like telephone or face-to-face surveys), where advance letters are used. A number of studies compared the use of participants' name in the letter's salutation (Dillman, Gorton Gallegos & Frey, 1976; Groves & Snowdon, 1987; Traugott, Groves & Lepkowski, 1987). None of these studies showed any influence of this kind of personalisation on response results. The meta-analysis by De Leeuw et al., (2007) used differences between level of personalization in advance letters of 29 studies (operationalized as: individually personalized, addressed to household, or not personalized at all) to show that this kind of personalization does not influence response rates in telephone surveys. Dillman, Smyth and Christian (2009) recommend that all communication with potential respondents be personally addressed to the recipient, although response gain is modest. However, even a modest increase in response rates could signify a substantial cut in costs if it would be followed by an equivalent reduction of sample size. So, even modest increases in response rates may be worth further examination.

What all these studies have in common, is that they do not study whether personalization affects all participants equally. The literature on advance letters shows that response rates of particular subgroups may be lowered by sending advance letters. Goldstein and Jennings (2002), using a listed sample of registered voters found that people between 18 and 29 years old were less likely to participate (17 percent) when they received a letter. They conclude that research should be done into the possibility that different subgroups need different letters. Likewise, Parks, Kennedy, and Hecht (1994) found that letters improve response of those people who are better responders to start with: women, whites and house owners. They state that advance letters "might actually increase differential noncooperation and produce data that are not representative of the population they are assumed to be drawn from". Camburn, Lavrakas, Battaglia, Massey and Wright (1995) find that people with unlisted telephone numbers respond differently to advance letters.

To the extent that personalizing letters may strengthen the effect of the advance letter by heightening the chance that it is read (Couper, Mathiowetz & Singer, 1995), it is possible that differences between subgroups in response behaviour as a result of advance letters are aggravated. Helgesen, Voss and Terpening (2002) provide evidence that the first step of gaining the respondent's attention via the advance letter is highly dependent on how thoroughly the respondent usually sorts through and reads their mail. Using a name on the letter may help getting that attention.

One attempt to differentiate the effect of personalisation for specific subgroups was made by Dillman et al., (2007). In this study, personalisation was compared in general public surveys, versus specific group surveys (like All-Terrain-Vehicle owners). It was found that of the five general public surveys, personalization treatments significantly augmented response rates in two of them. In the four specific group surveys, however, either no effect of personalisation was found, or in one case a reversal was found, where the personalised letters had a lower response than the non-personalised.

In the present study, examining differential reactions to personalisation is taken one step further: in line with research that shows that subgroups may be differentially susceptible to advance- or cover letters, we study whether subgroups in the general population react differently to personalisation. A large scale experiment (n=39,518) was conducted, based on the Dutch Labour Force Survey, a monthly CAPI survey. The LFS uses an address sample of 7000 addresses per month. For this experiment, the address sample was linked to information in communal registries. The availability of registry information on respondents as well as nonrespondents makes the study of response patterns in subgroups possible. The registries supplied the names of inhabitants and other relevant information, like ethnic origin, age, income and household composition. The standard LFS advance letter is addressed to 'inhabitants of *address, number, postal code, community name*'. The standard salutation is 'dear sir / madam'. The personalised experimental letter addressed the household core members by name, and adapted the salutation accordingly. This operationalisation is not the optimum in personalisation. Although the letter is printed on first class stationary in two colour print, the letter is not signed by hand and it has no precise date (but only the month and year), two factors research shows to be of importance (Dillman, 2000; Scott & Edwards, 2006). Dillman acknowledges however, that in large (government) surveys, the size of the samples makes optimal personalisation not realistic.

Link and Mokdad (2005) show that there are differences between subgroups in the recollection of seeing the advance letter. On average 61% of respondents remembered seeing it, but in non-white, young, low educated, and low income groups recollection could be as low as 48%. It is unclear whether 'seeing the letter' is equivalent to reading it, but to the extent that the advance letter is drawn up

so as to optimally help convince sample units to participate, not reading the letter could be the first step toward non-response. Couper et al., (1995) show that in about half of the households, one person sorts the mail prior to reading, and over 60% throw away some mail without opening it. Letters addressed to individuals who are targeted as respondents do not suffer from such problems (Groves & Couper, 1998). Addressing the letter by name may help drawing attention to the letter, and thereby possibly diminishing differences in reading behaviour between groups. However, the unit of observation in the LFS is the household. One letter is sent to address both members of the household core (if more than one is present). If one person sorts through the mail, as Couper et al., show, this may be another person than the one answering the door to the interviewer. To study whether the effect of personalisation is less in households with a two-person core, household composition is analyzed. If the named letter is indeed 'depersonalized' if more than one adult is present in the household, one would expect a larger effect of personalisation in single households.

As will be described below in more detail, linking addresses to inhabitants is not always straightforward. Where addresses contain multiple households, addressing by name is not possible. Also, numerous people fail to register partnership with the community, resulting in unclear family relations. As we strived to accurately name inhabitants, we were quite conservative, which resulted in 20% records that could not unequivocally be linked to names. These unlinked records may resemble unlisted subgroups in RDD dual frame studies. Parsons, Owens and Skogan (2002) found that the listed samples in their two studies were more likely to be white, older, and college educated, but less likely to be married than the unlisted sample. Link and Mokdad (2005) show that in RDD research, numbers that cannot be linked to addresses belong primarily to racial minorities and those aged 18 to 34. They warn for the possibility of introducing bias when response stimulating measures can only be applied to part of the sample. To the extent that naming response letters does have an effect on response rates, this possibility exists in this study too, and will be addressed.

3.2 Method

Advance letters of the Dutch CAPI Labour Force Survey were randomly assigned to addresses. The control condition (n=30,899) consisted of a non-personalised letter addressed to '*the inhabitants of*', with a standard salutation of '*dear sir / madam*'. In the personalised experimental version (n=8,619), the name or names of the

household core were derived from municipal registries and used for addressing the letter.

The Dutch Labour Force Survey is a continuous monthly CAPI survey. Each month a sample of addresses is selected from which during the data collection households, the sampling units, are identified. The target population of the LFS consists of the non-institutionalised population aged 15 years and older, residing in the Netherlands. The sampling frame is a list of all known occupied addresses in the Netherlands, which is derived from the municipal basic registration of population data. The LFS is based on a stratified two-stage cluster design of addresses. Strata are formed by geographical regions. Municipalities are considered as primary sampling units and addresses as secondary sampling units. All households residing at an address, up to a maximum of three, are included in the sample. All persons of 15 years and older in the household are interviewed, proxy is allowed. The LFS has a rotating panel design, with a face-to-face CAPI interview as the first wave, and four subsequent CATI waves. The study reported here concerns the first CAPI wave of a number of months in the 2007 and 2008 LFS.

The sampling frame, derived from the municipal registries, contains the names of the inhabitants. This information is used by Statistics Netherlands for finding telephone numbers. These numbers are used by field interviewers as an aid in contacting sample units, but the names are not used in the fieldwork procedure. The telephone numbers are the ones that can be found by automated search in the records of the Royal Dutch telephone company (KPN), owner of the fixed landlines. Around 35% of addresses can be linked to telephone numbers by this method. By intensive personal search, numbers can be found from other providers, for an additional 25% of addresses. The other 40% either has a shielded number, or an unregistered cell phone. The intensive search is put in for CATI surveys, but for the purpose of aiding field interviewers only the automated search was performed at the time of this research. In this study, the number of households with an unlisted landline (62%) exceeds the number of listed households (38%).

For this experiment, experimental and control addresses were linked with the names of the inhabitants, their sex, age, ethnic origin, household composition, and position in the household. Age, household composition and position in the household were used to determine if persons at an address formed a household, and who was/were household core member(s): one name for singles and single parents, two names for partners. Only one letter was sent per household. Because all household members of 15 years and older are eligible, and while the household core is allowed to respond by proxy for other household members, all addressed household members are potential respondents. If more than one household appeared to live at an address, or if household composition could not be determined, a standard advance letter was sent, addressed to 'the inhabitants of; address; postal code; town'. 80% of addresses could thus be furnished with one

or two names. Sex of the core members was used to formulate the appropriate salutation: dear sir; dear sir, dear madam; but also: dear sirs, etc. A translated advance letter is included in appendix 1.

The demographic information from the municipal registries was aggregated to form household level variables: mean age of the household core, sex composition (male(s), female(s), mixed), household composition (partners, one parent/ single, undetermined) and ethnic background (native Dutch, non-western and western foreigners, and mixed). Interviewer paradata informed us whether or not a fixed landline could be determined for the address. Linking addresses with Statistics Netherlands' area statistics made available information at postal code level, like urban density and income. The analyses in this study focused on subgroups that are mentioned in literature as reacting differently to advance letters: age groups (Goldstein & Jennings, 2002), gender and ethnicity (Parks et al., 1994), and listed telephone numbers (Camburn et al., 1995). Parks et al., also mention homeownership. As registries of homeownership were not yet available for 2008, monthly income per postal code is used here as a approximation. In addition, household composition is studied, to shed light on the effect of naming two persons in one letter.

During five months, ten percent of the LFS sample, and during two further months half of the sample, was assigned the condition 'with name'. Power calculations had determined that this substantial cell filling was needed to be able to reliably distinguish the relatively small response differences expected.

The experimental condition was assigned randomly to addresses. Statistics Netherlands' interviewers send out the advance letters themselves. All interviewers had advance letters of both conditions and were therefore aware of the experimental condition of each address. Addresses in both conditions were treated according to Statistics Netherlands' uniform fieldwork strategy. This strategy prescribes that all first calls must be made during the first half of the fieldwork period, that either the first or second visit should be in the evening or on Saturday, that non-contacts should be visited six times, and that visits should be spread across times and days. Incentives are not given, neither to respondents nor to interviewers. No refusal conversion is attempted. The field interviewers are civil servants employed by Statistics Netherlands, and are monitored rigorously on adherence to these rules.

The linking of addresses with names was done one month prior to fieldwork, so some households might have moved at the time of the fieldwork. Named advance letters would in that case be forwarded to the old inhabitants, and the new inhabitants would not have received a letter. Interviewers were instructed to show the standard letter to the new inhabitants in that case, and to make a note in their fieldwork administration of the event. In spite of regular reminders of this instruction, only four mentions were made of this, two in each condition.

One to two weeks after the initial interview, a sample of respondents and non-respondents (n=3,607) was contacted for a (re-)interview, a standard procedure for the Labour Force Survey. One of the questions asked, was whether the respondent (of the re-interview) read the advance letter. These results are used to analyse if naming letters led to better reading.

3.3 Results

Table 3.3.1 shows response and cooperation results for the experimental group with personalised letter ('Name'; n=8,619) and control group with unpersonalised letter ('No name'; n=30,899). It shows results for the entire sample (all), including those cases for which no name(s) could be determined, and results for those units for which one or more names could be determined (name found). The former results give an indication of the gain in response or cooperation that would be achieved by introducing the measure into standard fieldwork procedure; the latter results show a purer image of the effect of introducing names.

3.3.1 Response²⁾ and cooperation with and without name on advance letter

Advance letter	No name			Name			χ^2
	%	n	SD	%	n	SD	
Response; all	60.6	28,995	2.9	61.5	7,994	5.4	$\chi^2_{(1)} = 2.08, ns$
Cooperation; all	65.9	26,662	2.9	66.1	7,439	5.5	$\chi^2_{(1)} = 0.08, ns$
Response; name found	61.2	23,848	3.2	62.6	7,007	5.8	$\chi^2_{(1)} = 4.41, p < .05$
Cooperation; name found	66.0	22,114	3.2	66.7	6,574	5.8	$\chi^2_{(1)} = 1.15, ns$

As can readily be seen, introducing names on advance letters has no general influence on response and cooperation. In the group where names can be determined, using them has a positive slight but significant influence on response, but not on cooperation³⁾. The effect on response proved to be caused by a higher number of cases that were returned as unprocessed by the interviewer in the

²⁾ Response rate is defined as the number of complete and partial interviews with reporting units divided by the number of eligible reporting units in the sample. This is AAPOR RR2 (American Association for Public Opinion Research, 2008). Cooperation is defined as complete and partial response of contacted eligible cases (AAPOR COOP2).

³⁾ Two-sided tests of significance were used, to allow for the possibility that in some groups cooperation would be lowered by personalisation.

'no name' condition (1.5% vs. 0.3%, $\chi^2_{(6)} = 61.4$, $p < .001$)⁴⁾. As cooperation is the more obvious dependent variable in these analyses however, it will be used from here on.

Reading the advance letter

The re-interview of respondents and non-respondents shows that 85% of respondents and 83% of non-respondents (claim to) have read the advance letter. There are significant differences between subgroups in the number of persons reading the letter: Respondents over 40 years of age read them more than younger respondents (81, 81, 87 and 89% for the four age groups respectively, $\chi^2_{(6)} = 29.63$, $p < .001$); in households of non-western ethnic origin the letter is read less than in households of mixed ethnic origin and native Dutch households (86, 73 and 83% for natives, non-western and mixed households respectively, $\chi^2_{(4)} = 26.38$, $p < .001$); in single parent households the letter is read less than in either single households or households with two adults in the household core. The letter is read least in households of which composition could not be determined (86, 85, 81, and 73% for partners, singles, single parents and undetermined households, respectively, $\chi^2_{(6)} = 28.55$, $p < .001$). Fewer people living in low income neighbourhoods read the letter than people in high income neighbourhoods (83, 82, 88 and 87% for income groups <€1600pm, €1600–1900, €1900–2300 and >€2300pm, $\chi^2_{(6)} = 16.79$, $p < .01$). No differences were found between male households, female households and mixed sex households in reading behaviour.

Naming letters did not lead to a higher percentage of people reading the letter (83% of the named letter was read, versus 85% of the unnamed letter). The difference was not significant.

The effect of naming in subgroups

In order to analyse whether lack of effect of personalisation could be the result of positive effects in some groups, countered by negative effects in other groups, the reaction of different subgroups to naming of letters was studied. Table 3.3.2 shows results of naming in different age, ethnic, gender and income groups and of different household composition.

Judging from these results, the fact that no higher cooperation rates were found when letters were named, was not a resultant of differential cooperation effects in

⁴⁾ If interviewers are unable to handle their entire workload, as a result of illness or other unforeseen circumstances, they consult their regional supervisors to decide which addresses to return. This result shows that supervisors prioritized the experimental addresses.

these specific groups. In general, when names could be found, cooperation rates were slightly higher when letters were named, but in neither group the difference was significant, according to Pearson's χ^2 analyses.

3.3.2 Cooperation with and without name by Age, Ethnicity⁵⁾, Gender, Income and Household composition

	All		Name found					
	no name		name		no name		name	
	%	n	%	n	%	n	%	n
Age								
≤ 25 years of age	70.9	1,260	71.1	376	69.8	755	73.5	274
26-40	66.6	9,039	66.5	2,514	66.4	6,924	67.7	2,122
41-65	65.2	16,492	65.5	4,617	65.5	14,727	65.7	4,284
> 65 years of age	65.8	2,479	66.5	680	66.1	2,326	67.0	652
Ethnicity								
Dutch-natives	67.2	22,699	67.4	6,364	67.2	19,791	67.8	585
Other ethnic origin	56.0	2,583	54.1	722	55.5	1,995	54.7	598
Mixed ethnic origin (Dutch-other)	64.0	3,987	65.5	1,101	64.0	2,946	66.3	884
Gender								
Male(s)	65.0	4,641	65.5	1,267	65.1	4,151	66.5	1,165
Female(s)	63.8	5,896	63.5	1,640	63.5	5,411	63.9	1,532
Mixed	66.6	18,733	67.0	5,280	66.9	15,170	67.6	4,635
Income								
≤ €1600 p.m.	64.1	8,418	62.7	2,280	64.0	6,538	65.0	1,883
€1601-1900	64.9	6,815	65.6	1,876	64.9	5,537	65.3	1,592
€1901-2300	66.8	6,646	69.6	1,773	66.8	5,497	70.1	1,558
>2300 p.m.	68.6	5,771	67.2	1,599	69.2	4,762	68.9	1,401
Household composition								
Partners	67.0	14,137	67.7	4,246	67.1	14,030	67.7	4,217
Singles / single parents	64.7	9,941	64.5	2,515	64.1	7,748	64.9	2,205

n = total of cooperation and non-cooperation.

In the ethnic groups, slightly higher cooperation rates in the named letter condition were found for the native Dutch, and the mixed group, but slightly lower rates for the people of other ethnic origin. Again, differences failed to reach significance, however. When analysing ethnicity, using COOP3 (AAPOR 2008) is a more suitable measure. It measures cooperation of those that are contacted and able to cooperate. Nonresponse due to language problems is in this definition not considered to be non-cooperation. Using this definition, cooperation of households

⁵⁾ The majority of 'other ethnic origin' exists of people of non-western ethnic origin (according to the Statistics Netherlands' definition, these are persons originating in African, Latin-American and Asian Countries [excluding Indonesia and Japan] and Turkey). A small number of people were of western ethnic origin. Although their response behaviour is somewhat different from the group of non-western ethnic origin, collapsing their results with that of the non-western group did not change results.

of non-Dutch ethnic origin was 4.1% lower when an advance letter with name was used ($\chi^2_{(2)} = 3.31, p = .06$). Analyses of household composition (partners, single parents, singles), and urban density showed no relation at all with naming of letters.

Another group that merits further investigation is the group of people with unlisted landline numbers. Having a listed telephone number is highly predictive of response behaviour and even the most predictive explanatory factor for contactability as well as cooperation in a study of the Dutch Survey of Living Conditions (Bethlehem & Schouten, 2004). In the present research, the response rate in the group with the listed numbers was 64.1%, compared to 55.7% in the unlisted group. The unlisted group had a higher noncontact rate, more often language problems and a lower cooperation rate. Unlisted people react differently to advance letters (Camburn et al., 1995) and may well react differently to personalisation as well. Table 3.3.3 shows cooperation rates for named and unnamed letters, in the entire sample, and for those households where names could be determined.

3.3.3 Cooperation with and without Name by Listed or Unlisted telephone number

	All				Name Found			
	no name		name		no name		name	
	%	n	%	n	%	n	%	n
Listed telephone number	67.4	9,738	68.6	3,811	67.6	5,853	69.2	2,438
Unlisted telephone number	65.3	16,718	63.6	3,597	63.7	8,678	63.9	1,934

A logistic regression for the entire sample with Name (2), Listed (2) and a Listed x Name interaction as explanatory variables again failed to show a significant effect of naming, but showed a main effect of Listed Telephone (Odds Ratio = 1.10, $p < .001$, 95% CI 1.04–1.16) indicating that the odds ratio for cooperation is 1.10 larger in the Listed group than in the unlisted group. A significant Listed x Name interaction (Odds Ratio = 1.14, $p < .05$, 95% CI 1.02–1.27), showed that naming had a positive effect (+1.2%) for listed addresses, but a negative effect (–1.7%) for unlisted ones.

When analysing the addresses where names could be determined (independently of whether they were used), again a main effect of Listed Telephone was found (Odds Ratio = 1.13, $p < .001$, 95% CI = 1.07–1.20), that was also qualified by a Name x Listed Telephone interaction (Odds Ratio = 1.13, $p < .05$; 95% CI 1.03–1.27), here signifying that cooperation was higher with the named letter for the listed group, but no difference was found for the unlisted group. No significant three way

interactions between naming, listed phone number and any of the other variables were found.

Who are the ones whose name cannot be determined?

Differences in results between the total group, and the subgroup where names could be determined, are an indication that whether or not names can be determined, is not distributed randomly among subgroups. Indeed, highly significant differences were found in demographic make-up of the persons of whom no name can be determined, versus those with registered household composition. The persons of whom no name could be determined were far more often of foreign origin, or lived in households of mixed ethnic origin ($\chi^2_{(2)} = 739.6$, $p < .001$), were younger than 40 years of age ($\chi^2_{(3)} = 1717.22$, $p < .001$), were predominantly single or, not surprisingly, lived in households of unregistered composition ($\chi^2_{(3)} = 14,235.9$, $p < .001$), lived in apartment flats ($\chi^2_{(1)} = 435.1$, $p < .001$) and more often in regions of high urban density than in the country ($\chi^2_{(6)} = 501.0$, $p < .01$).

3.4 Summary and discussion

This study was set up to examine whether personalising advance letters by adding names and appropriate salutation, affects the survey cooperation of subgroups in the general population differently, in analogy to findings that subgroups react differently to advance letters. Differential reactions could be an explanation for the mixed findings in the literature on personalisation of advance or cover letters. Advance letters of the Dutch CAPI Labour Force Survey were randomly assigned to addresses. The control condition consisted of a non-personalised letter addressed to *'the inhabitants of'*. In the personalised experimental version, the name or names of the household core were derived from municipal registries and used for addressing the letter. Information about the inhabitants' gender was used to formulate the appropriate salutation. A re-interview addressed the issue whether the advance letter was read by more households when the household was addressed by name. The subgroups studied were the same ones the literature indicated as differentially reacting to advance letters, i.e., different age, ethnic, gender, and income groups, and groups with or without a listed telephone number. Results show that there was hardly any difference in the overall level of cooperation following the personalised letter. Subsequent inspection of subgroup cooperation showed that it is unlikely that lack of overall results is the resultant of

differential reactions of subgroups to advance letters. In almost all age-, ethnic-, gender- and income groups, personalisation led to a small increase in cooperation, but the difference failed to reach significance. The hypothesized depersonalisation in case more than one person was addressed, was not found to be an issue; in households with a one-person core the difference between named and unnamed letters was not larger than in two-person core households. The only differential effect was found for the groups with listed versus unlisted landlines, in the sense that listed households reacted positively to personalisation, while cooperation in the unlisted households was unaffected. The higher cooperation in the group with listed phones, although significant, was slight. Inspection of the reasons people gave when refusing to participate, showed that privacy concerns were significantly more often expressed when letters were personalized.

The role of advance letters and the possible role of personalisation therein is smaller in interviewer-mediated surveys than in mail surveys, although advance letters in interviewer-mediated surveys do exert an independent influence on the outcome (Groves & Snowden, 1987). One possible explanation for the absence of an effect of naming could be the generally high response in face-to-face survey research. If the response rate is high, a ceiling effect could exist that limits the potential of the personalization treatment. The same could be caused by a survey request by a respected institution as Statistics Netherlands. Neither seems to be the case in this study, however. Not only are the response rates not that high (about 61%), but also, other interventions in the Statistics Netherlands LFS showed that response rates can be raised substantially. An experiment with unconditional incentives in the form of postal stamps, included in the advance letter, for example, showed that response rates were raised by 8 percentage points (Wetzels, Schmeets, Van den Brakel & Feskens, 2008; Feskens, Hox, Schmeets & Wetzels, 2008). It is quite probable however, that the interviewer mitigates the effect of personalization. It would be worthwhile to study possible differential reactions to naming in mail or web surveys.

The interviewers in this experiment were not blind to the conditions. This could in principle have exerted an influence on the results. However, once the letters were sent, the information was no longer available for them, unless they made an express effort to copy the information. In the bulk of their workload, it is highly unlikely that they would remember which address was in what condition.

The results showed that the only significant difference between subgroups in the effect of naming was found for listed versus unlisted telephone numbers. However, the amount of listed numbers was relatively small. With more effort, more numbers could have been found. A part of the numbers that are classified as being unlisted, would have been classified in the other category if more effort had been made. Who is listed or not is always a question of definition of the list, and is highly dependent on the registries available for search. The group that was found in the automated search is a group bound to diminish even further in size with the higher

penetration of cell phones and internet telephone. Whether this would mean that the slightly positive effect of naming in this group would disappear altogether, should be addressed in time.

The call-back survey (n=3,607) revealed that, contrary to expectation, the named letter was not read better than the unnamed one. In both conditions about 16% of respondents and nonrespondents appear not to have read the letter. It may be that the mechanism underlying not reading is different in the two groups. In the named group, letters may have been forwarded in case of households having moved in between the drawing of the sample and the fieldwork. In the unnamed group the letter may have been discarded as bulk mail, but the net result is the same. In both conditions the letter was read very well, compared to other findings (e.g., Groves & Snowden, 1987; The Australian Bureau of Statistics, 2002).

Link and Mokdad (2005) and Couper et al., (1995) warn of the danger of introducing bias when response stimulating measures can only be applied to part of the sample. In this study, a substantial part of the sample could not be addressed by name, either because an address consisted of multiple households, or because the family relations were unclear. These households were not spread evenly among demographic groups. On the contrary, these households were predominantly peopled with those who are traditionally under-represented in survey results as a result of low response rates: the young, the single, the highly urban, the apartment dwellers and the people of non-western ethnic origin. Had the result of the personalised letter been unequivocally positive, this circumstance would have had an influence on sample composition, with an increased potential for non-response bias. Now we find the situation that in part of the sample, the intervention has a positive effect on response, but in another part of the sample either has no effect or even a negative effect. Although the net effect of the measure neared nil in this experiment, such an outcome could still be desirable in term of bias reduction, if the stimulated part of the sample would coincide with the underrepresented part of the sample. In this case, however, it did not. Positive reactions were seen in the part of the sample that is already over-represented: the persons with a listed telephone, while negative reactions or no reactions were found in the unlisted group. In terms of response rates only, the results suggest that it could be advisable to use a differentiated fieldwork approach, in which unnamed advance letters are sent to addresses with unlisted telephones, and named letters to the listed ones. In terms of bias control, that would be unadvisable though. The results may well signify that, even though net response did not change, the sample composition worsened as a result of addressing households by name.

Future research into the effect of response stimulating measures should be mindful of the possibility that subgroups are differentially influenced and be equally mindful of the possibility that bias is introduced by well intentioned interventions.

4.

**Tailored fieldwork
design to increase
representative
household
survey response**

An experiment in the Survey of

Consumer Sentiment

4.1 Introduction¹⁾

For years survey practitioners have struggled to attain high response rates as a safeguard against biased survey results. A number of circumstances forces us to rethink this strategy. Response rates in household surveys are getting lower (De Leeuw & De Heer, 2002). More effort is required, so the costs of getting acceptable response rates rises (Starick & Steel, 2012). Also, response rates are not necessarily good indicators of non-response bias (Curtin, Presser & Singer, 2000; Keeter, Miller, Kohut, Groves & Presser, 2000; Groves & Peytcheva, 2008; Heerwegh, Abts & Loosveldt, 2007).

Non-response may have different implications for different variables within one survey. The mechanisms causing non-response may be different for different groups. This implies that survey designs need to minimize potential bias across various domains of the key survey variables. Recent research addresses these issues. Groves (2006) advises to replace the blind pursuit of high response rates by informed pursuit, guided by knowledge of the relation between response-stimulating measures, the groups sensitive to them, and their influence on survey estimates.

Groves and Heeringa (2006) use the term 'responsive design' for survey designs where the status and the treatment of sample units are made dependent on an estimate of their contribution to the final survey result, relative to the costs of obtaining that result. Characteristic for the approach is that analyses of costs and errors calculated during fieldwork may lead to decisions and design alterations in mid-course (Groves et al., 2006; Mohl & Laflamme, 2007; Gambino, Laflamme & Wrighte, 2010). Responsive survey designs are especially useful in settings where little is known about the sample beforehand or little information about the effectiveness of treatments is available from historic data. Sometimes, however, there is information available on sample units from registers or prior panel rounds. Also, ongoing surveys may yield information about the response propensities of groups of sample units. Such information can be used to design a tailored or differential approach before the survey starts.

Several authors have studied the use of prior knowledge in designing differential designs. Wagner (2008) introduced the terms adaptive and dynamic design to describe differential survey designs tailored to the characteristics of sample units. Previous experience with similar sample units in similar surveys provides insight

¹⁾ This chapter was published as Luiten, A., and Schouten, B., (2013). Tailored fieldwork design to increase representative response. An experiment in the Survey of Consumer Satisfaction. *Journal of the Royal Statistical Society, series A*, 176, 169–191. The experiment in this paper was part of the RISQ Project, financed by the 7th Framework Programme (FP7) of the European Union. Cooperation Programme, Socio-economic Sciences and the Humanities, Provision for Underlying Statistics.

in how to treat each sample unit. Adaptive design allows treatment to vary with time, using rules specified before data collection. Peytchev et al., (2010) similarly describe how experience in previous panels is used to prioritize sample units with a low predicted response propensity to diminish non-response bias.

Like responsive designs, adaptive and dynamic designs may also base their design on paradata such as interviewer observations on housing units or neighbourhood characteristics, and on process and administrative data produced as auxiliaries to survey data collection. Examples are: timing and outcome of call attempts, the nature of the interaction with household members, how long the interviews took, the reluctance of the interviewee, and the mode of communication (Couper, 1998; Couper & Lyberg, 2005; Lepkowski et al., 2010; Durrant, D'Arrigo and Steele, 2011). Whether designs are altered during fieldwork, or whether they are tailored to specific subgroups before fieldwork commences, what these approaches have in common is a differential fieldwork strategy, aimed at minimizing non-response bias and survey costs, while at the same time trying to maintain survey response at a level necessary for precise survey estimates (Schouten, 2010). While calculating the response rate is a relatively straightforward task (e.g., AAPOR, 2008), calculating costs and bias is far less so (Groves, 2004). Bethlehem (2002) defines non-response bias as the ratio of the covariance between the survey variable and the response propensity to the mean propensity. If there is no correlation between a target variable and response behaviour, the estimator is approximately unbiased. On the other hand, the stronger the relation between a target variable and response behaviour, the larger the bias. The size of the bias also depends on the amount of nonresponse.

However, one may encounter difficulties when using the formula proposed by Bethlehem (2002) to determine bias in a survey. The difference between respondents and non-respondents may be unknown for any number of target variables. Also, different sample estimates within the same survey can be subject to different non-response biases, making it difficult, if not impossible, to design a fieldwork strategy for minimizing overall bias.

These considerations led Schouten, Cobben and Bethlehem (2009) to propose an alternative quality measure, the Representativity Indicator or R-indicator, that measures the similarity between the response and the sample of a survey. The response obtained in a survey is defined to be representative if the individual response propensities are equal for all units in the population. Let ρ_x denote the response propensity function for variable X , say age, i.e. $\rho_x(x)$ is the probability of response of a population unit with value $X = x$. X in general is a vector of relevant auxiliary variables, e.g. age and sex. The distance between two vectors of response propensities ρ_1 and ρ_2 is expressed using a function d

$$d(\rho_1, \rho_2) = \sqrt{\frac{1}{N} \sum_u (\rho_{1,i} - \rho_{2,i})^2}, \quad (1.1)$$

where N is the population size, U the population and i the unit.

Given (1.1), the indicator of representativeness, or R-indicator, is defined as the distance between ρ_x and the survey response rate ρ .

$$R(X) = 1 - 2d(\rho_x, \rho_0) = 1 - 2S(\rho_x) \quad (1.2)$$

d is the standard deviation S of the response propensities for different values of X . A transformation in (1.2) is made so that $R \in [0, 1]$. A value of 1 represents perfect representative response, while a value of 0 indicates the largest possible deviation from representative response.

The R-indicator describes the representativeness of the response given the whole of a vector of auxiliary variables. However, it is crucial to know which (vector of) X and which category within X is responsible for the deviation of representativeness when developing an adaptive or tailored fieldwork design, or when monitoring fieldwork in view of a responsive design (see e.g. Schouten, Luiten, Loosveldt, Beullens & Kleven, 2010). To that end, so called partial R-indicators can be employed. Unconditional and conditional partial R-indicators are distinguished. Unconditional partial R-indicators describe the impact of each variable separately, while conditional partial R-indicators adjust the impact of one variable for the impact of other variables.

Schouten, Cobben and Bethlehem (2009) describe how the R-indicator can be used as a tool for comparing different surveys, surveys over time, or different data collection strategies and modes. The present study was set up as part of a large international research program, the RISQ project (Representativeness Indicators for Survey Quality, www.risq-project.eu), aimed at developing R-indicators and studying their use in monitoring and controlling fieldwork. See Schouten and Shlomo (2010), Schouten, Cobben and Bethlehem (2009), Schouten, Morren, Bethlehem, Shlomo and Skinner (2009), Shlomo, Skinner, Schouten, Bethlehem and Zhang (2009), and Shlomo, Skinner, Schouten, Carolina and Morren (2009) for details.

In this paper, we describe how to get a more representative sample by using a tailored survey design. In an experimental setting, a standard uniform survey design was compared with a tailored adaptive design. Paradata from previous Consumer Sentiment Surveys (SCS) and information on sample units available in registers were used to predict the contact and cooperation propensities and at home patterns of sample units in a new wave of the SCS. The tailored design sought to reduce the variability in response propensities of socio-demographic and socio-economic groups. It did so by stimulating response of sample units with low response propensity, while curbing those with high response propensity. This was done by assigning sample units to different modes (web, mail) in an initial

approach, and by differentiating the timing and number of CATI contact attempts, and the interviewers assigned to specific sample units in the follow-up approach. Two constraints were important in developing the design: fieldwork should cost no more than the standard SCS, and the response rate had to be maintained. A major consideration in the design was to get a representative response in each step of the fieldwork: the first (web / mail) wave, the CATI contact phase and the CATI cooperation phase. That meant that we would sometimes curb the contact chance of sample units with a high contact propensity, while at the same time stimulating the same group with a low cooperation propensity. In section 4.2 we outline the design of the experiment. In section 4.3 we describe the results of the experiment in terms of the response rates attained, the representativeness of the response in control and experimental groups, and the costs. Section 4.4 discusses the findings.

4.2 Method

We used the Survey of Consumer Sentiment (SCS) as a vehicle for the experiment. The SCS is an ongoing cross-sectional CATI survey, conducted among 1500 households of whom a listed telephone number can be found. Questions may be asked of any person in the household core (head of household or partner). The questionnaire takes about eight minutes to complete. Questions are related to sentiments about the household's economic situation and expenditure. Fieldwork is conducted in the first ten workdays of each month. This experiment was conducted in October and November 2009, alongside the regular SCS, during the same 10 day fieldwork periods, with a similar sampling method and sample size, and the same interviewers. The regular SCS served as the control group for the experimental manipulations.

To achieve better representativeness at the same costs, we chose a mixed mode design for the experiment, in which a mail and/or web round was followed by a CATI follow-up of non-respondents. Mail and web questionnaires cost less to administer than CATI questionnaires, and can reach respondents who are otherwise hard to contact and/or convince to cooperate. As it is not feasible to conduct a mixed mode design as well as the CATI follow-up within ten days, we did the web/mail part of the survey a fortnight before the first ten days of the month traditionally reserved for the SCS. Sample units received an advance letter with a web link and/or a mail questionnaire. One week later, we sent a reminder. Another week later, we started the CATI follow-up.

The fieldwork strategy of the experiment was based on the response propensities of sample units in two datasets. First, historic SCS data were used to identify groups with low, medium and high contact and cooperation propensities in this telephone survey. The dataset contained paradata about the response behaviour of about 18,000 sample units. We determined for all sample units whether they were contacted and cooperated, how many attempts were needed, and at what time these attempts were made. The propensity to respond in either a web or mail mode was gauged from the paradata of another survey: the Safety Monitor 2008 (Kraan, Van den Brakel, Buelens & Huys, 2009). In 2008 the previously single mode annual CAPI survey was redesigned to a mixed mode web, mail, CATI, CAPI design with a net sample size of 62,803 respondents. First the sample units were invited to complete a web questionnaire. A mail questionnaire was available on request. Groups with a high propensity to cooperate in the SCS turned out to also have a high propensity to cooperate in web surveys, while the opposite was true for groups with a low cooperation propensity in the SCS. Web response in the Safety Monitor of the group with a high propensity to cooperate in the SCS was 31.3% , while web response in the group with a low propensity to cooperate in the SCS was 4.8%. On the other hand, the mail response was relatively high in the group with low CATI cooperation propensity (13.5%), against 6.4% in the group with the high CATI cooperation propensity. Our conclusion was that we needed a web questionnaire to cut the costs of the tailored design, but also a mail questionnaire to get cooperation from households with the lowest cooperation propensities.

The web/mail wave

With the aim of representativeness in the first wave in mind, sample units with a low cooperation propensity received a mail questionnaire, sample units with a high cooperation propensity received an invitation to the web survey, while the middle groups were given a choice. The historic Safety Monitor data showed that we should not expect a substantial web response from the group with the lowest cooperation propensity, but could expect a relatively high mail response. This group mainly consists of elderly people, often without access to the web (Statistics Netherlands, 2011), and (first generation) ethnic minorities. We expected that the shorter, simpler advance letter of the mail-only condition and the short, simple paper questionnaire could persuade this difficult group to participate. The group with the highest cooperation propensity hardly used the mail option in the Safety Monitor. Because we could expect a relatively high response in the web mode from this group, and because of costs considerations, we decided not to send this group a paper questionnaire. The groups between these two extremes were given the choice, and received an invitation to the web survey as well as a questionnaire on paper.

The telephone wave

Non-response from the first wave was followed up by CATI. We attempted to stimulate cooperation and contact for groups with low cooperation and contact propensities, and curb those for groups with high cooperation and contact propensities.

To influence the chance of making contact, we defined different call schedules for the different contact propensity groups. Groups with a high contact propensity were primarily called during the day and were started later in the fieldwork period. Apart from freeing valuable capacity for evening calls, this was also cost effective, as daytime shifts are paid 20% less than evening shifts. Households with the lowest contact propensity, on the other hand, were to be called in every shift (morning, afternoon and evening), every day of the fieldwork period. The group with the low middle propensity was called in the evening for the first two contact attempts. Subsequent attempts were made alternating between day and evening. The group with the high middle contact propensity received the same default treatment as the control group, the regular SCS.

The rationale for these call schedules was based on analyses of paradata of historic SCS data and CAPI surveys (Van Veen, 2004; Luiten, Schouten & Cobben, 2007). The group with a high contact propensity consisted largely of elderly people, who can be reached during the day and usually need only one to two contact attempts, hence the decision to call during the day and start fieldwork later. The strategy for the group with the lowest contact propensity, to call every day in every shift, obviously optimizes chance of contact. The third group consisted largely of working households with younger children. There was a greater chance of contacting them in the evening, but if the first two attempts failed, we would spread the calls. We manipulated the assignment of sample units to specific interviewers in order to influence the probability of cooperation. Based on SCS paradata, interviewers were classified according to their response rates achieved in 2008 and 2009 (respective response rates: 82%, 76% 72% and 66%). The best interviewers called the households with the lowest cooperation propensity. The interviewers with the lowest response rates called those with the highest propensity. And the group in between called on the middle group. The hypothesis was that low cooperation propensity would be stimulated, and high propensity curbed. If their workload permitted, interviewers could always call 'easier' addresses, but never 'more difficult' ones. The assignment of groups of addresses to groups of interviewers was handled by the CATI management system. See the Blaise CATI guide (Westat, 2004) for details on creating differential call schedules and allocating specific interviewers to specific addresses.

The selection of auxiliary variables for the tailored design

The objective in this experiment is to improve the representativeness of the survey response. But for which variables do we want the response to be representative? Auxiliary variables may relate to response behaviour, to one or more of the key survey variables, or to the main publication domains. By the latter we mean subpopulations that appear as marginals in publication tables and other publication statistics. When the assessment of response representativeness is used to compare multiple surveys, then it is necessary to select variables that relate to response behaviour, and are generally available in many surveys (Schouten, Cobben & Bethlehem 2009). For use in tailored survey designs, it is important that variables relate either to the key survey variables or to the main publication domains (Bethlehem and Schouten, 2009).

Both SCS and experimental samples were linked to the Social Statistical Database of Statistics Netherlands. This database is an integrated register based on registrations of all kinds of subjects. It contains administrative information on individuals, households, jobs, benefits, pensions and income. The sample addresses were matched on the basis of a precise combination of address, house number, and date of contact. The variables used for this experiment are related to the key variables of the SCS. Table 4.2.1 summarizes them.

The registers contain information on individuals. As the SCS is a household survey, the individual level information was aggregated to household core level (head of household and partner). So, the variables ethnic group and sex have a category to indicate a mixture of the categories on the personal level (e.g., mixed native-ethnic minority). Some information is only available at the postal code level, which is a quite narrow geographical area around the sample unit's house.

4.2.1 Registry data linked to the Survey of Consumer Sentiment

Variable	Categories
Household level	
Ethnic Group	Native Dutch, Moroccan, Turkish, Suriname / Netherlands Antilles, other non-western, other western, mixed and unknown. For the present analyses aggregated to native, ethnic minority, mixed and unknown
Sex	All male, all female, mixed, unknown
Average age of household core	15-30;31-44;45-65; over 65, unknown
Type of Household	Single, partners without children, partners with children, single parents, unknown
Postal code area level	
Degree of urbanization	Very strong, strong, moderate, low, not urban, unknown
Percentage non-western non-natives	Very high, high, average, low, very low, unknown
Average monthly income	Quartiles

Each variable has a category 'information not available'. This has to do with linking sample units to registers. Registers are never entirely up to date: people move, dwellings are built or demolished, and unregistered people may lead to unavailable information at the individual, household or postal code level. Rather than treating these absent data as missing values, they are incorporated as meaningful values. The amount of absent data for each category is about 5%, with the exception of ethnic group, where it amounts to almost 11%.

Defining groups with differential contact and cooperation propensities

We determined which groups are over or under-represented in the historical SCS data by calculating partial R-indicators. Contact and cooperation propensities were calculated separately because measures to stimulate contact may be different from measures to stimulate cooperation. See appendix 2 for these partial R-indicators. These propensities were then projected on the sample units for the experiment. Cooperation was defined according to AAPOR definition COOP2 (AAPOR 2008) as the number of complete and partial interviews divided by the number of interviews (complete plus partial) plus the number of non-interviews that involve the identification of and contact with an eligible respondent (refusal and break-off plus other). Contact was defined according to AAPOR definition CON1, which assumes that all cases of indeterminate eligibility are eligible.

We used a sum score to determine if the expected contact and cooperation propensity was low, medium or high. For example, the partial R-indicators showed that elderly households, low income households, households of non-Dutch origin, households in a neighbourhood with a high percentage of people of non-Dutch origin, and singles were less likely to participate than the other households. A household would receive a 'risk point' for each of these socio-demographic groups it belonged to. The more risk points, the lower the cooperation propensity. I.e., a low income elderly household had a lower cooperation propensity than a high income elderly household.

