KU LEUVEN

FACULTEIT SOCIALE WETENSCHAPPEN

The use of paradata to assess survey representativity

Cracks in the nonresponse paradigm

Promotor: Prof. Dr. Geert Loosveldt
Onderzoekseenheid:
Centrum voor Sociologisch Onderzoek [CeSO]

Proefschrift tot het verkrijgen
van de graad van
Doctor in de Sociale Wetenschappen
aangeboden door
Koen BEULLENS

2013

KU LEUVEN

FACULTEIT SOCIALE WETENSCHAPPEN

The use of paradata to assess survey representativity

Cracks in the nonresponse paradigm

Koen BEULLENS

Proefschrift tot het verkrijgen van de graad van
Doctor in de Sociale Wetenschappen

Nr. 245

Samenstelling van de examencommissie:

Prof. Dr. Jan Van den Bulck (voorzitter)
Prof. Dr. Geert Loosveldt (promotor)
Prof. Dr. Jelke Bethlehem [Universiteit van Amsterdam, NL]
Prof. Em. Dr. Jaak Billiet
Prof. Dr. Dirk Heerwegh
Dr. Ineke Stoop [Sociaal en Cultureel Planbureau, NL]

2013

Gepubliceerd door:
Faculteit Sociale Wetenschappen - Onderzoekseenheid: Centrum voor Sociologisch Onderzoek [CeSO],
KU Leuven, Parkstraat 45 bus 3601 - 3000 Leuven, België.

D/2013/8978/15

Symbols

r_i	Response outcome for sample unit i, where $r \in \{0, 1\}$
ρ_i	Response propensity for sample unit i, where $0 < \rho < 1$
n	Total sample size (respondents and nonrespondents)
N	Population total
n_r	Total number of respondents in the sample
n_{nr}	Total number of nonrespondents in the sample
$\bar{r}$	Response rate under the fixed response model
$\bar{\rho}$	Response rate under the random response model
$\bar{\rho}_1, \bar{\rho}_2, \ldots$	Response rate after first, second, ... contact attempt
y	Target survey variable, only available among respondents
aux	Auxiliary variable, available for both respondents and nonrespondents
$a\bar{u}x_f$	Mean for auxiliary variable aux for the full sample
$a\bar{u}x_r$	Mean for auxiliary variable aux for respondents only
$a\bar{u}x_{nr}$	Mean for auxiliary variable aux for nonrespondents only
$\bar{y}_f$	Mean for target variable y for the full sample
$\bar{y}_r$	Mean for target variable y for respondents only
$\bar{y}_{nr}$	Unobservable mean for target variable y, nonrespondents only
w_i	Weight score for sample unit i, usually $w_i = 1/\rho_i$ or $w_i = \bar{\rho}/\rho_i$
$\bar{y}_{unw}$	Unweighted mean for target variable y, respondents-only ($=\bar{y}_r$)
$\bar{y}_w$	Weighted mean for target variable y, respondents-only
p	Number of available auxiliary variables
ψ	Variance inflation factor under the fixed response model
F	Variance inflation factor under the random response model
k_i	Number of contact attempts for unit i
E	Total number of contact attempts during fieldwork
C	Number of possible nonresponse outcome categories
H	Number of categories of auxiliary variable ($h = 1, 2, \ldots, H$)
c_{fix}	Fixed cost of a survey
c_{var}	Variable survey costs (cost per completed interviewer)
$\bar{c}_{eff}$	Average cost per effective sample unit
λ	Re-selection probability (constant in blind fieldwork strategy)

Acronyms

ESS	European Social Survey
ESS1 . . . 6	First round of the ESS (2002) up to the sixth round (2012)
ESS3-BE	Belgian part of the third round (2006) of the ESS
FHS	Flemish Housing Survey
GPS	General Population Survey
CST	Core Scientific Team of the ESS
SSD	Social Statistical Database
BE	Belgium
BG	Bulgaria
CH	Switzerland
CY	Cyprus
DK	Denmark
EE	Estonia
ES	Spain
FI	Finland
FR	France
GR	Greece
HR	Croatia
HU	Hungary
IL	Israel
NL	The Netherlands
NO	Norway
PL	Poland
PT	Portugal
RU	Russian Federation
SE	Sweden
SI	Slovenia
SK	Slovakia
UA	Ukraine
UK	United Kingdom

List of Tables

List of Figures

Contents

Acknowledgement

Ofschoon deze dissertatie grotendeels tot stand is gekomen door overleg tussen mezelf en mijn computer, zijn er toch heel wat mensen die rechtstreeks of onrechtstreeks een belangrijke bijdrage hebben gehad. Vooreerst zou ik graag mijn promotor Geert Loosveldt willen bedanken voor de goede begeleiding, de feedback en vooral voor de vrolijkheid. Daarnaast mag ik van het geluk getuigen zeer leuke collega's te hebben gehad, zowel voor inhoudelijke discussies alsook voor louter cammeraadschappelijke belangen. Vandaar een welgemeende 'dank u' aan Jorre, Eva, Sara, Sanne, Joke, Jeroen, Hideko, Dries, Leen, Nathalie, Michaël, Bart, Victor, Marie-Sophie, Veronique, Ahu, Koen, Dmitriy, Marc, Jos, Christien, Marina, Sofie, Martine, Elke, en last but not least, Jaak.