We did a similar exercise with chance of contact. Young households, singles or partners without children, households in highly urban areas, ethnic minority households, and households living in neighbourhoods with a high percentage of ethnic minorities appeared to have a low contact propensity. Again, the more 'risk points' the lower the contact propensity. Based on these analyses, each sample unit was classified as having a high, medium high, medium low or low contact propensity and having a high, medium high, medium low or low cooperation propensity. Table 4.2.2 shows non-contact and cooperation rates for these groups in historic SCS data.

4.2.2 Cooperation and contact rates in groups with low, medium and high contact and cooperation propensity

Propensity	Cooperation	Propensity	Contact
Low cooperation propensity	56.5	Low contact propensity	88.5
Low medium cooperation propensity	67.5	Low medium contact propensity	92.5
High medium cooperation propensity	72.5	High medium contact propensity	94.2
High cooperation propensity	78.5	High contact propensity	95.7

Fieldwork in the control group

The regular SCS is a single mode – telephone only – survey. No information of the characteristics of the households is available beforehand. All households have an equal probability of being selected in the day batch, although households with whom appointments are made are prioritized. About 80% of the fieldwork is performed during the evening shifts. During daytime shifts, an interviewer is present to call appointments made for daytime. He or she uses spare time to phone other numbers. Interviewers are assigned to the SCS on the basis of availability, not ability or experience.

Supervisors determine on a daily basis whether the work progresses well and whether it makes sense to call an address additional times. The decision is based on the overall response rate. An advance letter is sent some days prior to commencing fieldwork, same as in the experimental group. No incentives are given or promised, and no attempt is made to convert refusals in the regular survey or the experiment.

4.3 Results

Response

Table 4.3.1 shows response results for the regular Consumer Sentiment Survey and the experiment.

In both months, the number of response cases was slightly higher in the experimental group, but the difference was not significant. Truly ineligible cases (0.8% in the control group and 0.5% in the experimental group) were collapsed with cases of unknown eligibility.

4.3.1 Response results in the SCS and experimental group

	SCS		Experiment	
	n	%	n	%
Results				
Ineligible	225	7.5	144	4.8***
Non-contact	196	6.5	183	6.1
Not present during fieldwork period	73	2.4	62	2.1
Not able (ill, dementia)	115	3.8	122	4.1
Language problems	40	1.3	26	0.9
Refusal	467	15.6	548	18.3*
Response	1,884	62.8	1,915	63.8
response WEB-PAPI			1,081	36.0
response CATI			834	27.8

*** p < .001.

* p < .05.

The ineligible cases turned out not to be households. The cases of unknown eligibility have disconnected telephone numbers, or numbers that do not belong to the sample address. As these cases are not followed up, they are called ineligible, but are counted as non-response. The collapsed amount of ineligibility was significantly less in the experiment than in the control group ($\chi^2_{(1)} = 19.37$, $p < .000$). As the percentage of ineligibility is typically stable within the SCS across months²⁾, the lower percentage of ineligible addresses in the experimental group can only be attributed to mail or web participation of households that we would otherwise not have been able to reach, because their number was disconnected. Analysis of the paradata of the historic SCS shows that disconnected numbers are found mostly in households with a high non-contact and non-cooperation propensity. Sending a mail questionnaire to these addresses contributed substantially to a better representative response. Surprisingly, the number of refusals was higher in the experiment than in the control group ($\chi^2_{(1)} = 4.23$, $p < .05$). In the section on experimental manipulations (page 73) we will elaborate on this result. No differences were found in the other non-response categories.

Predicted contact and cooperation

Before interpreting the results of the experimental manipulation, we evaluated if the estimated cooperation and contact propensities proved to be predictive of the actual outcomes. Table 4.3.2 shows cooperation and contact rates of propensity groups in the regular SCS. The prediction proved to be quite accurate: so the higher

²⁾ Mean percentage of ineligible sample units in 2009 is 8.4, SD = 1.3.

the predicted contact propensity, the higher the actual contact rates and the higher the predicted cooperation propensity, the higher the actual cooperation rate.

4.3.2 Cooperation and contact rates in groups with low, low medium, high medium and high contact and cooperation propensities in the SCS and control group

Propensity	Cooperation	n	Propensity	Contact	n
low cooperation propensity	62.7	630	low contact propensity	84.2	814
low medium cooperation propensity	68.4	493	low medium contact propensity	94.5	455
high medium cooperation propensity	75.3	674	high medium contact propensity	95.7	896
high cooperation propensity	79.2	1,100	high contact propensity	96.9	732

Representativeness

In this section we examine the effect of the adaptive design on measures of representativeness, the R-indicator and partial R-indicators. Table 4.3.3 shows the value of the R-indicator for the response compared to the sample, as well as the R-indicator for each step in the fieldwork process: the representativeness of the eligible part of the sample, of those contacted from those eligible, of those able to cooperate from those contacted³⁾, and of those actually cooperating from those able to cooperate. As the table shows, the R-indicator of each subsequent step is higher in the experiment than in the control group, with the exception of 'being able to cooperate'. Only for the R-indicator of response do the confidence intervals not overlap, however, and the one-sided null hypothesis $\zeta_0: control - pilot \geq 0$ is rejected at the 5% level.

4.3.3 R-indicators and 95% confidence interval for eligible, contacted, able and cooperating cases and overall response in SCS and experiment

	SCS			Experiment		
	n	R	(ci)	n	R	(ci)
Sample	3.000			3,000		
Eligible	2.774	0.84	(0.813-0.865)	2,856	0.88	(0.856-0.905)
Contacted	2.578	0.83	(0.801-0.856)	2,673	0.87	(0.842-0.895)**
Able	2.350	0.86	(0.832-0.881)	2,463	0.85	(0.831-0.877)
Cooperation	1.184	0.87	(0.842-0.896)	1,915	0.89	(0.862-0.911)
Response	1.184	0.77	(0.743-0.799)	1,915	0.85	(0.821-0.872)*

* p <.05.

** p <.10.

³⁾ Sample units able to cooperate speak the language sufficiently, and are not too ill.

Analysis of the partial R-indicators shows how the experimental manipulations affected sample composition, see table 4.3.4. Partial indicators ideally have values equal to zero. Large unconditional partial indicators show that a variable has a strong impact on representativeness. Negative values indicate under-representation, positive values over-representation. Large conditional partial indicators correspond with a large impact even after conditioning on the other auxiliary variables. Contrary to the situation for the unconditional partial indicators, positive or negative sign cannot be assigned to conditional partial indicators. This is because the sign may be different for each subclass of X. In some subclasses a certain age of the head of the household may have a positive effect on response while in others it may have a negative effect.

The following section illustrates the use of partial R-indicators in evaluating the effects of the experimental manipulation. We will not go into detail for all variables and all columns, but will illustrate the interpretation with the variables sex and age. No estimator for the variance of a partial R-indicator has as yet been developed⁴⁾, so we cannot draw strong conclusions about the extent of the deviation from the representative response. The analysis illustrates, however, that all auxiliary variables in the experiment deviate less from representativeness than in the control group in the unconditional analysis, and also mostly so in the conditional one.

Table 4.3.4 shows the results of the analysis of unconditional partial R-indicators for each step in the fieldwork process, and the conditional partial R-indicators for response only. For each auxiliary variable, the italic value is the composite contribution of the variable to representativeness; the other values describe the positive or negative contribution of the categories of the variable. For example, in the Response columns, the unconditional partial R-indicators show that the variable with the largest deviation from representativeness is sex, both in the control group and, to a lesser extent, in the experiment.

The category level information shows that households that are all male and all female are underrepresented, while mixed gender households are overrepresented. As single sex households are mostly single households, and often either young or old, conditioning upon the other auxiliary variables removes part of the impact of the variable, although sex is still the variable with the largest deviation in the control group.

⁴⁾ The methodology to calculate confidence intervals has become available after the article was published.

4.3.4 Unconditional and conditional partial indicators for SCS and experiment

	Unconditional R-indicators SCS				Unconditional R-indicators Experiment				Conditional R-indicators		
	eligible	contact	coop- eration response	response	response web/ mail	eligible	contact	coop- eration response	response	response SCS	response experiment
Age	59 ¹⁾	52	35	58	62	43	33	21	36	24	13
< 30	-26	-41	17	-25	-10	-18	-12	14	-10	25	3
30-44	-17	-13	14	1	-34	-11	-23	12	0	5	8
45-64	13	8	7	29	3	9	15	-8	22	21	6
65>	26	22	-26	-10	43	20	10	-7	-10	6	1
no information available	-42	-17	-2	-43	-28	-30	-8	3	-25	2	1
Sex	142	108	41	209	118	105	71	58	134	31	12
Male(s)	-18	-37	-1	-43	13	-5	-30	25	-3	40	6
Mixed	21	27	7	54	2	15	20	-9	30	27	1
Female(s)	-2	-8	-13	-38	-0	-1	-5	-4	-28	26	8
no information available	-51	-27	5	-53	-28	-37	-9	-0	-34	0	0
Household composition	51	49	18	88	53	40	38	29	52	22	18
Single	1	-28	0	-37	4	-0	-27	10	-33	6	14
Partners, with children	15	20	2	32	37	18	14	-16	17	4	5
Partners, no children	11	20	7	37	-32	3	18	-3	25	4	4
Single parent	-16	-5	-16	-28	3	-13	-7	22	4	4	11
no information available	-45	-29	-5	-57	-20	-33	-11	4	-26	31	1
Ethnic group	57	33	13	71	35	42	15	17	43	17	15
Native Dutch	14	10	-6	15	1	12	5	-1	13	8	7
Mixed	8	3	8	23	21	7	3	-8	9	3	2
Foreign	-19	-16	7	-38	-2	-16	-11	14	-22	18	14
no information available	-51	-27	5	-53	-28	-37	-9	-0	-34	0	0
Income in quartiles	51	29	14	67	49	37	14	14	54	21	28
<1600	-1	-4	-5	-21	-10	3	2	-5	-23	13	24
1600-1900	5	4	-9	-7	-13	-1	-7	-7	-11	5	15
1900-2300	4	3	1	12	-4	5	5	-0	14	2	8
2300>	10	7	9	31	36	9	4	11	30	22	28
no information available	-49	-28	4	-53	-28	-35	-10	-1	-34	0	2

4.3.4 Unconditional and conditional partial indicators for SCS and experiment (end)

	Unconditional R-indicators SCS				Unconditional R-indicators Experiment				Conditional R-indicators		
	eligible	contact	coop- eration	response	response web/ mail	eligible	contact	coop- eration	response	response SCS	response experiment
Urban density	18	30	16	32	28	31	15	26	24	14	14
very strongly urban	-6	-21	9	-19	5	-6	-13	18	0	0	8
strongly urban	-11	-6	-1	-16	14	4	2	-0	-0	10	1
medium urban density	6	-0	7	13	-9	6	4	6	10	5	3
low urban density	9	11	-9	12	5	5	4	-16	-2	2	6
No urban density	2	16	-6	10	-4	7	4	-5	5	1	1
no information available	4	5	-0	5	-20	-28	-4	1	-21	1	2
% non-western foreigners in area	47	38	35	60	33	34	18	18	25	15	5
Less than 5%	14	17	-4	22	14	12	9	-7	8	2	1
5-10%	9	3	-2	8	6	8	-1	1	6	1	0
10-20%	-3	-4	-16	-15	-9	-1	-15	7	0	7	0
20% and more	-12	-23	30	-13	-12	-11	-3	14	-6	3	1
no information available	-42	-26	-1	-51	-24	-29	-3	-4	-22	10	1

¹⁾ Unconditional and conditional R-indicators * 1000.

Inspection of the category level variables shows that this is caused by the under-representation of single males. In the experiment, deviation from representativeness of sex has become smaller, and especially the males are far better represented. The unconditional R-indicators show that men in the experiment cooperated relatively well in the web / mail first round, and were also overrepresented in overall cooperation. They were still underrepresented in contact, however. This results in a nearly perfect representation of men in the response of the experiment.

Another illustration concerns age, a category explicitly targeted in the design. Young households were stimulated in the contact phase, while elderly households were stimulated in the cooperation phase. The partial R-indicators show that the adaptive design was successful in augmenting representativeness of this variable. The unconditional R-indicators at the (italic) variable level for contact, cooperation, and finally response, are lower for the experiment than the control group. Unconditionally, the under-representation of the young households is lower in the experiment than in the control group (-12 vs. -41). The conditional R-indicators show that the young households hardly differ from perfect representation in the final response. Cooperation of the elderly households was -26 in the control group but more representative (-7) in the experiment for the unconditional R-indicators.

Maximum bias

Schouten, Cobben and Bethlehem (2009) show that for any survey item y , the R-indicator can be used to approximate an upper bound to the non-response bias, in the case that y covaries maximally with the available (vector of) X . They use this upper bound to evaluate the impact under worst-case scenarios and to derive acceptable values for the R-indicator. The bias of y variables that are not fully explained by X , may be smaller or larger. The maximum bias provides intuition as to how R-indicators relate to bias and is most useful in surveys with many Y variables. The upper bound of the bias is approximated by

$$\frac{|B(\hat{y})|}{S(y)} \leq \frac{S(\hat{\rho})}{\hat{\rho}} = \frac{1-R(\hat{\rho})}{2\hat{\rho}} = B_m(\hat{\rho}, y). \quad (1.3)$$

The maximum bias in the control group is 0.18, in the experimental group it is 0.12. This means that the non-response bias is anticipated to be at most 18% of the standard deviation of any item in the control group and 12% in the experiment. The lower maximum bias in the experiment indicates that the more representative response influenced the estimates. Whether the experimental estimates were less biased than those of the control group cannot be ascertained with certainty, however.

Experimental manipulations

The experiment consisted of three manipulations: adding a mode, manipulation the chance of contact in the CATI part, and manipulation of chance of cooperation, again in CATI. This paragraph describes the effect of these measures on the subsequent distribution of response. Response, cooperation and contact rates are used to illustrate the effect of the manipulations on representativeness.

Adding a mode

The number of ineligible cases was significantly lower in the experimental condition due to the web/mail first round of data collection, which contributed significantly to the better representativeness of the experimental response. As table 4.3.4 shows, males and single parents were better represented as a result of the added mode. Adding mail as a mode resulted in a very high cooperation of elderly people in the first round. This did not lead to over-representation, however, due to the curbing measures that were taken in the subsequent CATI round. Because of the low predicted web participation in the groups with low cooperation propensity, a mail questionnaire was added to the design. Table 4.3.5 shows that this measure succeeded in securing a fairly balanced first round response. The

response of the high cooperation propensity group that was given the web option only, lagged behind the high medium propensity group. The latter had been given a choice of mode (Odds ratio = 1.635, SE = .101, $p < .001$). It even lagged marginally significantly behind the groups with low medium (Odds ratio = 1.215, SE = .108, $p < .10$) and low cooperation propensity (Odds ratio = 1.215, SE = .109, $p < .10$). Compared with the web/mail first round of the Safety Monitor however, where the response of the high propensity group was 38%, while the response of the low propensity group was 18%, the variability in response across groups is substantially less. When given the choice, 81% of households chose the mail questionnaire. The higher preference for the mail option is found repeatedly in research (e.g., Shih & Fan, 2007; Millar & Dillman, 2011).

4.3.5 Response on either web or mail questionnaire by cooperation propensity

	Web/mail response	n
Low cooperation propensity	35.1	304
Low medium cooperation propensity	35.1	326
High medium cooperation propensity	42.1	224
High cooperation propensity	30.8	227

Manipulating chance of contact

The higher R-indicator for the contact phase shows that the manipulations of contactability were successful in attaining a more representative contacted sample. Table 4.3.6 illustrates these findings with the contact rates for SCS, compared to total contact rates for the experiment and the contact rates for the CATI part of the experiment separately. The table shows that contact rates were somewhat higher in the experiment than in the SCS for the lower contact propensity groups, and somewhat lower for the high propensity group. A logistic regression analysis on contact rate with propensity group as factor showed a significant interaction between propensity group and experimental condition (Wald₍₃₎ = 10.39, $p < .05$). Although the variability in contact rate in the experiment was reduced, compared to the SCS control group, we failed to get representative contact in the CATI part of the experiment. After the first wave web/mail response, the remaining group of non-respondents in the group with the lowest contact probability lagged behind considerably in contact rate, even with one call in every shift, every day.

4.3.6 Contact rate per propensity category for the SCS, the experiment and the CATI-part of the experiment

Contact propensity	SCS	n	Experiment total	n	Experiment CATI	n
	Low contact propensity	84.2	640	87.1	657	79.4
Low medium contact propensity	94.5	858	96.6	951	94.8	610
High medium contact propensity	95.7	415	93.7	443	91.2	317
High contact propensity	96.9	794	95.3	804	91.7	459

Manipulating chance of cooperation

The chance of cooperation was manipulated by having the best interviewers call addresses with the highest chance of refusal, while the addresses with the highest chance of cooperation were called by the least successful interviewers. Analysis of the fieldwork verified that the fieldwork strategy was applied as planned and that the mean level of interviewer capacity was comparable in the experiment and the SCS.

Although the number of cooperating sample units was slightly higher in the experiment, the cooperation rate was somewhat lower ($\chi^2_{(1)} = 4.23, p < .05$). The R-indicator for cooperation, table 4.3.3, showed that there was hardly any difference in the distribution of participation for the experiment and control group. Table 4.3.7 illustrates this finding with the cooperation rates per propensity group for the experiment, its CATI part, and the SCS. Like the findings concerning contact, cooperation in the experiment is higher for the two groups with the lower cooperation propensity and lower for the two groups with the higher cooperation propensity. A logistic regression on cooperation rate with propensity group as factor showed a significant interaction between propensity group and experimental condition (Wald₍₃₎ = 10.21, $p < .05$). The interaction signified that having the best interviewers call the hardest cases did not bring about the expected rise in cooperation. But having the lesser interviewers call the easy cases brought about a significant decline in cooperation in this group. The difference was not enough, however, to bring about the desired change in cooperation representativeness.

Some light is shed on the issue of why the best interviewers were not able to secure a higher cooperation rate by studying table 4.3.8. This table shows response results for experiment and SCS by cooperation propensity.

4.3.7 Cooperation rate by cooperation propensity for the SCS, the CATI part of the experiment and the experiment total

Cooperation propensity	SCS		Experiment total		Experiment CATI	
	SCS	n	Experiment total	n	Experiment CATI	n
Low cooperation propensity	62.7	630	65.1	619	43.8	392
Low medium cooperation propensity	68.4	493	71.4	639	52.8	415
High medium cooperation propensity	75.3	674	72.8	744	50.3	418
High cooperation propensity	79.2	1,100	74.7	995	62.8	691

4.3.8 Response by cooperation propensity for the experiment and SCS

	Experiment			SCS				
	low	low medium	high medium	high	low	low medium	high medium	high
	%							
Not able (ill, not present)	12.3	6.9	2.5	2.7	12.1	7.3	4.5	3.5
Language problems	2.4	1.1	0.5	0.0	4.3	1.0	0.6	0.2
Refusal	15.7	13.3	20.8	21.2	14.4	16.8	16.9	15.5
n	619	639	744	995	630	493	674	1,100
Cooperation rate COOP2	65.1	71.4	72.8	74.7	62.7	68.4	75.3	79.2
Cooperation rate COOP3	78.4	81.1	76.2	76.9	78.2	76.4	79.8	82.5

The first cooperation rate in table 4.3.8, also shown in table 4.3.7, shows again that, as predicted, the cooperation rate is higher when the estimated participation propensity is higher, in both experiment and the control group. Prediction of participation propensity was based on the calculation of cooperation according to COOP2 (AAPOR 2008), as cooperation of contacted eligible sample units. However, as table 10 shows, prediction of cooperation appears to be heavily correlated with the ability to participate, and with the existence of language problems. In the experiment, the percentage of sample units unable to participate ranges from 2.7% in the group with high cooperation propensity to 12.3% in the group with low cooperation propensity, and language problems range from absent to 2.4%. The control group shows similar patterns. The second cooperation rate in this table shows cooperation of eligible, contacted, and able sample units (COOP3). With this calculation, the propensity differences all but disappear. If the only difference between groups in the level of cooperation is related to the ability to cooperate, a different intervention is needed, for example using translated questionnaires and bi-lingual interviewers.

Costs

One of the aims of this experiment was to raise data quality while maintaining or ideally lowering costs. We took two potentially costs saving measures: the web round, and more day-time interviewing. The latter because interviewers at Statistics Netherlands receive 20% more pay for working in the evenings. Mail questionnaires are about 50% cheaper than CATI interviewing, but in a mixed mode experiment, a part of the sample will be addressed in both modes, thereby adding to the costs, unless the mail response is substantial.

To compare the costs of the experiment with that of the control group, we considered the costs of observation and data processing, notably, postage and printing costs for the advance letters, reminders, and paper questionnaires, (including labour and machine depreciation), data entry for the paper questionnaires, and the interviewer time, differentiated by shift. Table 4.3.9 shows the total costs and the items contributing to total costs.

4.3.9 Itemised total costs for SCS and experiment

	SCS			Experiment		
	n	rate	€	n	rate	€
Postage advance letters	3,000	0.36	1,071	1,032	0.36	368
Postage advance letters + mail questionnaire				1,968	0.69	1,358
Reminders				2,318	0.36	828
Printing costs mail questionnaire				1,968	2.16	4,244
Interviewer hours evening rate	334	36	12,024	159	36	5,724
Interviewer hours daytime rate	105	30	3,150	86	30	2,573
Data entry hours				48	33	1,571
Total			16,245			16,665

As table 4.3.9 shows, the costs of the experiment are marginally higher (2.6%) than those of the SCS in those two months.

4.4 Summary and discussion

We described a tailored fieldwork strategy to obtain a better representative sample at comparable response and costs levels. The results showed that the tailored fieldwork strategy was successful in maintaining the level of response,

while significantly augmenting representativeness, even within the very short fieldwork period of the SCS. A longer fieldwork period would have provided more possibilities to vary the number and spacing of calls, and to make use of the paradata becoming available during fieldwork. The tailored design was slightly more expensive. We will comment on the expense first, and then discuss the findings regarding representativeness.

The experiment was somewhat more costly than the regular SCS due to a number of circumstances. First, 81% of people who were given the choice between a web and mail questionnaire, chose mail. This meant that more money had to be spent on data entry than expected. Second, although the paper questionnaire, including extra postage and subsequent data entry, only costs about half of what a CATI sample unit in the control group cost, there were many sample units who did not respond in the cheaper mode, and had to be called in the more expensive CATI mode. This resulted in a higher per unit costs. The experiment was designed to incorporate a larger number of (cheaper) day calls, targeting groups with a high contact propensity. Although 11 percent more daytime calls were made in the experiment than in the control group (35% versus 24%), this difference was not enough to offset the mechanisms described above.

The addition of the paper questionnaire was the obvious cause of the relatively high costs. We could easily have achieved lower costs by using only a web approach, followed by a CATI non-response follow up. We did not do this because we sought representativeness within each step of the fieldwork. For the same reason we stimulated cooperation in some of the same groups in which we curbed contact. For example, an elderly person who refuses is a different person than an elderly person who cannot be contacted and may have a differential influence on potential bias. By adding the mail questionnaire we succeeded in obtaining a far more balanced first wave response than if we had used a web only approach. Representativeness was measured with the R-indicator, which measures the distance between mean response level and the response of subgroups defined by the auxiliary variables present in the research. Partial R-indicators were first used to examine which groups were under- and over-represented in the SCS, and later to study the impact of the experimental manipulations on representativeness within auxiliary variables. The choice of auxiliary variables is paramount in designing tailored designs. First, it is imperative that variables are known for all sample units. Second, they need to be related to key variables of the survey and main domains of interest in publications. The broader the subject of a survey, the more general the auxiliary variables need to be (Bethlehem & Schouten, 2009). The choice of auxiliary variables also influences the conclusions that can be drawn as to representativeness. Representativeness is not an absolute given, but depends on the auxiliary variables in the model. A survey could have a very high R-indicator

and still contain biases on variables for which correlating paradata or other auxiliary variables are not available.

The auxiliary variables chosen in this experiment all relate to the key variables of the SCS. The finding that the response composition was more representative with regard to these variables can be generalized to variables that were not part of the design. Schouten and Cobben (2012) show that the design in this experiment was successful in reducing non-representative response on other variables than those used to differentiate subgroups, specifically, ownership of a company car, business type of the person in the household with the largest job, and job number and sizes in the household. Although these variables were not used in the tailored survey design, they are associated with the selected design variables age, ethnicity, income, type of household and urbanization. If tailored or other adaptive survey designs are to be promising extensions of sampling designs, then the indicator values should also be better for variables that were not involved in the adaptation. Statistics Netherlands is allowed by law to use registers to link to survey results. However, the number of register variables is limited. If matching with register data is impossible, or if the available variables are not related to key variables, the only option is to resort to paradata like observations of sample units and /or their environment that are expected to relate strongly to the main survey variables. A key question is whether the higher representativeness found in the experimental group is due to the experimental manipulation of response propensity, the introduction of a second mode, or the longer fieldwork period. The manipulations are partly confounded and the independent effects of each of the treatments cannot be disentangled completely. Undoubtedly, adding a mode helped in improving representativeness: we have shown that the number of households that could not be approached as a result of disconnected telephone numbers was significantly reduced in the experimental group, bringing in households with a low response propensity. The mode offered was differentiated according to response propensity, as a result of which the first wave response was fairly balanced for the four propensity groups. Consequently, the sample for the CATI re-approach was also balanced. The differential response in the CATI wave for the four propensity groups is therefore also the result of the experimental CATI manipulations. Additional support for the contention that the mere introduction of a second mode does not in itself result in a better representative response, is found in the analysis of several redesigns of Statistics Netherlands surveys, where (CAPI) uni-mode designs were replaced by mixed mode designs. In all of these redesigns, adding a mode led either to a slight reduction in representativeness or to a comparable level, but never to augmented representativeness (Banning, Cobben & Leufkens, 2011; Cobben, 2011). The longer fieldwork period did not lead to higher overall contact rates, compared with the regular SCS. Although in the experimental group higher contact rates were realised for the groups with the lowest contact propensity,

lower rates were attained for the groups with the highest contact propensity, thereby reducing variability in contact rates across the groups. The longer fieldwork period in itself cannot account for the differentiation in these results. The representativeness of the experimental group was augmented especially as a result of more representative eligible and contacted cases. The manipulation of cooperation had less impact. This result may have been influenced by the introduction of a web / mail first round, filtering away the 'easiest' respondents, leaving the interviewers to deal with a relatively uniform difficult group of initial non-respondents. In other words, the cooperation propensity of the remaining group may have been different from the expected propensity. With increasing experience in mixed mode survey methodology it will be possible to gauge the influence of different modes on cooperation propensity of different groups. Another explanation is to be found in the definition of cooperation we used in this study. When cooperation was defined conditional on the ability to cooperate, cooperation propensity was not predictable with the auxiliary variables available in this study. As a result of this finding, Luiten and Cobben (2010, see also chapter 5) analysed a large database that consisted of numerous surveys with a variety of topics, lengths, modes and sampling types, and containing an extensive number of auxiliary variables. Again, cooperation conditional on ability to cooperate could not be predicted. That is not to say that cooperation cannot or should not be influenced, but rather that cooperation has other underlying dimensions than socio-economic or demographic correlates. Present attempts to find and incorporate paradata that relate both to response propensity and to substantive variables may fill this gap (Schouten, 2010; Kreuter et al., 2010).

In this paper we set out to show that it is possible to attain a more representative sample while keeping response and costs levels the same. We found that we could. Far more research is needed, however. We need more experience with design variations to find out if it can be done even better, and learn what works best for which groups. The field of adaptive design is only just starting.

Groves (2006) set in motion awareness amongst survey practitioners that we need to think in terms of non-response bias as much as in terms of response rates. This can only be accomplished when survey designs are aimed at reducing bias, which in turn means that sample units should not be treated in a uniform fashion. Adaptive, tailored, or responsive survey designs are a means to accomplish this end. R-indicators may help in drafting these designs by selecting subsets of cases that need extra attention, in monitoring the fieldwork and in gauging the maximum bias in a given survey. Future research should focus on determining which groups are susceptible to which treatments, and how differential treatment relates to the reduction of non-response bias.

5.

Predicting

contactability and

cooperation

in survey research

5.1 Introduction

Nonresponse is a common feature of sample surveys. A part of the sampled elements cannot be contacted, refuses cooperation or does not participate in the survey for other reasons. It is important to distinguish between these types because they require different measures to reduce nonresponse and may differently affect survey estimates.

There are many ways to classify nonresponse according to cause. This complicates comparing nonresponse for different surveys. The American Association for Public Opinion Research (AAPOR, 2008) and Lynn, Beerten, Laiho and Martin (2002) have published comprehensive lists of definitions of possible survey outcomes.

The response process can be seen as a hierarchical (nested) process with three stages. First, the eligibility of the sample elements is determined¹⁾. Eligible elements are approached. Once contacted, this results in cooperation, refusal or another form of non-interview. AAPOR (2008) distinguishes multiple forms of (non-)cooperation, that are differentiated by the inclusion or exclusion of being able to participate. Definitions COOP1 and COOP2 define cooperation as response of those contacted, while COOP3 and COOP4 define cooperation as response of those contacted and able to participate. Inability to participate may be the result of long-time illness, or language problems.

A host of literature is devoted to the identification of different correlates of nonresponse. Most authors thereby focus on non-contact and refusal. Extensive overviews are written by Goyder (1987), Groves and Couper (1989) and Stoop (2005). When it comes to correlates of (non)contact, the literature is quite unambiguous. At-home patterns of sample persons, combined with physical impediments in the environment and call patterns of interviewers, determine whether contact will be made in a given fieldwork period, and how much effort is needed to do that. At-home patterns are related to the time a respondent spends at home, as a result of working hours and other out of doors activities. The at-home patterns of other household members are relevant as well, as they may inform the interviewer when the sample person may be available or make an appointment. Interviewer call patterns must tailor respondent's at-home patterns in order to contact them. Physical impediments like gates, entry phones or intercoms, doormen, and locked communal entrances may make contacting sample persons difficult or even impossible.

The literature on correlates of survey cooperation is less straightforward. Goyder (1987) concluded that the reading of the literature on socio-demographic

¹⁾ This may not always be the first step during fieldwork; sometimes eligibility can only be determined after contact.

correlates reveals an un-integrated corpus. Cobben (2009, p. 44) concludes that “the findings reported on survey cooperation are not consistent”. This conclusion is confirmed by Stoop, Billiet, Koch and Fitzgerald (2010, p. 122) who state that “there are no simple mechanisms that lead some groups to participate less and others more. Indeed, there is very little empirical evidence as to which socio-demographic and socio-economic factors are related to survey cooperation and the evidence that does exist is usually weak or mixed”.

5.1.1 Auxiliary variables related to survey cooperation in previous research

	1	2	3	4a	4b	5	6	7a	7b	8	9
Age	n	y	y	y	y	y	y	y	y	n	n
Gender	n	n	y	n	n	n	n			y	n
Education	n					n	y	n	n	y	
Labour Force Status	y		n	y	y		n	n	n	y	y
Income		y		y	n		y			y	
SES			y	y	y						
Receiving social benefit								n			n
Value house								y	n		y
Home owner			y	n	n		n	n	n		
Car owner						y					
Maintenace dwelling										y	
Household composition	y		y	y	n	n	n	y	y	y	n
Dependent children in household						n	n	y	y	n	
Ethnic background		n	y				y	y	n	y	n
Urban density		y	y	n	n	n	n	y	y	y	n
% non-western in neighbourhood										n	n
Dwelling type			y	y	n	n		y	n	n	
Religion				n							
Residential stability				n		y		y	n		
Landline							y			n	n
Region											y
Health						y					
Lenght interview	n										
Subject of survey			y								
Mode	y			y							
Contact with neighbours	y										
Citizenship				n							
Cultural participation										y	
Hours away from home								n			

Note 1: y: present in tested model and found to have a significant ($p < .05$) relation with cooperation. n: present in tested model, but no significant relation.

Note 2: 1. Nicoletti & Peracchi, 2005; 2. DeMaio, 1980; 3. Goyder, 1987 literature review; 4a. Goyder, 1987 Bivariate analysis; 4b. Goyder, 1987 Multivariate analysis; 5. Durrant & Steele, 2010; 6. Abraham, Maitland ea, 2006; 7a. Groves and Couper, 1998 Bivariate analysis; 7b Groves and Couper, 1998 Multivariate analysis; 8. Stoop, 2005; 9. Cobben, 2009.

A by no means exhaustive literature review, reflected in table 5.1.1, shows the large number of sample unit characteristics that has been studied in relation to survey cooperation specifically. The table summarizes research by Nicoletti and Peracchi (2005), DeMaio, (1980), Goyder (1987), Durrant and Steele (2009), Abraham, Maitland and Bianci (2006), Groves and Couper (1989), Stoop (2005) and Cobben (2009). From Goyder's (1987) work both his literature overview and his own research is incorporated. In addition, both Goyder and Groves and Couper show bivariate as well as multivariate results, both of which are incorporated in table 5.1.1.

The table merely notes whether a variable was studied, and whether a relation with cooperation was found, irrespective of the nature of that relation. Even without the latter information, the lack of consistency is apparent. Including the direction of the relation would exacerbate the picture.

Evident in table 5.1.1 is that the use of a multilevel model opposed to a bivariate model has a marked influence on the correlates found. In 1987 Goyder noted that 'the most productive development for the research seems to lie not with additional case studies of bivariate effect (although high-quality nonresponse data are always valuable) but with a design stressing replication, multivariate analysis, and explicit modelling of the importance of contact as the intervening variable preceding the response decision'.

There may be multiple reasons why research findings are mixed. More than in establishing contact, cooperation is dependent on a complex of influences. The length of the survey, the topic, the sponsor, the contents of the advance letter, the absence or presence of an interviewer, all this and more will influence the decision to participate. Topic may for example change the direction of the relation between cooperation and auxiliary variables (Goyder, 1987), depending on the respondent's involvement with the topic. In general population surveys, topic may be less relevant, as multiple topics are generally touched upon. These multiple design influences may make comparison of literature findings difficult. Furthermore, research on nonresponse is often limited by the absence of meaningful information on nonrespondents.

In this chapter we strive to augment the existing empirical evidence concerning correlates of cooperation. The immediate cause for this endeavour were the findings in Luiten and Schouten (2013; see chapter 4), where we tried to predict cooperation in the Survey of Consumer Sentiments. We wanted to develop a tailored fieldwork strategy in which sample persons that were predicted to be hard to convince would be approached by the best interviewers. It was found that cooperation could only be predicted under certain definitions of cooperation. If 'not able to cooperate' was included in the definition, cooperation could be predicted, but if cooperation was defined as cooperation from those who are able to cooperate (COOP1 versus COOP3, AAPOR, 2008), we could not.

The status of 'unable to cooperate' is unclear in the body of literature. Goyder (1987) and Groves and Couper (1998) treat this outcome as refusal. Others, e.g., Durrant and Steele (2009) and Stoop (2005) treat unable sample units as ineligible, and remove them from the basis in the calculation of contact and cooperation rates. Cobben (2009) explicitly treats ability to cooperate as a separate step in the response process, and models cooperation on ability to cooperate. These findings led us to re-address the issue of correlates of cooperation, in which we follow Cobben (2009), and define cooperation as COOP3: cooperation from sample units who are contacted and able to cooperate. To do this, we studied seven Statistics Netherlands surveys that were linked to registry information, with a total of 203,556 sample units. This linking made a large number of auxiliary variables available that informs with equal precision on characteristics of respondents and nonrespondents. The surveys were of different lengths (from 8 minutes to one hour), contents, sample units (persons versus addresses), modes (CAPI, CATI and mixed mode) and fieldwork strategies (with and without nonresponse follow-up). The large number of records may augment the chance of finding stable correlates and allows the inclusion of interactions. We will concentrate the discussion of findings on correlates of contact and cooperation.

5.2 Method

The seven surveys were linked to the Social Statistical Database of Statistics Netherlands. This database is an integrated register of registrations of all kinds of subjects. It contains administrative information on individuals, households, jobs, benefits, pensions and income. The sample addresses were matched on the basis of a precise combination of address, house number, and date of contact. The registers contain information on individuals. For household surveys, the individual level information was aggregated to household core level (head of household and partner). So, the variables ethnic group and sex have a category to indicate a mixture of the categories on the personal level (e.g., mixed native-ethnic minority). Some information is only available at the postal code level, which is a quite narrow geographical area around the sample unit's house. The presence of a registered landline was incorporated as auxiliary variable. This characteristic is consistently found to be heavily correlated with contact and cooperation propensity (Bethlehem & Schouten, 2004; Cobben & Schouten, 2007; Cobben 2009). This relates on the one hand to greater ease of contacting sample units if their telephone number is known. On the other hand it reflects significant differences between characteristics

of sample persons who do and those who do not have a registered landline: landline owners are under-represented in young persons, ethnic minorities, urban dwellers, higher incomes, males, and people who do not live in apartments. These are all characteristics that interact with contact and cooperation propensity as well. In Table 5.2.1 the auxiliary variables available for the analysis are summarized. Not all auxiliary variables were available for all surveys and not all categories within auxiliary variables were always applicable. Paradata that were observed by interviewers, like the kind of dwelling, were not available for telephone surveys. 'Mixed' categories were not applicable for person samples.

5.2.1 Auxiliary variables available from registries

Variable	Categories
Household level	
Ethnic Group	Native, non-western ethnic minorities, western ethnic minorities, mixed (in household surveys)
Gender	All male, all female, mixed
Average age of household core	15-30;31-44;45-65; over 65
Number of persons in household	1, 2, 3, 4, 5 and more
Type of Household	Single, partners without children, partners with children, single parents
Percentage of employed members of household core	0, 50 (in household surveys), 100%
Percentage of members of household core with social benefit (excluding pension)	0, 50 (in household surveys), 100%
Type of residence	Single-family dwelling; apartment; housing for pensioners; villas, mansions and residences; other and unknown
Registered landline	Yes, no
Rent or ownership of residence	Rent, ownership, unknown
Postal code area level	
Province	The 12 Dutch provinces
Degree of urbanization	Very strong, strong, moderate, low, not urban
Percentage non-western non-natives	Very high, high, average, low, very low
Average worth of housing	Quartiles
Average monthly income	Quintiles

All surveys are conducted by Statistics Netherlands, but differ on a number of aspects. In table 5.2.2 the main characteristics are summarized.

5.2.2 Main characteristics of the surveys analysed

	Acronym	Year	n	Sample unit	Mode
Survey					
Labour Force Survey	LFS	2005	16,963	Address	CAPI
Labour Force Survey	LFS	2008	67,282	Address	CAPI
Survey of Amenities and Services Utilisation	ASU	2007	9,188	Address	CAPI
Survey of Consumer Sentiments	SCS	2005	17,502	Address	CATI
Survey of Living Conditions and Housing	SLCH	2009	73,224	Person	Mixed Mode (MM)
Health Survey	HS	2005	15,411	Person	CAPI
Dutch Parliamentary Election Survey	DPES	2006	3,986	Person	CAPI with MM follow-up

For the Survey of Living Conditions and Housing in 2009 28,212 persons were approached by CAPI (those without listed telephone number); 41,098 by CATI and 3,913 with a web survey. For the Dutch Parliamentary Election Study persons were approached by CAPI; noncontacts were re-approached by CATI and mail. After collinearity analyses, it was decided that 'number of persons in the household' could not be used in one multivariate analysis with household composition, and 'income of the neighbourhood' not with 'real estate value'. Both variables had two or more variance proportions of over .50 for all or most surveys. Urbanicity and province had separate contributions and were both maintained.

5.3 Results

Table 5.3.1 shows contact and (non) cooperation rates for the surveys in this study by auxiliary variable. Contact rates are from eligible and processed sample units; cooperation rates from contacted units that were able to cooperate (COOP3). To facilitate comparison with literature findings, we first analysed these variables bivariate. Table 5.3.2 gives the values of Cramèr's V per survey for contact and cooperation. Generally, the bivariate relation with auxiliary variables is higher for contact than for cooperation. Even with the large number of sample units in the analyses, many relations with cooperation are not significant. Relations with contact are stronger for CAPI surveys than for the SCS, a CATI survey, while the opposite is true for relations with cooperation. The sample unit's dwelling and the presence of a landline have the strongest bivariate relation with contact. The strength of the relations with cooperation depends on the survey.

Inspection of bivariate results per survey shows that these findings replicate earlier findings: noncontact rates are consistently highest for men, singles, young persons, non-western minorities, urban areas, people living in apartments, and people who do not have a fixed landline. People and households on social allowance are also hard to reach. In most surveys noncontact is highest if the entire household core works, and for people in the lowest income group. Noncontact rates are lowest in rural areas, for households with mixed sex partners, for households with four or more persons, and for partners without children. They tend to be lowest for higher income groups, living in expensive properties, in neighbourhoods with a small amount of ethnic minorities, for Dutch natives, and households where one of the partners works.

Bivariate results for cooperation are less clear cut. They show that:

- Cooperation is lowest in the eldest age group in five out of seven surveys and cooperation is highest in the youngest age group, again in five surveys.
- Cooperation is lower for people with unlisted numbers in all surveys but one. The latter survey, the housing survey, had a heavy oversampling of underprivileged neighbourhoods, and took special measures to obtain an acceptable level of response in them. The result is that this survey repeatedly shows different trends.

5.3.1 Contact and cooperation rates by auxiliary variables and surveys

	Noncontact							Cooperation						
	ASU	SCS	LFS08	LFS05	HS	DPES	SLCH	ASU	SCS	LFS08	LFS05	HS	DPES	SLCH
Gender														
mixed	1.9	4.2	3.3	3.0	na	na	na	76.9	74.3	71.6	75.1	na	na	na
men	11.3	11.1	14.6	16.5	4.2	3.6	7.2	74.2	65.7	77.1	75.4	78.0	80.4	70.8
women	6.4	6.1	8.9	8.3	3.1	2.2	4.9	74.1	62.4	72.9	74.3	77.8	80.3	74.4
Age														
<30	13.2	9.0	13.7	12.7	4.3	5.1	9.9	78.0	74.9	78.8	80.8	83.1	80.0	74.3
30-44	6.5	7.3	7.2	6.5	5.2	3.2	8.3	77.5	76.5	72.9	75.5	74.6	82.9	72.1
45-64	3.3	4.6	4.3	5.0	2.6	2.1	4.6	75.8	73.5	71.0	74.0	74.1	80.8	72.8
>=65	1.8	3.9	2.7	3.2	1.8	1.1	1.7	73.8	60.9	72.6	70.7	77.5	75.2	71.8
Ethnicity														
Dutch	3.7	5.0	5.2	5.9	3.0	2.3	4.1	75.9	71.0	72.5	74.5	77.5	80.6	72.2
non-western	10.6	10.6	13.6	11.6	8.6	9.4	14.7	74.0	65.4	77.8	75.4	82.0	78.5	75.6
western	8.8	10.0	10.2	8.4	5.1	3.8	6.9	72.7	56.2	75.2	85.4	78.4	79.4	72.6
mixed	3.1	5.6	5.4	3.3	na	na	na	78.1	73.3	69.6	76.6	na	na	na
Household composition														
single	8.3	8.0	13.3	13.7	9.6	8.1	9.4	73.7	62.3	73.4	74.8	75.9	78.0	71.7
partners, with children	3.0	4.7	4.6	4.4	2.9	1.4	4.1	75.3	72.3	72.6	73.5	75.8	79.7	73.2
partners, no children	1.3	3.6	2.6	2.6	1.9	1.7	3.4	78.7	76.5	72.6	76.1	79.8	81.8	73.2
single parent	6.5	5.9	6.7	6.4	5.5	4.9	7.7	75.2	71.9	71.2	75.2	75.8	80.4	73.2
Number of persons														
1	8.3	8.0	13.3	13.7	9.6	8.1	9.4	73.7	62.3	73.4	74.8	75.9	78.0	71.7
2	3.6	4.9	5.0	3.7	3.4	1.9	4.9	75.2	72.1	72.4	75.3	75.4	79.3	72.9
3	2.2	4.0	3.7	2.8	3.0	1.9	5.1	76.4	74.6	70.5	73.3	77.5	79.5	72.5
4	1.2	3.7	2.3	2.5	1.8	1.5	3.2	79.3	76.3	73.2	76.0	79.8	81.7	73.7
>=5	2.0	3.1	2.6	3.3	1.9	2.8	4.2	80.9	78.8	74.6	73.6	82.2	84.8	74.2
Owner / renter														
owner	-	5.2	-	-	2.4	1.7	4.2	-	74.8	-	-	79.1	83.0	73.4
renter	-	5.8	-	-	5.9	5.0	7.1	-	66.2	-	-	75.6	75.5	71.9
unknown	-	-	-	-	-	-	7.1	-	-	-	-	-	-	72.6
% of household core with job														
0	3.3	4.8	-	7.4	3.0	3.2	4.3	73.4	63.9	-	72.1	79.6	78.1	72.2
50	2.7	3.7	-	2.9	na	na	na	74.1	73.5	-	73.0	na	na	na
100	5.8	6.6	-	6.8	4.6	2.7	7.0	78.1	75.6	-	76.6	75.7	81.8	72.9
% of household core with social benefit														
0	4.4	5.5	-	6.1	3.6	2.6	5.6	76.4	70.5	-	76.3	78.4	80.5	72.5
50	1.2	3.8	-	3.8	na	na	na	71.3	73.9	-	72.9	na	na	na
100	7.5	7.1	-	9.6	5.0	7.0	8.3	73.9	70.4	-	70.0	72.8	78.3	73.7
Dwelling														
single-family dwelling	2.4	-	3.2	3.4	2.5	1.4	-	76.5	-	72.7	76.5	80.5	81.7	-
apartment building	9.8	-	14.6	12.7	9.5	9.2	-	72.5	-	71.9	71.0	77.8	82.4	-
old people's flat	2.5	-	4.1	3.5	1.2	0.0	-	76.7	-	76.0	73.9	76.2	73.9	-
villa, mansion	1.8	-	2.4	3.6	1.3	0.9	-	83.3	-	78.1	77.9	80.1	80.8	-
other	2.7	-	3.3	3.0	2.9	1.9	-	74.2	-	67.3	74.2	56.1	65.7	-

5.3.1 Contact and cooperation rates by auxiliary variables and surveys (end)

	Noncontact							Cooperation						
	ASU	SCS	LFS08	LFS05	HS	DPES	SLCH	ASU	SCS	LFS08	LFS05	HS	DPES	SLCH
Landline														
yes	1.8	na	3.4	3.2	1.9	1.4	-	77.6	na	73.5	76.5	79.2	81.5	-
no	6.0	na	8.8	10.5	6.9	8.1	-	74.8	na	71.9	72.6	75.4	76.0	-
Urbanicity														
highly urban	9.6	8.6	12.7	11.6	7.8	6.9	10.9	75.7	68.6	72.0	73.1	76.7	77.4	72.3
2	4.2	5.0	6.8	5.9	3.9	2.9	5.8	74.7	71.2	72.2	74.0	75.2	83.2	73.2
3	4.6	5.0	4.8	4.0	3.1	2.7	3.7	75.8	71.6	71.8	74.9	79.3	81.1	74.3
4	1.3	5.3	3.2	3.7	1.9	1.3	2.5	75.5	71.6	72.7	76.7	79.5	78.3	73.1
highly rural	2.2	4.0	2.5	3.8	2.3	1.4	2.1	79.3	70.1	74.6	77.3	79.1	80.9	70.7
Province														
Groningen	3.5	4.9	4.7	10.6	3.1	6.3	4.1	78.2	69.6	74.0	72.5	77.6	79.8	74.7
Friesland	1.6	5.0	2.3	4.1	2.1	0.6	2.5	77.6	73.5	77.1	78.5	83.1	80.8	74.5
Drenthe	1.1	3.4	2.6	4.8	4.2	1.6	2.2	79.4	70.4	73.9	77.5	76.7	82.5	72.9
Overijssel	2.9	4.4	5.0	6.4	3.4	1.5	4.0	75.2	69.5	74.3	77.7	80.0	82.2	72.5
Flevoland	4.7	6.2	5.6	-	6.8	4.6	4.5	75.7	76.7	75.0	-	80.4	86.8	74.4
Gelderland	4.3	4.7	5.6	4.2	3.6	1.4	5.0	77.0	70.7	74.4	76.9	81.1	79.5	73.1
Utrecht	4.0	6.6	4.8	4.6	2.1	3.0	4.5	75.6	70.9	72.6	76.4	79.6	78.3	73.3
Noord-Holland	8.1	6.5	11.5	10.5	5.9	5.7	10.4	75.6	70.5	72.1	71.8	75.4	83.2	70.6
Zuid-Holland	5.2	6.5	8.0	6.4	4.6	2.9	7.4	73.7	68.4	67.7	73.0	74.0	77.9	71.6
Zeeland	3.6	3.3	4.6	-	4.6	2.2	3.9	73.8	64.7	70.9	-	78.2	81.6	70.6
Noord-Brabant	2.8	4.9	3.3	3.7	2.0	2.1	3.3	77.3	71.0	74.6	75.8	79.1	77.5	73.9
Limburg	3.2	4.2	3.3	3.4	2.2	2.6	3.0	75.5	78.1	76.1	78.0	78.9	86.1	77.0
% non-western immigrants														
< 5%	2.7	4.9	3.7	4.2	2.4	1.3	2.8	77.5	70.6	73.3	75.5	78.4	79.9	72.3
5-10%	5.4	5.8	6.3	6.5	4.1	4.5	5.0	73.6	72.2	73.3	75.4	77.6	82.5	73.1
10-20%	6.0	6.5	8.0	7.6	4.5	5.2	7.0	74.8	72.5	71.6	75.2	77.0	81.7	72.5
20-40	7.1	6.7	11.8	10.8	7.4	5.8	9.5	70.6	69.7	70.3	72.3	75.7	79.9	73.0
>40%	11.3	7.8	14.6	11.5	8.3	7.2	14.5	72.9	64.2	71.3	73.0	78.4	77.9	73.8
Income in quintiles														
lowest quintile	8.1	4.6	8.0	8.3	5.0	3.3	8.1	71.8	63.1	71.7	73.1	77.7	77.3	73.7
2nd	6.8	4.5	7.1	6.5	3.7	3.7	7.0	72.4	69.9	71.2	73.1	78.2	77.5	71.9
3rd	3.4	5.6	5.7	6.2	3.9	2.9	5.3	75.2	72.0	72.7	74.7	76.9	79.9	72.3
4th	2.8	5.8	5.3	4.9	2.9	1.8	4.6	78.3	74.1	73.5	76.3	77.6	82.6	73.2
highest quintile	2.5	6.2	5.3	5.3	2.8	3.0	3.8	79.2	74.2	74.0	77.2	79.3	84.1	72.3
Mean value of dwellings in neighbourhood														
1st quartile <=91	6.8	7.2	10.2	10.0	6.9	5.6	8.0	72.7	65.5	71.8	73.7	77.0	79.5	73.2
2nd quartile 91-122	4.9	4.2	6.3	6.7	3.4	3.4	6.3	75.4	69.5	71.2	73.3	77.4	80.5	73.2
3rd quartile 122 -160	3.4	5.4	4.7	4.7	2.9	1.5	4.7	77.5	73.5	72.4	75.1	78.2	77.4	72.4
4th quartile > 160	2.6	5.2	3.8	4.0	2.3	1.9	4.2	77.5	72.5	75.1	77.2	78.6	83.2	71.8

Note: 'na' not applicable; categorie not possible in survey; '-': variable not linked to survey.

5.3.2 Values of Cramèrs V for contacted and cooperating sample units per survey

	contact					cooperation								
	ASU	SCS	LFS05	LFS08	HS	DPES	SLCH	ASU	SCS	LFS05	LFS08	HS	DPES	SLCH
Gender	.17	.10	.20	.18	.03	.04 **	.05	.03 *	.11	.01 ns	.04	.00 ns	.00 ns	.04
Age	.14	.07	.10	.13	.06	.08	.13	.03 *	.15	.06	.05	.10	.06 **	.02
Ethnicity	.10	.06	.09	.11	.09	.11	.16	.02 ns	.05	.04 **	.04	.03 **	.01 ns	.03
Household composition	.14	.08	.18	.18	.15	.15	.12	.05	.13	.03 *	.01 *	.05	.04 ns	.02
Number of persons in household	.13	.08	.18	.18	.14	.14	.10	.05	.13	.02 ns	.03	.06	.04 ns	.02 **
Owner / renter	-	.01 ns	-	-	.09	.09	.06	-	.09	-	-	.04	.09	.01 **
% of household core with job	.07	.05	.06	-	.04	.01 ns	.05	.05	.13	.05	-	.05	.05 **	.01 ns
% of household core with social benefit	.06	.03 **	.06	-	.02 *	.07	.04	.03 *	.02 *	.05	-	.04	.02 ns	.01 *
Dwelling	.16	-	.18	.21	.15	.19	-	.06	-	.06	.04	.16	.11	-
Landline	.10	-	.15	.11	.13	.17	-	.03 **	-	.04	.02	.04	.06 **	-
Urbanicity	.13	.06	.13	.15	.10	.12	.15	.03 ns	.02	.04 **	.02	.04	.05 ns	.03
Province	.09	.05	.11	.12	.08	.09	.11	.04 ns	.06	.06	.07	.07	.07 ns	.04
% non-western minorities	.12	.04	.11	.15	.10	.12	.17	.05	.03 **	.03 ns	.02	.02 ns	.03 ns	.01 ns
Income in quintiles	.11	.03 **	.05	.04	.04	.04 ns	.07	.07	.09	.04	.02	.02 ns	.07 **	.02 *
Mean real estate value in neighbourhood	.08	.04	.10	.10	.09	.09	.06	.05 **	.07	.04	.03	.02 ns	.06 *	.01 *

Note: all values $p < .001$ unless otherwise indicated: ** $p < .01$, * $p < .05$.

'-' : auxiliary variable not linked to this survey.

- Findings concerning ethnic background are mixed: cooperation is highest for native Dutch in one survey only; in three surveys it is highest for non-western minorities, in one survey for western minorities and in two surveys highest for mixed ethnic households. Findings concerning the ethnic composition of the neighbourhood are somewhat more consistent: in six out of seven surveys, lowest level of cooperation was found in one of the two categories with the highest percentage of non-western minorities. In four surveys, highest cooperation was found in neighbourhoods with the highest number Dutch natives.
- Cooperation is highest in households living in the most expensive houses (5 out of 7) and lowest in the least expensive ones (4 out of 7); parallel to this finding, cooperation is highest if houses are owned.

Cooperation is highest in households where the entire core works (five out of six surveys in which this information is known), and lowest in households where no one works (again 5 out of 6). Parallel to this finding, cooperation is highest if no one in the household core is on social benefit (4 out of 6 studies).

5.3.3 Logistic regression of auxiliary variables on contact (yes/no)

	ASU		SCS		LFS05		p		
	Wald	log odds	p	Wald	log odds	p			
Gender (mixed is referent)	18.3		***	18.0		***	73.0	***	
men		0.31	***		0.55	**		0,38	***
women		0.41	***		0.81	ns		0,73	ns
Age (youngest is referent)	42.6		***	130.3		***	32.4	***	
30-45		1.08	ns		0,98	ns		1.31	**
45-65		1.95	***		1.84	***		1.51	***
>65		3.47	***		2.84	***		2.50	***
Ethnicity (Dutch is referent)	7.9		*	25.2		***	11.0		*
non-western		0.66	*		0,58	***		0,87	ns
western		0.75	ns		0,61	*		1.59	*
mixed		0.86	ns		0,79	*		1.23	ns
Household composition (single is referent)	20.7		***	47.2		***	26.9		
partners no children		0.95	ns		1.24	ns		1.26	ns
partners with kinderen		2.46	*		2.31	***		1.87	***
single parent		1.39	ns		1.60	**		1.55	**
other			ns		0,77	ns		1.77	*
Owner/renter (renter is referent)	no X			ns			no X		
owner									
% with job in household core (all job is referent)	8.6		*	ns			ns		
0%		1.47	*						
50%		0.82	ns						
% with allowance in (all allowance is referent)	ns			ns			ns		
0%									
50%									
Dwelling (single-family dwelling is referent)	20.9		***	no X			51.0		***
appartement building		0.54	***					0,54	***
old people's flat		0.62	ns					1.22	ns
villa, mansion		0.92	ns					0,69	ns
other		0.60	ns					1.10	ns
Landline (yes is referent)	21.9		***	no X			103.9		***
no		0.49	***					0,46	***
Urban density (Rural is referent)	30.5		***	27.3		***	22.3		***
very urban		0.72	ns		0,58	***		0,94	ns
2		1.18	ns		0,88	ns		1.30	ns
3		0.77	ns		0,80	ns		1.44	*
somewhat rural		2.44	**		0,75	*		1.32	*
Province (Groningen is referent)	32.6		**	ns			112.3		***
Friesland		1.21	ns					1.31	ns
Drenthe		1.67	ns					0,75	ns
Overijssel		0.63	ns					0,75	ns
Flevoland		0.42	ns					-	
Gelderland		0.39	**					1.20	ns
Utrecht		0.71	ns					1.31	ns
Noord Holland		0.44	*					0,62	*

LFS08	HS			DPES			SLCH		
	Wald	log odds	p	Wald	log odds	p	Wald	log odds	p
84.1	***	6.4	*	6.1	*	88.5	***		
0,59	***								
0,83	*	1.30	*	1.67	*	1.39	***		
296.7	***	48.0	***	ns		767.5	***		
1.15	**	1.01	ns			1.17	***		
1.79	***	1.83	***			2.07	***		
3.38	***	3.21	***			6.53	***		
46.7	***	13.6	**	ns		228.7	***		
0,76	***	0,64	***			0,51	***		
0,91	ns	0,87	ns			0,76	***		
0,71	***	na				na			
208.3	***	161.0	***	24.4	***	766.3	***		
1.68	***	2.33	***	3.66	***	1.80	***		
2.78	***	5.19	***	2.61	***	3.60	***		
1.76	***	2.46	***	1.40	ns	1.93	***		
ns		1.19	ns	ns		1.02	ns		
no X		5.9	*	ns		8.0	*		
		1.29	*			1.03	ns		
						0,91	*		
no X				ns		ns			
		7.3	1.33	**					
no X		ns		ns		ns			
222.1	***	25.3	***	24.9	***	no X			
0,51	***	0,59	***	0,29	***				
0,71	ns	1.80	ns	1.00	ns				
0,90	ns	1.21	ns	1.10	ns				
0,71	**	0,69	*	0,54	ns				
132.8	***	54.1	***	36.2	***	no X			
0,63	***	0,47	***	0,27	***				
26.7	***	ns		ns		89.4	***		
0,74	***					0,53	***		
0,82	*					0,68	***		
0,84	*					0,77	**		
1.01	ns					0,92	ns		
312.1	***	57.2	***	ns		200.6	***		
1.47	*	1.08	ns			1.04	ns		
1.03	ns	0,44	*			0,89	ns		
0,69	**	0,69	ns			0.80	ns		
0.70	*	0.40	**			0,87	ns		
0,59	***	0,62	ns			0.60	***		
1.00	ns	1.57	ns			1.00	ns		
0,48	***	0,59	ns			0,52	***		

5.3.3 Logistic regression of auxiliary variables on contact (yes/no) (end)

	ASU		SCS		LFS05				
	Wald	log odds	p	Wald	log odds	p	Wald	log odds	p
Zuid Holland		0.75	ns				1.30		ns
Zeeland		0.43	ns				-		
Noord Brabant		0.70	ns				1.36		ns
Limburg		0.61	ns				1.62		ns
% non-western minorities (<5% is referent)	ns			ns			ns		
5-10%									
10-20%									
20-40%									
>40%									
Mean value of real estate (most expensive is referent)	ns			23.0		****	ns		
1e quartile					1.14	ns			
2e quartile					1.61	****			
3e quartile					1.10	ns			
Nagelkerke R ²	.15			.06			.18		

Note: 'no X': variable not linked to survey; 'na', category not applicable.

**** p < .001; ** p < .01; * p < .05.

The referent for owner/renter in the SLCH is 'unknown' the first category is owner, the second is renter.

The referent for gender in HS, DPES and SLCH is 'men'.

- Cooperation is lowest in apartment buildings (3 out of 6), and highest in villas (3 out of 6) or single family dwellings (2 out of 6).
- Concerning the number of persons in the household, cooperation is generally higher, the higher the number of persons present. In five out of seven surveys a perfect linear relationship between cooperation and number of persons in the household was found. In terms of type of household, in 5 out of 7 studies cooperation was highest in households with partners and children. Cooperation was lowest in single households for 4 out of 7 studies.
- Highest cooperation was found in the more rural regions for 5 out of 7 studies, and lowest cooperation in the more urban regions, for 4 out of 7.

Subsequently, a series of multivariate stepwise logistic models was fitted. Separate models were fitted for each survey²⁾. Table 5.3.3 shows results for contact, table 5.3.4 for cooperation.

²⁾ We refrained from overall analyses on the complete set of data. Although that would enable studying interactions with sample type (person or household), it would necessitate weighting of the data, as underprivileged groups are oversampled in the SLCH and elderly are undersampled in the LFS. This we deemed out of scope for the present purpose.

LFS08	HS		DPES		SLCH			
	log Wald odds	p	log Wald odds	p	log Wald odds	p		
	0,75	**	0,81	ns		0,76	*	
	0,64	**	0,44	*		0,58	**	
	1.14	ns	1.28	ns		1.04	ns	
	1.07	ns	1.18	ns		0,92	ns	
	39.3	***	ns		10.0	*	94.5	***
	0,91	ns			0,41	**	0,86	*
	0,86	**			0,54	*	0,74	***
	0,73	***			0,64	ns	0,68	***
	0,75	***			0.70	ns	0,56	***
	ns		ns		ns		ns	
	.11		.17		.21		.17	

The multivariate analyses changes the interpretation of relations between auxiliary variables and chance of contact and cooperation for a number of variables. To aid the interpretation of the change in findings, bivariabe (bv) and multivariate (mv) results are summarized in table 5.3.5.

5.3.4 Logistic regression of auxiliary variables on cooperation (yes/no)

	ASU		SCS			LFS05			
	Wald	log odds	p	Wald	log odds	p	Wald	log odds	p
Gender (mixed is referent)	ns			ns			ns		
men									
women									
Age (youngest is referent)	ns			65.2		***	48.9		***
30-45					1,04	ns		0,67	***
45-65					0,92	ns		0,64	***
>65					0,61	***		0,52	***
Ethnicity (Dutch is referent)	ns			31,6		***	38.2		***
non-western					0,85	ns		1.27	**
western					0,49	***		2.65	***
mixed					0,99	ns		1.20	**
Household composition (single is referent)	ns			91.7		***	22.1		***
partners no children					1.42	***		0,77	***
partners with kinderen					1.42	***		0,85	**
single parent					1.34	**		1.04	ns
other					0,46	***		1.01	ns
Owner/renter (renter is referent)	no X			24.7		***	no X		
owner					1.22	***			
% with job in household core (all job is referent)	24.0		***	6,5		*	ns		
0%		0,77	***		0,95	ns			
50%		0,78	**		0,87	*			
% with allowance in (all allowance is referent)	ns			ns			31.2		***
0%								1.40	***
50%								1.21	*
Dwelling (single-family dwelling is referent)	19.4		**	no X			18.8		**
appartement building		0,88	ns					0,82	***
old people's flat		1.08	ns					0,97	ns
villa, mansion		1.47	**					0,97	ns
other		0,88	ns					0,78	*
Landline (yes is referent)	5,9		*	no X			27.4		***
no		0,87	*					0,80	***
Urban density (Rural is referent)	ns			18.6		**	ns		
very urban					1.23	**			
2					1.25	***			
3					1.18	**			
somewhat rural					1.11	*			
Province (Groningen is referent)	ns			51.7		***	38.7		***
Friesland					1.17	ns		1.41	ns
Drenthe					1.05	ns		1.29	ns
Overijssel					0,95	ns		1.27	ns
Flevoland					1.3	ns		-	
Gelderland					1.01	ns		1.23	ns
Utrecht					0,96	ns		1.16	ns
Noord Holland					1.01	ns		0,95	ns

LFS08	HS			DPES			SLCH					
	Wald	log odds	p	Wald	log odds	p	Wald	log odds	p	Wald	log odds	p
215.7			***	ns			ns			114.7		***
	2.19		***									
	1,83		***							1.23		***
178.4			***	59.1		***	8.5		*	14.6		**
	0,7		***		0,66	***		1,15	ns		0.90	***
	0,62		***		0,65	***		1,08	ns		0.94	*
	0,66		***		0,72	***		0.80	ns		0.96	ns
113.8			***	16.9		***	ns			44.3		***
	1.60		***		1.46	***					1.24	***
	1.17		**		1.10	ns					1.02	ns
	0,93		*		na						na	
127.8			***	ns			ns			27.5		***
	1.73		***								1.10	***
	1.75		***								1.08	**
	0,98		ns								1.00	ns
	na										0.80	***
no X				ns			20.5		***	16.3		***
								1.54	***		1.02	ns
											0.93	**
no X				8.4		**	ns			9.9		**
					1.19	**					0.92	***
					-						na	
no X				5.3		*	ns			ns		
					1.00	ns						
114.4			***	431.1		***	76.8		***	no X		
	0,95		ns		1.10	ns		1.44	**			
	1.23		*		0.90	ns		1.33	ns			
	1.23		***		0,84	ns		0,78	ns			
	0,69		***		0,22	***		0.30	***			
34.5			***	30.1		***	ns			no X		
	0,89		***		0,77	***						
25.7			***	40.6		***	19.5		**	43.9		***
	1.05		ns		0,72	***		0,62	**		1.18	***
	0,98		ns		0,63	***		0,89	ns		1.19	***
	0,91		*		0,75	***		0,77	ns		1.23	***
	0,91		*		0,78	**		0,62	**		1.14	***
244.1			***	5.2		*	ns			100.4		***
	1.21		**		1.45	*					1.02	ns
	1		ns		0,95	ns					0.95	ns
	1.03		ns		1.20	ns					0.88	*
	1.07		ns		1.20	ns					0.93	ns
	0,99		ns		1.35	*					0.92	ns
	0,85		**		1.12	ns					0.89	ns
	0,85		**		0,94	ns					0.78	***

5.3.4 Logistic regression of auxiliary variables on cooperation (yes/no) (end)

	ASU		SCS			LFS05			
	Wald	log odds	p	Wald	log odds	p	Wald	log odds	p
Zuid Holland					0,92	ns		1.01	ns
Zeeland					.80	ns		-	
Noord Brabant					1.00	ns		1.13	ns
Limburg					1.51	***		1.34	ns
% non-western minorities (<5% is referent)	13.5		**	ns			ns		
5-10%		0,82	*						
10-20%		0,89	ns						
20-40%		0,75	**						
>40%		0,89	ns						
Mean value of real estate (most expensive is referent)				ns			12.2		**
1e quartile								0,86	*
2e quartile								0,82	**
3e quartile								0,87	**
Nagelkerke R ²	.01			.05			.02		

Note: 'no X': variable not linked to survey; 'na', category not applicable.

*** p < .001; ** p < .01; * p < .05.

The referent for owner/renter in the SLCH is 'unknown' the first category is owner, the second is renter.

The referent for gender in HS, DPES and SLCH is 'men'.

5.3.5 Summary of bivariate (bv) and multivariate (mv) relations between contact and cooperation and auxiliary variables per survey

	Contact											
	ASU		SCS		LFS05		LFS08		HS		DPES	
	bv	mv	bv	mv	bv	mv	bv	mv	bv	mv	bv	mv
Gender	***	***	***	***	***	***	***	***	***	*	**	*
Age	***	***	***	***	***	***	***	***	***	***	***	ns
Ethnicity	***	*	***	***	***	*	***	***	***	**	***	ns
Household composition	***	***	***	***	***	***	***	***	***	***	***	***
Number of persons in household	***		***		***		***		***		***	
Owner / renter	-	-	ns	ns	-	-	-	-	***	*	***	ns
% of household core with job	***	*	***	ns	***	ns	-	-	***	**	ns	ns
% of household core with social benefit	***	ns	**	ns	***	ns	-	-	*	**	***	***
Dwelling	***	***	-	-	***	***	***	***	***	***	***	***
Landline	***	***	-	-	***	***	***	***	***	***	***	***
Urbanicity	***	***	***	***	***	***	***	***	***	ns	***	ns
Province	***	***	***	ns	***	***	***	***	***	***	***	***
% non-western minorities in neighbourhood	***	ns	***	ns	***	ns	***	***	***	ns	***	ns
Income in quintiles	***		**		***		***		***		ns	ns
Mean real estate value in neighbourhood	***	ns	***	***	***	ns	***	ns	***	ns	***	ns

Note: *** p < .001; ** p < .01, * p < .05.

'-' : auxiliary variable not linked to this survey.

LFS08		HS		DPES		SLCH					
Wald	log odds	p	Wald	log odds	p	Wald	log odds	p	Wald	log odds	p
	0,7	***		0,83	ns					0.80	***
	0,88	ns		0,94	ns					0.85	*
	0,99	ns		1.12	ns					0.94	ns
	1.12	ns		1.12	ns					1.13	ns
20.7		***	ns			ns			ns		
	1.06	ns									
	0,96	ns									
	0,89	**									
	0,88	*									
88.5		***	ns			13.4		**	ns		
	0,75	***					0,75	*			
	0,77	***					0,78	*			
	0,86	***					0,66	***			
.03			.08			.06			.01		

Cooperation

SLCH		ASU		SCS		LFS05		LFS08		HS		DPES		SLCH	
bv	mv	bv	mv	bv	mv	bv	mv	bv	mv	bv	mv	bv	mv	bv	mv
***	***	*	ns	***	ns	ns	ns	***	***	ns	ns	ns	ns	***	***
***	***	*	ns	***	***	***	***	***	***	***	***	***	***	*	***
***	***	ns	ns	***	***	***	***	***	***	***	***	***	ns	ns	***
***	***	***	ns	***	***	*	***	*	*	***	ns	ns	ns	***	***
***	***	***	ns	***	***	ns	***	***	***	***	***	ns	ns	***	***
***	*	-	-	***	***	-	-	-	-	***	ns	***	***	***	***
***	ns	***	***	***	*	***	ns	-	-	***	***	***	ns	ns	***
***	ns	*	ns	*	ns	***	***	-	-	***	ns	ns	ns	*	ns
-	-	***	***	-	-	***	***	***	***	***	***	***	***	***	-
-	-	***	*	-	-	***	***	***	***	***	***	***	***	ns	-
***	***	ns	ns	***	***	***	ns	***	***	***	***	***	ns	***	***
***	***	ns	ns	***	***	***	***	***	***	***	***	*	ns	ns	***
***	***	***	***	***	ns	ns	ns	***	***	ns	ns	ns	ns	ns	ns
***	***	***		***		***		***	***	ns		***		*	
***	ns	**	ns	***	ns	***	***	***	***	ns	ns	*	**	*	ns

Value of real estate was no longer related to contact in six of the seven surveys, while percentage of non-western minorities in the neighbourhood was no longer related in five surveys. The percentage of household core members with a job did not have a bivariate relation to contact in one survey, and in three more surveys neither a multivariate relation. Number of household members or on social support was no longer related in four surveys. Other variables that sometimes were not related to contact were Age, Ethnicity, Urbanicity, Province, and Owner/renter. Gender, Household composition, Dwelling, and Landline were related to contactability in all surveys. The effect of Gender indicated that mixed gender households had a higher contact rate than same sex households. Obviously, the variable gender is contaminated by other influences. Same sex households are, even in the Netherlands, mostly single households. Stoop (2005) shows that single men are mostly young, while single females are mostly elderly. Households with partners and children were generally easiest to contact, and always easier to contact than single households. The contact rate of other categories, like partners without children, and single parents, were different from singles in four surveys only.

The effect of Dwelling consistently showed that households in apartment buildings were harder to contact than households in single family dwellings. There were no differences in contactability between single family dwellings and the other categories.

Households with a fixed landline were easier to contact in all surveys. Having a fixed landline is also the variable with the most consistent influence on cooperation. It was related to cooperation in four out of five surveys. In all, cooperation was higher for households with a fixed landline.

Consistent too was the role of Owner/renter: in three out of four surveys the relation was significant, in all surveys indicating that cooperation was higher for owners. The only other variable with somewhat consistent findings was Value of real estate: in three of the seven surveys the relation was significant, indicating that cooperation was highest in the most affluent neighbourhoods.

Dwelling is related to cooperation in all (CAPI) surveys. The interpretation is however far from univocal. In one survey, the LFS05, people in apartment buildings cooperated less than people in single family dwellings. In the ES, the opposite is seen. And in three surveys there is no difference between the two categories. In some surveys, the highest cooperation is found for households living in a villa, but in three other surveys cooperation of this category is lower than that of single family dwellers.

In six out of seven surveys, the variable Age was multivariately related to cooperation. In three of those, the youngest age group had the highest cooperation rate. In the SCS the youngest group had a higher cooperation rate than the eldest

group, in the SLCH the rate was higher than the two middle groups, but not the eldest group, and in one survey there were no differences between the age groups. Gender was no longer significantly related to cooperation in multivariate analyses in five surveys.

Mixed findings were the pattern for all other variables. Cooperation in the households where no one hold a job was lower than in households where the whole core had a job, but higher in one other case, and not significant in others. The cooperation of native Dutch was highest in one survey, in three surveys it was the cooperation of non-western minority households, and in one survey that of the western minorities. In two other surveys no effect of Ethnicity was found. In three surveys the cooperation of households consisting of partners was higher than that of single households, but in one survey the relation was the opposite, while in three surveys there was no relation. Urban density was related to cooperation in five out of seven surveys, but in two of them, cooperation was lowest in rural areas, and in no single case was cooperation lowest in the most urban areas. Province had a separate influence on cooperation, but again, results were different for every survey.

Although these analyses show statistically significant results, the results are highly volatile. In fact, the relations are extremely weak, as is shown in an overview of the fit of the various models, expressed in the Nagelkerke pseudo R^2 (Nagelkerke, 1991). The Nagelkerke pseudo R^2 can be interpreted in a way analogous to R^2 in classical regression analysis as the proportion of explained variation by the model. The last row of table 5.3.4 shows that, in spite of all the auxiliary variables available to explain cooperation, including variables that are related to the survey subject, the model fit is extremely low. One percent of variance only is explained for the Survey of Amenities and Services Utilisation and the Survey of Living Conditions and Housing; two percent for the LFS 2005, and three percent for the LFS 2008. Most variance in cooperation is explained in the Health Survey: eight percent. The fit for 'contacted' is generally higher, although far from perfect, see the last row of table 5.3.3. The highest percentage of explained variance is for the Dutch Parliamentary Election Study: 21 percent.

5.4 Summary and discussion

In this chapter we have joined in a long research tradition, in which we searched for socio-demographic characteristics that can predict contact and cooperation in survey research. With the rise of adaptive survey designs, where treatment depends partly on characteristics of sample persons, attention for these correlates

is rekindled. Like many authors before us, we concluded from the literature that correlates for contact can be found, but that the picture is fuzzy for socio-demographic correlates for cooperation.

More than in establishing contact, cooperation is dependent on a complex of influences. The length of the survey, the topic, the sponsor, the contents of the advance letter, and the absence or presence of an interviewer will all influence the decision to participate. Prior studies differed in all these aspects, which makes comparison difficult. Furthermore, studies differ in the extent to which the important nonresponse reason 'not being able to cooperate' is counted as non-cooperation. People who are too ill to cooperate, or do not speak the language are categorised as 'not able'. As these will frequently be elderly people or foreigners, including this category in the definition of cooperation will influence predictability. By studying seven Statistics Netherlands surveys with different designs, but with comparable fieldwork procedures and equal definition of cooperation, we were able to gain a better understanding of which characteristics contribute to contactability and cooperation. Both bivariate and multivariate analyses were performed.

In these analyses we used a rather crude method, by implicitly assuming independence across observations. In fact, observations are nested within interviewers. The analyses we used underestimate the standard errors of the statistics. This means, however, that even the very low estimations we found are exaggerations.

It was found that bivariate results largely replicate earlier findings concerning noncontact. Men, singles, young persons, non-western minority groups, people living in apartments, in urban areas, people who do not have a fixed landline, people on social allowance, people in areas with the lowest property values, people in areas with the highest percentage of non-western foreigners and households where the entire core works are hardest to reach. In multivariate analyses, the influence of the value of real estate, the percentage of non-western foreigners, the number of working people in the household, and the number of people on social allowance disappeared. The effects of age, ethnicity and urbanicity were no longer present in all surveys. The only findings that were multivariately consistent in all studies were that men, singles, people in apartment buildings and people without a fixed landline had a lower contact chance. The latter finding may reflect the greater ease of contacting sample persons if a telephone number is known by the interviewers.

In spite of the consistent results, the fit of the models, as witnessed by the amount of variance explained, was not high though: around fifteen percent, with a high of 21% for the Dutch Parliamentary Election Study, and a low of six percent for the Survey of Consumer Sentiment.

Bivariate results for cooperation were less consistent: no single characteristic was found to be consistently related to cooperation in all surveys. Tendencies were for cooperation to be highest in the youngest age group, the most affluent neighbourhoods, in households where all members work, in larger households, households with children, and in rural areas. Lowest cooperation tended to be found in the eldest age group, when people did not have a listed telephone number, in neighbourhoods with a high number of ethnic minorities, and in households where nobody had a job. When tested multivariately, the pattern is even less clear. The only consistent findings were that people with a fixed landline, home owners and people in the most affluent neighbourhoods cooperate more. But even these findings were not significant in all surveys.

The finding that people with a listed landline cooperate more is remarkable. In the introduction was described how having a fixed landline is related to characteristics like age and ethnicity, that are also related to contact and cooperation propensity. The multivariate analyses showed that having a fixed landline has a separate influence on contact and cooperation, even after controlling for the influence of the socio-demographic variables.

The most noteworthy result of the multivariate analyses of correlates of cooperation is the extremely low fit of the models, with explained variance as low as one percent for two surveys, two and three percent for two others.

Although numerous researches have focussed on social-demographic correlates of cooperation, few researchers would maintain that these characteristics are determinants of nonresponse. They are seen as 'indicators for underlying social and psychological constructs, explaining the likelihood of nonresponse' (Stoop, 2005), as the producers of a set of psychological dispositions that affect the participation decision (Groves, Cialdini & Couper, 1992), or as indicators of centrality within society (Goyder, 1987). Stoop (2005) concluded that 'giving an overview of individual and household correlates of nonresponse may not advance knowledge on nonresponse, as the results so far indicate that background variables do not explain response behaviour'. Our analyses confirm that conclusion, and extend it. At least for survey cooperation, the 'may not' in the previous sentence can be supplanted by 'do not'.

Predictability is a necessary condition for the development of tailored fieldwork strategies that aim to determine beforehand which treatment to give to which sample unit. These findings therefore have implications for the development of such strategies. To a certain extent it will be possible to predict who will be hard to contact and who will not be able to participate, and take appropriate measures. Concerning cooperation, this prediction is not possible with the information at hand. Non-cooperation is by far the largest category of nonresponse. In the surveys studied in this chapter, around 20% of the eligible sample refuses, while between 2.5% and 6.3% is not contacted. Being able to predict cooperation and to fine-tune

measures to prevent refusal would be a valuable and cost effective supplement to the survey practitioner's tool kit. Future research needs to focus on finding meaningful correlates of cooperation. The study of paradata may offer a new impetus to finding these correlates.

6.

**Fieldwork strategy
at Statistics**

Netherlands

6.1 Introduction

In 2001 Statistics Netherlands employed their hitherto free lance field interviewers. During the years that interviewers worked free lance, they were given some very general guidelines on how to behave in the field, but generally followed their own initiatives. The way interviewers were remunerated, by paying for response cases, with a token compensation for nonresponse, may have induced bias, by stimulating interviewers to pursue the 'low hanging fruit'. Legislation forbade us however to enforce stricter rules and to monitor interviewers individually.

Employing the interviewers changed all that. A uniform fieldwork strategy was developed and introduced in 2003. Interviewers were monitored to see if they adhered to the new rules. A uniform fieldwork strategy guarantees that sample units have a similar chance of being contacted, while at the same time minimizing the fieldwork costs. The fieldwork strategy comprises that all sample units are visited at least once in the first half of the fieldwork period, either the first or the second visit must be on 'prime time' (after 5 pm, or on Saturday¹⁾), noncontacts must be visited six times, and visits must be spread over time, days and weeks within the fieldwork period. The first three contact attempts are to be face-to-face. After that, contact attempts by phone are allowed, but not required²⁾. Calling cards are left behind after the first three visits to alert the respondent to the interviewer's visit. This strategy was based on findings in survey literature about the relative contact rates of various contact times and call schedules (Barton, 1999; Campanelli, Sturgis & Purdon, 1997; Foster, 1997; Snijkers, Hox & De Leeuw, 1999). Social exchange theory (Dillman, 1978), suggested the calling cards as a means to express the sincerity of the survey request. Stressing face-to-face as the preferred contact mode was suggested by the higher response rates attained in this mode (Collins, Sykes, Wilson & Blackshaw, 1988; De Leeuw & Van Zouwen, 1988; Nicolaas & Lynn, 2002). The number of six contact attempts was determined after analyses that showed that up till the sixth visit significant changes were found in the number the additional contacts and responses were limited after this number, but extremely costly.

Introduction of the fieldwork strategy and monitoring program gave a huge boost to the response rates. The response rates for the Health Survey improved gradually with almost ten percentage points from 59% to 68% in the period from 2001 to

¹⁾ Sunday would undoubtedly be a good day to find people at home, but Statistics Netherlands interviewers do not work on Sunday. Neither do they work after 9 pm or prior to 9 am. Both are admittedly good times to contact people, but not necessarily good times to secure an interview.

²⁾ Most SN surveys are mixed mode nowadays, and sample persons with known telephone numbers will be called for a telephone interview. The analyses in these chapters stem from the pre-mixed mode area, where this rule was still valid.

2007³⁾, and for the Labour Force Survey with nine percentage points from 56% to 65%. The largest response gain came from diminishing noncontact rates: over eight percentage points less for the LFS, and six percent less for the Health Survey. But cooperations rates increased too, three and four percentage points respectively. These results made eager for more. Huge gains were made, but more was possible. Analyses showed that not all fieldwork rules were sufficiently adhered to by interviewers. These findings gave rise to a research project to find out why interviewers did not follow rules that were so patently sensible. Chapter 7 is a reflection of that work. One of the findings in that project was that interviewers needed to be convinced that the strategy worked. This appeared not to be an easy task. We could of course point to the higher response rates, but interviewers were keen to report that that the fieldwork strategy could work in general, but that in their specific circumstances entirely different methods led to better results. This paragraph describes some of the analyses that were done over the years to show that this strategy leads to the best results in the shortest time. It also shows some of the pitfalls in this kind of analyses.

In an early logistic regression analysis, the effect of a number of fieldwork rules on contact rates was assessed in the workload of the 2004 and 2005 Labour Force Survey and Health Survey, with a total of 213,000 records. A complete model was fitted, with four fieldwork elements (spreading over time periods, spreading over weekdays, first visit on prime time and first visit in the first half of the fieldwork period), population density in five categories, and sample (household or person) as factors, plus their interactions. Adhering to fieldwork strategy was coded as 0 (no adherence) or 1 (adherence). In successive steps, non-significant interactions were removed from the model. Table 6.1.1 shows the final model.

From these results it seemed clear that spreading of calls had the largest impact on contact rates, especially spreading over time. It surprised us that visiting at prime time seemed to diminish the odds of getting contact. 'Spreading over time periods' was involved in interactions, indicating that transgressions of other fieldwork elements are much less serious if visits are well spread. For example, when visits are spread over time, but not over weekdays, contact rate is 0.4% less. When weekdays are well spread, but time periods are not, the loss is 8.7%. However, when both (weekdays and time periods) are not spread, contact rate is 19.8% less than if both are spread.