Graag wil ik bij deze ook Silvie, mijn familie en vrienden bedanken voor de ondersteuning en het geloof om hierin te slagen.

Introduction

In order to form an accurate picture of society, the economy, the labor market, health situations or housing conditions, a sample is drawn from the target population and all selected individuals are asked questions. If everything goes as planned, a list of all members of the target population is available from which a random sample can be drawn. Each selected household or individual participates and provides appropriate answers. Unfortunately, errors may occur in many stages of the data production process, violating the necessary conditions for unbiased inference. These potential sources of adverse influence have been categorized into four distinct types of survey error: coverage error, sampling error, nonresponse error and measurement error (see, among others, Biemer & Lyberg, 2003; Biemer, 2010; Groves, 2004; Groves & Lyberg, 2010). The first three types relate to the selection of sample units or observation units. Measurement error is associated with the quality of the recorded answers after (successful) selection.

In recent years, nonresponse error has received more attention because of an increasing inability to contact sample units or the growing unwillingness of these sample units to participate (Singer, 2006). Survey researchers seem to agree that there is an international trend of declining response rates (Atrostic, Bates, Burt, & Silberstein, 2001; de Leeuw & de Heer, 2002; Rogers, Murtaugh, Edwards, & Slattery, 2004; Curtin, Presser, & Singer, 2005; Brick & Williams, 2013). This issue of increasing nonresponse has fostered the awareness among researchers that nonresponse can contaminate survey outcomes. During the last decade, many researchers have devoted more time and effort to assess the errors in survey estimates

and have outlined possible strategies to combat nonresponse. Despite increased efforts in the field, response rates have not really improved.

Whether or not conclusive evidence is available to confirm a continuous decline in response rates, it is difficult to find household surveys that reach a response rate of, for example, more than 90%. In the European Social Survey, which serves as an empirical reference throughout this dissertation, response rates are usually in the range of 45% to 75% (Stoop, Billiet, Koch, & Fitzgerald, 2010). In some of the participating countries, extensive fieldwork efforts have been made, aimed at achieving an acceptable level of response (for example 70%), but unfortunately these have failed, despite a substantive refusal conversion program or low noncontact rates. In countries such as France, Germany, Belgium, the Netherlands, and Switzerland, researchers are experiencing serious difficulties in finding a satisfactory number of respondents for surveys.

Nonresponse in surveys not only has a detrimental effect on the data volume that can be used for analyses, but also survey researchers acknowledge the fact that nonresponse may have a systematic nature: particular groups or individuals may be more likely to participate than others. Whenever variables or parameters of interest are related to this responsiveness (response propensity), bias may arise. As surveys usually have an observational (as opposed to experimental) status, the relationship between any two survey variables is not in anyone's control, implying that the independence of two survey variables is somewhat exceptional. If one is willing to accept that survey participation is one such variable, many substantive and meaningful associations may be expected between survey response and other target survey variables. In other words, survey bias because of nonresponse may be the rule rather than the exception.

In this regard, Little and Rubin (1987, 2002) describe three different scenarios, applicable to a whole range of problems due to missing data, such as unit nonresponse, attrition, or item nonresponse. First, Missing Completely At Random (MCAR) applies to a situation where missing data is merely a result of a stochastic process: there are no (hidden) structures whatsoever in the data that relate to the absence of information. In such a

case, missing data or nonresponse only affects the statistical power of the dataset: less data means higher standard errors. A simple (though possibly expensive) solution is to increase the sample size. In fact, this MCAR scenario assumes independence between a survey variable and (non)response which is, given the observational nature of survey data, somewhat unlikely. Second, the Missing At Random (MAR) mechanism assumes that nonresponse is systematic but known. One may think of some age groups that are more/less inclined to participate. In such a case, missing information will not only lower the power of the dataset, but may also induce bias. Provided that the age group is the only variable that contributes to the differences in response propensities, solving the problem of bias can be achieved by adjustments such as weighting. Reshuffling the field efforts between groups so that they all attain the same response rate is another option (also termed 'balanced response rates'). The assumption of MAR is probably also very fragile. Mostly, (social) behavior such as survey (non)response is multicausal, in practice insufficiently explained by one or only a few variables. The least favorable, but most realistic scenario is the third one. Missing Not At Random (MNAR) implies that missing data or nonresponse is systematic and (partially) unknown. Satisfactory solutions to deal with this scenario are not obvious: increasing the sample size does not alter the bias, whereas weighting might only partially solve the problem.