³⁾ The introduction of mixed mode data collection from 2007 does not allow comparison of fieldwork response with later years, as in the mixed mode design only the most difficult cases who have not responded by web and have no listed telephone number are visited Face-to-Face.

6.1.1 Logistic regression on contact by fieldwork strategy, sample and urbanicity

Factor	B	S.E.	Wald	df	p	Exp(B)
Spreading over weekdays	0.8	.17	24.0	1	***	2.3
Spreading over time periods	1.5	.15	103.1	1	***	4.7
1 st or 2 nd visit on prime time	-0.4	.05	80.4	1	***	0.6
1 st visit in 1 st half of fieldwork period	0.6	.09	43.0	1	***	1.7
Sample kind	-0.2	.04	14.4	1	***	0.9
Population density			329.3	4	***	
Population density (1)	-1.1	.08	175.2	1	***	0.3
Population density (2)	-0.6	.09	55.0	1	***	0.5
Population density (3)	-0.3	.09	7.8	1	**	0.8
Population density (4)	-0.1	.10	1.3	1	ns	0.9
Weekdays x time periods	-1.0	.10	98.0	1	***	0.4
Weekdays x prime time	0.2	.09	5.6	1	*	1.2
Weekdays x 1 st half	0.4	.13	10.2	1	**	1.5
Time periods x prime time	0.5	.10	26.0	1	***	1.7
Time periods x 1 st half	0.4	.14	7.3	1	**	1.5
Population density x weekday			11.8	4	*	
Population density x weekday (1)	-0.2	.14	2.2	1	ns	0.8
Population density x weekday (2)	0.1	.15	0.9	1	ns	1.2
Population density x weekday (3)	-0.2	.16	1.1	1	ns	0.8
Population density x weekday (4)	-0.1	.17	0.2	1	ns	0.9
Constant	1.3	.11	151.9	1	***	3.8
Nagelkerke R ² = .23						

Note: *** p < .001, ** p < .01, * p < .05.

The interactions in this analysis suggested that adhering to fieldwork elements is additive: results are much better when two elements are applied than just one. Therefore, a measure was calculated, representing the amount in which fieldwork strategy was followed (from 0 elements to 4 elements). Sequences that ended after the first visit or where an appointment was made for a second visit were also excluded. The result is depicted in table 6.1.2.

6.1.2 Response and contact rates per number of correctly applied fieldwork elements by urban density

Correct fieldwork elements	Highly urban		Urban		Moderately urban		Moderately rural		Highly rural	
	response	contact	response	contact	response	contact	response	contact	response	contact
0	38	70	52	82	58	86	63	89	67	90
1	42	73	53	83	59	87	62	89	64	90
2	46	78	58	88	63	92	66	93	69	94
3	56	92	63	96	66	96	68	98	71	98
4	60	95	68	98	67	98	71	98	72	99
Difference	22	25	16	16	9	12	8	9	5	9

This analysis seemed to support the idea that response and contact rates were substantially higher if fieldwork strategy was adhered to in all its aspects. The sanction for not following strategy was especially high in urban areas, with a 25 percentage point difference between addresses where all elements of the fieldwork strategy were followed, and addresses where none were followed. What we failed to see initially however, was what the negative sign for 'visiting at prime time' should have alerted us to: the influence of the addresses that are hard to contact. The more difficult an address, the more visits have to be made, and the higher the chance that visits will not be well spread, for example. Also, it appeared that especially interviewers in highly urban areas were prone to visit their addresses during the evening. As urban areas have the lowest contact rates, this resulted in a negative log odds.

Complicated analyses may obscure what actually happens over the course of a sequence of visits. So, with the lessons above in mind, I will show rule by rule what the effect is on – mostly – contact rates. These analyses were performed on the 2008 workload of the LFS and Health Survey, with a total of 65,093 records. Where informative, urban density will be part of the analysis, as urbanicity has a marked influence both on contact rates as on the way interviewers behave in the field.

6.2 One of the first two visits should be after 5 pm or on Saturday

Table 6.2.1 shows contact rates by time period (weekday day, weekday evening, weekend) by sample type in the first visit. 'Evening' is defined as the time period between 5 pm and 9 pm.

6.2.1 Contact rate by visiting time and sample. First visit

	Address sample		Person sample	
	contact %	n	contact %	n
Weekday day	45	34,990	51	14,503
Weekday evening	57	8,321	70	3,115
Weekend	51	2,259	55	855

As table 6.2.1 shows, weekday evening visits have a higher chance of contact than weekday visits. The chance of contact on Saturday lies in between. A small minority

of interviewers chooses the 'prime time' periods with higher chance of contact for their first visit.

Table 6.2.2 shows these findings in more detail and shows results for different urbanicity areas, again for the first visit. In both sample types and in urban as well as rural communities and those in between, the pattern is the same: the later in the day, the higher the chance of contact. The only difference is in the level: contact rates are higher for person samples, and contact rates are higher for less urban sample units. The difference between a morning visit and a visit after seven pm is around 20 percentage points, and can be as high as 27 percentage points. As the difference between address samples and person sample is merely one of level but not of pattern, they will be collapsed in further analyses.

6.2.2 Contact rate by time of contact, sample type and urbanicity; first visit

	Address sample				Person sample			
	< 12 h	12-17 h	17-19 h	19-21 h	< 12 h	12-17 h	17-19 h	19-21 h
Highly urban	36	37	43	49	41	46	61	66
Urban	39	44	55	58	45	50	70	72
Moderately urban	43	48	63	67	47	55	72	71
Moderately rural	43	51	62	65	49	54	71	76
Highly rural	48	54	65	68	50	59	71	72

The chance of contact is high in the evening for the first visit, but that is not what is required of the interviewer: one of the first two visits should be on prime time.

Table 6.2.3 shows contact rates of first and second contact attempts by time of the attempt; weekday, evening or weekend. Contact rate is calculated on interviewer instigated attempts only, visits as a result of prior appointments are not included. The cumulative contact rates show that, if in any of the two visits an evening call was attempted, the contact rate after two attempts was substantially higher than if the two visits were made during the day, although the evening-day combination is less successful than other combinations. The day-evening combination is the most popular sequence, with 46% of cases handled in this way. The day-day combination is popular as well, though, with 28% of cases. If we include the weekend calls here, which are almost exclusively made during the day, the percentage of daytime-daytime calls is 37%. Although some of the less successful strategies fall within the rules of the fieldwork strategy (all weekend – weekday combinations), I focus here on the cases that do not conform to fieldwork strategy, i.e., the 'day-day' combination.

6.2.3 Contact rate for first and second attempt by time of day

	% contact	n
1st attempt		
weekday	46	49,012
evening	60	11,285
weekend	52	3,082
2nd attempt		
d-d	42	8,917
d-e	57	14,633
d-w	46	2,135
e-d	34	2,315
e-w	41	332
e-e	45	174
w-d	39	776
w-e	58	583
w-w	36	94
weekday	40	12,031
evening	55	16,984
weekend	45	2,567

Note: Spontaneous visits only.

d-d = weekday-weekday, d-e = weekday-evening, d-w = weekday-weekend, etc.

The interpretation of the impact of the various behavioural options is influenced by the fact that there are differences between areas in the distribution of day and evening visits. Table 6.2.4 shows that the amount of evening visits at the first attempt is larger and the amount of weekday attempts smaller, as areas get more densely populated. The amount of weekend visits does not vary with urban density. In as much as contact rates are lower in more densely populated areas, this influences the interpretation of chance of success of daytime – evening combinations.

6.2.4 Percentage of weekday, evening and evening visits by urban density; first visit

	Weekday	Evening	Weekend
Highly urban	71	24	5
Urban	76	19	5
Moderately urban	78	17	5
Moderately rural	81	15	4
Highly rural	83	12	5

While daytime visits have a lower contact probability than evening calls, they have a higher probability of being resolved in the same call as the contact is established.

When contact is made during daytime in the first call, an appointment for the interview is made in 28% of cases. When the contact is made in the evening, the percentage of appointments is 39%. In weekends, 44% of contacts result in appointments. The higher percentage of appointments during evening hours applies mainly to the hours of 5 to 7 pm, there is generally no difference between the number of appointments made later at night and the number made during the day.

In this section we have found that the fieldwork rule that one of the first two visits should be at prime time is sound advice, as the chance of contact is substantially higher if the rule is followed.

6.3 First visits should be in the first half of the fieldwork period

The rationale for this rule is that interviewer should reserve sufficient time to visit the required number of times, to spread over time periods and weeks, and to offer respondents time to make an appointment. Fieldwork periods for the LFS and Health Survey are one month, Table 6.3.1 shows contact rates for addresses where this rule is followed versus where it is not followed by urban density.

6.3.1 Final contact rate by timing first contact (first or second half fieldwork period) by urbanicity

	First half	Second half
Highly urban	88	82
Urban	93	90
Moderately urban	95	94
Moderately rural	97	95
Highly rural	98	95
All	94	91

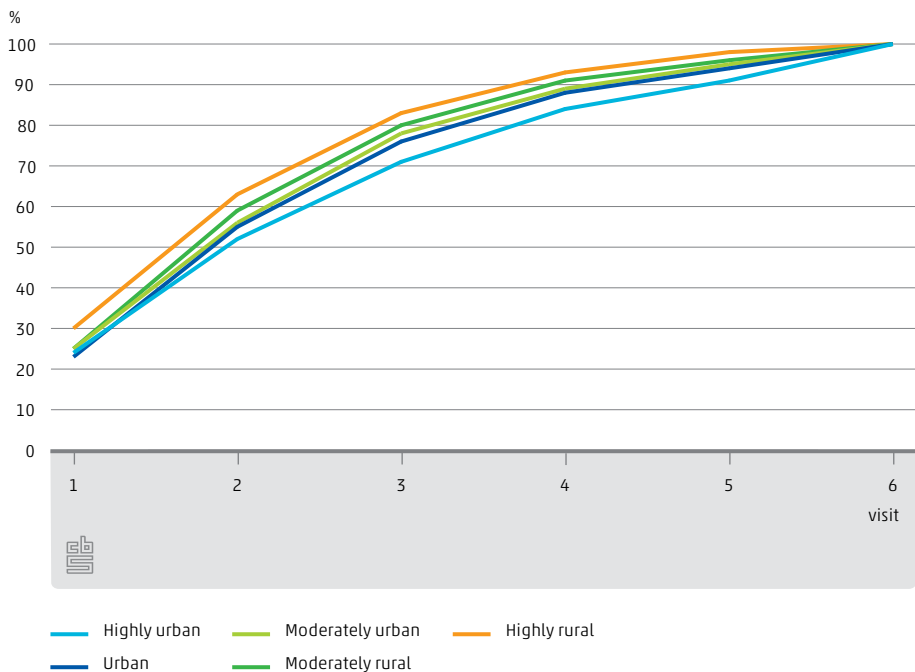
Over all, contact rates are three percentage points higher for addresses where this rule is followed, than when it is not followed. The difference is larger for highly urban areas, which makes sense as generally more visits need to be made to get in contact with the sample person or household (a mean of 1.7 visits is necessary for first contact in highly rural areas, against 2.0 for highly urban areas). Of course, the

difference between the first half of the fieldwork period and the second half may be only one day. If the first visit is made in the first week of the fieldwork period, final contact rates are slightly higher for the most urban areas: 90% and 94% respectively.

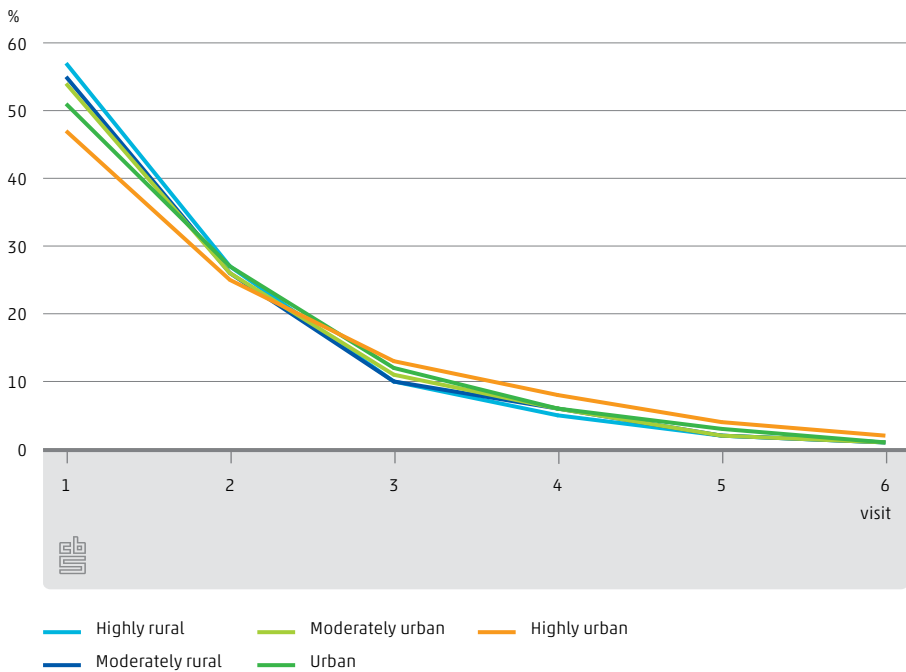
6.4 Noncontacts should be visited six times

By this rule is meant that the interviewer should visit six times (a seventh time is allowed if an appointment is made in the sixth visit). No more and certainly no less. The wisdom of this rule can be seen in figure 6.4.1 and figure 6.4.2. Figure 6.4.1 depicts the percentage of responses that is gained in each visit (from the number of responses that is eventually attained). Figure 6.4.2 depicts the number of first contacts in each visit.

6.4.1 Cumulative percentage of response per contact by urbanicity



6.4.2 Percentage of first contacts from all contacts by call number and urbanicity



From figure 6.4.1 can be concluded that a substantial number of responses is attained in the last visit. From highly urban to most rural areas, the last visit brings in respectively nine, six, five, four and two percent. So, especially in the urban areas, the sixth visit is very important. Figure 6.4.2 shows that even in the sixth visit first contacts are procured, although not many: two percent of contacts in highly urban areas, and one percent in other areas.

Cooperation rate is lower for sample units that are contacted after more effort. The cooperation rate for sample units that are contacted in the first two visits does not differ (70% and 69% respectively), but from the third visit, cooperation diminishes significantly with each contact attempt: 67%, 62%, 58% and 55% cooperation for sample units that are contacted for the first time in the third to sixth visit respectively ($F(5,58222) = 53.5, p < .001$). As the definition of cooperation that is used throughout this manuscript is cooperation from those contacted sample persons that are able to cooperate (COOP3, AAPOR 2008), this trend does not reflect insufficient time late in the fieldwork period. The finding makes sense, of

course. Had these persons been eager to cooperation, they could have contacted the interviewer any number of times, from the information on the calling cards, or could have been present at the suggested appointment the interviewer proposes after the third visit.

In spite of the lessened enthusiasm of the later contacts, it can still easily be seen that arresting the effort before the agreed six visits costs responses, and more so in the more urban areas.

6.5 Visits should be spread over time, days and weeks of the field-work period

Table 6.2.3 already gave a first glimpse of the obstinacy of the subject 'spreading over time'. The least successful combination of times for the first two visits were indeed the daytime-daytime or weekend-weekend combinations, but one of the most successful combination is the evening-evening combination. And although the day-evening combination is very successful, the well spread evening-day combination is not.

A pitfall in studying spreading patterns when you try to predict the successful outcome of the third visit is that the chance of a favourable outcome is higher if the first two visits were sub-optimal. The pattern with the highest chance of success in the third visit, for example, is 'weekday-weekday-evening'; that does not mean of course that this is the way to go. Table 6.5.1 shows contact rates for day, evening and weekend sequences of spontaneous third visits where no prior contact was made, ordered from most successful to least successful.

The best way to spread over times is to follow up a non-successful day time visit with a visit later in the day. That goes for early evening visits too: a visit between 5 pm and 7 pm that did not result in contact, should be followed up by a visit after 7 pm. After an unsuccessful evening visit, the best follow-up visit is again an evening visit (on another day), and even after two unsuccessful evening visits, the best chance is again an evening visit, although by this time chances are slim that the household will ever be contacted.

6.5.1 Contact rates of call sequences in third attempt

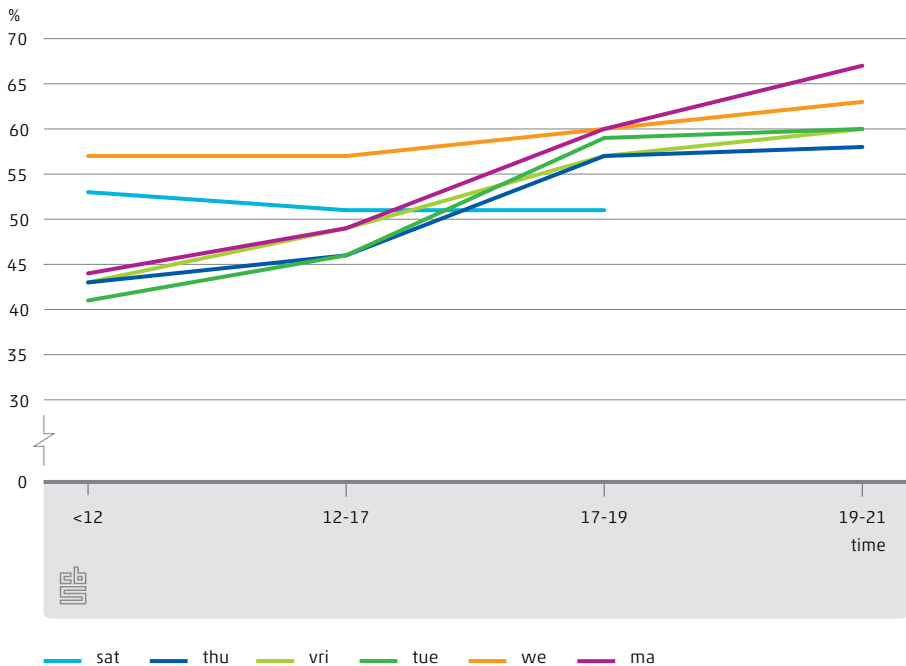
Third attempt	% contact	n	Third attempt	% contact	n
d-d-e	54	2,901	e-d-d	35	635
w-d-e	49	231	d-e-w	34	477
d-e-e	46	2,761	d-w-d	33	473
w-e-e	46	97	w-d-d	33	170
d-d-w	46	376	e-w-d	32	78
d-w-e	46	595	e-d-w	31	115
d-w-w	46	61	d-e-d	30	2,933
e-d-e	41	737	e-e-d	26	391
w-d-w	38	60	w-e-d	26	113
d-d-d	38	1,741	weekday	33	6,581
e-w-e	38	98	evening	48	7,955
e-e-e	35	447	weekend	38	1,226

Note: sequences with N < 50 are removed.

d-d-d = weekday day-weekday day-weekday day, e = evening, w = weekend.

To complicate matters, there are marked interactions between chance of contact, time of visit, and day of the week. Figure 6.5.2 shows that Monday nights have a ten percentage points better chance of contact than Thursday nights, for example, while Wednesday is a good day to try daytime visits, with chances of contact that are as high as Thursday nights.

6.5.2 Contact rates by weekday and time; First visit



Apart from spreading over times and days, interviewers should also spread their visits over the weeks of the fieldwork period, in order to contact people who are away from home for a length of time. Table 6.5.3 shows chance of contact in the second spontaneous visit, by number of days between the first and second visit.

Chances are slightly higher if interviewers wait at least one week for the next trial. The highest chance of success is when the second visit is done on the same day as the first. This finding is contaminated with timing, however, as the second visit on the same day will always be later. As the finding that contact chance is highest when two evening visits are done, this finding equally shows that it is not necessarily bad not to spread over days.

6.5.3 Chance of contact of the second spontaneous visit, by number of days between the first and second visit

Number of days between 1st and 2nd visit	Contact
0	60
1	34
2	33
3	34
4	34
5	35
6	35
7	34
8-13	36
14+	38

6.6 Summary and discussion

In this chapter I have evaluated a number of aspects of the fieldwork strategy interviewers should adhere to. It could be unequivocally shown that it makes sense to start work early, to start visiting on prime time early on, and to visit no less than six times in case of noncontact. The rule that visits should be spread over time, days and weeks of the fieldwork period proved to need qualification: not under all circumstances is spreading over time the best action, as it is better in terms of contact to visit twice during evening hours. Spreading over days is equally not always best, as the best action is to visit twice on the same day.

However, although it would be very efficient to visit all households several times during evening hours, this is on the one hand not possible, and on the other hand

not wise. Statistics Netherlands interviewers generally work 20 hours a week. That means they should work every day of the week from 5 pm to 9 pm to be able to do that. The interviewer corps would dwindle spectacularly if we would propose such a regime. More importantly, we have found evidence that a number of day time visits are necessary to preclude biased results. Chapter 8 shows how interviewers who work exclusively during evening hours produce biased Labour Force results. Daytime visits are necessary, but should be handled with discretion. The strategy that is implicit in the fieldwork rules, to first visit during the day, and in the second during the evening, is a good one. In order to free valuable evening time for persons who will never be at home during the day, a differential fieldwork strategy will be the next step, in which interviewers will be informed which sample units can safely be visited during the day, and for which sample units they should avoid that.

7.

**Understanding
factors leading to
interviewers'
non-compliance
with fieldwork rules**

7.1 Introduction

Most survey organisations performing fieldwork will have formulated contact procedures for their interviewers, stating the minimum number of visits to each sample unit, and the timing of those visits. For example, all countries participating in the European Social Survey instruct their interviewers to *"...make a personal visit on a minimum of 4 occasions, at different times of the day and spread across the fieldwork period before you classify the address/household/individual as unproductive. At least one of these personal visits should be in the evening ... and one at the weekend"* (ESS, 2010). These contact procedures guarantee that sample units have a uniform chance of being contacted. Interviewers who put in less effort for difficult cases, for example by not visiting during evening hours in some neighbourhoods, or not visiting a sufficient number of times, may bias substantial results by causing differences in distribution of respondents (Couper & Groves, 1992; Purdon, Campanelli & Sturgiss, 1999; Wang, Murphy, Baxter & Aldworth, 2005; West & Olson, 2010). A second reason why contact procedures should be followed is that they are designed to lead to the highest chance of contact (Barton, 1999; Campanelli, Sturgis, & Purdon, 1997; Foster, 1997; Snijkers, Hox, & De Leeuw, 1999, Stoop, Billiet, Koch & Fitzgerald, 2010). The third reason is that contact is established with the smallest number of visits, which obviously has positive implications for the costs of fieldwork.

For these reasons, Statistics Netherlands has likewise formulated a fieldwork strategy for field interviewers, aimed at optimising the chance of response, while at the same time minimising costs of face-to-face fieldwork. Specifically, interviewers are supposed to visit all sample units at least once in the first half of the fieldwork period, to plan either the first or the second visit on 'prime time' (after 5 pm hours, or in the weekend), to visit noncontacts six times, and to spread their visits over time and days within the fieldwork period. The first three contact attempts are to be face-to-face. After that, contact attempts by phone are allowed, but not required. After each of the first three unsuccessful visits, calling cards are left at the address, with on each card increasing information to facilitate the respondent's contacting the interviewer.

Analyses of fieldwork paradata show that this fieldwork strategy is successful in heightening chance of response, while lessening the number of visits necessary to attain an end result (Luiten, 2006, see also chapter 6).

The importance of fieldwork strategy for results makes that interviewers are monitored constantly on adherence (De Vree, De Bie & Luiten, 2006), and adherence to strategy is an important aspect of interviewer evaluation. In spite of this, most aspects of fieldwork strategy are followed only partly. First visits in the first half of the fieldwork period are well adhered to, but all (measurable)

other aspects are only followed in about half the cases. Interviewers are either not willing, or not able to follow the strategy in the other half. The Statistics Netherlands interviewers are not alone in this. Stoop et al., (2010) show that interviewers in a number of countries participating in the European Social Survey likewise do not adhere to the prescribed fieldwork strategy to a substantial degree, resulting in a sub-optimal contact rate.

Before deciding if more coercion would lead to the desired result of better adherence, it was felt that a better understanding of interviewers' reasons for not adhering was needed. Not much is known about the determinants of rule following in interviewers, but a body of research exists in other fields, e.g., concerning compliance of fishermen to quota rules (Hønneland, 1999; Raakjaer Nielsen, 2003), of farmers to agro-environmental rules (e.g., Winter & May, 2001), of common pool resource management (Jenny, Hechavarria Fuentes & Mosler, 2007), the maintenance of law and order (e.g., Elffers, Van der Heijden & Hezemans, 2003) and human resource management (Thompson & Heron, 2005). Literature in these fields and more general psychological and sociological research into determinants of rule compliance (e.g., Cialdini & Goldstein, 2004; Thibaut, Friedland & Walker, 1974; Tyler, 2006) suggest that a number of dimensions underlie the extent to which rules are followed.

To aid the reader in keeping track of these dimensions, table 7.1.1 summarizes what will be described in the remainder of this paragraph.

7.1.1 Dimensions potentially underlying interviewers' rule compliance

Dimension	Subdimension	Interpretation
Traditional authority	Coercion	Likelihood of detection, severity of sanction
	Obligation to follow rule	Sense of obligation
Procedural justice		Involvement of subjects in the decision on the regulation
Interactional justice	Attention for rule	Attention given by management
	Knowledge of rule	
Distributive justice		Remuneration, bonus, distribution of difficulties
Contextual factors	Compatibility	Compatibility with personal life
	Work pressure	Subjective feeling of work pressure
Legitimacy of rule		Does the rule make sense
Social norms		Influence of other interviewers
Personal norms		Normative disposition to follow the rules
Routine		Number of years worked as interviewer
Region		Interviewer's regional manager
Urban density		Urban density of area where interviewer generally works

The first studies into the mechanisms underlying rule compliance tried to understand compliance as the assessment of the benefits of compliance or non-compliance, the likelihood of detection and the severity of the sanction. Numerous

empirical studies and theoretical contributions have provided evidence that this traditional model – also called rational choice model (Cornish & Clarke, 1986), economic model (Jenny et al., 2007) or instrumental model (Raakjaer Nielsen, 2003) – is too simplistic to predict human behaviour (e.g., Kahneman & Tversky, 1979; Thibaut et al., 1974). The tactics relating to sanctioning and surveillance in the traditional model are what Cialdini and Goldstein (2004) call harsh tactics. They are the tools of authority based on hierarchy. Soft tactics on the other hand rely on authority based on the expertise and credibility of the institution issuing a regulation. Emans et al., (2003) showed that supervisors whose compliance-gaining repertoire included the use of both soft and harsh tactics were most likely to elicit compliance with their requests.

The soft tactics Cialdini and Goldstein (2004) describe, are related to the concept of legitimacy. If people comply with the requests of authorities because they feel that their decisions are right and ought to be followed, they may do so because they feel that the authority that issued the rules is legitimate (Tyler, 2006), because the way the rules are issued is transparent and just, or because the rule itself is judged to make sense (Raakjaer Nielsen, 2003). Because of legitimacy, people feel that they ought to comply with decisions and rules, following them voluntarily out of obligation rather than out of fear of punishment (Tyler, 2006).

The next dimension is that of procedural justice (DeCremer & Tyler, 2005; Thibaut et al, 1974; Thompson & Heron, 2005; Tyler, 2006), also called procedural fairness. Although the precise definition of the concept differs across authors, all stress the importance of the involvement of subjects in the decision on the regulation.

Interactional justice concerns the transparency and effectiveness of communication, and the respect of the organisation towards their people (Thompson & Heron, 2005; Jenny et al., 2007; Kickul, Neuman, Parker & Finkle, 2001; Flaherty & Moss, 2007). If sufficient attention has been given to the rule, the communication has been effective and people know the rule and know what is expected of them.

Distributive justice is concerned with fairness and the fair distribution of pleasure and pain (DeCremer & Tyler, 2005, Barling & Phillips, 1993, Jenny et al, 2007).

Cialdini and Goldstein (2004), Jenny et al., (2007), Raakjear Nielsen (2003) and Tyler (2007) also refer to social and personal norms as influencing the amount of compliance to rules. Social norms may tell us what is typically approved of, as well as what is typically done; the influence of peers is of relevance here. Personal norms are the moral values or personal standards with which people regulate their behaviour. These may be expressed as a general normative disposition to follow the rules.

Jenny et al., (2007) recognize the importance of contextual factors in rule following, specifically the extent to which rules are compatible with a person's private life and the extent to which routine and habits may play a role. A contextual influence that in qualitative work done in preparation of this study

(Luiten, 2008) was found to be specifically of influence for interviewers was work pressure. This may for example hinder the amount of work that can be done in the evening, or the amount of spreading over time.

An important contextual influence on interviewers is formed by their regional managers. Once a month, interviewers come together with their regional managers to discuss work progress. The 13 regional managers receive monthly monitoring information on their interviewers, including adherence to fieldwork strategy. Twice a year, interviewers are evaluated by the regional managers. Response rates, contact rates and adherence to fieldwork strategy are important aspects of this evaluation. The evaluation may determine whether interviewers proceed to a next salary scale, or get an additional bonus. Regional managers differ among themselves in the importance they ascribe to the fieldwork strategy and they tend to be more lenient to interviewers with higher response rates.

A further contextual influence is exerted by the level of effort an interviewer has to undertake to reach a sample unit, which is typically higher in more urban areas. The added effort could also influence the extent to which fieldwork directions are followed. If more visits are necessary, the chance that a visit is not spread over time increases, for example. Urbanicity was therefore added as another contextual influence on the interviewers' work.

Routine is another contextual factor identified in literature (Jenny et al., 2007). The interviewers working for Statistics Netherlands have in the majority of cases worked as interviewer for a very long time. More than half have been interviewer for over ten years, and a substantial number of them for up to 25 years. During most of that time, they were paid in piece for each response case, but were virtually left to their own devices as to how to achieve that goal. Two circumstances changed this situation: due to changes in legislation, the interviewers had to be engaged by Statistics Netherlands, and were henceforward paid a salary based on the hours of their contract (mostly 20 – 24 hours a week). Secondly, diminishing response rates and heightening costs necessitated more stringent rules concerning fieldwork strategy, leading to the design of the present fieldwork strategy in 2003. For the first time, interviewers could be obliged to work on times they would not spontaneously favour, i.e., in the evening and on Saturday. This led to what DeCremer and Tyler (2005) call a mixed-motive situation, in which self-interest is not always consistent with that of the organisation. Length of the interviewers' employment is used to operationalize routine.

It may be expected that interviewers who were employed after the fieldwork strategy was already in operation comply better than interviewers who had to change their behaviour after 20 years of doing things their own way. On the other hand, some of the longer active interviewers were involved in the design of the fieldwork, which may have induced them to comply better with the new

demands than their uninvolved colleagues (DeCremer & Tyler, 2005; Thibaut et al., 1974;Thompson & Heron, 2005; Tyler, 2006).

These dimensions may influence compliance in their own right, but may also interact. Thompson and Heron (2005) and Kickul et al., (2001) find interactions between procedural justice and interactional justice, Thibaut et al., (1974) between procedural justice, distributive justice and the likelihood of detection of transgression. Cialdini and Goldstein (2004) find that social and personal norms are only of influence if they are saliently present in conscience, i.e., if sufficiently attention is given to the rule, not only at the time of introduction of the rule, but also afterwards.

In this paper we investigate why interviewers do not comply to fieldwork rules. Noncompliance may lead to nonresponse, unnecessary costs and biased results. 14 possible influences on rule compliance in general were identified, and their respective influences on interviewer noncompliance are studied.

7.2 Method

To study whether the dimensions identified influence compliance to rules, the interviewers' work in 2010 and half of 2011 on the Labour Force Survey and the General Health Survey was analysed for rule transgression. The dataset consisted of 80,892 records. Mean number of addresses per interviewer was 270 (SD 123). Interviewers with a workload of less than 50 addresses were removed from analysis.

A (paper) questionnaire was devised to measure the extent to which the dimensions described above influence interviewer behaviour. The questionnaire was introduced to the interviewers as evaluation of the fieldwork strategy. The questionnaire was sent to all face-to-face interviewers (n=237). Interviewers were paid one hour to fill in the questionnaire, but participation was voluntary. Response was 95% after two rappels. See appendix 3 for the questions, ordered according to dimension; this was not the precise order for the interviewers though. Apart from the questions concerning the dimensions described above, questions were added about contract hours, other paid jobs and education. One question determined the extent to which interviewer thought they followed each rule and a number of questions went into an issue that constitute the core of any possibility to analyse interviewer's behaviour in the field, i.e., the way they handle the administration of the call records. Specifically, we were interested in the extent to which all visits and telephone call attempts were recorded and the accuracy of filling in the requested information.

Questions were developed to capture each of the dimensions described in the introduction. Mostly, multiple questions were devised to tap a dimension, but this was not always feasible. Sometimes one question was deemed to be sufficient to adequately measure a dimension; for example, routine was determined by one question on the number of years the interviewer had worked for Statistics Netherlands. Sometimes more than one question on a subject would make interviewers aware that the questionnaire had more motives than a mere evaluation of the fieldwork strategy.

After data entry of the paper questionnaires, the data were examined for missing values. Missing values were generally few, and were imputed with neutral values.

Dependent variables

For this paper, we concentrate on four dependent variables: (1) the percentage of cases in which the first or second visit was on prime time, (2) in which non-contacts were visited six times, (3) in which visits were spread over time slots, and (4) in which the third calling card was used. The choice for this particular calling card was made because of the high percentage of interviewers using the first and second card. Influence of the psychological dimensions on the rule that the first visit should be in the first half of the fieldwork period could also not be studied as a result of too little interviewer variance (the rule was observed in over 95% of cases).

Dependent variables were of two kinds: objective measures, calculated from the interviewer contact sheets, and one subjective measure, the interviewer's own estimate of the number of times she follows a specific rule.

Interviewer contact sheets are BLAISE questionnaires, to be filled in after each contact attempt, either face-to-face or by telephone. The contact forms inform us, among other things, of the day of the visit, the timeslot (in five time slots), the mode (face-to-face or telephone), who initiated the contact (interviewer or respondent), and what the result of the contact attempt was (no contact; contact with sample unit, but no appointment or end result; contact with appointment; contact with another person than sample unit; end result). In this way we were able to determine whether addresses were visited six times in case of noncontact, whether one of the first two visits was at prime time, whether visits were spread over timeslots (before 12h, from 12-17h, from 17-19h, from 19-21h and after 21h) and / or over days and whether the first visit was in the first half of the fieldwork period. We could also determine if deviation from a rule was initiated by the respondent, for example by making an appointment in a prior visit. In that case no transgression was calculated. Because there are only four productive time slots (calls should be made before nine o'clock at night), only the first four

(spontaneous) calls to an address were used for this analysis. Rule following was calculated for each interviewer on the work they did in 2010 and 2011 on the Labour Force Survey and the Health Survey. Interviewers with a workload of less than 50 addresses were not included in the analysis. A mean number of 283 addresses per interviewer (SD = 112) was analysed. The dependent variable for each rule was the percentage of cases in which it was followed.

Not all rule following could be observed in the field: the number of times interviewers use each calling card in the designated situation cannot objectively be gauged. Because of that, the interviewers' own estimate of the percentage of cases in which she follows the rule was used.

Dimensions, sub-dimensions and scale properties

For most dimensions scale properties could be calculated, see below. Generally, the values of Cronbach's alpha are moderately strong to strong. Concerning the lower reliabilities of some of the scales, note that most indices are based on a small number of questions, with sometimes skewed distributions. For short scales, the reliability is satisfactory (De Leeuw & Hox, 2009), with some exceptions that will be described. Table 7.2.1 shows (sub)dimensions, the number of items that measure the dimension, and the Cronbach's alpha. Some dimensions were not scalable, for example as a result of differing answering formats. This will be described below.

7.2.1 Cronbach's alpha and number of items for (sub)dimensions

	n items	Cronbach's α
Authority: coercion	4	0.78
Authority: obligation	5	0.66
Legitimacy: prime time	5	0.65
Legitimacy: calling cards	3	0.60
Legitimacy: spreading over time	3	0.67
Legitimacy: visiting 6 times	3	0.72
Interactional justice: attention	8	0.84
Interactional justice: knowledge	8	0.76
Distributive justice	3	0.14
Context: compatibility	5	0.58

Four items concerned the sub-dimension 'Coercion' of the dimension 'Traditional Authority'. The questions measured the extent to which interviewers feel that transgressions are spotted by monitoring, the extent to which they feared bad evaluations, to which they feel that Statistics Netherlands and their regional manager expects them to follow the fieldwork strategy, and to which they follow

fieldwork strategy because that is how they were taught to do it. See the appendix for the precise wording of the questions. A sum score was calculated from these questions, with higher scores indicating more perceived authority. This procedure to calculate the importance of dimensions was used for all dimensions, with some exceptions, noted below. Five questions measured the sub-dimension 'Obligation' Legitimacy of the rules, in the meaning of 'does the rule make sense' was measured by three questions for each rule, one determining general agreement with the rule on a five point scale from totally agree to totally disagree, and two measuring the extent to which the rule influences percentage of non-contacts and number of necessary visits. For the rule on prime time, separate questions were asked for evening visits and visits on Saturday, totalling the number of questions for this rule to five.

The sub-dimension 'Attention' within 'Interactional justice', was measured by eight questions concerning the attention by both Statistics Netherlands and regional manager to the contents of the fieldwork strategy, and the advantages and disadvantages for Statistics Netherlands and the interviewer. The effectiveness of the communication and transparency of the rules was measured by eight questions that gauged interviewers' knowledge about the rules.

Distributive justice is about the fair distribution of 'pleasure and pain'. Three question went into the evaluation of remuneration, of bonus distribution, and of distribution of difficult addresses over interviewers in a region. Cronbach's alpha of this scale was unacceptably low (0.14). On closer inspection, the three questions showed no correlation at all. Furthermore, preliminary analyses showed that evaluation of bonuses and distribution of difficult addresses did not have any relation with any of the dependent variables, while remuneration did. It was decided to do the remainder of the analyses with the one question only.

Two questions were formulated to tap into the dimension 'Social norms'. One determined the extent to which information about performance and behaviour of other interviewers influenced the interviewer, the second whether other interviewers call the interviewers to account. The dimension turned out to be not scalable, because in one of the two questions only two answer categories were used. Because of the theoretical importance of the dimension, it was decided to keep the dimension anyway.

The dimension 'Personal norms' was operationalized by one question only, agreement with the statement 'I follow fieldwork strategy because I want to do my work as best I can'. Inspection of the answer distribution made us suspect that interviewers did not interpret the question as intended, but that they in fact interpreted the question as meaning 'I always do my job as best I can' to which practically all interviewers agreed obviously. The weak operationalisation made further analyses not possible.

The contextual dimension of compatibility was operationalized as the influence of a rule on the private life of the interviewer. Five questions were asked, about evening visits, visits on Saturday, using the calling cards, spreading of visits and visiting six times.

The dimension 'Procedural justice' measures the extent to which interviewers were involved in developing the fieldwork strategy, and their evaluation of that extent, in two questions. Because of the differing answer formats of the questions, no scale properties could be calculated.

The sub-dimension 'Work pressure' was ascertained in 17 questions. The questions measured the discrepancy between the actual work status and the desired one, between the number of actual work hours and the subjectively felt number of hours, the amount of free time planned during a week, and the intrusions by respondents. Because of the very different answer formats of the various questions, among those answers to open questions, the dimension was not scalable, but analysis of the intercorrelations showed that there was sufficient ground to consider these questions as belonging to one dimension.

Routine was operationalized as the number of years the interviewer has worked for Statistics Netherlands, in five categories: less than one year, one to two years, two to five years, five to ten years, more than ten years. Because only eight interviewers appeared to be in the second category, the first and second were taken together. Dummy variables were created, with last category (more than 10 years of experience) as reference group.

12 Dummy variables were created to incorporate region(al manager) in the analysis. The region closest to the overall mean on the variable 'visits on prime time' was chosen as referent. The covariate 'Urbanicity' was incorporated as continuous variable, calculated from the mean urbanicity of the interviewers' addresses¹⁾.

Interactions

Because of the large number of independent variables, not all possible interactions were calculated. Rather, we concentrated on the interactions indicated in earlier research: procedural justice x interactional justice (Kickul et al., 2001; Thompson & Heron, 2005) and social norms x interactional justice (Cialdini & Goldstein, 2004). Because of the relevance these authors attribute to interactional justice, other interactions with this dimension were calculated as well. In addition, a number

¹⁾ This would be a meaningless statistic if interviewers work in both rural and extremely urban areas. In practice, the interviewers' addresses are fairly clustered.

of interactions deemed plausible in view of the subject matter were added, specifically interactions with routine, and the interaction between legitimacy of rules and authority.

The sum total of dimensions, covariates and proposed interactions amount to 32 variables. The number of responding interviewers with a workload of over 50 amounted to 219 however, a number not sufficient for regression analysis with 32 variables (Field, 2005, pp 172–174). To circumvent this problem we tried to limit the number of factors by univariately determining if factors existed that did not relate to any of the dependent variables, and to determine the necessity of including the interactions.

A series of standard multiple regressions was performed on the four dependent variables and the independent variables mentioned above. Analysis was performed using SPSS regression and SPSS explore for evaluation of assumptions. In view of the proposed interactions, the independent variables were centred and standardized (Aiken & West, 1991).

Results of the evaluation of assumptions led to transformations of the dependent variables to reduce skew and improve normality, linearity, and homoscedasticity of the residuals. A square root transformation was used on the measure 'visiting less than six times' and a logarithmic transformation on 'visiting on prime time'. The measures of spreading over time and use of the third calling card were skewed, but did not benefit from transformation. They were not transformed. Analysis of Cook's and Mahalanobis' distances showed no outliers among the cases.

7.3 Results

7.3.1 R statistics and evaluation of change in models

	Less than 6x			Prime time			Spreading over time			3rd calling card		
	R ²	Adj R ²	Sig. F Change	R ²	Adj R ²	Sig. F Change	R ²	Adj R ²	Sig. F Change	R ²	Adj R ²	Sig. F Change
Urbanicity	0.03	0.03	**	0.07	0.07	***	0.10	0.09	***	0.00	0.00	ns
Model 1 + dimensions	0.11	0.06	ns	0.23	0.19	***	0.21	0.16	**	0.16	0.11	***
Model 2 + interactions	0.15	0.06	ns	0.24	0.17	ns	0.26	0.19	○	0.20	0.13	ns
Model 3 + region	0.20	0.07	ns	0.31	0.20	○	0.34	0.23	*	0.26	0.13	ns

*** p < .001 ** p < .01 * p < .05 ○ p < .10.

For the dependent variable 'visiting less than six times' the final model was restricted to one factor only (knowledge of the rule), and the covariate urban density. Although the beta values showed one significant interaction, the increase in explained variance by adding interactions, see table 7.3.1, was so low that it was not included in the final model. Region was also not included in the final model for this variable. For the variable 'visiting on prime time' three factors were included in the final model (legitimacy of the rule, routine and authority-obligation), as well as urban density. No significant interactions were found. One region had a significant beta value, but in view of the absence of significant change in the model, region was not included in the final model. For the variable 'spreading over time', urban density and four factors were included in the final model: interactional justice, authority-obligation, routine and legitimacy of the rule. As with 'less than 6 visits' one interaction was significant (social norms x routine), but now, the change statistic was (marginally) significant and the interaction was incorporated, together with the lower level factors. Region as well was added, because of the marginally significant beta values for two of the dummy variables, and the significant change statistic. The model for the variable 'use of the 3rd calling card' consisted of the four factors interactional justice, both measures of authority, and legitimacy of the rule. One significant interaction and a significant beta value for one region failed to influence the amount of explained variance to a significant amount. Neither interactions nor region were incorporated in the final model for this variable.

7.3.2 Beta values for independent variables, interactions and covariates in four models

	Less than 6x		Prime time		Spreading		3rd calling card	
	β	p	β	p	β	p	β	p
Model 1								
(constant)		***		***		***		***
urban density	0.18	**	0.27	***	0.31	***	0.01	
Model 1 + dimensions								
interactional justice	0.03		0.06		0.13	*	0.11	○
distributive justice	0.06		0.01		0.09		0.10	
compatibility	0.05		0.04		0.08		-0.02	
authority - coercion	0.06		0.10		0.06		0.16	*
authority - obligation	0.06		0.21	**	0.16	*	0.15	*
routine 2	0.09		0.13	○	0.18	*	-0.09	
routine 3	-0.11		-0.03		0.08		-0.05	
routine 4	-0.06		0.02		0.05		-0.01	
social norms	0.01		0.02		0.05		0.02	
legitimacy rule	0.06		0.22	**	0.13	*	0.24	**
knowledge rule	0.16	*	0.10					

7.3.2 Beta values for independent variables, interactions and covariates in four models (end)

	Less than 6x		Prime time		Spreading		3rd calling card	
	β	p	β	p	β	p	β	p
Model 2 + interactions								
int justice * soc norms	0.04		0.07		0.02		-0.18	*
int justice * routine 2	0.16	*	0.03		0.05		0.01	
int justice * routine 3	-0.03		0.03		-0.03		0.06	
int justice * routine 4	-0.00		0.02		0.00		-0.11	
social norms * routine 2	0.07		-0.05		0.09		0.12	
social norms * routine 3	0.06		0.03		-0.11		0.03	
social norms * routine 4	0.10		-0.03		-0.19	*	0.03	
Model 3 + region								
region1	0.11		-0.07		-0.01		0.14	
region2	0.04		-0.05		0.03		0.13	
region3	0.07		0.10		-0.13		0.05	
region4	-0.05		-0.19	*	0.08		0.19	*
region5	0.13		0.01		-0.19	o	0.04	
region6	0.03		0.04		-0.09		0.08	
region7	0.12		0.04		0.01		0.01	
region8	-0.00		-0.10		0.02		0.11	
region9	0.02		0.01		-0.18	o	-0.07	
region11	-0.08		0.08		0.01		0.04	
region12	0.08		0.05		-0.08		0.03	
region13	-0.05		-0.05		0.03		0.03	

*** p < .001 ** p < .01 * p < .05 o p < .10.

The factors, interactions and covariates for the final reduced models were entered in one step. Table 7.3.3 shows R statistics for the final models, tables 7.3.4 to 7.3.7 show beta values.

7.3.3 R statistics for the final models

	R	R ²	Adj R ²
Less than 6 visits	.26	0.07	0.06
Visits on prime time	.45	0.20	0.18
Spreading over time	.56	0.32	0.23
Using 3rd calling card	.39	0.15	0.12

7.3.4 Beta values and test of significance for explanatory variables of 'visiting 6 times'

	β	t	p
(Constant)		29.56	***
Knowledge of rule	0.18	2.72	**
Urban density	0.16	2.45	*

*** p < .001 ** p < .01 * p < .05 o p < .10.

Table 7.3.4 shows that visiting less than six times in case of noncontact can, after controlling for urban density, only be understood as a function of the knowledge of interviewers of concept of noncontact. Interviewers who have an accurate understanding of the concept visit noncontacts the required number of times in 82% of cases. Interviewers who make the maximum number of mistakes in the vignettes do that in 57% of cases. Interviewers in urban regions follow the rule to a greater extent (in 76% of cases) than their rural colleagues (58%). The model is fairly weak, however. Only 7% of variability (6% adjusted) can be accounted for with these factors.

Table 7.3.5 shows the final model for the variable 'visiting at prime time'. Two factors play a significant role in the extent to which this rule is followed: the amount of perceived obligation and the evaluation of the legitimacy of the rule. Again, the covariate urban density is significant.

7.3.5 Beta values and test of significance for explanatory variables of 'visiting at prime time'

	β	t	p
(Constant)		23.12	***
Authority (obligation)	0.25	3.92	***
Routine < 2 years	0.12	1.83	o
Routine 2-5 years	-0.02	-0.36	
Routine 5-10 years	0.00	0.01	
Legitimacy of rule	0.24	3.83	***
Urban density	0.30	4.64	***

*** p < .001 ** p < .01 * p < .05 o p < .10.

To help interpretation, the values of these dimensions were collapsed into three categories: high, medium and low. Interviewers who feel to a large extent that they are obliged to follow the rule follow it in 79% of cases, while interviewers with a low level of perceived obligation follow it in 66% of cases. Interviewers who feel the rule makes sense follow it in 78% of cases, while interviewers who doubt

the legitimacy of the rule follow it in 66% of cases. Again, in highly urban regions, the rule is better adhered to than in rural areas (82% vs. 68%). This model explains 20% (18% adjusted) of variability in interviewer behaviour.

Table 7.3.6 shows the model for the variable 'spreading visits over time'. Explained variability is 32% (23% adjusted). Significant influences are found for routine, with recent interviewers performing less well than the most experienced ones (46% versus 54%), and perceived legitimacy of the rule (47% if not versus 54% if yes). The effect for routine is classified by a an interaction with social norms, indicating that social norms do not play a role in the amount this rule is followed for the most senior interviewers, but play a role for the junior interviewers: if they value the opinion of their peers, their performance is at the same standard as the senior interviewers (53%), but if they do not, their performance is less (38% rule following for the interviewers who claim to be not at all influenced by their peers, and 44% for interviewers with an intermediate evaluation of the importance of their peers). In contrast to the other variables, where the influence of the regional manager disappears if urban density is introduced in the model, the opposite is seen in this model: the influence of urban density disappears when the region is introduced. The influence of region is not strong though: only two regions are marginally different from the reference category. These two regions are also the two largest cities in the country. The result means that spreading over time is problematic in all urban areas, but specifically in the two largest cities.

7.3.6 Beta values and test of significance for explanatory variables of 'spreading over time'

	β	t	p
(Constant)		17.64	***
Interactional justice	0.09	1.41	
Authority - obligation	0.12	1.82	o
Routine < 2 years	-0.21	-3.05	**
Routine 2-5 years	-0.07	-1.12	
Routine 5-10 years	-0.04	-0.62	
Social norms	0.02	0.21	
Legitimacy rule	0.14	2.29	*
Urban density	0.13	1.42	
Social norms * routine <2	0.15	1.99	*
Social norms * routine 2-5	-0.09	-1.27	
Social norms * routine 5-10	-0.12	-1.67	o
Region 1	-0.00	-0.03	
Region 2	0.08	0.96	
Region 3	-0.12	-1.33	
Region 4	0.07	0.86	

7.3.6 Beta values and test of significance for explanatory variables of 'spreading over time' (end)

	β	t	p
Region 5	-0.19	-1.94	o
Region 6	-0.09	-1.02	
Region 7	0.02	0.25	
Region 8	0.01	0.09	
Region 9	-0.17	-1.91	o
Region 11	0.01	0.12	
Region 12	-0.07	-0.91	
Region 13	0.02	0.25	

*** p < .001 ** p < .01 * p < .05 o p < .10.

Table 7.3.7 shows results for the variable 'using the 3rd calling card'. 15% of variability (12% adjusted) is accounted for with the factors. In the final model, the most influential factor is perceived legitimacy of the rule, where interviewers who agree with the rule comply in 86% of cases, while interviewers who do not, comply in 64% of cases. Both measures of authority influence the extent to which interviewers adhere to this rule. Interviewers who perceive a low level of traditional authority adhere in 68% of cases, in 74% of cases for medium perceived authority, and 86% for high perceived authority. Although the correlation between both measures of authority is low ($R = .29$), the pattern is highly comparable for perceived obligation: 66%, 76% and 82% respectively.

7.3.7 Beta values and test of significance for explanatory variables of 'using 3rd calling card'

	β	t	p
(Constant)		27.61	***
Interactional justice	0.10	1.56	
Authority - coercion	0.17	2.52	*
Authority - obligation	0.16	2.25	*
Routine < 2 years	-0.09	-1.29	o
Routine 2-5 years	-0.05	-0.67	
Routine 5-10 years	0.01	0.16	
Legitimacy of rule	0.23	3.50	**

*** p < .001 ** p < .01 * p < .05 o p < .10.

7.4 Summary and discussion

In an effort to understand the reason why interviewers do not always follow proven sensible fieldwork rules, a questionnaire was sent to all Statistics Netherlands' field interviewers. Questions were designed to tap the dimensions literature suggests underlay (non)compliance with rules in different subject areas. By relating interviewers' rating on these dimensions to measures of rule compliance, it was determined if, and to what extent, any of the proposed dimensions was relevant to explain interviewer behaviour. Four rules were studied: visiting six times in case of noncontact, visiting on prime time in one of the first two visits, spreading visits over time, and using the third calling card after the third futile call. Compliance to the first three rules was studied using interviewers' call records of their workload on two major surveys. The fourth rule was measured by a subjective evaluation by the interviewers.

Results show that the variability between interviewers in the extent rules are followed is indeed related to some, but not all, of the dimensions described in the literature. After controlling for urban density, relevant for most measures, an additional 8% to 16% of variance was accounted for by the dimensions. The first conclusion to be drawn from this research is that different mechanisms underlay compliance to different rules. For the rule that noncontacts should be visited six times, only knowledge of the definition of noncontact could explain differences between interviewers. To a large extent it can be maintained that noncontacts that are not visited six times are probably not really noncontacts, but should have been classified in another nonresponse category.

A dimension that plays a role in each of the other three measures is the extent to which the interviewer feels a rule makes sense, the operationalisation of 'legitimacy of the rule'. If perceived legitimacy is low, the interviewer will be less likely to display the desired behaviour, although few interviewers are so bold as to never do it. In the absence of perceived legitimacy, applying coercion is no substitute, as witnessed by the absence of interactions between legitimacy and authority in all measures. Authority, in the sense of feeling obliged to follow a rule, does play a role in two of the measures: visiting on prime time and use of the third calling card. Although traditional authority does not generally play a direct role in the amount of transgression, it does play an indirect role, in that the feeling of obligation is influenced by the level of traditional authority ($\beta = .28, p < .001$). Feeling of obligation is also influenced by routine: the more routine, the less sense of obligation to follow rules. Compared to the interviewers with over ten years of experience, interviewers with less than two years experience ($\beta = .16, p < .05$) and interviewers with two to five years of experience ($\beta = .13, p = .05$) feel more obliged to follow the rules. Routine does play a direct role as well though.

It figures in three of the four final models, although its role is only significant in spreading visits over time. From the initial full models it seemed that interactional justice, the amount of attention given to a rule, played a role in two of the measures. Both failed to reach significance in the final models though.

Routine was introduced as a factor, with two possible outcomes in mind: interviewers with more routine have had more exposure to fieldwork rules because of the regular attention given to them in monthly progress meetings, and the twice yearly performance interviews. On the other hand, the most experienced interviewers have been set in their ways for a long time before the introduction of this fieldwork strategy and may be reluctant to give them up. Results showed that routine played a role in one rule only: spreading over time which the more experienced interviewers did better. Whether this reflects more compliance is doubtful, however. As mentioned above, spreading visits well is a complicated task, which less experienced interviewers may find too demanding. Instead of giving more attention to this rule or applying more coercion, interviewers would probably be better off with a tool to help them accomplish spreading their visits well.

Social norms, the influence of the fellow interviewers, were found to play a modest role, where especially less experienced interviewers indicated to be influenced by their colleagues to follow the rules. Social relations can however also create disincentives to comply (Tyler, 1990). If the experienced interviewers do not comply to the rules, this may be a powerful incentive for novices to do likewise.

A number of dimensions that literature indicated as important precursors to rule compliance did not play a role at all in explaining interviewer behaviour. Noteworthy is the lack of influence of procedural justice, the amount of involvement of the interviewers in developing the rules. Literature indicates that this is one of the most important determinants of compliance (DeCremer and Tyler, 2005; Raakjear Nielsen, 2003; Thompson and Heron, 2005; Tyler, 1990). Additional analyses showed that interviewers who had been involved in drawing the fieldwork strategy did not think the rules made better sense than their uninvolved colleagues. Presumably, however, in order to be satisfied with the outcome of involvement subjects should not only to be involved, but also to be heard. The extent to which this has been the case is not clear; the questionnaire only ascertained the amount of involvement, and satisfaction with the amount, not with the outcome.

Surprising was the lack of influence of context compatibility: the amount a rule influences private life. Although 73% of the interviewers indicated that evening visits had a negative influence on their private life, and 61% stated the same about visits on Saturday, their evaluation did not influence compliance to the rule concerning visits on prime time, although it may influence the total number of evening visits interviewers make. Concerns for private life could also be a reason not to present the third calling card, as this is the only one to contain the cell

phone number of the interviewer. Only a small number of interviewers rate the cards as intrusive, however (8%). In fact, most interviewers propose to introduce their telephone number earlier in the fieldwork.

Another circumstance that interviewers found important in qualitative research but was not found to be of influence in the quantitative analysis was workload. Although some interviewers indicated in the open questions that they stopped visiting noncontact addresses when the allocated number of work hours was depleted, the operationalisation chosen here, subjective work pressure, failed to show any influence. Objective work pressure is equal for all interviewers, as the number of addresses in the monthly workload of each interviewer is calculated as a function of the number of work hours, the distance the interviewer lives from the allocated addresses, the number of foreseen visits, and the predicted response in her work area.

Two factors were introduced as covariates in this analysis: urban density and regional manager. Urban density proved to be an important factor. Although it did not play a role in the calling card rule, it explained about a third of the variance in visiting on prime time and spreading over time, and almost half in six visits to noncontact. These three rules were followed to a larger extent in urban regions than less urban or rural regions. This again may reflect a focus of interviewers and regional managers on response rates, disregarding the costs of untimely visits. Regional managers indicate that they hesitate to admonish interviewers with high response rates. The lower amount of visits to noncontacts in rural areas on the other hand, may well be inspired by the time and effort visits to widespread addresses cost. After controlling for urban density, and after introducing the dimensions, the regional managers had hardly any additional influence. The differences between regional managers that interviewers indicate, for example in the amount of attention given to the fieldwork rules, or the severity with which she upholds compliance, are accounted for in the dimensions.

Understanding the reasons why interviewers do not comply to rules is the first step towards interventions to heighten compliance. Troubling in that respect is the low opinion that interviewers have of the working of some of those rules in attaining the desired high contact rate with the lowest possible amount of visits. The first obvious argument may be that they may in fact be right in their scepticism. This is, however not the case. As is shown in chapter 6, interviewers who follow these rules attain a higher contact rate against lower costs²⁾. One possible explanation for a sceptical attitude towards the fieldwork strategy is that interviewers are taught to think in terms of response rates, and they are correct in thinking that there is no straight link between these rules and response rate. Interviewers with good

²⁾ The effect of the calling cards cannot be ascertained.

doorstep skills may succeed in attaining relatively high response results even if their contact rate is relatively poor. In the same vein, interviewers may feel that certain addresses may be better approached during daytime, i.e., households with elderly respondents. However, even though contact propensity of these households during the daytime is relatively high, it is still less than their contact propensity during evening hours (Luiten et al., 2007). Contact attempts to elderly households should also be spread over time periods, and include evening visits. Nevertheless, as a result of this inventory Statistics Netherlands has adapted the fieldwork strategy and allows two daytime visits to households with elderly residents. Evening visits become again compulsory after two futile daytime visits to these households.

Although lack of perceived legitimacy is not overcome by applying coercion, authority does play a role in securing a higher level of compliance for two of the rules: visiting at prime time and using calling cards. It is important for the organisation that issues the rules to keep stressing that they value the rules, and to support that message by constant attention and monitoring (Cialdini & Goldstein, 2004). The absence of the effect in the rule about visits to noncontacts is understandable in view of the deviant definitions of noncontact that exist. The lesser importance in the rule about spreading over time may also be related to uncertainty about the definition of 'spreading', witnessed in the large discrepancy between the objective and subjective measures of spreading. In addition, it may well be that optimal spreading of visits over time is too demanding a task. In a pilot in which 12 interviewers committed themselves to fully adhere to the fieldwork strategy during one month (Luiten, 2008), they succeeded in reaching almost 100% compliance for all rules, except for spreading over time, where the prior level of around 50% continued to be found. Here too, however, the interviewers' estimation was that they performed well.

The analysis of the effect of calls and call schedules, and with that the possibility to ascertain whether fieldwork rules are adhered to, depends entirely on the honesty of interviewers when filling in the contact sheets. That is another reason why interviewer management should not depend on mere coercion to enforce: it could lead to falsification of contact sheets, which would be very hard to detect. The fact of finding relatively many transgressions is in that sense reassuring: as yet there is no reason to suspect interviewers of falsifying. The way forward is to better communicate the importance of contact rates and avoiding unnecessary costs, and the role of call schedules therein. This is possibly the first step towards securing a higher level of perceived legitimacy.

The present research is different from other research in a number of ways, which may account for the dimensions and interactions that literature indicates to be important, but that are not replicated here. The dependent variables reflect objectively measured interviewer behaviour, instead of lab results (e.g., Thibaut et

al., 1974), respondents' avowal of compliance (Hessing, Elffers & Weigel, 1988), or randomized response (e.g., Elffers et al., 2003). The number of factors in this research is large, but still only scrapes the surface of understanding interviewers' behaviour. In spite of the influence of some of the dimensions of rule compliance, and the additional influence of urbanicity, most variance is unaccounted for. Variance can also be introduced by the nature of the addresses. If interviewers prefer visiting elderly people or difficult neighbourhoods during daytime, this would influence compliance rates. The influence of the nature of the address on fieldwork strategy is explored in the next chapter. Even though interviewers' compliance cannot be entirely explained, this research led to insights that can be used to improve fieldwork compliance, and with that result in higher contact rates and lower fieldwork costs.

8.

**Do interviewers
introduce bias by
not following
fieldwork strategy?**

**The interplay between
interviewer behaviour,
sample unit characteristics and
non-response bias**

8.1 Introduction

The job of field interviewer is no easy one. The interviewer deals with reluctant respondents on a daily basis and to find these respondents, she is expected to roam the streets at ungodly hours, and perhaps visit neighbourhoods where her natural inclinations would not normally lead her. It should come as no surprise if she at times would be reluctant to do so. Couper and Groves (1992) suggest that the lower response rates in some metropolitan areas may have to do with reluctance on the part of the respondent to open the door to unknown persons, but also with the interviewer's reluctance to enter these areas.

Social environmental characteristics of the area, socio-demographic attributes of the respondent, at home patterns, accessibility of the respondent's dwelling and characteristics of the interviewer like her experience, gender, age and workload all influence contact and response rates (Blohm, Hox & Koch, 2007; Blom, De Leeuw & Hox, 2011; Durrant, Groves, Staetsky & Steele, 2009; Martin & Beerten, 1999).

Timing and number of calls to sample units interact with respondent characteristics and at home patterns and determine if and when contact is made with the sample unit. Literature indicates that there are significant differences between interviewers in contact rates that have to do with calling patterns (Barton, 1999; O'Muircheartaigh & Campanelli, 1999; Purdon, Campanelli & Sturgis, 1999; Stoop, 2005; Stoop, Billiet, Koch & Fitzgerald, 2010). As calling patterns can and should be influenced by characteristics of the sample unit, research tries to control for these characteristics, but continues to find differences in contact rates between interviewers (Blohm et al., 2007; Blom et al., 2011, Martin & Beerten, 1999, West & Olsen, 2010).

Differences between interviewers in the timing and number of calls could lead not only to differences in contact rate, but also to differences in the distribution of respondents included in the net sample (Purdon et al., 1999; Wang, Murphy, Baxter & Aldworth, 2005; West & Olson, 2010). In order to minimize the chance that interviewer preference for calling patterns unduly influences the chance that sample units will be contacted, most survey organisations will issue standard rules on how often and when face-to-face interviewers should contact respondents. All countries participating in the European Social Survey instruct their interviewers to *"...make a personal visit on a minimum of 4 occasions, at different times of the day and spread across the fieldwork period before you classify the address/ household/individual as unproductive. At least one of these personal visits should be in the evening ... and one at the weekend"* (ESS, 2010). These contact procedures guarantee that sample units have a uniform chance of being contacted, while at the same time optimising the chance of contact. For these reasons, Statistics Netherlands has likewise formulated a fieldwork strategy for field interviewers,

aimed at optimising the chance of response, while at the same time minimising costs of face-to-face fieldwork. Specifically, interviewers are supposed to visit all sample units at least once in the first half of the fieldwork period, to plan either the first or the second visit on 'prime time' (after 5pm, or on Saturday), to visit noncontacts six times, and to spread their visits over time and days within the fieldwork period. The first three contact attempts are to be face-to-face. After that, contact attempts by phone are allowed, but not required. Calling cards are left behind after the first three visits to alert the respondent to the interviewer's visit. This strategy was based on findings in survey literature about the relative contact rates of various contact times and call schedules (Barton, 1999; Campanelli, Sturgis & Purdon, 1997; Foster, 1997; Snijkers, Hox & De Leeuw, 1999). Social exchange theory (Dillman, 1978) suggested the calling cards as a means to express the sincerity of the survey request. Stressing face-to-face as the preferred contact mode was suggested by the higher response rates attained in this mode (Collins, Sykes, Wilson & Blackshaw, 1988; De Leeuw & Van der Zouwen, 1988; Nicolaas & Lynn, 2002). The number of six contact attempts was determined after analyses that showed that the additional contacts and responses were limited after this number, but extremely costly.

Analyses of fieldwork paradata show that this fieldwork strategy is successful in increasing chance of contact, while decreasing the number of visits necessary to attain an end result (Luiten, 2006, see also chapter 6). The importance of fieldwork strategy for results makes that interviewers are monitored weekly (De Vree, De Bie and Luiten, 2006), and adherence to strategy is an important aspect of interviewer evaluation. In spite of this, most aspects of fieldwork strategy are followed only partly. First visits in the first half of the fieldwork period are well adhered to, but all other aspects are only followed in about half the cases. Interviewers are either not willing, or not able to follow the strategy in the other half. Stoop et al., (2010) show that this may be common interviewer behaviour: interviewers in a number of countries participating in the European Social Survey do often not adhere to the prescribed fieldwork strategy. This results in sub-optimal contact rates. Kennickell (2003) also showed that rules of a prescribed fieldwork strategy were not followed to an important degree. Pickery and Loosveldt (2002) remark that standard contact rules usually demand serious efforts from the interviewers. While some interviewers are not motivated enough to execute the standard rules, other interviewers will do more than what is expected. Consequently, the chances of contacting respondents will be subject to interviewer effects.

In chapter 7 a large scale study among interviewers is described, meant to determine if dimensions related to rule compliance in general play a role in the extent to which interviewers comply with fieldwork rules. Dimensions like perceived authority of the manager, knowledge of the rule, and perceived legitimacy of the rule were shown to play a role, as well as the interviewer's

experience and the urbanicity of her usual work area. Workload, in terms of the number of hours working as interviewer, did not appear to be related to rule compliance, nor was the number of hours working besides interview work. In spite of the significant relations, it was found that most variance in rule transgression was still unaccounted for. Besides, it was found that although there were differences between interviewers in the amount of transgression, all interviewers transgressed to some extent. This means that in some cases they complied with the rules, while in others they did not. The suggestion put forward in chapter 7 is that within-interviewer variation in rule compliance may be related to characteristics of the sample unit.

The question is whether interviewer rule transgression leads to biased results. If there are no differences between respondents and nonrespondents, inefficiency only leads to higher fieldwork costs. But if allocation of effort is due to non-random factors which in turn may be related to substantive variables, this could lead to bias in survey estimates. A number of researchers have proposed this concern and have tried to ascertain if indeed interviewer effects due to differential fieldwork effort is related to bias. Groves and Couper (1998) found little evidence that interviewers were less likely to make evening calls, controlling for environmental characteristics such as urban status, population density, crime rates, percentage of minorities and dilapidated units in the neighbourhood. Wang, Murphy, Baxter and Aldworth (2005) found that there are differences between areas in the amount of evening visits, which result in initial differences between early and later respondents. After a small number of calls, however, the amount of bias converges to the final level. West and Olson (2010) find some evidence that interviewers tend to attract respondents similar to themselves. Kennickell (2003) found that interviewers applied less effort to sample units living in mobile homes or apartment buildings, living in buildings with a gate keeper, in neighbourhoods with many elderly people, with many people not proficient in English, and in neighbourhoods with high incomes. More effort was given to neighbourhoods with a high percentage of Hispanics. Differential effort may lead to demographic bias. Kennickell (2004) describes how, after efforts to tailor fieldwork effort to response propensity, interviewers applied their effort differently than before, favouring sample units in less affluent neighbourhoods who had received an incentive. Kennickell (2004) hypothesizes that this may have had to do with the expectation that these sample units would have a high chance of ending in response. These authors show that there is some evidence that interviewers systematically apply differential effort to different sample units, with the risk of introducing bias, but a clear picture does not emerge.

This research readdresses the issue whether interviewers who do not follow fieldwork strategy systematically disfavour classes of sample units, with a number of extensions to previous research: we were able to study bias directly,

by linking the sample with a registry that approximates the substantial variables of the Labour Force Survey, and second, we have a better understanding of the characteristics of the sample persons or households by using registry information and interviewer paradata.

Previous research focussed on characteristics of neighbourhoods for their analysis. In this study we were able to link addresses to registry information, which made available detailed information on a number of relevant household characteristics, specifically, household composition, age, gender, ethnic background of each household member, and availability of a fixed landline. Additional information was available on a fine gridded postal code level, specifically, income, the percentage of non-western non-native population, and the urban density of the area were thus determined. Interviewer paradata informed us of dwelling type and whether there was an intercom present.

Differential effort on the part of the interviewer may lead to an unbalanced composition of the sample, which may lead to biased results. Unbalanced response can be weighted with auxiliary variables to population distribution, but this cannot be done for substantive variables. One of the most important substantive variables of the Labour Force Survey is obviously the labour force status of household members, in three categories: working, not working but part of the labour force (e.g., unemployed, but looking for work), and not working and not part of the labour force (e.g., pensioners). In order to study whether interviewers introduce bias by not following fieldwork strategy, we would need to know the labour force status of non-respondents as well as respondents. A registry with this kind of information does not exist, however. An approximation could be found in a registry that lists the most important source of income for all members of the household (work, benefit, pension), or specifies that a household member is a student. We used this data to study the question if interviewers introduce bias by not following fieldwork strategy.

In this study we first analyse whether sample unit characteristics and neighbourhood characteristics known to be related to response rates are related to interviewer behaviour, specifically following fieldwork rules. In a second analysis we study whether interviewers differ in the way they are affected by sample unit and neighbourhood characteristics. Finally, we analyse if fieldwork behaviour is related to bias in substantive variables.

8.2 Method

Interviewer call schedules were calculated from the interviewer contact sheets. These inform us, among other things, of the day of the visit, the timeslot, the mode of contact, who initiated the contact (interviewer or respondent), and what the result of the contact attempt was. Interviewers are obliged to update the schedules after each workday and to upload the updates daily to inform management of their progress. The interviewers' 2008 workload on the Labour Force Survey and the Health Survey was used for this analysis. Interviewers with a workload of less than 50 addresses were not included in the analysis. A total number of 64,403 addresses was available for analysis, 46,121 LFS addresses and 18,282 Health Survey addresses. A mean number of 258 addresses per interviewer ($SD = 79$) was analysed. All but a few interviewers worked on both surveys. The LFS is an address sample, while the Health Survey is a person sample.

By studying the contact sheets, it was determined whether interviewers followed fieldwork strategy, i.e., whether one of the first two calls was later than 5 pm or in the weekend, whether six visits were made to households where contact was not yet established, and the extent to which visits were spread over time periods, days and weeks of the fieldwork period. In determining if fieldwork strategy was followed, only cold calls were analysed. Transgressions from fieldwork strategy as a result of appointments were not counted as such.

Auxiliary information about all households was obtained by linking addresses to communal registries available to Statistics Netherlands. In this way, the household composition, age, gender, and ethnic background of each household member was determined, as well as the presence of a fixed landline. Because household composition heavily influences contactability, even in a person sample, this information is collapsed into a profile of the household core¹⁾, which could contain 'mixed' categories of gender and ethnic background. Additional information was available on postal code area level, which form fine grids of on average 15 households each. Specifically, income, the percentage of non-western immigrants, and the urban density of the area were thus determined. The last source of information on the household was the interviewer, who determined the dwelling type and whether there was an intercom present. Table 8.2.1 summarizes the auxiliary variables available for each household.

296 interviewers worked on the two surveys during this period. The interviewers in this study are employed by Statistics Netherlands, and their remuneration depends on the number of hours they work, not on the response rate they achieve. There is

¹⁾ Head of household and partner.

therefore no incentive to work harder for cases with a high contact or cooperation propensity, as might be the case if interviewers are paid per response or per sample unit.

8.2.1 Auxiliary variables available from registries

Variable	Categories
Household level	
Ethnic Group	Native, non-western ethnic minorities, western ethnic minorities, mixed
Gender	Male, female, mixed
Average age of household core	Under 30;31-44;45-65; over 65
Type of Household	Single, partners without children, partners with children, single parents
Young child in household	Yes, no
Percentage of employed members of household core	0, 50, 100
Most important source of income	Employment, allowance (disabled, unemployment, social security, other), pension, student
Type of residence	Apartment; other
Intercom present	Yes, no
Registered landline	Yes, no
Postal code area level	
Degree of urbanization	Very strong, strong, moderate, low, rural
Percentage non-western non-natives	< 5%; 5-10%, 10-20%, 20-40%, >40%
Average monthly income	Quartiles

Of the interviewer we know the number of years of experience as a Statistics Netherlands interviewer, and her education in three categories (lower vocational, middle vocational, higher). Experience is shown to be related to contact and cooperation rates in a number of studies, although the direction of the relation is not the same in all these studies (Blohm et al., 2007; Purdon et al., 1999; Martin & Beerten, 1999; Wang et al., 2005). The classification of education was based on a survey amongst interviewers, performed in 2011. By that time, some 30% of interviewers were no longer working. Their educational attainment level is therefore not known. They were classified as 'education unknown'. This made visible whether interviewers who stopped working differed from their colleagues in their fieldwork behaviour

In the analysis whether household characteristics contribute to rule transgression in interviewers, the dependent variable is dichotomous (yes or no transgression), which indicates using multilevel logistic regression models with respondents nested within interviewers to control for differences between target households. Interviewer effects are modelled conditional on the target household's characteristics. Using multilevel modelling also incorporates the clustering in the sample caused by having respondents nested within interviewers (Hox, 2010).

Three fieldwork behaviours will be described: visiting after 5 pm in the first visit, visiting after 5 pm in either the first or the second visit and visiting six times in case of noncontact. Although visiting after 5 pm in the first visit is not actually prescribed in the fieldwork strategy, it is an interesting measure, as it tells us something about how interviewers behave without prior knowledge of the sample unit, although experienced interviewers may have a fair idea what to expect in certain neighbourhoods. 18% of addresses was visited in the evening at the first visit. 72% of addresses was visited in the evening at either the first or second visit, and 59% of non-contacted cases was visited the required six times. Most of this variance is within interviewers; with very few exceptions, all interviewers transgress some of the time.

8.3 Results

Which behaviour with which household

To analyse whether interviewers systematically differentiate between sample units in their fieldwork strategy, a number of multilevel logistic regression models was fitted in R (Bates, 2010), with household characteristics as first level fixed factors, and interviewer as second level random (intercept) factor. These analyses show how interviewers generally behave, the next section will study individual differences between interviewers in these findings.

The dependent variable in each analysis was whether or not a specific behaviour was shown. Four variables will be described: visiting after 5 pm in the first visit, visiting after 5 pm in either the first or the second visit, visiting six times in case of noncontact, and behaving according to fieldwork strategy in the first three visits (i.e., first visit in the first half of the fieldwork period, one of the first two visits at prime time, and spreading over time periods). First, an intercept only model was calculated that showed the amount of interviewer variation in the model. Next, factors relating to neighbourhood and household characteristics were introduced, and subsequently the factors relating to interviewer characteristics (experience and education). The full model, including dummy variables for the categorical factors is expressed as follows:

$$\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \beta_0 + \beta_1\text{ethnic}2_{ij} + \beta_2\text{ethnic}3_{ij} + \beta_3\text{ethnic}4_{ij} + \beta_4\text{gender}2_{ij} + \beta_5\text{gender}3_{ij} + \beta_6\text{age}2_{ij} + \beta_7\text{age}3_{ij} + \beta_8\text{age}4_{ij} + \beta_9\text{household}2_{ij} + \beta_{10}\text{household}3_{ij} + \beta_{11}\text{household}4_{ij} + \beta_{12}\%working2_{ij} + \beta_{13}\%working3_{ij} + \beta_{14}\text{residence}2_{ij} + \beta_{15}\text{intercom}2_{ij} + \beta_{16}\text{landline}2_{ij} + \beta_{17}\text{urban}2_{ij} + \beta_{18}\text{urban}3_{ij} + \beta_{19}\text{urban}4_{ij} + \beta_{20}\text{urban}5_{ij} + \beta_{21}\%nonwestern2_{ij} + \beta_{22}\%nonwestern3_{ij} + \beta_{23}\%nonwestern4_{ij} + \beta_{24}\text{income}2_{ij} + \beta_{25}\text{income}3_{ij} + \beta_{26}\text{income}4_{ij} + \mu_{0j}.$$

The first four columns in table 8.3.1 show how sample unit characteristics and interviewer differences influence the extent to which the first visit was done during evening hours. The first column describes the intercept only model, the second column describes the addition of household and neighbourhood characteristics, and the third column shows the full model with interviewer characteristics added. A likelihood ratio (LR) test for the difference between models showed that the difference between the intercept only model and the model with sample unit characteristics was highly significant with LR = 4812 on 29 d.f., $p < .0001$. The LR for the difference between the sample unit model and the interviewer characteristics model was 15 on 5 d.f., $p < .01$.

The variance partitioning coefficient (VPC) adapted for logit models (Steele, 2009, p.17), shows that in the intercept only model 34% of variance is interviewer variance. The VPC is expressed as $\sigma_{\mu}^2 / (\sigma_{\mu}^2 + \sigma_e^2)$, where σ_e^2 equals $\pi^2 / 3$ for a logit model. Adding neighbourhood and sample unit characteristics to the model, and in a second step interviewer characteristics, reduces the interviewer variance with a small amount only: 0,7% by adding sample units, and an additional 1,1% by introducing interviewer characteristics.

So, both sample unit characteristics and interviewer characteristics influence whether interviewers will make their first visit during evening hours. Table 8.3.1 shows that the higher the urban density, the more first attempts were during the evening. Independent of urban density, these first evening attempts were mostly done in more affluent areas. In general, more experienced interviewers visit less in the evening at the first attempt. Relations were also found with known landline (fewer first visits after 5 pm if the number is known) and with survey type (fewer with the person sample). The relations found with sample unit characteristics were all on the level of the neighbourhood; none of the household characteristics were found to be related to first visits after 5 pm, with the exception of the number of household core members working. As interviewers have no way of knowing this in advance, this finding may reflect the earlier finding that more first evening visits are made in more affluent neighbourhoods.

For the first visit, the interviewer is free to follow her own discretion as to the timing of it. For the second visit, however, a rule applies. If the first visit was performed during day time, as most visits are, the second one must be done either in the evening, or on Saturday. Were interviewers to follow this rule, we would not expect to find any relation with sample unit characteristics. The fourth, fifth and sixth column in table 8.3.1, however, show that relations do exist, although they are not strong.

Again, the interviewer variance in the intercept only model is 34%, according to the VPC. Adding sample unit characteristics reduces this variance with 0,4% only. Adding interviewer characteristics did not significantly alter the model, according to the likelihood ratio.

As in the first analysis, urbanity and neighbourhood income were related to visits at prime time, in the same direction. New in this analysis is the influence of age of the household core, indicating that interviewers transgress the rule more, the older the household core, and especially when the household core is elderly. Sample units living in an apartment building were visited more during evenings than sample units in other dwellings.

The third variable examined is whether or not an address is visited six times in case of noncontact. Interviewer variance in the constant model is 42%. Adding sample unit characteristics and interviewer characteristics reduces this variance by 0,6% and 3,4% respectively. LR tests showed that both models were significantly different from the intercept only model: LRs were 626 on 28 d.f., $p < .001$ and 338 on 6 d.f., $p < .001$, respectively.

Columns 7 to 9 in table 8.3.1 show that, more than in the prior analyses, sample unit characteristics play a role in the decision to continue visiting. Like before, we see that more rule transgression takes place in less urban and rural areas. Generally, more effort is given to harder cases, the ones in areas with many non-native inhabitants, living in apartment buildings, (single) males, and households where the entire household core works. An exception to this tendency is that more effort is given to households with a known landline. By far the largest amount of variance is explained however by type of survey: hardly any non-contacted sample units in the person sample were visited six times. Interviewers with more experience followed the rule more.

Individual differences

We have found that interviewers, other things being equal, generally follow rules more strictly with harder cases. In this section we investigate if this behaviour is shared by all interviewers. In order to allow the effect of household and neighbourhood on rule following to vary across interviewers, a series of random slope models was fitted in R (Bates, 2010). Because interpretation of random slope models with all household and neighbourhood characteristics would be extremely complicated, the vector of characteristics was condensed into propensity scores, one for contact propensity, and one for cooperation propensity. The propensity scores were calculated in logistic regressions with the auxiliary variables described in table 8.2.1. The propensity scores give an indication of the predicted ease of contacting a sample unit and their subsequent predicted cooperation. The model with a random slope coefficient for (contact or cooperation) propensity allowing its effect on rule following to vary across interviewers is thus simplified to

$$\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \beta_0 + \beta_1 \text{propensity}_{ij} + \mu_{0j} + \mu_{1j} \text{propensity}_{ij}$$

8.3.1 Relation between fieldwork behaviour, sample unit characteristics and interviewer characteristics

	First visit in evening		
	Intercept-only (s.e.)	Sample unit characteristics (s.e.)	Interviewer characteristics (s.e.)
Intercept	-1.95 (0.08) ***	-1.87 (0.11) ***	-1.61 (0.61) **
Survey (LFS = reference)			
health Survey		-0.07 (0.03) *	-0.07 (0.03) *
Dwelling (other = reference)			
flat / appartement building		0.05 (0.05)	0.05 (0.05)
Intercom present (no = reference)			
yes		-0.01 (0.05)	-0.01 (0.05)
Urbanicity (Highly urban is reference)			
2		-0.03 (0.04)	-0.02 (0.04)
3		-0.10 (0.05) *	-0.10 (0.05) *
4		-0.17 (0.05) ***	-0.16 (0.05) ***
rural		-0.38 (0.05) ***	-0.38 (0.05) ***
Percentage of non-western non-native inhabitants (0-5% is reference)			
5-10%		0.02 (0.04)	0.02 (0.04)
10-20%		0.01 (0.04)	0.01 (0.04)
20-40		0.06 (0.05)	0.06 (0.04)
>40%		0.04 (0.06)	0.04 (0.06)
Mean income of neighbourhood (lowest quartile is reference)			
2nd quartile		0.06 (0.03)	0.06 (0.03)
3rd quartile		0.05 (0.04)	0.05 (0.04)
highest quartile		0.11 (0.04) **	0.11 (0.04) **
Ethnic background (native Dutch = reference)			
mixed		-0.04 (0.04)	-0.04 (0.04)
non-western non-native		0.01 (0.05)	0.01 (0.05)
western non-native		-0.01 (0.05)	-0.01 (0.05)
Gender (male(s) = reference)			
mixed		0.04 (0.06)	0.04 (0.06)
female(s)		-0.01 (0.04)	-0.01 (0.04)
Age (<30 = reference)			
30-44		-0.01 (0.05)	-0.01 (0.04)
45-64		-0.02 (0.04)	-0.02 (0.06)
>=65		-0.06 (0.06)	-0.06 (0.04)
Young child in household (no = reference)			
yes		-0.00 (0.04)	-0.00 (0.04)
Landline number known (no = reference)			
yes		-0.09 (0.03) **	-0.09 (0.03) **
Household composition (single = reference)			
partners, no children		-0.03 (0.06)	-0.03 (0.06)
partners, with children		-0.03 (0.07)	-0.03 (0.07)
single parent		-0.02 (0.05)	-0.02 (0.05)
Percentage of household core members with income from labour (0 = reference)			
50%		0.04 (0.04)	0.04 (0.04)
100%		0.07 (0.04) *	0.07 (0.03) *

First or second visit in evening

Six visits to noncontacts

Intercept-only (s.e.)	Sample unit characteristics (s.e.)	Interviewer characteristics (s.e.)	Intercept-only (s.e.)	Sample unit characteristics (s.e.)	Interviewer characteristics (s.e.)
1.25 (0.08) ***	1.44 (0.12) ***	1.79 (0.63) **	0.64 (0.10)	0.55 (0.26)*	0.78 (0.85)
	0.05 (0.04)	0.05 (0.04)		-6.53 (1.18) ***	-6.50 (1.17) ***
	0.22 (0.06) ***	0.22 (0.06) ***		0.34 (0.16) **	0.34 (0.16) *
	-0.09 (0.06)	-0.09 (0.06)		0.09 (0.13)	0.09 (0.13)
	-0.03 (0.05)	-0.03 (0.05)		-0.33 (0.14) *	-0.34 (0.14) *
	-0.14 (0.06) *	-0.14 (0.06) *		-0.53 (0.17) **	-0.55 (0.17) **
	-0.14 (0.06) *	-0.14 (0.06) *		-0.55 (0.20) **	-0.55 (0.20) **
	-0.32 (0.07) ***	-0.32 (0.07) ***		-0.58 (0.23) *	-0.62 (0.23) **
	0.03 (0.05)	0.02 (0.05)		0.33 (0.16) *	0.33 (0.16) *
	0.01 (0.05)	0.01 (0.05)		0.20 (0.15)	0.21 (0.15)
	0.08 (0.06)	0.08 (0.06)		0.25 (0.15)	0.27 (0.15) ○
	0.10 (0.07)	0.10 (0.08)		0.49 (0.19) **	0.51 (0.15) **
	0.09 (0.04) *	0.09 (0.04) *		0.06 (0.12)	0.06 (0.12)
	0.12 (0.05) **	0.12 (0.05) **		0.20 (0.14)	0.20 (0.14)
	0.13 (0.05) **	0.12 (0.05) **		0.14 (0.16)	0.16 (0.16)
	-0.06 (0.05)	-0.06 (0.05)		-0.04 (0.18)	-0.04 (0.18)
	-0.02 (0.06)	-0.02 (0.06)		0.12 (0.14)	0.12 (0.14)
	0.03 (0.07)	0.03 (0.07)		0.09 (0.16)	0.09 (0.16)
	-0.10 (0.07)	-0.10 (0.07)		-0.24 (0.21)	-0.23 (0.21)
	-0.05 (0.05)	-0.05 (0.05)		-0.30 (0.12) **	-0.30 (0.12) **
	-0.03 (0.05)	-0.02 (0.05)		-0.11 (0.13)	-0.12 (0.13)
	-0.14 (0.05) **	-0.15 (0.05) **		-0.14 (0.14)	-0.14 (0.14)
	-0.40 (0.08) ***	-0.40 (0.08) ***		-0.22 (0.25)	-0.22 (0.25)
	-0.05 (0.05)	-0.05 (0.05)		0.13 (0.19)	0.14 (0.19)
	-0.08 (0.03) *	-0.08 (0.03) *		0.62 (0.11) ***	0.61 (0.11) ***
	-0.02 (0.08)	-0.02 (0.08)		-0.52 (0.21) *	-0.54 (0.21) *
	-0.09 (0.08)	-0.09 (0.08)		-0.79 (0.26) **	-0.80 (0.26) **
	-0.04 (0.06)	-0.04 (0.06)		-0.18 (0.19)	-0.20 (0.19)
	0.04 (0.06)	0.04 (0.06)		0.19 (0.19)	0.20 (0.19)
	0.01 (0.04)	0.01 (0.04)		0.27 (0.12)*	0.27 (0.12)*

8.3.1 Relation between fieldwork behaviour, sample unit characteristics and interviewer characteristics (end)

	First visit in evening		
	Intercept-only (s.e.)	Sample unit characteristics (s.e.)	Interviewer characteristics (s.e.)
Interviewer experience (< 2 years = reference)			
2-7 years			-0.48 (0.26) ◊
> 7 years			-0.65 (0.17) ***
Interviewer education (lowest = reference)			
middle			0.28 (0.60)
highest			0.13 (0.61)
unknown (interviewer no longer in service)			-0.00 (0.61)
σ^2 interviewer (SD)	1.71 (1.31)	1.66 (1.29)	1.58 (1.26)
Deviance	51,270	46,459	46,433

*** p < 0.001 ** p < 0.01 * p < 0.05 ◊ p < 0.1.

where i represents the respondent, and j the interviewer. Dependent variable in each analysis is again whether or not a desired behaviour was shown. In this section we concentrate on visiting on prime time (after 5 pm or on Saturday) in one of the first two visits, and visiting noncontacts the required number of times. Table 8.3.2 shows the results of the random slopes analyses with contact propensity.

8.3.2 Random slopes model of the influence of contact propensity on rule following

	Visiting on prime time	Visiting noncontacts
Fixed effects		
intercept	3.02 (0.25) ***	8.26 (0.57) ***
contact propensity	-1.90 (0.25) ***	-8.71 (0.61) ***
Random effect variances		
interviewer level	6.47 (2.54)	24.5 (4.95)
sample unit level	4.72 (2.17)	21.3 (4.63)
VPC interviewer	57.8	51.3
VPC sample unit	42.2	48.7

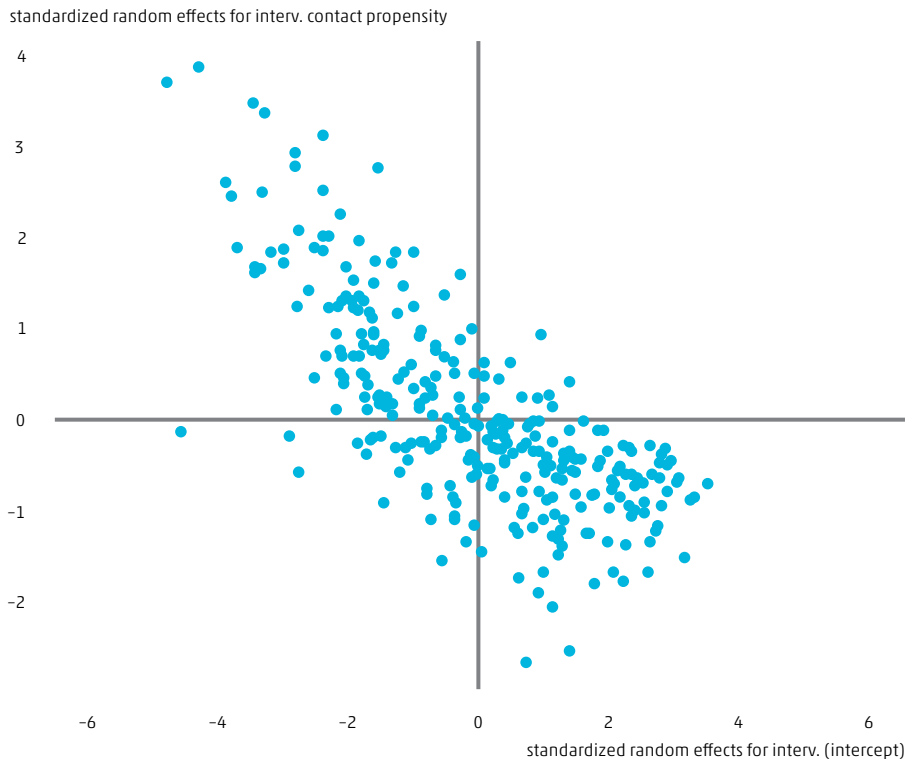
The first model fits interviewer differences in visits on prime time given contact propensity. Test of the random slope model against a random intercept model indicated that allowing a random slope had a significant influence on the model.

First or second visit in evening			Six visits to noncontacts		
Intercept-only (s.e.)	Sample unit characteristics (s.e.)	Interviewer characteristics (s.e.)	Intercept-only (s.e.)	Sample unit characteristics (s.e.)	Interviewer characteristics (s.e.)
		0.18 (0.27)			0.52 (0.34)
		-0.08 (0.17)			0.72 (0.23) **
		-0.21 (0.62)			-0.40 (0.81)
		-0.26 (0.63)			-0.46 (0.82)
		-0.51 (0.62)			-1.08 (0.81)
1.71 (1.31)	1.68 (1.30)	1.65 (1.29)	2.34 (1.53)	2.28 (1.51)	1.98 (1.41)
30,634	27,903	27,898	4,357	3,418	3,393

Figure 8.3.3 shows the standardized random effects of intercepts and slopes for each interviewer. The intercepts represent differences between interviewers in the percentage of cases in which visits are made on primetime in either of the first two visits. The slopes represent differences between interviewers in the extent to which contact propensity influences that percentage. As figure 8.3.3 shows, four groups can be distinguished: the largest group, consisting of almost half of the interviewers, has a positive intercept, combined with a negative slope. In this group, interviewers adhere more than average to the prime time rule, and they do so especially with addresses with characteristics that give them a low contact propensity. The second largest group, consisting of 109 interviewers, combines a negative intercept with a positive slope. This group adheres less than average to the rule, and on top of that, they tend to adhere even less with harder cases. Especially the interviewers with the steepest slopes in this group have to be monitored extensively, as they tend to shirk from difficulties.

The lower left side of figure 8.3.3 represents interviewers with negative intercept and negative slope. They adhere less than average to the rule, but they put in more effort for difficult addresses. The last group, which consists of only 13 interviewers, adheres very well to the rule. The slopes for these interviewers are all close to zero, which means that they do not significantly differentiate between addresses with high and lower contact propensity.

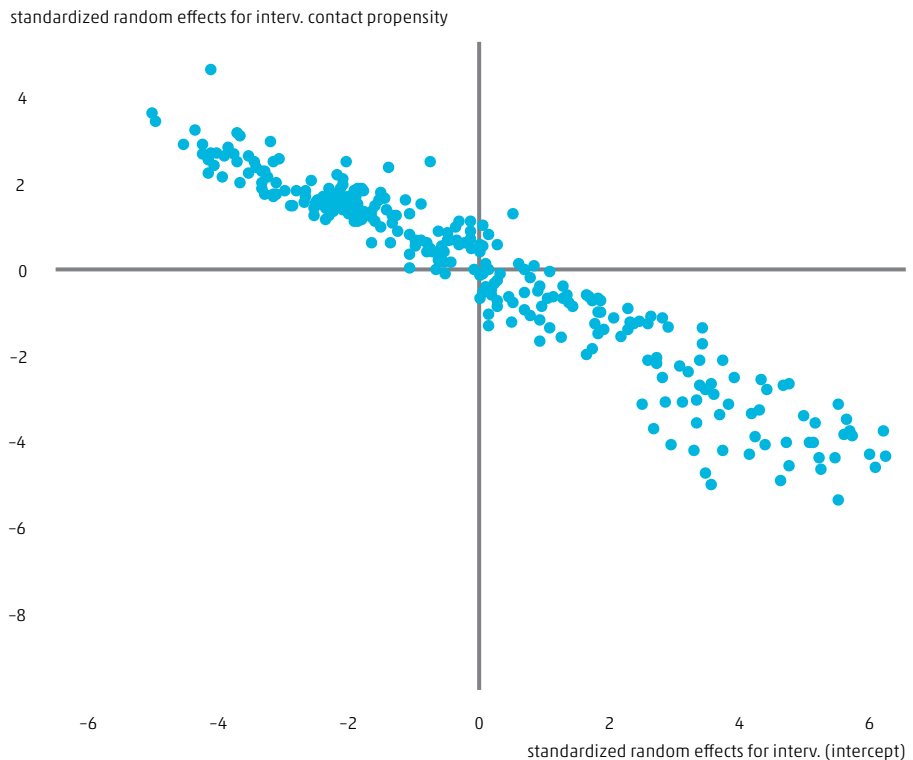
8.3.3 Interviewer slopes versus interviewer intercepts for the influence of contact propensity on visiting on prime time



Of course, interviewers may let their fieldwork behaviour not only be influenced by perceived contact propensity, but also by perceived cooperation propensity. Analysis showed this to be not the case for the rule concerning visits on prime time, however. There was no correlation at all between following the rule and cooperation propensity, neither with cooperation conditional on contact (COOP1, AAPOR, 2008), nor with cooperation conditional on contact and being able to cooperate (COOP3). The random slope models with either cooperation measure were not significantly different from the random intercept models. The second column in table 8.3.2 describes individual differences in the relation between contact propensity and visiting noncontacts the required number of times. The random slopes model indicates that there are differences between interviewers in the amount to which their fieldwork behaviour is influenced by contact propensity of their addresses. Figure 8.3.4 shows the standardized random effects of intercepts and slopes. It can be seen that, with a few exceptions, most interviewers belong to one of two groups: one with negative intercepts and positive slopes, and one with positive intercepts and negative slopes. The

extremes of these groups exist on the one hand of interviewers who seldom visit their noncontacts the required number of times, but when they do, they do it with households with a high contact propensity. They inhabit the upper left side of figure 8.3.3. The lower right side exists of interviewers who more than averagely visit their noncontacts six times, and when they do not, they err with households with a very high contact propensity. Their experience may have taught them that if these kinds of households are not contacted within the first couple of visits, chances are slim that they will be contacted at all. We will come back to this in the discussion.

8.3.4 Interviewer slopes versus interviewer intercepts for the influence of contact propensity on visiting noncontacts six times



Again, a second model was fitted with both measures of cooperation propensity. Like before, cooperation propensity did not relate to interviewer behaviour, and random slope models did not add to the random intercept model that merely

indicated that there are differences between interviewers in the amount of adherence to this fieldwork rule.

Fieldwork strategy and substantive variables

One of the most important questions we set out to answer in this research was whether interviewers introduce bias in the measurement of substantive variables by the way they perform the fieldwork. In the analyses above we approached this issue indirectly, by studying auxiliary variables. As is explained in the introduction, a registry of labour force status does not exist, but we do know whether household core members had income from labour at the time of the survey. We used this information as a proxy for labour force status, and with that performed a more direct analysis of the impact of interviewer behaviour on substantive variables for the LFS data only. The dependent variable in these analyses is the aggregated mean discrepancy between the percentage of working people in each interviewer's workload, and the percentage found in their response. In the sample, the mean percentage of people receiving income from labour in the household core was 63.6%²⁾, the mean percentage in the response was 64.1%. The unweighted aggravated discrepancy between sample and response over all interviewers was 0.5 percentage points, with a range from -17.2 to 10.6, and a standard deviation of 2.9 percentage points. For 79 percent of interviewers the number of working people in their response did not diverge more than one standard deviation from the expected number. 11% of interviewers had a negative discrepancy of over one standard deviation, of which 2% of over two standard deviations. The negative discrepancy means that these interviewers found significantly less working persons than they should have. On the other end of the spectrum, 9.5% of interviewers found less non-working persons than they should have, of which 1.4% with a difference of over two standard deviations.

A (stepwise) multivariate regression analysis related the amount and direction of the discrepancy of the percentage of expected and realized working people to interviewers' fieldwork behaviour. Interviewers with a contact or cooperation rate that was higher or lower than expected on the basis of the contact or cooperation propensities of their sample units also had discrepancies in the number of working people they found ($\beta = .14$, $t = 2.28$, $p < .05$ and $\beta = .19$, $t = 3.31$, $p < .01$ for contact and cooperation respectively). A higher than expected contact or cooperation rate was related to a positive discrepancy, i.e., finding relatively too many working persons, and too few non-working persons, while low contact and cooperation

²⁾ Pensioners were counted as having no labour.

rates were related with finding too few working persons. Interviewers with a higher than expected contact rate were much more inclined to visit (interim) noncontacts the required number of times ($\beta = .42$, $t = 8.30$, $p < .0$), and visit more often during the required evenings ($\beta = .11$, $t = 2.27$, $p < .05$). On the whole, these interviewers also used a somewhat lower than expected number of visits to attain this result ($\beta = -.22$, $t = -2.34$, $p < .05$). It should be stressed that these interviewers who do their utmost to attain the highest possible contact and cooperation rates may introduce as much bias by an overrepresentation of working people, as their colleagues who do not try hard enough and end up with an underrepresentation.

8.4 Summary and discussion

Fieldwork rules for interviewers are formulated with, among other things, the goal of giving each sample person or household an equal chance of being contacted and thus to respond to the survey query. In this research we set out to examine if interviewers who do not follow fieldwork rules bias substantive results. The first question we examined was whether there is a relation between the effort interviewers put in, and characteristics of sample persons or households and their neighbourhoods. If interviewers favour 'easier' cases, for example by only visiting more affluent neighbourhoods during evening hours, this would potentially bias results.

It was found that especially characteristics of the neighbourhood are related to visiting during evening hours. In their first attempt, when visiting during evening hours is not prescribed, interviewers visit households in more affluent neighbourhoods and in urban regions more often during the evening. These households are also those where the entire household core works. Sample units in the health survey and sample units with a known telephone number are visited somewhat less during the evening in the first attempt. In the second attempt, when evening visits are obligatory, there is still a strong influence of urbanicity; the influence of affluence of the neighbourhood is less strong but still existent. In the least affluent neighbourhoods, less evening visits are performed. The only sample unit characteristic that influences evening visits is the age of the household: the older the household core, the more often the interviewer does not follow the fieldwork rule.

Sample unit characteristics as well as neighbourhood characteristics play a role in the number of visits interviewers pay to non-contacted sample units. Generally, more visits are made to households with an elevated chance of not being

contacted: singles, males, households where the entire core works, in urban neighbourhoods with a high percentage of non-western non-Dutch population, and living in apartment buildings. The amount of spreading over time did not appear to be related to characteristics of the household or their neighbourhood. These findings show that, independent of the area in which an interviewer works, she will generally put in more effort for cases that are potentially harder to reach. That means that the interviewer practises a form of adaptive or tailored design (Groves & Heeringa, 2006; Wagner, 2008; Luiten & Schouten, 2013), in which more effort is given to the harder cases.

Although it was found that characteristics of neighbourhood and sample units play a role in the appliance of fieldwork rules, most variance was caused by the interviewer. Interviewer experience was found to be related to some extent with two of the fieldwork behaviours studied: the timing of the first visit and the number of visits to noncontacts. Interviewer education and the continued employment as interviewer two years later were not related at all. That means that most interviewer variance is unaccounted for. Interviewer gender, found to play a role in other research (e.g., Blom et al., 2011), could not be incorporated in the analysis, as almost all Statistics Netherlands interviewers are female. Other interviewer characteristics, like personal norms, the perceived legitimacy of the rules and personal circumstances play a role as well, as is shown in chapter 7. Differences between interviewers in the amount of evening work or the number of visits they make to non-contacted households could in principle also be related to the interviewer workload. Interviewers with a large number of working hours per week, and subsequently a relatively large number of addresses will have a smaller proportion of their time available for evening work. Most SN interviewers work 20 hours a week, the maximum number of hours is 24. The amount of addresses the interviewer can handle within that period is calculated precisely, based on the envisaged number of visits, response propensity of the interviewer sample, and the time needed to travel (Van Berkel & Vosmer, 2006). Workload thus does not play a role in the amount of rule following in Statistics Netherlands interviewers, as is shown in chapter 7, but may well play a role in other circumstances.

The second research question was whether the finding that in general more effort is given to harder cases, is true for all interviewers. It was found that this was the case for a bit more than half of the interviewers, both for visiting at prime time, and visiting noncontacts the prescribed number of times. There is on the other end, however, a substantial group of interviewers who put in less effort to the hardest cases. These are interviewers who should be monitored closely, as they may potentially have a detrimental effect on data quality. A minority of interviewers did not differentiate at all between sample units in their behaviour. One group of interviewers almost always visited their noncontacts the required number of times, but when they did not, they transgressed with households with a high contact

propensity. It may be that their experience taught them that if these addresses are not contacted within a few visits, they will most probably not be contacted within the fieldwork period. It may on the other hand also be that these addresses should not have been classified as noncontact. Chapter 7 shows that an erroneous interpretation by interviewers of the concept 'noncontact' underlies a large number of transgressions of the rule to visit the required number of times.

Interviewer behaviour was found to be influenced by contact propensity only, not by cooperation propensity. This finding is consistent with other findings that cooperation is hardly related to auxiliary socio-demographic variables like age, income and ethnicity (Luiten & Cobben, 2010; Luiten & Schouten, 2013; Hox, De Leeuw & Kreft, 1991; Hox & De Leeuw, 2002; Durrant & Steele, 2009); see also chapter 5. Of course, once contact is made, interviewers will have a better understanding of the cooperation propensity of the sample person. Interviewer behaviour that is induced by appointments and other contacts between sample person and interviewer are however not incorporated in this analysis.

Finally, we analysed whether interviewers introduced bias in substantial variables by their fieldwork behaviour, by contrasting the number of working household members in their sample with the number in a registry. It was found that on average there was a small difference between the number of working people in the response and in the registry. There were however interviewers with large discrepancies, in both directions: interviewers with relatively too few working people, and interviewers with too few non-workers. The discrepancy between response and sample was related to higher or lower contact and cooperation rates than could be expected on the basis of the characteristics of the sample units, and these in turn by a higher or lower than average number of evening visits and visits to non-contacted sample units.

Even though the average bias in this substantive variable is limited, that should be no reason not to strive for bias reduction in each and every interviewer. That means that some interviewers must be encouraged to work more during evening hours, and to put in the required number of visits to non-contacted sample units. Other interviewers must be encouraged to work more during the day, as this is the better time to find the elderly, not working sample persons. Failing to balance the number of evening and day time calls may lead to a bias in substantive results of the Labour Force Survey that cannot be adapted through weighting (Luiten, 2012). These analyses suggest that the extent to which interviewers introduce bias by the way they conduct the fieldwork, seems to be limited. If anything, most interviewers use a differentiated fieldwork strategy that may actually play a role in securing a more representative response, by giving more effort to the harder addresses, and less to the easier ones. The extent to which these findings can be generalised to other fieldwork organisations, with other interviewer contracts and other ways of interviewer remuneration, remains an issue for further research.

9.

Summary and

discussion

The chapters in this thesis treat widely different subjects. There are a number of common threads however: all have to do with nonresponse, its effect on the quality of results, and, apart from the first introductory chapter on measures to enhance response, all illustrate how paradata can be used to optimize survey design and survey management. Chapter 3 describes an experiment in which auxiliary data were used to address advance letters to sampled addresses in the Labour Force by the names of the household members, instead of to the usual 'to the inhabitants of...'. Information on the gender of the inhabitants was used to formulate the correct salutation on the letter. Research findings on the issue of personalisation of advance letters are mixed in the literature and one of the hypotheses in this chapter was that this could be the result of differential reactions of subgroups within the population. One such differential reaction was indeed found: households who did not have a listed telephone number reacted negatively to the naming of the letter, while others reacted positively. The net result was that no overall effect of personalisation was found in this experiment. The group who reacted negatively to the naming of the letter is traditionally a group that is underrepresented in survey response. The experimental manipulation may thus have introduced bias by widening the response gap between the overrepresented and underrepresented group. To make use of the stimulating effect of personalisation without exacerbating response bias, it would make sense to apply personalisation in some cases, but not in others. The unlisted households are most likely jealous of their privacy, as was witnessed also in the increase in refusals because of privacy concerns. Survey organisations could cater for the concern in this group by especially stressing the privacy measures that are taken. Sending different letters to different groups of people is an example of an adaptive survey design, also called dynamic designs (Wagner, 2008), responsive designs (Groves & Heeringa, 2006) or tailored designs (Luiten & Schouten, 2013). Adaptive survey designs are the topic of chapter 4. This chapter describes an experiment with the Survey of Consumer Sentiment (SCS), aimed at attaining a more balanced (representative) response composition by diminishing the gap in response rates between subgroups in the population. That means that some groups from which we expected a relatively low response rate received extra attention, while other groups with an expected high response rate received less attention than usual. A constraint in the development of the design was that the response rate should be at the same level or higher than the parallel traditional SCS, and fieldwork should not be more costly than the SCS. Differential treatments concerned the initial mode in which sample units were asked to cooperate (either web or mail or a choice between these alternatives), the timing and spacing of subsequent telephonic contact attempts to nonrespondents of the first wave, and the interviewer we assigned to an address.

In responsive designs (e.g., Groves & Heeringa, 2006) fieldwork runs its normal course in the first part of the fieldwork period. Then a short breathing space is introduced in which the provisional results are evaluated and decisions are made on the treatment of the remainder of the sample. In the case of ongoing surveys, like the SCS, a host of information about contact and cooperation propensity of groups sample units is known from previous rounds. That information can be used to differentiate treatment right from the beginning of the fieldwork. This is the approach we chose in this experiment.

The contact and cooperation propensity of sample persons were gauged from similar sample units in previous round, and were shown to be very predictive of the contact and cooperation propensity of the new sample persons. Persons with a low cooperation propensity received a mail questionnaire; nonrespondents were subsequently called by the best interviewers. Persons with a high cooperation propensity received an invitation to fill in a web questionnaire, and were called by interviewers with less persuasion proficiency. Persons with a low contact propensity were called every day of the fieldwork period, at least once in every interviewer shift; persons with a high contact propensity were called more during daytime hours, and were started later in the fieldwork period. Intermittent regimes were designed for sample persons with medium high and medium low contact and cooperation propensities.

The experiment was successful in reaching the goal of higher representativeness within the constraints of equal response rate and costs. Higher representativeness was attained by the manipulation of mode and the manipulation of contact. Manipulation of cooperation proved far less successful, however. The less proficient interviewers were very good in lowering the response rates of the people with the highest cooperation propensity, but the best interviewers were not able to heighten the response rates of those with the lowest propensity. Subsequent analyses showed that this finding had to do with the definition of cooperation used in this study; if the definition was restricted to persons who were both contacted and able to cooperate, it appeared all but impossible to predict cooperation from the socio-demographic variables we had available as auxiliary variables.

This finding has implications for adaptive survey designs as employed in this study; we can easily differentiate the call schedules for persons in telephone surveys according to contact propensity and with some more effort also for face-to-face surveys. Especially in studies where contactability is related to substantive variables, like the Labour Force Survey, Travel Surveys and Time Use Surveys, it is important to strive for contact representativeness. However, with the exception of some inner cities, noncontact is the minor nonresponse problem, with typically a few percentage points of noncontact. Refusals, on the other hand, are far more prevalent.

The question is whether we can maintain that if we cannot predict refusals, which is equivalent to saying that we have attained cooperation representativeness, we can rest easy in the confidence that there is no bias as a result of refusal. The answer to that question depends on the mechanisms underlying nonresponse bias, which may be different for each survey. Groves (2006) describes a number of these models. In a 'separate causes model' variation in response propensity is unrelated to variation in substantive variables. In that case the level of response nor its representativeness is an issue as bias is not caused by nonresponse. In a 'common cause' model, response and substantive variables are both influenced by a third variable, and thus co-vary. If this variable would be 'topic interest' we are not much helped by representative response, unless topic interest co-varies with the variables included in the vector of auxiliary variables used to calculate representativeness. In the case of the Survey of Consumer Sentiment, it is plausible that the common cause is captured by one or more of the auxiliary variables, e.g., income and variables that co-vary with income. If that is the case, more representative response is indeed an indication that the resulting data are less biased. Groves' third model is the 'survey variable cause' model. In this model, the variable of interest causes the response propensity. Groves gives the example of time-use surveys, where respondents' time away from home may be both a variable of interest in the survey, and the reason they are hard to contact. Another example may be the Labour Force Survey, when people refuse because they are not in the labour force, or are not contacted because they are at work. In this model it may be especially worthwhile to strive for representativeness in the category most likely causing the bias: in the case of time-use surveys one should strive for contact representativeness, in the labour force survey both contact representativeness and cooperation representativeness should be strived for. Although again, whether response is truly representative as to the variable causing the bias is an issue that needs careful consideration, and may not always be possible to prove. In the absence of firm evidence of the causal relation between nonresponse and bias, it is sensible to strive for representativeness in the sample in each of the most relevant response stages: contact representativeness and cooperation representativeness. In the SCS cooperation could only be predicted if cooperation was defined as 'those who cooperate from those who are contacted', but not if the definition was reduced to 'those who cooperate from those who are contacted and able to cooperate. Under the first definition (COOP1 [AAPOR, 2008]), what we actually predicted was who was able to cooperate, which is no higher magic. This finding led us to revisit the existing literature on the correlates of cooperation and socio-demographic characteristics. We suspected that differing definitions might be one of the reasons research findings are highly dispersed. This is one of the topics of chapter 5.

The research in chapter 5 is aimed at finding socio-demographic correlates of contact and cooperation in survey research. The literature on contact propensity is fairly consistent, but the picture is fuzzy for socio-demographic correlates of cooperation. The causes for noncontact lay in the combination of at-home patterns of the sample person with the call schedule of the interview, coupled in some cases with physical boundaries like gates, gatekeepers and apartment buildings that cannot be entered. But cooperation depends on a far more complex set of influences. The length of the survey, the topic, the sponsor, the contents of the advance letter, and the absence or presence of an interviewer, his or her characteristics and behaviour at the doorstep, and the interaction of interviewer and sample persons are but a few. Prior studies differed in all these aspects, which makes comparison difficult. By studying seven surveys with different designs, but with comparable fieldwork procedures and equal definition of cooperation, we were able to gain a better understanding of which characteristics contribute to contactability and cooperation.

The results largely replicated earlier findings concerning noncontact: men, singles, people in apartment buildings and people without a fixed landline had a relatively low contact chance. Results for cooperation were less consistent over the different surveys studied: no single characteristic was found to be consistently related to cooperation in all surveys. The only consistent finding was that people with a fixed landline, home owners and people in the most affluent neighbourhoods cooperate more. But even these findings were not significant in all surveys.

Even though univariate and multivariate relations between contact, cooperation and socio-demographic variables were highly significant, in fact the fit of the multivariate models was extremely low. The explained variance in cooperation for two surveys was no more than one percent for two surveys, two and three percent for two others. The fit of the models for contact were somewhat higher: around fifteen percent.

In chapter 4 I described an adaptive fieldwork strategy in which the differential design was decided upon before commencement of the fieldwork, based on predicted contact and cooperation propensities. The findings in chapter 5 that only a small part of contact propensity and a negligible amount of cooperation propensity can be predicted has implications for the design of adaptive fieldwork strategy. Determining the design at forehand without further adaptation to the situation found in the field limits the possible gains of adaptive designs.

This means that at a certain moment *during* fieldwork decisions need to be taken as to further action for each sample persons: if to pursue and how to pursue. That decision needs to be driven by knowledge gained in the course of data collection, in contrast to the (socio-demographic) variables underlying the initial design. And the decision needs to be driven by knowledge about correlates of important substantive variables in each survey. Finding such correlates, another kind of

paradata, will be easier in the field than in telephone surveys, but will by no means be easy. It would imply finding out as much as possible about the sample person, either from his environment or himself, gauging how this knowledge may relate to possible bias, and put in most effort for those sample units with the highest chance of biasing results.

Chapter 6 is the bridge to the second part of this thesis. In this part I focus on interviewer behaviour, its causes and its implications. Interviewers and especially differences between interviewers have been in the centre of a long research tradition. One research tradition focuses on aspects like interviewer experience, attitudes, personality and skills to understand differences between interviewers in contact and cooperation rates (e.g., Blom, De Leeuw & Hox, 2011; Durrant et al., 2010; Jäckle, Lynn, Sinibaldi & Tipping, 2013; Snijkers, Hox & De Leeuw, 1999). In spite of the extensive research on this topic, no clear picture emerges as to the correlates of interviewer success with either personality or attitudes.

Instead of focussing on interviewer characteristics, I focus on interviewer behaviour, and specifically on the strategy interviewers employ to contact sample persons. It has long been acknowledged that interviewers may have a profound influence on survey results, both on contact and cooperation rates and substantive variables. As a result of different call patterns, differences in contact rates exist between interviewers (e.g., Purdon, Campanelli & Sturgis, 1999). Several authors found that differences in timing and number of calls may lead to differences in response distribution (e.g., Wang, Murphy, Baxter & Aldworth, 2005; West & Olson, 2010). Non-random allocation of effort may lead to bias (Purdon et al., 1999; Couper & Groves, 1992). That is why fieldwork organisations will generally insist on standard rules for their field interviewers. Chapter 6 describes the rules SN prescribes, and shows for some of these rules (those that can be monitored) that following them will lead to the attainment of high contact rates with a low number of contact attempts. The only reason I was able to compare the results of prescribed fieldwork strategy with other strategies is that interviewers do not always follow that strategy. Although there are differences between interviewers in the extent to which they generally follow the prescribed rules, all interviewers transgress to some extent. Chapter 7 investigates the reasons interviewers have for not following the fieldwork rules, while chapter 8 focuses on the question if interviewers introduce bias by not following the rules.

It should come as no surprise that interviewers are sometimes reluctant to follow the rules: it would imply working several evenings a week. They are asked to visit neighbourhoods and addresses where they would perhaps not spontaneously visit and most certainly not after dark. Yet that is exactly what we ask them to do. However, that does not seem to be the major reason interviewers do not comply. Chapter 7 starts with a literature overview of reasons why people may not follow rules, from fields where the comparable problem exists that rules

should be followed in a situation where compliance is not constantly or easily monitored. Examples are fishermen and fish quota and drivers in traffic. The review uncovered 12 different reasons why people may or may not follow rules, like the consequences of transgression, the probability of detection, the perceived legitimacy of the rule, whether the rule is consistent with personal circumstances, if important others follow the rule and knowledge of the rule. Each of these reasons was captured in a number of questions, and these questions were administered to the field interviewers, mixed in a questionnaire of interviewers' evaluation of the fieldwork strategy.

Results showed that variation between interviewers in the extent to which they follow the guidelines is related to some of these dimensions. Different dimensions were important for different rules. Variation in the extent to which noncontacts were visited the required number of times could only be understood as a function of knowledge of the rule, as well as knowledge of the concept of 'noncontact'. As interviewers are monitored constantly, and receive feedback on their performance, without this resulting in correct interpretation of the concept and the rule, it was found that additional training would not suffice. Therefore, a redesign of the nonresponse categorization questions is planned, to aid interviewers with the important task of correctly defining the nonresponse category a sample person belongs in.

The extent to which interviewers consider a rule to be legitimate plays an important role. If interviewers do not believe a rule to make sense, then they will be less likely to follow it up. Although coercion can influence interviewers to follow certain rules, coercion will not work for rules that are found to be illegitimate. Other dimensions that were found to play a role were routine and social norms: especially newly starting interviewers are influenced by the behaviour of their peers, which may result in newly trained interviewers not following rules they were taught in basic training.

Even though interviewers' compliance could not be entirely explained, this research led to insights that can be used to improve fieldwork compliance, and with that result in higher contact rates and lower fieldwork costs. The results triggered increased investment in the understanding of the interviewers of the specific goals of the rules, and the theoretical and practical underpinning of the rules. The fieldwork strategy was rewritten to show and explain the goal and working of each rule. The concern for contact rates and costs was stressed. Sessions with interviewer groups were dedicated to the advancement of interviewer understanding.

The number of factors in the research in chapter 7 was large, but most variance in interviewer behaviour is still unaccounted for. Variance can also be introduced by the nature of the addresses. If interviewers prefer visiting elderly people or difficult

neighbourhoods during daytime, this would influence compliance rates. The influence of the nature of the address was explored in chapter 8.

The subject of chapter 8 is bias introduced by interviewers who do not comply to fieldwork guidelines. The possibility of bias was researched in two ways. First, I looked into patterns in the kind of addresses where rules were not followed. If interviewers structurally put in more or less effort for specific groups of people, this may introduce bias. For example, if interviewers put in less effort in difficult neighbourhoods, this could reflect on the number of unemployed people that are found in the Labour Force Survey. If they are reluctant to work evenings, this will have implications for the number of working people found. The second way to ascertain bias was by comparing in each interviewer's workload the number of working people in her response to a registration of people's main source of income (work, allowance, pension, etc.).

The research showed that interviewers do indeed differentiate between addresses in the effort they put in, but not in the way I feared: interviewers tend to put in more effort for the more difficult addresses. For example, households with an elevated chance of not being contacted, like singles, and households living in apartment buildings are visited more during evening hours, while for addresses in the countryside (where response rates are high), interviewers will stop visiting sooner, and will visit more during the day. That is, most interviewers apply a differentiated strategy, that may actually play a role in securing a more representative response.

However, a minority of interviewers does what I dreaded: they evade difficulties. Whether these interviewers introduced bias by their fieldwork behaviour was ascertained by contrasting the number of working household members in their sample with the number in a registry. Although the average amount of bias was very small, there appeared to be interviewers with large discrepancies, in both directions: interviewers with relatively too few working people, and interviewers with too few non-workers. The discrepancy between response and sample was related to higher or lower contact and cooperation rates than could be expected on the basis of the characteristics of the sample units, and these in turn by a higher or lower than average number of evening visits and visits to non-contacted sample units.

9.1 Discussion

In these chapters I showed examples of how paradata can be used to study survey errors and survey costs. I wrote about quality, in term of bias and representativeness. And I wrote about smart fieldwork, both in the sense of adapting fieldwork to characteristics of the sample person, and in the sense of performing optimal fieldwork. Costs were mentioned explicitly in chapter 4, where the adaptive fieldwork design showed promise as a means to save costs. Cost awareness as well as awareness of survey error underlies the extensive emphasis on following field procedures. Savings are to be expected if survey persons are approached at their optimal time, with their optimal mode, and if interviewers would not waste time by embracing sub-optimal fieldwork strategy. However, in the last few years the savings with these kinds of measures bleach in comparison to the savings attained by the introduction of web data collection in the social surveys.

The introduction of web surveys and mixed mode data collection has implications for the findings in this manuscript. Much of the research described in chapter 2 about response enhancing measures for CAPI and CATI is currently repeated with web surveys in the main role. For example, the advance letter gains new importance in the absence of an interviewer but the necessary phrases to describe how to access the questionnaire on the web complicate and lengthen the letter, with possibly detrimental effect. Personalisation of the advance letter under such circumstances may be much more important than in the situation that an interviewer intervenes, as was the case in chapter 3 and is more reminiscent of the findings concerning cover letters for paper questionnaires. The research in chapter 3 should be repeated for the new design, where we may expect to find more complicated interactions with auxiliary variables, as a result of coverage and preference effects for specific modes.

Advance letters are but one of the examples where new research is needed to respond to the introduction of new modes and mixed mode data collection. The best way to optimize survey design in mixed-mode surveys remains far from clear and will occupy our attention for some years to come. Paradata will increasingly play a role as one of the tools available to aid that process.

10.

Samenvatting

(Summary in Dutch)

Deze dissertatie is geschreven als onderdeel van het onderzoeksprogramma 'non-respons en moeilijk waarneembare groepen', dat het Centraal Bureau voor de Statistiek (CBS) in 2003 is gestart in samenwerking met de universiteiten van Utrecht en Amsterdam. Binnen dit onderzoeksprogramma zijn eerder dissertaties geschreven door Cobben (2009) over correctie voor non-respons en door Feskens (2009) over moeilijk waarneembare groepen. Het derde onderdeel in het programma behelste 'non-respons vermindering'.

Onderzoek naar de reden waarom mensen niet meedoen met enquêteonderzoek (surveys) en het bedenken van manieren om het aandeel respondenten te verhogen is een bloeiende tak van het survey onderzoek. Ten tijde van de start van dit dissertatieonderzoek werden survey onderzoekers zich bewust van het feit dat het niet altijd zo hoeft te zijn dat een hogere respons een 'betere' respons is. Non-respons is problematisch om twee redenen: het vermindert de precisie van uitkomsten en het kan vertekening veroorzaken als mensen die meedoen met het onderzoek op cruciale variabelen anders zijn dan mensen die niet meedoen. De laatste jaren is duidelijk geworden dat er geen noodzakelijke relatie bestaat tussen de hoogte van de respons en de mate van vertekening. Het niet altijd zo is dat de resultaten van onderzoeken met een relatief lage respons leiden onder meer vertekening. Ook bleek dat onder sommige omstandigheden maatregelen om non-respons te verminderen meer kwaad doen dan goed, omdat ze het contrast tussen respondenten en non-respondenten verhogen.

Het is moeilijk te voorspellen onder welke omstandigheden non-respons vertekening veroorzaakt. Daarom is het verstandig om te streven naar zo hoog mogelijke responscijfers, maar men moet zich terdege bewust zijn van het gevaar dat door inspanningen om de respons te verhogen vertekening kan worden geïntroduceerd. Er zijn veel manieren waarop respons kan worden gestimuleerd. Hoofdstuk 2 vat ze samen. Hoofdstuk 3 beschrijft een experiment met een van de stimulerende maatregelen: het persoonlijk maken van de brief die mensen ontvangen als ze in de steekproef van een onderzoek vallen.

In verschillende hoofdstukken staat vertekening door non-respons expliciet of impliciet centraal. Hoofdstuk 4 beschrijft de mogelijkheid om vertekening te voorkomen door niet de hoogte van de respons te stimuleren, maar de samenstelling ervan. We hebben geprobeerd dit te doen door niet iedereen in de steekproef dezelfde behandeling te geven, maar door een gedifferentieerde benadering te ontwerpen voor verschillende groepen in de steekproef. Hoofdstuk 5 borduurt verder op dit thema, door te onderzoeken of we voor verschillende groepen kunnen voorspellen wat hun responsgedrag zal zijn.

Een belangrijke rol bij het voorkomen van non-respons en vertekening is weggelegd voor interviewers. Veldwerkorganisaties proberen interviewers optimaal te laten functioneren door ze te trainen en te monitoren en door ze richtlijnen mee te geven voor het doen van het veldwerk. Ervaring leert echter dat

er verschillen zijn tussen interviewers, zowel in de hoogte van de respons, de mate waarin ze richtlijnen opvolgen en de mate waarin ze vertekening veroorzaken. In hoofdstuk 6 worden de richtlijnen van het CBS voor de veldinterviewers beschreven en wordt aangetoond dat het opvolgen van de richtlijnen leidt tot betere resultaten tegen minimale kosten. In hoofdstuk 7 wordt onderzocht waarom interviewers de richtlijnen desondanks niet altijd opvolgen. Hoofdstuk 8 beschrijft een onderzoek naar de mate waarin interviewers vertekening veroorzaken door de manier waarop zij het veldwerk uitvoeren. Hoofdstuk 9, tenslotte, vat de resultaten samen.

Non-respons en vertekening zijn niet de enige rode draden in deze dissertatie. Om te kunnen zien of er sprake is van vertekening, wat de kenmerken zijn van non-respondenten, hoe interviewers zich gedragen en wat de gevolgen daarvan zijn, is gebruik gemaakt van zogenaamde paradata. Paradata zijn gegevens over het proces van het verzamelen in ruimste zin. Voorbeelden zijn het aantal keer en de tijden waarop een interviewer een bezoek heeft gebracht, observaties van interviewers over een adres, of over de interactie met een respondent, kenmerken van de stem of het taalgebruik van een interviewer. Het CBS heeft niet alleen de beschikking over al deze paradata, maar is in de unieke positie dat het toegang heeft tot externe registers die kunnen worden gekoppeld aan de steekproef. Het bijzondere daarvan is dat gedetailleerde individuele informatie voorhanden is over non-respondenten. Daardoor is heel precies te onderzoeken of en hoe respondenten afwijken van non-respondenten. Deze informatie kan worden gebruikt om het onderzoek af te stemmen op kenmerken en voorkeuren van mensen in de steekproef.

10.1 Maatregelen om de respons bij enquêteonderzoek te verhogen

Niet alle huishoudens of personen uit de steekproef zullen meedoen met de enquête. Een deel zal blijken niet tot de steekproef te behoren (ze zijn bijvoorbeeld verhuisd naar het buitenland, of overleden), een deel zal niet worden behandeld (bijvoorbeeld omdat de interviewer ziek wordt) en een deel zal niet in staat zijn om mee te doen (omdat ze de taal niet spreken, of omdat ze gedurende de hele veldwerkperiode niet aanwezig zijn). Dit zijn soorten non-respons waar de veldwerkorganisatie of de interviewer niet veel aan kan doen. Twee verdere soorten non-respons kunnen wel worden beïnvloed: weigeringen en non-contacts. Weigeringen en non-contacts hebben verschillende oorzaken, en

hebben verschillende invloed op de kwaliteit van de gegevens, afhankelijk van de mate waarin de non-respons samenhangt met het onderwerp van het onderzoek. Het CBS verzamelt gegevens op verschillende manieren: door telefonisch enquêtes af te nemen, door mensen te bezoeken ('face-to-face'), door papieren vragenlijsten op te sturen, en de laatste jaren ook door enquêtes via het internet. Al deze manier van gegevens verzamelen ('modes') vereisen speciale maatregelen, allereerst om contact te maken met de beoogde respondent, en vervolgens om de respondent¹⁾ te bewegen om mee te doen.

Hoofdstuk 2 beschrijft de maatregelen die een veldwerkorganisatie kan nemen om in elke mode zoveel mogelijk mensen te bereiken (bijvoorbeeld door vaak contactpogingen te doen, op voor de respondent zo gunstig mogelijke tijden, en deze goed te spreiden over de beschikbare tijd) en vervolgens over te halen. Dit laatste begint al voor de interviewer belt of aan de deur komt, met het schrijven van een pakkende brief die duidelijk maakt waarom de respondent benaderd is en mee moet doen. Vervolgens is een belangrijke rol weggelegd voor de interviewer. Door haar voorkomen, maar vooral door hoe zij zich aanpast aan de respondent en diens zorgen ('is het wel anoniem', 'duurt het niet te lang', 'ben ik niet te oud', etc.) kan zij hem overhalen om toch mee te doen. Er zijn verschillen tussen interviewers in die vaardigheid die door training en ervaring kunnen worden bijgeschaafd. Als de interviewer met het interview begonnen is kan dat nog steeds tot een voortijdig einde komen als het enquêtemateriaal niet in orde is: de vragenlijst moet niet te lang zijn, de vragen moeten logisch zijn en goed geformuleerd.

10.2 Het persoonlijk maken van aanschrijfbrieven verhoogt niet altijd de respons. Samenhang met demografische kenmerken in een grootschalig onderzoek

Een belangrijke rol is weggelegd voor de brief die respondenten krijgen met het verzoek om een vragenlijst in te vullen op internet of op papier, dan wel om

¹⁾ Strikt genomen zijn dit 'potentiële respondenten'. Voor zover dit niet tot verwarring leidt worden steekprofeenheden (de huishoudens, adressen of personen die in de steekproef zijn getrokken), respondenten genoemd totdat we zeker weten dat ze dat niet zijn.

het bezoek of telefonische benadering door een interviewer aan te kondigen. Hoofdstuk 3 beschrijft een onderzoek naar het effect van het persoonlijk maken van de brief door er een naam op te zetten. Naar deze ingreep is al veel onderzoek gedaan, maar de resultaten zijn gemengd. In dit experiment wordt onderzocht of dit wordt veroorzaakt doordat verschillende groepen in de samenleving op een andere manier op een dergelijke brief reageren.

Het onderzoek is uitgevoerd bij de Enquête Beroepsbevolking (EBB). Voor de EBB wordt een steekproef van adressen getrokken, waarbij de brief wordt gericht aan 'de bewoners van dit adres'. Voor het onderzoek is gebruik gemaakt van registerinformatie om de correcte naam en sekse van de bewoners van elk adres te achterhalen en deze te gebruiken in de aanschrijving en de aanhef van de brief. Het bleek dat het plaatsen van een naam op de brief geen algemeen effect had op de mate waarin de brief werd gelezen, en ook niet op de mate waarin mensen met het onderzoek meededen. Wel bleek dat er inderdaad verschillen waren tussen subgroepen in hun reactie op een dergelijke persoonlijke brief. Mensen die hun telefoonnummer niet registreren hadden een *lagere respons* na een dergelijke persoonlijke brief, terwijl mensen met een geregistreerd telefoonnummer een hogere respons hadden. De verschillen waren niet groot, maar in dit onderzoek werd het daadwerkelijke overhalen door een interviewer gedaan. In het geval dat de brief de enige communicatie is tussen de veldwerkorganisatie en de respondent, zoals bij web surveys, is het verschil waarschijnlijk veel groter. Een negatief effect in de ene groep, maar een positief effect in de andere kan een wenselijke uitkomst zijn, als het positieve effect plaatsvindt bij groepen die normaal gesproken ondervertegenwoordigd zijn in de respons, en het negatieve effect bij groepen die oververtegenwoordigd zijn. Het netto resultaat zou dan zijn dat de respons een betere afspiegeling vormt van de populatie. In dit experiment was dat niet het geval: het negatieve resultaat werd gevonden in een groep die toch al ondervertegenwoordigd is in de respons, en vice versa.

10.3 Representatieve respons door gedifferentieerd veldwerk. Een experiment in het Consumenten Conjunctuur Onderzoek

In hoofdstuk 3 bleek dat een uniforme maatregel die was bedoeld om de respons te stimuleren tot gevolg kan hebben dat de samenstelling

van de respons verslechtert. In hoofdstuk 4 beschrijf ik hoe de respons gedifferentieerd kan worden gestimuleerd, zo dat groepen die over het algemeen ondervertegenwoordigd zijn in de respons worden gestimuleerd, en groepen die oververtegenwoordigd zijn wat worden afgeremd. Het doel van dit experiment was om een meer representatieve samenstelling van de respons te bewerkstelligen, terwijl de respons minimaal op hetzelfde niveau moest blijven. Vaak wordt een dergelijke gedifferentieerde behandeling in fasen uitgevoerd tijdens het veldwerk. Het CBS weet echter door ervaring al veel over het responsgedrag van mensen. Voor dit experiment hebben we dan ook van te voren besloten welke benaderstrategie optimaal is voor een huishouden. De steekproef van het CCO werd verrijkt met achtergrondkenmerken. Op basis van die kenmerken werd een schatting gemaakt van de kans op contact en de kans op coöperatie, aan de hand van de kansen in vroegere CCO-reeksen. Met die schattingen werd vervolgens een uiterst sophisticated design ontworpen.

Het doel was om een meer representatieve respons te krijgen dan in de parallel lopende reguliere CCO. Het is geen kunst om tot betere resultaten te komen als je onbeperkte financiën hebt, of onbeperkt de tijd. Daarom werd er als randvoorwaarde gesteld dat het resultaat een betere afspiegeling moest vormen van de steekproef, maar dat het veldwerk niet meer mocht kosten, de respons minimaal even hoog moest zijn, de veldwerkperiode even lang, en de interviewers even goed. Om geld vrij te maken voor duurdere ingrepen later in de veldwerkperiode werd het CCO herontworpen tot een mixed mode design met een web- en schriftelijke versie van de CCO-vragenlijst als eerste ronde. Of mensen een web, dan wel een schriftelijke versie kregen lieten we ook afhangen van geschatte responskansen. In het telefonisch deel deden we ingrepen om de contactkans en de kans dat mensen meedoen te manipuleren. Mensen waarvan we vermoedden dat we veel moeite zouden moeten doen om ze te bereiken werden meteen vanaf de eerste dag gebeld, in elke dienst en op elke dag. En dagelijks kregen ze de hoogste prioriteit om gebeld te worden. Mensen met een hoge kans op contact (bijvoorbeeld 65-plussers) werden in de eerste week van het veldwerk niet gebeld, en verder veel overdag. Dit maakte tijd en ruimte vrij voor het extra bellen van anderen. Kans op coöperatie werd gemanipuleerd door te variëren met de inzet van de interviewers. De interviewers werden in vier groepen verdeeld (van steengoed tot gewoon goed). Hoe kleiner de geschatte kans dat een persoon mee zou doen, hoe beter de interviewer die hem moest benaderen.

Wij bleken in staat de representativiteit significant te verhogen binnen de randvoorwaarden. Zowel het toevoegen van een mixed mode design en het stimuleren van contact droeg bij aan een betere representativiteit, uitgedrukt met de zogenaamde R-indicator. Alleen de ingreep met de inzet van de interviewers had niet het beoogde resultaat. Dit had te maken met het feit dat wie mee gaat

doen niet te voorspellen bleek aan de hand van socio-demografische kenmerken. In het volgende hoofdstuk wordt nader op deze bevinding ingegaan.

10.4 Het voorspellen van contact- en coöperatiegeneigdheid in enquêtes

Het ontwerpen van een gedifferentieerd design is erg geholpen met voorspelbaarheid. In hoofdstuk 5 beschrijf ik onderzoek naar socio-demografische kenmerken die samenhangen met contact en coöperatie in enquêteonderzoek. Uit literatuuronderzoek bleek dat er eensgezindheid was over de samenhang met de kans op contact, maar dat bevindingen uiterst gevarieerd waren ten aanzien van kans op coöperatie.

Coöperatie wordt dan ook door meer aspecten beïnvloed dan contactkans. De lengte van het onderzoek, het onderwerp, door wie het wordt gehouden, wat er in de aanschrijfbrief staat, of er een interviewer is, zal allemaal bepalen of iemand al dan niet mee gaat doen. Onderzoek verschilt op al deze dimensies, en daardoor is het lastig om onderzoeksresultaten te vergelijken. Ook verschillen onderzoeken in de definitie van coöperatie. Door zeven verschillende CBS enquêtes te onderzoeken konden een aantal variatiebronnen uitgeschakeld worden. De designs van de enquêtes waren verschillend, maar de veldwerkprocedures waren hetzelfde, en de definitie van coöperatie ook.

Mannen, alleenstaanden, mensen die in flatgebouwen wonen en mensen die geen geregistreerd telefoonnummer hebben, werden in alle onderzoeken significant minder bereikt. In sommige onderzoeken kwamen daar ook jongeren, stedelingen en niet-westerse minderheden bij. Ondanks de consistente bevindingen, werd maar gemiddeld 15% van de variantie in contactkans verklaard door socio-economische achtergrondkenmerken.

De resultaten met betrekking tot coöperatie waren minder consistent: er was geen enkel kenmerk dat in alle onderzoeken samen bleek te hangen met coöperatie. De enige min of meer consistente bevinding was dat mensen met een geregistreerd telefoonnummer, huiseigenaren en mensen in rijkere buurten vaker meedoen met enquêteonderzoek. Het bleek dan ook dat het zelfs met een grote hoeveelheid beschikbare achtergrondkenmerken niet mogelijk was te voorspellen wie er mee gaat doen: voor sommige onderzoeken werd niet meer dan 1% van de variantie verklaard door deze kenmerken.

Deze bevindingen hebben implicaties voor de ontwikkeling van gedifferentieerde benaderstrategieën. Het is mogelijk te voorspellen wie er moeilijk te bereiken zal

zijn, en ook wie er niet in staat zal zijn mee te doen, en daar maatregelen voor te nemen. We blijken echter niet in staat van te voren te voorspellen wie er al dan niet mee gaat doen. Andere manieren zullen moeten worden ontwikkeld om hier meer zicht op te krijgen. Paradata die tijdens het veldwerk worden verzameld zullen hierin een belangrijke rol spelen.

10.5 Veldwerkstrategie van het Centraal Bureau voor de Statistiek

Veldwerkorganisaties willen dat hun interviewers efficiënt werken en een zo hoog mogelijke respons verwezenlijken. Tegelijk moeten ze ervoor waken dat interviewers vertekening veroorzaken door bijvoorbeeld vooral aandacht te geven aan adressen of personen waarvan ze verwachten dat die tot respons zullen leiden. Een van de manieren om dit te bewerkstelligen is het uitvaardigen van richtlijnen over het uitvoeren van veldwerk. Ook het CBS heeft dergelijke richtlijnen opgesteld voor de veldinterviewers. In hoofdstuk 6 beschrijf ik welke richtlijnen dat zijn, waarom deze zijn geformuleerd en wat het gevolg is van het niet opvolgen van de regels voor het percentage mensen dat bereikt wordt en het aantal bezoeken dat nodig is om een respondent te bereiken.

De veldwerkstrategie voor CBS-interviewers schrijft voor dat alle steekproef-eenheden minimaal één keer bezocht moeten worden in de eerste tien dagen van de veldwerkstrategie, dat een van de eerste twee bezoeken na 5 uur 's middags moet vallen, dan wel in het weekend, en dat personen waarmee nog geen contact is gelegd zes maal bezocht moeten worden. Na de eerste drie bezoeken waarin niemand thuis wordt getroffen, laten interviewers een bezoekaartje achter met contactinformatie. Bezoeken moeten worden gespreid over tijden, dagen en weken van de veldwerkperiode. De eerste drie contactpogingen moeten face-to-face zijn, daarna kan gepoogd worden telefonisch contact te verkrijgen. Nadat deze veldwerkstrategie werd geïntroduceerd in 2003, samen met een programma om te monitoren of interviewers zich eraan houden, gingen de responscijfers aanzienlijk omhoog: met bijna 10 procentpunten voor de Gezondheidsenquête, en negen procentpunten voor de EBB, vooral als gevolg van verbeterde contactpercentages.

In hoofdstuk 6 wordt van een aantal elementen van de veldwerkstrategie aangetoond wat het effect is van het al dan niet opvolgen. Die analyses zijn alleen mogelijk omdat de strategie niet altijd wordt opgevolgd. Er is verschil tussen

interviewers in de mate waarin ze de richtlijnen volgen, maar alle interviewers overtreden ze in meer of mindere mate.

De analyses maken duidelijk dat de veldwerkregels op goede gronden zijn uitgevaardigd: ze opvolgen leidt tot snellere en hogere respons, en daarmee tot een reductie van kosten. De enige regel die enige kwalificatie behoeft is ten aanzien van het spreiden over tijden: als een bezoek in de avond niet tot contact leidde, is de grootste kans op contact toch bij een volgend avondbezoek. Ook spreiden over dagen is niet altijd de beste manier van werken: de hoogste kans op succes is bij een avondbezoek op dezelfde dag als een vergeefs bezoek overdag.

10.6 **Waarom interviewers zich niet houden aan de regels van de veldwerkstrategie**

In hoofdstuk 7 beschrijf ik onderzoek naar de redenen die interviewers hebben voor het niet opvolgen van de veldwerkrichtlijnen. De analyses naar het effect van de veldwerkstrategie hebben aangetoond dat het volgen van de strategie altijd tot de beste resultaten leidt in de snelste tijd. Interviewers moeten dus andere redenen hebben om de regels niet op te volgen.

Een literatuuronderzoek toonde 12 verschillende redenen waarom mensen regels al dan niet opvolgen, zoals de consequenties van het niet opvolgen, de pakkans, of ze vinden dat een regel zin heeft, of de regel strookt met persoonlijke omstandigheden, of anderen in de omgeving de regel volgen, de kennis van de regel, enzovoort. Van elk van deze dimensies zijn een aantal vragen gemaakt die aan de interviewers zijn voorgelegd, gemengd met een vragenlijst die hun oordeel over de veldwerkstrategie inventariseerde. Het onderzoek toonde aan dat variatie tussen interviewers in de mate waarin ze de richtlijnen volgen samenhangt met een aantal van deze dimensies.

Het bleek dat de mechanismen die samenhangen met het opvolgen van regels voor elk van de bestudeerde regels verschillend was. Het niet opvolgen van de regel dat non-contacten zes maal moeten worden bezocht kon verklaard worden door gebrek aan kennis van de regel, en vooral door kennis van het concept 'non-contact'.

Een grote rol speelt de mate waarin een interviewer een regel legitiem vindt: als ze een regel als niet legitiem beoordelen (waarmee ze aangeeft dat ze niet gelooft dat de regel zin heeft), is ze minder geneigd deze op te volgen. Hoewel dwang interviewers kan bewegen bepaalde regels te volgen, geldt dat niet als

de regel als niet legitiem wordt beoordeeld. Invloed werd ook gevonden van routine en sociale normen: vooral pas beginnende interviewers richten zich op hun collega's. Als deze collega's de veldwerkregels niet volgen kan dat tot gevolg hebben dat de nieuwe interviewer dit ook niet doet.

Naar aanleiding van dit onderzoek heeft de veldwerkorganisatie van het CBS de benaderstrategie aangepast door bij elke regel aandacht te besteden aan de reden ervan, aangevuld met bewijs dat de regel werkt.

10.7 Veroorzaken interviewers vertekening door zich niet te houden aan de veldwerkstrategie? De samenhang tussen interviewergedrag, kenmerken van de steekproefeenheid en vertekening

Hoofdstuk 8 onderzoekt of interviewers die zich niet aan de veldwerkstrategie houden, vertekening veroorzaken. Dat is op twee manieren onderzocht. In de eerste plaats is gekeken of er systeem is in het soort adressen waar interviewers de regels niet volgen. Als interviewers structureel meer of minder moeite doen voor bepaalde groepen mensen kan dat vertekening veroorzaken. Als interviewers bijvoorbeeld minder moeite doen in moeilijke wijken, dan kan dat invloed hebben op het aantal werklozen dat wordt waargenomen in de EBB. Als ze niet graag in de avond werken, zal het aantal werkenden niet goed worden waargenomen.

De tweede manier kijkt rechtstreeks naar de vertekening, door het aantal werkenden in de steekproef van elke interviewer de vergelijken met een registratie van de belangrijkste bron van inkomsten van de personen in de steekproef (werk, uitkering, enzovoort).

Het onderzoek toonde aan dat interviewers inderdaad verschil maken in de moeite die ze doen voor soorten adressen, maar op een andere manier dan verwacht: juist bij de moeilijke adressen doen interviewers meer moeite: gaan vaker, en meer in de avond, terwijl bij de adressen op het platteland interviewers eerder geneigd zijn om eerder op te houden met bezoeken, en vaker overdag te gaan.

Interviewers hanteren van nature een gedifferentieerde benadering, in een richting die we toejuichen.

Er is echter een minderheid die wel doet waar ik bevreesd voor was: moeilijkheden uit de weg gaan. Deze interviewers vertekenen de resultaten door te weinig werkende personen waar te nemen. Deze groep werd gebalanceerd door interviewers die een hogere respons hadden dan was te verwachten op basis van de kenmerken van de steekproefeenheden in hun portie adressen. Deze interviewers vertekenen resultaten de andere kant op en vinden te weinig mensen zonder werk.

10.8 Conclusie

De hoofdstukken in deze dissertatie tonen voorbeelden van het gebruik van paradata in de studie van fouten in en kosten van survey-onderzoek. De komende jaren zullen in het teken staan van web surveys en mixed mode surveys en van manieren waarop survey designs gedifferentieerd kunnen worden aan de hand van kenmerken en gedrag van respondenten. Paradata zullen daarbij een belangrijke rol spelen.

References

- Abraham, K.G., Maitland, A., & Bianci, S.M. (2006). Nonresponse in the American Time Use Survey. Who is missing from the data and how much does it matter? *Public Opinion Quarterly*, 70, 676–703.
- Aiken, L.S., & West, S.G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, Sage.
- American Association for Public Opinion Research (2008). *Standard definitions: Final dispositions of case codes and outcome rates for surveys* (5th ed.). Lenexa, Ka: Author.
- Arends-Tóth, J., Meertens, V., Engelen, F., & Kroeze, D. (2008). *Klanttevredenheidsonderzoek 2008* [Customer satisfaction survey 2008]. Internal report, Statistics Netherlands.
- Australian Bureau of Statistics (2002). *Advance Letters. An ABS approach*. Retrieved from [http://www.nss.gov.au/nss/home.NSF/533222ebfd5ac03aca25711000044c9e/f8e9db491a6e19a8ca2571ab002471b1/\\$FILE/ASC%2016%20-%20Advance%20Letters_24-7-2002.pdf](http://www.nss.gov.au/nss/home.NSF/533222ebfd5ac03aca25711000044c9e/f8e9db491a6e19a8ca2571ab002471b1/$FILE/ASC%2016%20-%20Advance%20Letters_24-7-2002.pdf)
- Baker, R., & Couper, M. (2007). *The impact of screen size, background color and navigation button placement on response in web surveys*. Paper presented at the General Online Research Conference, Leipzig, Germany.
- Banning, R., Cobben, F., & Leufkens, K. (2011). *Kwaliteitsmeting van EBB kernvariabelen in eerste peiling 2010* [Measuring quality of substantive LFS variables, first wave 2010]. Internal report, Statistics Netherlands.
- Barling, J., & Phillips, M. (1993). Interactional, formal and distributive justice in the workplace: An exploratory study. *Journal of Psychology*, 127, 649–656.
- Barton, J. (1999). Effective calling strategies for interviewers on household surveys, *Survey Methodology Bulletin*, 44, 14–26.
- Bates, D. (2010). *lme4: Mixed-effects modeling with R*. Retrieved from <http://lme4.r-forge.r-project.org/LMMwR/lrgprt.pdf>

Becks, S. (2008). *Don't take "NO" for an answer. Het effect van het introductiegedeelte in telefonische surveys* [The effect of the introductory part of telephone surveys] (Master's thesis). Hogeschool Zuyd, Heerlen, the Netherlands.

Berlin, M., Mohadjer, L., Waksberg, J., Kolstad, A., Kirsch, I., Rock, D., & Yamamoto, K. (1992). An experiment in monetary incentives. *Proceedings of the Section on Survey Research Methods, American Statistical Association* (pp. 393–398). Retrieved from http://www.amstat.org/sections/srms/proceedings/papers/1992_062.pdf

Bethlehem, J.G. (2002). Weighting non-response adjustment based on auxiliary information. In R.M. Groves, D. Dillman, J. Eltinghe & R. Little (Eds.), *Survey non-response* (pp. 265–287). New York, NY: Wiley.

Bethlehem, J.G. (2009). *Applied survey methods, a statistical perspective*. Hoboken, NJ: Wiley.

Bethlehem, J.G., & Biffignandi, S. (2012). *Handbook of web surveys*. Hoboken, NJ: Wiley.

Bethlehem, J.G., & Schouten, B. (2004). *Nonresponse analysis of the integrated Survey on Living Conditions (POLS)* (Discussion paper 04004). The Hague: Statistics Netherlands. Retrieved from www.cbs.nl/NR/rdonlyres/D0EED1C3-7255-4858-9EE6-18629EAAE31/0/discussionpaper04004.pdf

Bethlehem, J.G., & Schouten, B. (2009). *Representativeness indicators for measuring and enhancing the quality of survey response* (RISQ deliverable 9). Retrieved from www.risq-project.eu/papers

Beukenhorst, D. (1999). *Vertekening door non-respons in het WBO* [Bias through nonresponse in the Housing Survey]. Heerlen: Statistics Netherlands.

Beukenhorst, D. (2008). *Het inzetten van anderstalige interviewers* [The use of foreign-language interviewers]. Heerlen: Statistics Netherlands.

Biemer, P.P., Chen, P., & Wang, K. (2013). Using level-of-effort paradata in non-response adjustments with application to field surveys. *Journal of the Royal Statistical Society, Series A*, 176, 147–169.

Blohm, M., Hox, J.J., & Koch, A. (2007). The influence of interviewers' contact behavior on the contact and cooperation rate in face-to-face household surveys. *International Journal of Public Opinion Research*, 19, 97–111.

Blom, A.G., & Blohm, M. (2007). *The effects of first contact by phone: evidence for the European Social Study*. Paper presented at the 18th international workshop on Survey Household Nonresponse, Southampton, United Kingdom.

Blom, A.G., De Leeuw, E.D., & Hox, J.J. (2011). Interviewer effects on nonresponse in the European Social Survey. *Journal of Official Statistics*, 27, 359–377.

Bogen, K. (1996). *The effect of questionnaire length on response rates. A review of the literature*. Paper presented at the annual conference of the American Association for Public Opinion Research, Salt Lake City, Utah. Retrieved from <http://www.amstat.org/sections/srms/Proceedings/y1996f.html>

Böhme, R. (2004). *The effect of motivating elements on response strategies in online surveys*. General Online Research Conference, Essen, Germany. Retrieved from <http://www.websm.org/db/12/894/Bibliography/The%20Effect%20of%20Motivating%20Elements%20on%20Response%20Strategies%20in%20Online%20Surveys/>

Camburn, D., Lavrakas, P.J., Battaglia, M.P., Massey, J.T., & Wright, R.A. (1995). Using advance respondent letters in random-digit-dialling telephone surveys. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, Alexandria, VA: American Statistical Association, 969–974.

Campanelli, P., Sturgis, P., & Purdon, S. (1997). *Can you hear me knocking: An investigation into the impact of interviewers on survey response rates*. London, United Kingdom: The Survey Methods Centre.

Carton, A. (1999). *Een interviewernetwerk: Uitwerking van een evaluatieprocedure voor interviewers* (Doctoral dissertation) [An interviewer network: Details of an interviewer evaluation procedure]. Catholic University Leuven, Belgium.

Carton, A., & Pickery, J. (2010). *Worden interviewers beter in het behalen van respons naarmate ze meer ervaring opdoen?* [Are more experienced interviewers better at getting responses?]. Retrieved from <http://www4.vlaanderen.be/dar/svr/publicaties/Publicaties/webpublicaties/2010-11-05-webartikel2010-16-interviewer.pdf>

Champion, D.J., & Sear, A.M. (1969). Questionnaire response rate: a methodological analysis. *Social Forces*, 47, 335–339.

Cialdini, R.B. (2001). *Influence. Science and practice*. Boston, MA: Allyn and Bacon.

Cialdini, R.B., & Goldstein, N.J. (2004). Social influence: compliance and conformity. *Annual Review of Psychology*, *55*, 591–621.

Cobben, F. (2009). *Nonresponse in sample surveys. Methods for analysis and adjustment* (Doctoral dissertation). University of Amsterdam, Statistics Netherlands.

Cobben, F. (2011). *Responsgedrag in de Gezondheidsenquête 2010* [Response behaviour in the Health Survey 2010]. Internal report, Statistics Netherlands.

Cobben, F., & Schouten, B. (2007). *A follow-up with basic questions of nonrespondents to the Dutch Labour Force Survey*. Discussion paper, Statistics Netherlands.

Collins, M., Sykes, W., Wilson, P., & Blackshaw, N. (1988). Nonresponse: The UK experience. In R.M. Groves, Biemer, P.P., Lyberg, L.E., Massey, J.T., Nicholls, W.L., & Wacksberg, J. (Eds.), *Telephone Survey Methodology* (pp. 213–232). New York, NY: Wiley.

Conrad, F.G., Broome, J.S., Benkí, J.R., Kreuter, F., Groves, R.M., Vanette, D., & McCain, C. (2013). Interviewer speech and the success of survey invitations. *Journal of the Royal Statistical Society, Series A*, *176*, 191–211.

Cooper, J. (2007). Factors affecting return rates to online surveys. In M. Trotman (Ed.) *Proceedings of the 5th Association for Survey Computing International Conference* (pp. 93–102). Berkley, United Kingdom: ASC.

Cornish, D.E., & Clarke, R.V. (1986). *The reasoning criminal: Rational choice perspectives on offending*. New York, NY: Springer-Verlag.

Couper, M.P. (1998). Measuring survey quality in a CASIC environment. *Proceedings of the Survey Research Methodology Section, American Statistical Association* (pp. 41–49). Retrieved from <http://www.amstat.org/Sections/Srms/Proceedings/y1998f.html>

Couper, M.P., Baker, R.P., Clark, C.Z., Martin, J., Nicholls, W.L., O'Reilly, J.M. (1988). *Computer assisted questionnaire information collection*. New York, NY: Wiley.

Couper, M.P., & Groves, R.M. (1992). The role of the interviewer in survey participation. *Survey Methodology*, *18*, 263–277.

Couper, M.P., & Kreuter, F. (2013). Using paradata to explore item level response times in surveys. *Journal of the Royal Statistical Society, Series A*, 176, 271–287.

Couper, M.P., & Lyberg, L. (2005). The use of paradata in survey research. *Proceedings of the 55th session of the International Statistical Institute, Sydney, Australia*. Retrieved from http://isi.cbs.nl/iamamember/CD6-Sydney2005/ISI2005_Papers/1635.pdf

Couper, M.P., Mathiowetz, N.A., & Singer, E. (1995). Related households, mail handling and returns to the 1990 Census. *International Journal of Public Opinion Research*, 7, 172–177.

Couper, M.P., Traugott, M., & Lamia, M. (2001). Web survey design and administration. *Public Opinion Quarterly*, 65, 230–253.

Crawford, S.D., Couper, M.P., & Lamia, M. (2001). Web surveys. Perceptions of burden. *Social Science Computer Review*, 19, 146–162.

Cuppen, M., Van der Laan, P., & Van Nunspeet, W. (2011). *Re-engineering Dutch social surveys: From single-purpose surveys to an integrated design*. Paper presented at the ISI World Statistics Congress, Dublin, Ireland.

Cuppen, M., & Te Riele, S. (2008). De voorbereiding van het veldwerk: het testtraject [Fieldwork preparation: the test phase]. In H. Schmeets, (Ed.), *Het Nationaal Kiesersonderzoek 2006. Opzet, uitvoering en resultaten* [The Dutch Parliamentary Election Study 2006. Design, execution and results] (pp. 83–92). Heerlen: Statistics Netherlands.

Curtin, R., Presser, S., & Singer, E. (2000). The effects of response rate changes on the index of consumer sentiment. *Public Opinion Quarterly*, 64, 413–428.

DeCremer, D., & Tyler, T.R. (2005). Managing group behaviour: the interplay between procedural justice, sense of self and cooperation. In M.P. Zanna (Ed.), *Advances in experimental social psychology* (pp. 151–218). New York, NY: Academic Press.

De Groot, B., & Luiten, A. (2001). *Weigeringsconversie in het veld: een exploratief onderzoek* [Refusal conversion in the field: an exploratory investigation]. Internal report, Statistics Netherlands.

De Leeuw, E.D. (2005). To mix or not to mix. Data collection modes in surveys. *Journal of Official Statistics*, 21, 233–255.

De Leeuw, E.D., Callegaro, M., Hox, J., Korendijk, E., & Lensvelt-Mulders, G. (2007). The influence of advance letters on response in telephone surveys. *Public Opinion Quarterly*, 71, 413–443.

De Leeuw, E.D., & De Heer, W. (2002). Trends in household survey non-response: A longitudinal and international perspective. In R.M. Groves, D. Dillman, J. Eltinghe & R. Little (Eds.), *Survey non-response* (pp. 41–54). New York: Wiley.

De Leeuw, E.D., & Hox, J.J. (1996). The effect of the interviewer on the decision to cooperate in a survey of the elderly. In S. Laaksonen (Ed.), *International perspectives on Nonresponse. Proceedings of the Sixth International Workshop on Household Survey Nonresponse* (pp. 46–52). Helsinki: Statistics Finland.

De Leeuw, E.D., & Hox, J. (2009). *International interviewer questionnaire (IQUEST). Development and scale properties*. Department of Methodology and Statistics, Utrecht University. Retrieved from <http://surveymethodology.eu/media/files/warsaw/IQUESTreport.pdf>

De Leeuw, E.D., & Van der Zouwen, J. (1988). Data quality in telephone and face-to-face surveys: A comparative analysis. In R.M. Groves, P. Biemer, L. Lyberg, J. Massey, W. Nicholls III & J. Waksberg (Eds.), *Telephone survey methodology* (pp 283–299). New York, NY: Wiley.

DeMaio, T.J. (1980). Refusals: Who, where and why. *The Public Opinion Quarterly*, 44, 223–233.

De Vree, M., De Bie, S., & Luiten, A. (2006). *Output monitoring at Statistics Netherlands*. Paper presented at the IFD&TC conference, Montreal, Canada.

Dijkstra, J.W., & Smit, J.H. (1999). *Persuading reluctant recipients in telephone surveys*. Paper presented at the International Conference on Survey Nonresponse, Portland, OR.

Dillman, D.A. (1978). *Mail and telephone surveys: The total design method*. New York, NY: Wiley.

Dillman, D.A. (2000). *Mail and internet surveys: The tailored design method* (2nd ed.). New York, NY: Wiley.

- Dillman, D.A., Clark, J.R., & West, K.K. (1995). Influence of an invitation to answer by telephone on response to census questionnaires. *Public Opinion Quarterly*, *58*, 557–568.
- Dillman, D.A., Gorton Gallegos, J., & Frey, J.H. (1976). Reducing refusal rates for telephone interviews. *Public Opinion Quarterly*, *40*, 66–78.
- Dillman, D.A., Lesser, V., Mason, B., Carlson, J., Willits, F., Robertson, R., & Burke, B. (2007). Personalization of mail surveys on general public and other populations: Results from nine experiments. *Rural Sociology*, *72*, 632–646.
- Dillman, D.A., Smyth, J.D., & Christian, L.M. (2009). *Internet, mail and mixed-mode surveys. The tailored design method*. Hoboken, NJ: Wiley.
- Dillman, D.A., Tortora, R.D., Conradt J., & Bowker, D. (1998). *Influence of plain vs. fancy design on response rates for web surveys*. Paper presented at the Joint Statistics Meetings, Dallas, Texas. Retrieved from <http://www.websm.org/uploadi/editor/influence%20of%20plain.pdf>
- Durrant, G.B., D'Arrigo, J., & Steele, F. (2011). Using paradata to predict best times of contact, conditioning on household and interviewer influences. *Journal of the Royal Statistical Society, Series A*, *174*, 1029–1049.
- Durrant, G.B., D'Arrigo, J., & Steele, F. (2013). Analysing interviewer call record data. *Journal of the Royal Statistical Society, Series A*, *176*, 251–269.
- Durrant, G.B., Groves, R., Staetsky, L., & Steele, F. (2009). Effects of interviewer attitudes and behaviors on refusal in household surveys. *Public Opinion Quarterly*, *74*, 1–36.
- Durrant, G.B., & Kreuter, F. (2013). Editorial: the use of paradata in social survey research. *Journal of the Royal Statistical Society, Series A*, *176*, 1–3.
- Durrant, G.B., & Steele, F. (2009). Multilevel modelling of refusal and noncontact in household surveys: evidence from six UK government surveys. *Journal of the Royal Statistical Society, Series A*, *172*, 361–381.
- Elffers, H., Van der Heijden, P., & Hezemans, M. (2003). Explaining regulatory non-compliance. A survey study of rule transgression for two Dutch instrumental laws, applying the randomized response method. *Journal of Quantitative Criminology*, *19*, 409–439.

Emans, B.J.M., Munduate, L., Klaver, E., & Van de Vliert, E. (2003). Constructive consequences of leaders' forcing influence styles. *Applied Psychology, 52*, 36–54.

European Social Survey (2010). *Project instructions for interviewers, round five*. Retrieved from http://www.europeansocialsurvey.org/index.php?option=com_content&task=view&id=119&Itemid=155

Feskens, R.C.W. (2009). *Difficult groups in survey research and the development of tailor-made approach strategies* (Doctoral dissertation). University of Utrecht, Statistics Netherlands.

Feskens, R.C.W., Hox, J.J., Schmeets, J.J.G., & Wetzels, W. (2008). Incentives and ethnic minorities: Results of a controlled randomized experiment in the Netherlands. *Survey Research Methods, 2*, 159–165.

Field, A. (2005). *Discovering statistics using SPSS*. London, United Kingdom: Sage.

Flaherty, S., & Moss, S.A. (2007). The impact of personality and team context on the relationship between workplace injustice and counterproductive work behavior. *Journal of Applied Social Psychology, 37*, 2549–2575.

Fosen, J., Kleven, Ø., Lagerstrøm, B.O., Luiten, A., & Wetzels, W. (2010). *Indicators and data collection control. Report from the pilots in Statistics Norway and Statistics Netherlands* (RISQ project deliverable). Retrieved from <http://www.risq-project.eu/publications>

Foster, K. (1997). The effect of call patterns on non-response bias on household surveys. *Survey Methodology Bulletin, 41*, 37–47.

Fouwels, S., Janssen, B., & Wetzels, W. (2006). *Experiment internetwaarneming bij personen en huishoudens. Onderzoek naar infrastructuur, logistiek en respons* [Internet experiments with persons and households. Research into infrastructure, logistics and response]. Internal report, Statistics Netherlands.

Fouwels, S., Janssen, B., & Wetzels, W. (2006). *Experiment mixed-mode waarneming bij de VMR. Toepassing van internet, face-to-face en telefonische waarneming* [Experiment with mixed-mode data collection in the National Security Monitor. Application of internet, face-to-face and telephone]. Internal report, Statistics Netherlands.

- Fowler, F.J., & Mangione, T.W., (1991). *Standardized survey interviewing: Minimizing interviewer-related error*. Newbury Park, CA: Sage.
- Fox, R.K., Crask, M.R., & Kim, J. (1988). Mail survey response rate: A meta-analysis of selected techniques for inducing response. *Public Opinion Quarterly*, 52, 467–491.
- Galešić, M. (2005). *Effects of questionnaire length on quality of responses in web surveys*. Paper presented at the ESF Workshop on internet survey methodology, Dubrovnic, Croatia.
- Gambino, J., Laflamme, F., & Wrighte, D. (2010). *Responsive design at Statistics Canada: Development, implementation and early results*. Paper presented at the 21st International Workshop in Household Survey Non-response, Nürnberg, Germany.
- Goldstein, K.M., & Jennings, M.K. (2002). The effect of advance letters on cooperation in a list sample telephone survey. *Public Opinion Quarterly*, 66, 608–617.
- Göritz, A.S. (2006). Cash lotteries as incentives in online panels. *Social Science Computer Review*, 24, 445–459.
- Goyder, J.C. (1982). Further evidence on factors affecting response rates to mailed questionnaires. *American Sociological Review*, 47, 550–553.
- Goyder, J.C. (1987). *The silent minority: Nonrespondents on sample surveys*. Cambridge, United Kingdom: Polity Press.
- Griffin, D.H., Fischer D.P., & Morgan, M.T. (2001). *Testing an internet response option for the American Community Survey*. Paper presented at the 56th annual conference of the American Association of Public Opinion Research, Montreal, Quebec, Canada.
- Groves, R.M. (1990). Theories and methods of telephone surveys. *Annual Review of Sociology*, 16, 221–240.
- Groves, R.M. (2004). *Survey errors and survey costs*. Hoboken, NJ: Wiley.
- Groves, R.M. (2006). Non-response rates and non-response bias in household surveys. *Public Opinion Quarterly*, 70, 646–675.
- Groves, R.M., Cialdini, R.D., & Couper, M.P. (1992). Understanding the decision to participate in a survey. *Public opinion Quarterly*, 56, 475–495.

Groves, R.M., & Couper, M.P. (1998). *Nonresponse in household interview surveys*. New York, NY: Wiley.

Groves, R.M., Fowler, F.J., Couper, M.P., Lepkowski, J.M., Singer, E., & Tourangeau, R. (2004). *Survey methodology*. Hoboken, NJ: Wiley.

Groves, R., & Heeringa, S. (2006). Responsive design for household surveys: Tools for actively controlling survey errors and costs, *Journal of the Royal Statistical Society, Series A*, 169, 439-457.

Groves, R.M., & Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias. A meta-analysis. *Public Opinion Quarterly*, 72, 167-189.

Groves, R.M., & Snowden, C. (1987). The effects of advanced letters on response rates in linked telephone surveys. *Proceedings of the annual meeting of the American Statistical Association*, 633-638. Retrieved from http://www.amstat.org/sections/SRMS/Proceedings/papers/1987_113.pdf

Harvey, L. (1987). Factors affecting response rates to mailed questionnaires: A comprehensive literature review. *Journal of Market Research Society*, 29, 341-353.

Heberlein, T.A., & Baumgarten, R.B. (1978). Factors affecting response rates to mailed questionnaires: A quantitative analysis of the published literature. *American Sociological Review*, 43, 447-462.

Heerwegh, D. (2000). Overtuigingsstrategieën in telefonische surveys [Persuasion strategies in telephone surveys]. *Tijdschrift voor de Sociologie*, 21, 383-406.

Heerwegh, D. (2005). *Web surveys. Explaining and reducing unit nonresponse, item nonresponse and partial nonresponse* (Doctoral dissertation). Catholic University Leuven, Belgium.

Heerwegh, D., Abts, K., & Loosveldt, G. (2007). Minimizing survey refusal and noncontact rates: Do our efforts pay off? *Survey Research Methods*, 1, 3-10.

Helgesen, J. G., Voss, K. E., & Terpening, W. D. (2002). Determinants of mail-survey response: Survey design factors and respondent factors. *Psychology and Marketing*, 19, 303-328.

Hessing, D.J., Elffers, H., & Weigel, R.H. (1988). Exploring the limits of self-reports and reasoned action: an investigation of the psychology of tax evasion behavior. *Journal of Personality and Social Psychology*, *54*, 405–413.

Hønneland, G. (1999). A model of compliance in fisheries: theoretical foundations and practical application. *Ocean and Coastal Management*, *42*, 699–716.

Hox, J.J., De Leeuw, E.D., & Kreft, I.G.G. (1991). The effect of interviewer and respondent characteristics on the quality of survey data: A multilevel model. In P.P. Biemer, R.M. Groves, L.E. Lyberg, N.A. Mathiowetz & S. Sudman (Eds.), *Measurement errors in surveys* (pp 439–463). New York, NY: Wiley.

Hox, J.J. (2010). *Multilevel analysis: techniques and applications* (2nd ed.). New York, NY: Routledge.

Hox, J.J., & De Leeuw, E.D. (2002). The influence of interviewers' attitude and behavior on household survey nonresponse: An international comparison. In R.M. Groves, D.A. Dillman, J.L. Eltinge & R.J.A. Little (Eds.), *Survey nonresponse* (pp. 103–118). New York, NY: Wiley.

Jäckle, A., Lynn, P., Sinibaldi, J., & Tipping, S. (2013). The effect of interviewer experience, attitudes, personality and skills on respondent co-operation with face-to-face surveys. *Survey Research Methods*, *7*, 1–15.

Jenny, A., Hechavarria Fuentes, F., & Mosler, H.J. (2007). Psychological factors determining individual compliance with rules for common pool resource management: the case of a Cuban community sharing a solar energy system. *Human Ecology*, *35*, 239–250.

Johnson, R., & Price, Y. (1988). *The relationship of the Jackson personality traits of US Census Bureau current survey interviewers to work performance and turnover*. Washington, DC: US Bureau of the Census.

Kahneman, D. & Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica*, *57*, 263–291.

Keeter, S., Miller, C., Kohut, A., Groves, R.M., & Presser, S. (2000). Consequences of reducing nonresponse in a national telephone survey. *Public Opinion Quarterly*, *64*, 125–148.

- Kennickell, A.B. (2003). Reordering the darkness: Application of effort and nonresponse bias in the Survey of Consumer Finances. *Proceedings of the Survey Research Methods Section, American Statistical Association*, 2119–2126.
- Kennickell, A.B. (2004). Action at a distance: Interviewer effort and nonresponse in the SCF. *Proceedings of the Survey Research Methods Section, American Statistical Association*, 3774–3781.
- Kersten, H.M.P., & Bethlehem, J.G. (1984). Exploring and reducing the nonresponse bias by asking the basic question. *The Statistical Journal of the United Nations Commission for Europe*, 2, 369–380.
- Kickul, J.R., Neuman, G., Parker, C., & Finkl, J. (2001). Settling the score: The role of organisational justice in the relationship between psychological contract breach and anticitizenship behavior. *Employee Responsibilities and Rights Journal*, 13, 77–93.
- Korte, K., & Dagevos, J. (2011). *Survey Integratie Minderheden (SIM)* [Survey Integration Minorities]. Den Haag, the Netherlands: Sociaal Cultureel Planbureau. Retrieved from http://www.scp.nl/Onderzoek/Bronnen/Beknopte_onderzoeksbeschrijvingen/Survey_integratie_minderheden_SIM
- Kraan, T., Van den Brakel, J., Buelens, B., & Huys, H. (2009). *Social desirability bias, response order effects and selection effects in the new Dutch Safety Monitor*. Paper presented at the Federal Committee on Statistical Methodology Research Conference, Washington, DC.
- Kreuter, F. (2013). *Improving surveys with paradata: Analytic uses of process information*. Hoboken, NJ: Wiley.
- Kreuter, F., Olsen, K., Wagner, J., Yan, T., Ezzati-Rice, T.M., Casa-Cordero, C., ... Raghunathan, T.E. (2010). Using proxy measures and other correlates of survey outcomes to adjust for non-response: Examples from multiple surveys. *Journal of the Royal Statistical Society, Series A*, 173, 389–407.
- Lepkowski, J., Axinn, W., Kirgis, N., West, B., Kruger Ndiaye, S., Mosher, W., & Groves, R. (2010). Use of paradata in a responsive design framework to manage a field data collection. *National Survey of Family Growth Survey Methodology Working Papers*, 10– 012. Retrieved from <http://www.psc.isr.umich.edu/pubs/pdf/ng10-012.pdf>

Link, M.W., & Mokdad, A. (2005). Advance letters as a means of improving respondent cooperation in random digit dial studies. A multistate experiment. *Public Opinion Quarterly*, 69, 572–587.

Luiten, A. (2006). *Interviewen op zaterdag. Responskans en datakwaliteit* [Interviewing on Saturdays. Response probability and data quality]. Internal report, Statistics Netherlands.

Luiten, A. (2006). *Lifting response. The influence of fieldwork strategy and refusal avoidance training on response rates*. Paper presented at the 17th International Workshop on Household Survey Nonresponse, Omaha, NE.

Luiten, A. (2008). *Determinants of interviewers' rule following behaviour*. Paper presented at the 19th International Workshop on Household Survey Nonresponse, Ljubljana, Slovenia.

Luiten, A. (2009). *Responsverhogende maatregelen bij sociale statistieken. Methodenreeks, thema benaderingsstrategieën*. [Measures to heighten response in social statistics] Retrieved from <http://www.cbs.nl/NR/rdonlyres/8E7C1E1E-F6C8-4956-B993-3D58DE622B13/0/200905x37pub.pdf>

Luiten, A. (2011). *Response bias due to interviewer behaviour*. Paper presented at the LFS methodology workshop, Wiesbaden, Germany.

Luiten, A. (2012). *De relatie tussen schommelingen in EBB doelvariabelen en uitbating van het veldwerk 2011–2012* [The relation between fluctuations in LFS substantive variables and fieldwork operation 2011–2012]. Discussion Paper PPM-20120815-JRBS: Statistics Netherlands.

Luiten, A., Campanelli, P., Klaassen D., & Beukenhorst D. (2008). *Advance letters and the language and behaviour profile*. Paper presented at the 19th International Workshop on Household Survey Nonresponse, Ljubljana, Slovenia.

Luiten, A., & Cobben, F. (2010). *Predicting cooperation in survey research*. Paper presented at the 21st International Workshop in Household Survey Non-response, Nürnberg, Germany.

Luiten, A., & Schouten, B. (2013). Tailored fieldwork design to increase representative response. An experiment in the Survey of Consumer Satisfaction. *Journal of the Royal Statistical Society, series A*, 176, 169–191.

Luiten, A., Schouten, B., & Cobben, F. (2007). *Experiments in balancing response, representativeness and costs in a CATI survey*. Paper presented at the 18th International Workshop on Household Survey Nonresponse, Southampton, United Kingdom.

Lynn, P. (2003). PEDAKSI: Methodology for collecting data about survey non-respondents, *Quality and Quantity*, 37, 239–261.

Lynn, P. (2012). *Mixed-mode survey data collection (including web): where next?* Paper presented at the ESSnet workshop on mixed mode data collection for social surveys, Wiesbaden, Germany.

Lynn, P., Beerten, R., Laiho, J., & Martin, J. (2002). Towards standardisation of survey outcome categories and response rate calculations. *Research in Official Statistics*, 1, 61–84.

Lynn, P., Clarke, P., Martin, J., & Sturgis, P. (2002). The effects of extended interviewer efforts on nonresponse bias. In R.M. Groves, D.A. Dillman, A. Eltinge, & R.J.A. Little (Eds.), *Survey nonresponse* (pp. 135–149). New York, NY: Wiley.

Martin, J., & Beerten, R. (1999). *The effect of interviewer characteristics on survey response rates*. Paper presented at the International Conference on Survey Nonresponse, Portland, OR.

McFarlane Smith, J. (1972). *Interviewing in Marketing and Social Research*. London: Routledge.

Mehta, R., & Sivadas, E. (1995). Comparing response rates and response content in mail versus electronic mail surveys. *Journal of the Market Research Society*, 7, 429–40.

Millar, M., & Dillman, D. (2011). Improving response to web and mixed-mode surveys. *Public Opinion Quarterly*, 75, 249–269.

Mohl, C., & Laflamme, F. (2007). Research and responsive design options for survey data collection at Statistics Canada, *Proceedings of ASA Joint Statistical Meeting*, 2962–2968. Retrieved from <http://www.amstat.org/sections/srms/proceedings/y2007/Files/JSM2007-000421.pdf>

Morton-Williams, J. (1993). *Interviewer approaches*. Aldershot, United Kingdom: Dartmouth.

Nagelkerke, N.J.D. (1991). Miscellanea. A note on a general definition of the coefficient of determination, *Biometrika*, 78, 691 – 692.

Neubarth, W., Kaczmirek, L., & Hellwig, O. (2007). *Pure text vs. HTML in e-mail invitations*. Paper presented at the General Online Research Conference, Leipzig, Germany.

Nicolaas, G., & Lynn, P. (2002). Random-digit dialing in the UK: Viability revisited. *Journal of the Royal Statistical Society, Series A*, 165, 297–316.

Nicoletti, C., & Peracchi, F. (2005). Survey response and survey characteristics: Microlevel evidence from the European Community Household Panel. *Journal of the Royal Statistical Society, Series A*, 168, 763–781.

O'Muircheartaigh, C., & Campanelli, P. (1999). A multilevel exploration of the role of interviewers in survey non-response. *Journal of the Royal Statistical Society, Series A*, 162, 437–446.

Oksenberg, L., & Cannell, C. (1988). Effects of vocal characteristics on nonresponse. In: R.M. Groves, P.P. Biemer, L.E. Lyberg, J.T. Massey, W.L. Nichols & J. Waksberg (Eds.), *Telephone Survey Methodology* (pp. 257–273). New York, NY: Wiley.

Parks, R.B., Kennedy, J.M., & Hecht, L.F. (1994). A demographic analysis of the impact of presurvey letters on cooperation rates in urban neighborhoods. *Proceedings of the 62nd Annual Meeting of the American Association for Public Opinion Research*, Danvers, MA.

Parsons, J., Owens, S., & Skogan, W. (2002). Using advance letters in RDD surveys: Results of two experiments. *Survey Research*, 33, 1–2.

Peytchev, A., Riley, S., Rosen, J., Murphy, J., & Lindblad, M. (2010). Reduction of non-response bias in surveys through case prioritization. *Survey Research Methods*, 4, 21–29.

Pickery, J., & Loosveldt, G. (2002). A multilevel multinomial analysis of interviewer effects on various components of unit nonresponse. *Quality and Quantity*, 36, 427–437.

Pondman, L.M. (1998). *The influence of the interviewer on the refusal rate in telephone surveys* (Doctoral dissertation). Amsterdam, the Netherlands.

Purdon, S., Campanelli, P., & Sturgiss, P. (1999). Interviewers' calling strategies on face-to-face interview surveys. *Journal of Official Statistics*, *15*, 199–216.

Raakjaer Nielsen, J. (2003). An analytical framework for studying compliance and legitimacy in fisheries management. *Marine Policy*, *27*, 425–432.

Reese, S.D., Danielson, W.A., Schoemaker, P.J., Chang, T.K., & Hsu, H.L. (1986). Ethnicity of interviewer effects among Mexican-Americans and Anglos. *Public Opinion Quarterly*, *50*, 563–572.

Schaefer, D.R., & Dillman, D.A. (1998). Development of a standard e-mail methodology: Results of an experiment. *Public Opinion Quarterly*, *62*, 378–397.

Schmeets, H. (2005). Slecht bereikbare allochtonen, autochtone weigeraars [Hard-to-contact members of ethnic minorities, native Dutch refusers]. In H. Schmeets & R. Van der Bie (Eds.), *Enquêteonderzoek onder allochtonen. Problemen en oplossingen* [Surveys involving members of ethnic minorities. Problems and solutions] (pp. 81–87). Heerlen: Statistics Netherlands.

Schmeets, H. (2010). Increasing response rates and the consequences in the Dutch Parliamentary Election Study 2006. *Field Methods*, *22*, 391–412.

Schouten, B. (2007). *A follow-up of nonresponse in the Dutch Labour Force Survey*. Discussion paper, Statistics Netherlands.

Schouten, B. (2010). *Dynamic adaptive survey designs: From paradata to data*. Paper presented at the 21st Workshop on Household Survey Non-response, Nürnberg, Germany. Retrieved from <http://cbsh1sps/sites/TmoMeth/WaarPers/03%20Waarnemingsmethoden%20personen/2010/NR%20workshop%20paper%20-%20Schouten.pdf>

Schouten, B., & Cobben, F. (2012). *Non-representative survey response: to balance, to adjust or both?* Discussion paper, Statistics Netherlands.

Schouten, B., Cobben, F., & Bethlehem, J. (2009). Indicators for the representativeness of survey response, *Survey Methodology*, *35*, 101 – 113.

Schouten, B., Luiten, A., Loosveldt, G., Beullens, K., & Kleven, Ø. (2010). *Monitoring and changing data collection through R-indicators and partial R-indicators* (RISQ deliverable). Retrieved from www.risq-project.eu/papers

Schouten, B., Morren, M., Bethlehem, J., Shlomo, N., & Skinner, C. (2009). *How to use R-indicators?* (RISQ deliverable). Retrieved from www.risq-project.eu/papers

Schouten, B., & Shlomo, N. (2010). *Indicators for representative response*. Paper presented at the 5th European Conference on Quality in Official Statistics, Helsinki, Finland.

Scott, P., & Edwards, P. (2006). Personally addressed hand-signed letters increase questionnaire response: a meta-analysis of randomised controlled trails. *BMC Health Services Research*, 6, 111. Retrieved from www.biomedcentral.com/1472-6963/6/111

Shih, T-H., & Fan, X. (2007). Response rates and mode preferences in web-mail mixed-mode surveys: A meta-analysis. *International Journal of Internet Science*, 2, 59-82.

Shlomo, N., Skinner, C., Schouten, B., Bethlehem, J., & Zhang, L.C. (2009). *Statistical properties of representativity indicators* (RISQ deliverable). Retrieved from www.risq-project.eu/papers

Shlomo, N., Skinner, C., Schouten, B., Carolina, T., & Morren, M. (2009). *Partial indicators for representative response* (RISQ deliverable). Retrieved from www.risq-project.eu/papers

Singer, E. (2006). Introduction. Nonresponse bias in household surveys. *Public Opinion Quarterly*, 70, 637-645.

Smith, C.B. (1997). Casting the net: Surveying an internet population. *Journal of Computer Mediated Communication*, 3, 1.

Smith, T. W. (1983). The hidden 25 percent: An analysis of nonresponse on the 1980 General Social Survey. *The Public Opinion Quarterly*, 47, 386-404.

Snijkers, G., Hox, J.J., & De Leeuw, E.D. (1999). Interviewers' tactics for fighting survey nonresponse. *Journal of Official Statistics*, 15, 185-198.

Snijkers, G., & Morren, M. (2010). *Improving web and electronic questionnaires: the case of the audit trail*. Paper presented at the 5th European Conference on Quality in Official Statistics, Helsinki, Finland.

Starick, R., & Steel, J. (2012). *Does increased effort lead to a less representative response? Selected case studies from the Australian Bureau of Statistics*. Paper presented at the 6th European Conference on Quality in Official Statistics, Athens, Greece.

Statistics Netherlands (2011). *Media en ICT. Gebruik televisie, krant, PC en internet*. [Media and ICT. Use of television, paper, PC and internet]. Retrieved from <http://statline.cbs.nl/StatWeb/publication/?DM=SLNL&PA=70655ned&D1=76-77,90&D2=37-41,52&D3=a&HDR=T&STB=G1,G2&VW=T>

Steele, F.A. (2009). *Module 7: Multilevel models for binary responses (concepts)*. Centre for multilevel modelling. Retrieved from www.cmm.bris.ac.uk/lemma

Stoop, I., & Koch, A. (2005). *Survey efforts, survey climate and response rates in the ESS*. Paper presented at the 16th Workshop on Household Survey Nonresponse, Tallberg, Sweden.

Stoop, I. (2005). *The hunt for the last respondent (Doctoral dissertation)*. The Hague, the Netherlands: Sociaal Cultureel Planbureau.

Stoop, I., Billiet, J., Koch, A., & Fitzgerald, R. (2010). *Improving survey response. Lessons learned from the European Social Survey*. Chichester, United Kingdom: Wiley.

Sutton, A., & Hopkins Burke, K. (2006). *Online response patterns*. Paper presented at the General Online Research Conference, Bielefeld, Germany.

Thibaut, J., Friedland, N., & Walker, L. (1974). Compliance with rules: some social determinants. *Journal of Personality and Social Psychology*, 30, 792–801.

Thompson, M., & Heron, P. (2005). The difference a manager can make: organizational justice and knowledge worker commitment. *International Journal of Human Resource Management*, 16, 383–404.

Traugott, M.W., Groves R.M., & Lepkowski, J.M. (1987). Using dual frame designs to reduce nonresponse in telephone surveys. *Public Opinion Research*, 51, 522–539.

Tyler, T. (1990). *Why People obey the law*. New Haven, Yale University Press.

Tyler, T. (2006). Psychological perspectives on legitimacy and legitimation. *Annual Review of Psychology*, 5, 375–400.

Van Berkel, C., & Vosmer, M. (2006). *Benodigde adresgebonden uren (FtF)* [Necessary address-related hours (FtF)]. Internal report, Statistics Netherlands.

Van der Vaart, W., Ongena, Y., Hoogendoorn, A., & Dijkstra, W. (2006). Do interviewers' voice characteristics influence cooperation rates in telephone surveys? *International Journal of Public Opinion Research*, 18, 488–499.

Van Veen, P. (2004). *Contact and participation probabilities in CAPI and CATI surveys* (Master's thesis, University of Utrecht). The Hague: Statistics Netherlands.

Vosmer, M., & Engelen, F. (2008). Het herzien van de vragenlijst van het NKO 2006 [Revision of the questionnaire of the Dutch Parliamentary Election Study 2006]. In H. Schmeets, (Ed.), *Het Nationaal Kiezersonderzoek 2006. Opzet, uitvoering en resultaten* [The Dutch Parliamentary Election Study 2006. Purpose, execution and results] (pp. 53–61). Internal report, Statistics Netherlands.

Wagner, J. (2008). *Adaptive survey design to reduce non-response bias* (Doctoral dissertation). Michigan, MI.

Wang, K., Murphy, J., Baxter, R., & Aldworth, J. (2005). *Are two feet in the door better than one? Using process data to examine interviewer effort and nonresponse bias*. Presented at the Federal Committee on Statistical Methodology Research Conference, Washington DC. Retrieved from <http://www.fcsm.gov/05papers/WangAldworthetalVIB.pdf>

West, B.T. (2013). An examination of the quality and utility of interviewer observations in the National Survey of Family Growth. *Journal of the Royal Statistical Society, Series A*, 176, 211–227.

West, B.T., & Olson, K. (2010). How much of interviewer variance is really nonresponse error variance? *Public Opinion Quarterly*, 74, 1004–1026.

Westat (2004). *Blaise CATI guide*, Rockville, MD: Autor.

Wetzels, W., Janssen, B., & Schooten, M. (2007). *Mixed-mode enquêtering bij personen en huishoudens 2005–2007: Een overzicht* [Mixed-mode surveying of persons and households 2005–2007: An overview]. Internal report, Statistics Netherlands.

Wetzels, W., Schmeets, J.J.G., Van den Brakel, J., & Feskens, R.C.W. (2008). Impact of prepaid incentives in face-to-face surveys: A large-scale experiment with postage stamps. *International Journal of Public Opinion Research*, 20, 507–516.

Wetzels, W., Schmeets, H., Feskens, R., & Van den Brakel, J. (2008). Het gebruik van attenties in enquêteonderzoek [The use of incentives in surveys]. In H. Schmeets, (Ed.), *Het Nationaal Kiesersonderzoek 2006. Opzet, uitvoering en resultaten* [The Dutch Parliamentary Election Study 2006. Design, execution and results] (pp. 63–76). Heerlen: Statistics Netherlands.

White, A., Martin, J., Benneth, N., & Freeth, S. (1997). Improving advance letters for major government surveys. In A. Koch, & R. Porst (Eds.), *Nonresponse in survey research* (pp. 151–171). Mannheim, Germany: ZUMA.

Winter, S.C., & May, P.J. (2001). Motivation for compliance with environmental regulations. *Journal of Political Analysis and Management*, 20, 675–698.

Worthen, B.R., & Valcarce, R.W. (1985). Relative effectiveness of personalized and form covering letters in initial and follow-up mail surveys. *Psychological Reports*, 57, 735–744.

Yammarino, F.J., Skinner, S.J., & Childers, T.L. (1991). Understanding mail survey response behavior. A meta-analysis. *Public Opinion Quarterly*, 55, 613–639.

Yu, J., & Cooper, H. (1983). A quantitative review of research design effects on response rates to questionnaires. *Journal of Marketing Research*, 20, 36–44.

Appendix 1

Advance letter used in the SCS experiment



Centraal Bureau voor de Statistiek

<i>Controlgroup</i>	<i>Personalized letter Example 1</i>	<i>Personalized letter Example 2</i>
The occupants of	Mrs. E.J. Jansen Street, nr Postal code, City	Mr. D.C. Pietersen and Mrs. E.J. Jansen Street, nr Postal code, City
Dear sir/madam,	Dear madam,	Dear sir, dear madam

It is my pleasure to invite you to participate in an important study of the Central Bureau of Statistics (CBS): the Labour Force Survey. This CBS research is an indispensable source of figures on labour, care tasks, education, unemployment, disability, and (pre) pensioning.

Each month, CBS draws a sample of about 5000 addresses of all Dutch addresses. This time, your address was selected. For the quality of CBS statistics, it is very important that as many selected people as possible participate in this study. So, it is very important that you should participate. You represent many other inhabitants in the Netherlands.

Shortly, a CBS employee will visit you and ask your cooperation. Our employee will ask questions about all members of your household that are 14 years of age, or older. The duration of the interview depends on the number of household members: per member we ask for about 10 à 15 minutes of your time

In all our studies, your privacy is guaranteed absolutely. Statistics are made by combining your information with that of all other participants, and with figures from other sources. On the back of this letter you can read more about that.

Should you have questions about this letter, or about CBS in general, our employee will be happy to answer these. You can also refer to our website: www.cbs.nl. It is

also possible to contact the CBS contact centre in Heerlen: (045) 570 73 74. The contact centre is available from Monday to Friday between 8.00 and 17.00 hours.

I would like to thank you for your cooperation.

Yours sincerely,
Director of Fieldwork,

(printed autograph)

(name)

On the back of the letter is printed:

In all our studies, your privacy is guaranteed. This is an obligation of CBS, that is put down in special legislation. To secure your data, CBS took a number of measures: CBS employees have to take an oath of secrecy, that, when breached will lead to legal proceedings. People's answers are separated from information about their names and addresses as soon as possible. Computations are made on extremely well secured computers, that are impossible to access for unauthorized personnel. The law guarantees that your data will only be used for making statistics. No institution can demand access to data that CBS collects. In CBS publications, personal information is never recognizable.

CBS does not only collect information itself, but receives data from other institutions. For example, the information in de Communal Registries, the centres for Work and Income, Social services, and the salary administration of a great number of companies. We automatically combine the information you provide, with information we receive from other sources. With this combined information, CBS makes statistics about the Dutch society. This allows us to work as economically as possible.

Appendix 2

Unconditional and conditional R-indicators for historic SCS data

	Unconditional R-indicators			Conditional R-indicators		
	contact	cooperation	response	contact	cooperation	response
Age	31	62	47	18	35	17
< 30	-23	3	-8	5	12	0
30-44	-13	29	16	13	27	2
45-64	8	18	22	5	29	12
65>	14	-51	-38	8	58	14
Sex	33	53	66	7	1	7
Male(s)	-28	-16	-35	3	0	3
Mixed	17	30	38	0	0	0
Female(s)	-8	-40	-41	2	0	2
Household composition	35	61	75	10	13	19
Single	-17	-40	-52	1	2	5
Partners, with children	9	12	15	4	2	5
Partners, no children	14	32	40	3	1	5
Single parent	-7	-1	-2	1	6	12
Ethnic group	27	24	31	8	9	14
Native Dutch	25	17	19	2	1	4
Foreign	-11	-14	-23	5	7	16
Mixed	1	10	8	1	0	1
Income in quartiles	11	71	64	6	30	25
<1600	5	-55	-47	1	51	33
1600-1900	4	9	3	0	17	9
1900-2300	9	39	38	0	19	18
2300>	2	22	20	2	2	1
Urban density	37	28	31	9	9	10
Very strongly urban	-14	-8	-18	4	0	4
Strongly urban	3	3	4	1	3	4
Medium urban density	3	5	6	0	0	0
Low urban density	3	6	6	0	0	0
No urban density	7	-5	2	1	4	1
No information available	-33	-25	-23	0	0	0
% non-western foreigners in area	12	16	18	2	10	8
Less than 5%	7	-1	4	0	1	1
5-10%	-4	4	4	0	1	1
10-20%	-5	7	2	0	3	2
20% and more	-5	-2	-6	0	0	0
No information available	-6	-14	-16	0	6	3

(1) Unconditional and conditional R-indicators * 1000.

Appendix 3

Interviewer questionnaire

Question	Answer Categories
Background information	
How long have you been working as an Statistics Netherlands (SN) interviewer?	< 1 yr; 1-2 yr; 2-5 yr; 5-10 yr; >10 yr
How many hours per week do you work as SN interviewer?	Open
Do you have other paid work?	Yes + nr of hours; No
Highest education?	8 categories
Work pressure	
Do you feel you put in more hours each week than you should according to your contract?	No; yes + nr of hours
How often do you work evenings (after 17h) per week?	Never, once, 2x; 3x; 4x, more than 4x
→ Why not more often?	Open
→ Why so often?	Open
If you work evenings, how many hours per evening?	<=1; 1-2; 2-3; 3-4; other
If it were up to you, when would you work evenings?	All the time; right at 1 st visit; after failed day call; after two failed day calls; after 3 or 4 failed day calls; never
How many times a month do you work Saturdays?	Never; 1x pm; 2x pm; 3x pm; 4x pm
→ Why not more often?	Open
→ Why so often?	Open
If it were up to you, when would you work Saturdays?	Never; right at 1 st visit; after failed weekday call; after 2 failed weekday calls; after 3 failed weekday calls; as much as possible
How do you spread your visits over weekdays?	Evenly; I prefer certain days; other
Do you feel your phone should be on at all times for respondents?	Yes; no
How often do respondents call you at inconvenient times?	Very often; often; sometimes; never
Do you feel your phone should be on at all times for your regional manager or the help desk?	Yes; no
If a respondent leaves a message on your voicemail, when do you call back?	As soon as possible, even when I'm not working; only during working hours; other
Apart from Sundays, do you structurally plan days or day parts when you do not work?	Always; usually; sometimes; never
Do you do interviewer work on non-work days (more answers possible)?	No; I am reachable by phone for respondents; I do my administration; I plan new calls; I call back respondents; other
How do you evaluate the travelling you have to do?	Very positive; fairly positive; neutral; fairly negative; very negative
Your regional manager sets you targets, among others for your response. How do you evaluate working with targets?	Very positive; fairly positive; neutral; fairly negative; very negative
How hard does your regional manager uphold these targets?	Very hard; fairly hard; hardly; not
How do you evaluate the weight of your laptop?	Not oppressive; fairly oppressive; very oppressive
How do you evaluate the work pressure of your job as interviewer?	Very high; high; not high / not low; low; very low
Social norms	
Does information on the achievements and way of work of your colleagues influence your behaviour?	Yes, very much; yes, somewhat; no

Interviewer questionnaire (continued)

Question	Answer Categories
<p>To what extent do you agree with the following:</p> <p>- I follow the FWS because the other interviewers in my region remark upon it if I don't?</p>	<p>Absolutely agree; agree; neither agree nor disagree; disagree; totally disagree</p>
<p>Personal norms</p> <p>To what extent do you agree:</p> <p>- I follow FWS because I always want to do my work as best I can?</p>	<p>Absolutely agree; agree; neither agree nor disagree; disagree; totally disagree</p>
<p>Distributional legitimacy</p> <p>How do you evaluate your salary?</p> <p>Exceptional interviewers may get a bonus. How do you judge the distribution of bonuses?</p> <p>→ Why unfair?</p>	<p>Good; fair; fairly low; clearly too low</p> <p>Just; not just not unfair; unfair; no opinion</p>
<p>When does an interviewer deserve a bonus?</p> <p>Have you ever received a bonus?</p>	<p>Open</p> <p>Yes; no; other</p>
<p>What is a difficult neighbourhood for you (more answers possible)?</p>	<p>A neighbourhood with many ethnic minorities; ...with many elderly; ...with many double income earners; ...with a low SES; other</p>
<p>How often do you have to work in a difficult neighbourhood?</p>	<p>(Almost) always; sometimes; never</p>
<p>Is the distribution of difficult addresses fair in your region?</p>	<p>Yes; no</p>
<p>Procedural legitimacy</p>	
<p>Have you been involved in designing the present fieldwork strategy (for interviewers working 7 years or more)?</p>	<p>Very much; somewhat; hardly; not</p>
<p>Where you satisfied with the level of your involvement?</p>	<p>Would have liked more involvement; involvement was just right; too much involvement</p>
<p>SN is thinking about a new fieldwork strategy. Do you feel involved in this?</p>	<p>Very much involved; somewhat involved; hardly; not; new fieldwork strategy? First time I hear about it.</p>
<p>Do you feel it is important that you are involved?</p>	<p>Yes, very important; somewhat important; no, not important</p>
<p>Interactional justice; transparency of the rules</p>	
<p>In general, how often attention is given to the fieldwork strategy by the regional manager?</p>	<p>1 x pm; 1x a quarter; 1 x every 6 months; 1 a year; only during basic training; not at all</p>
<p>In general, how often attention is given to the fieldwork strategy by SN?</p>	<p>1 x pm; 1x a quarter; 1 x every 6 months; 1 a year; only during basic training; not at all</p>
<p>In general, how often attention is given to the advantages of the FWS for SN by the RM?</p>	<p>1 x pm; 1x a quarter; 1 x every 6 months; 1 a year; only during basic training; not at all</p>
<p>In general, how often attention is given to the advantages of the FWS for SN by SN?</p>	<p>1 x pm; 1x a quarter; 1 x every 6 months; 1 a year; only during basic training; not at all</p>
<p>How often ...attention is given to the advantage of the FWS for interviewers by the RM?</p>	<p>1 x pm; 1x a quarter; 1 x every 6 months; 1 a year; only during basic training; not at all</p>
<p>How often ...attention is given to the advantage of the FWS for interviewers by SN?</p>	<p>1 x pm; 1x a quarter; 1 x every 6 months; 1 a year; only during basic training; not at all</p>
<p>How often... attention is given to the possible disadvantages of the FWS for interviewers by the RM?</p>	<p>1 x pm; 1x a quarter; 1 x every 6 months; 1 a year; only during basic training; not at all</p>

Interviewer questionnaire (continued)

Question	Answer Categories
How often... attention is given to the possible disadvantages of the FWS for interviewers by the SN?	1 x pm; 1x a quarter; 1 x every 6 months; 1 a year; only during basic training; not at all
How did you learn about the FWS	check all that applies
- by basic training	
- by the RM	
- by written instruction	
- by e-learning	
- in the monthly meetings	
- by colleague coaches?	
Do you interact with other interviewers outside of work? (filter)	Yes, often; yes sometimes; no
How do you evaluate the level of contact with your interviewer colleagues?	Too much contact; level is just right; too little contact
→ what kind of contact would you like with you colleagues? (misc. info)	Talk about the work; social contact; other
How do you evaluate the level of contact with your regional manager?	Too much; just right; too little
Interactional justice; knowledge of the rules	
To what extent do you think you master the FWS?	Well; fairly well; a bit; not at all
How many days after sending the advance letter should you visit the address?	3-5 days; 6-7 days; 8-9 days; 10-11 days; other
In which period should you visit an address for the first time?	Within the 1 st week of the fieldwork period; within the first two weeks of fieldwork period; within the first 3 weeks of fieldwork period.
How often should an address be visited in case of noncontact?	Maximally 6 times; 6 times; minimally 6 times; other
After which visit telephone is contact allowed?	After the 1 st ; the 2 nd ; the 3 rd ; the 4 th ; other
One of the first two visits should be in a preferential time slot. Which visit?	The 1 st ; the 2 nd ; the 1 st or the 2 nd ; the 1 st and the 2 nd .
What is the definition of preferential time slot?	Open
Visits should be well spread. Which of these visits are 'well spread':	All questions:
- Monday 9.30 - Monday 11.30	Yes; no
- Monday 9.30 - Monday 14.00	
- Monday 14.00 - Monday 18.00	
- Monday 18.00 - Monday 20.00	
- Monday 9.30 - Tuesday 14.00	
- Monday 9.30 - Tuesday 18.00	
- Monday 18.00 - Tuesday 18.00?	
What are reasons you do not visit six times in case of non-contact?	Check all that apply
- NA, I always visit 6 times	
- When others tell me the inhabitants are not present during fieldwork period	
- When I suspect that the house is uninhabited	
- When the address is far away	
- When I suspect it will end in refusal anyway	
- When the address is unreachable (porter / intercom/ gate)	

Interviewer questionnaire (continued)

Question	Answer Categories
<ul style="list-style-type: none"> - When I made an appointment at the end of the month that is broken - other, namely <p>We sometimes notice that interviewers do not always have the same notion of the concept of non-contact. Which of the following situations would you consider to be non-contacts?</p> <ul style="list-style-type: none"> - daughter of sample unit tells me that SU is absent during the entire fieldwork period - the neighbours inform me the household is absent during the entire fieldwork period - I visit in vain after an appointment - I cannot reach the address because of a porter - I cannot find the address - the overflowing mailbox clearly tells me that the house is empty - the sample unit tells me at the 1st or 2nd visit to come back, because they do not have time / are ill / have a death in the family. 	All questions: yes ; no
<p>Legitimacy of rule</p> <p>According to FWS the first or second visit should be at primetime. Do you agree with this rule?</p> <p>How do you evaluate the influence of evening calls on the following aspects:</p> <ul style="list-style-type: none"> - % contacts - % refusals - travelling time - travelling costs - number of necessary visits - private life 	<p>Absolutely agree; agree; neither agree nor disagree; disagree; totally disagree</p> <p>All questions:</p> <p>Very positive; fairly positive; neutral; somewhat negative; very negative</p>
<p>How do you evaluate the influence of Saturday calls on the following aspects:</p> <p>(see above) .</p>	<p>Very positive; fairly positive; neutral; somewhat negative; very negative</p>
<p>According to FWS you should leave behind an appropriate calling card after each visit without contact. Do you agree with this rule?</p>	<p>Absolutely agree; agree; neither agree nor disagree; disagree; totally disagree</p>
<p>How do you evaluate the influence of leaving calling cards on the following aspects:</p> <ul style="list-style-type: none"> - % contacts - % refusals - travelling time - travelling costs - number of necessary visits - private life 	<p>All questions:</p> <p>Very positive; fairly positive; neutral; somewhat negative; very negative</p>
<p>According to FWS, visits should be spread over days and times; Do you agree with this rule?</p>	<p>Absolutely agree; agree; neither agree nor disagree; disagree; totally disagree</p>
<p>How do you evaluate the influence of leaving calling cards on the following aspects: (see above)</p>	All questions:

Interviewer questionnaire (continued)

Question	Answer Categories
<p>According to FWS, non-contacts should be visited six times. Do you agree with this rule?</p> <p>How do you evaluate the influence of visiting six times on the following aspects: (see above)</p> <p>To what extent do you agree with the following:</p> <ul style="list-style-type: none"> - I follow the FWS because it gets me better results - ...because less visits are necessary as a result - The FWS was designed to get the highest response in the fastest and most efficient way. To what extent do you feel that the FWS succeeds in this? 	<p>Very positive; fairly positive; neutral; somewhat negative; very negative</p> <p>Absolutely agree; agree; neither agree nor disagree; disagree; totally disagree</p> <p>Very positive; fairly positive; neutral; somewhat negative; very negative</p> <p>Absolutely agree; agree; neither agree nor disagree; disagree; totally disagree</p>
<p>Traditional authority</p>	
<p>According to FWS the first or second visit should be at primetime. To what extent do you feel obliged to follow this rule.</p>	<p>Not at all; a bit; very much</p>
<p>According to FWS you should leave behind an appropriate calling card after each visit without contact. To what extent do you feel obliged to follow this rule.</p>	<p>Not at all; a bit; very much</p>
<p>According to FWS, visits should be spread over days and times. To what extent do you feel obliged to follow this rule.</p>	<p>Not at all; a bit; very much</p>
<p>According to FWS, non-contacts should be visited six times. To what extent do you feel obliged to follow this rule?</p>	<p>Not at all; a bit; very much</p>
<p>To what extent do you agree with the following:</p> <ul style="list-style-type: none"> - I follow the FWS because transgressions are easily spotted. - I follow the FWS because SN wants me to. - ... my RM finds it important. - ...that's what I was taught during basic training - ...I want to avoid a bad evaluation 	<p>Absolutely agree; agree; neither agree nor disagree; disagree; totally disagree</p>
<p>Miscellaneous</p>	
<p>What aspects could you ameliorate? (check all that apply)</p> <ul style="list-style-type: none"> - planning - less non-contacts - administration - gaining cooperation - less leading questions and better follow up questions - knowledge of surveys - handling the laptop - communicating with my regional manager - other 	
<p>How often do you discuss the following subjects with other interviewers:</p> <ul style="list-style-type: none"> - pleasant and unpleasant experiences - your non-contact percentage 	<p>All questions:</p> <p>Often; sometimes; rarely/never</p>

Interviewer questionnaire (continued)

Question	Answer Categories
<ul style="list-style-type: none"> - your percentage response - handling refusals - planning of calls and appointments - achieving your target - work pressure - following fieldwork strategy - other? 	
<p>How often do you discuss the following subjects with your regional manager:</p> <ul style="list-style-type: none"> - pleasant and unpleasant experiences - your non-contact percentage - your percentage response - handling refusals - planning of calls and appointments - achieving your target - work pressure - following fieldwork strategy - other? 	<p>All questions: Often; sometimes; rarely/never</p>
<p>What aspects could your regional manager ameliorate in herself?</p>	<p>Open</p>
<p>What aspects could SN ameliorate in communicating with interviewers?</p>	<p>Open</p>
<p>How do you evaluate the first calling card?</p>	<p>Positive; neutral; negative</p>
<p>Would you change something about the card or its use?</p>	<p>No; I don't want to use it; I want to put info about myself on it; other</p>
<p>How do you evaluate the 2nd calling card?</p>	<p>Positive; neutral; negative</p>
<p>Would you change something about the card or its use?</p>	<p>No; I want to use it sooner; I don't want to use it; I want to put info about myself on it; other</p>
<p>How do you evaluate the 3rd calling card?</p>	<p>Positive; neutral; negative</p>
<p>Would you change something about the card or its use?</p>	<p>No; I want to use it sooner; I don't want to use it; I want to put info about myself on it; other</p>
<p>Do you succeed in practice in spreading your visits over times and days?</p>	<p>(Almost) always; mostly; sometimes; mostly not; (almost) never.</p>
<p>What could be reasons not to spread?</p>	<p>Planning of well spread visits is hard; If I am around, I pass by, even if the visit is on the same day or time as a previous visit; planning is influenced heavily by other addresses.</p>
<p>Would a planning tool be helpful?</p>	<p>Yes, very much; yes, somewhat; no</p>
<p>What is the most important thing you strive for as an interviewer?</p>	<p>The highest possible contact; lowest possible refusal; highest possible cost efficiency; other</p>
<p>In which percentage of cases do you think you apply the following:</p> <ul style="list-style-type: none"> - spreading over days - spreading over time periods - spreading over the fieldwork period - visiting in the evening in one of the first two visits - visiting in the weekend in one of the first two visits - using the 1st card after the first visit with no contact - using the 2nd card after the 2nd visit with no contact - using the 3rd card after the 3rd visit with no contact 	

Interviewer questionnaire (end)

Question	Answer Categories
<ul style="list-style-type: none"> - using the 3rd card at another time than after the 3rd visit - trying to make a telephone contact after 3 visits - visiting six times in case of non-contact - make an appointment at first contact 	
<p>Some interviewers withdraw when they feel a refusal coming, and try again in a following visit. Do you use this technique?</p>	<p>Yes, some times; yes, regularly; yes, often; no</p>
<p>→ Why do you not use this technique</p>	<p>Open</p>
<p>→ in what percentage of cases do you think trying again leads to response?</p>	<p>Open</p>
<p>→ Under what circumstances does trying again lead to success?</p>	<p>When somebody else answers the door; when the timing is better; practically always; you can not foresee it; other</p>
<p>Do you ever go back when a refusal is actually uttered?</p>	<p>Yes, some times; yes, regularly; yes, often; no</p>
<p>→ in what percentage of cases does this lead to response?</p>	<p>Open</p>
<p>Does it happen that you pay more visits than you account?</p>	<p>Yes, often; yes, sometimes; no seldom or never</p>
<p>→ How often does it happen</p>	<p>With most addresses; a fair number of addresses per month; one or two addresses per month, not every month, but a couple of times a year; other</p>
<p>→ Why do you not account for all visits?</p>	<p>If I happen to drive past, I don't count that as a visit; Six visits is not enough; I cannot remember all visits; it is cumbersome to open the admin part of the questionnaire all the time; other</p>
<p>Does it happen that you try to contact the household by phone more often than you account?</p>	<p>Yes, often; yes, sometimes; no seldom or never</p>
<p>→ How often does it happen?</p>	<p>With most addresses; a fair number of addresses per month; one or two addresses per month, not every month, but a couple of times a year; other</p>
<p>→ Why do you not account for all telephone calls?</p>	<p>The possible number to account for is too small; I cannot recall all calls; it is cumbersome to open the admin part all the time, other. (Almost) 100% accurate; 90-100%; 80-90%; less than 80%.</p>
<p>How accurately do you think you fill in the admin block in general?</p>	<p>Open</p>
<p>Are you able to account all relevant information in the admin block? (if not, what not)</p>	<p>Open</p>
<p>End of the questionnaire. Any other remarks?</p>	<p>Open</p>