There are good reasons for accepting that MNAR is probably the best fitting assumption. Nonresponse in household surveys can be separated into diverse elements such as noncontacts, refusals, language barriers, and illness and these elements may all relate to the diversity of underlying patterns in everyday life. Noncontacts are probably busy people, with stressful jobs or many household tasks. Hard refusals may be thought of as related to grumpy individuals, who reject every kind of social engagement. Illness or language barriers not only affect the odds of participating in the survey, but are also reasons why individuals may answer survey questions differently. For this reason, some research has been carried out, examining the relationship between survey response and survey target variables:

personality traits (see, e.g., Saßenroth, 2010), the work-family balance and the use of time (see, e.g., Vercruyssen, Roose, & Van de Putte, 2011; Vercruyssen, Van de Putte, & Stoop, 2011; Maitland & Bianchi, 2006), social-economic background (see, e.g., Groves & Couper, 1998), housing or neighborhood conditions, social integration, and so forth. As knowledge about individual lives is particularly the (main) goal of survey research, the risk of nonresponse bias is a permanent and perhaps very complex threat.

The first chapter of this dissertation will examine the effect of nonresponse on the quality of survey estimates. The underlying hypothesis is that the MNAR scenario is most likely to be applicable to survey data in the presence of nonresponse. The chapter will deal with the dominant views on estimating survey statistics in the presence of nonresponse. Many researchers, often assisted by common statistical tools, treat nonresponse as an inconvenient interference, the harmful effects of which are usually acknowledged among most professional survey researchers and survey authorities, but for which a set of adequate methods and procedures to completely deal with the consequences is currently lacking. Although weighting procedures seem to be relatively popular, widespread, and commonly used in survey practice, they are not believed to entirely remedy the disadvantageous effects of nonresponse. As MNAR erodes the possibility of making valid inferences from surveys, survey researchers instead resort to less stringent measures. Reporting the response rate and how it is computed, the description of the weighting procedure, or the comparison of respondents and nonrespondents seem to be to standard elements of nonresponse documentation. This avoids the crucial step into the MNAR area, where test statistics or p-values become unreliable and scientific claims become as suspicious as a bouncing check.

This first chapter will try to measure the effects of nonresponse by observing the problem from different angles or perspectives, and using a variety of quality indicators. It is a conscious choice to apply a multitude of approaches, because nonresponse is by definition unobservable. Therefore, the effects of nonresponse need to be monitored in an indirect way, using detours, suboptimal data, and probably a variety of assumptions. For this

reason alone, nonresponse inevitably induces uncertainties that are very hard to quantify, marking the departure from the probability paradigm that is still currently used in survey research. This implies that the strong inferential framework offered by mathematical statistics should be replaced by a downgraded framework, only allowing exploratory data analyses based on survey data.

As the first chapter of the dissertation focuses on the output quality of a survey, the second chapter will deal with the production process of a survey. Indeed, survey response does not only involve (non)responding individuals or households. In addition, survey sponsors, researchers, and/or interviewers are survey agents with specific interests. Even the privacy regulations imposed by governments can be seen as a complicating factor when dealing with survey nonresponse. Hence, instead of focusing only on the differences between respondents and nonrespondents, this dissertation also seeks to look at the issue from the manufacturers' point of view. A respondent set can be seen as the result of a contact process that is directed and informed by the goals of the survey sponsor, fieldwork manager, interviewer, etc. This process approach assumes that these survey agents take decisions that may make prospective respondents participate in the survey.

In this respect, there are many instances of standard practices that have been developed in order to deal with survey nonresponse. Fieldwork agencies, responsible for the collection of survey data, are driven by the minimization of nonresponse rates. It is thereby assumed that reducing nonresponse also decreases the potentially related bias. Response rate maximization therefore still seems to be the dominant position in survey fieldwork. However, it is questionable that this practice really leads to better surveys. In particular, making response rate maximization the fieldwork objective may trigger the prioritization of easy cases, termed the 'low hanging fruit', suggesting that fieldwork efforts follow the line of least resistance. As a consequence, the risk increases that not all individuals in a sample have an equal probability of being included as respondents, clearly violating the ideal of a representative respondent set. The second

chapter of this dissertation will elaborate this line of thought. It will also urge reconsidering the almost institutionalized practice of maximizing the response rate in a survey. Also the option of smaller sample sizes will be explored.

As this dissertation seeks to refresh some ideas about the issue of nonresponse, both at the level of estimation and the level of the survey production process, it is important to have adequate data. Therefore, having process data (also termed paradata), is an important condition in order to make this evaluation. Such data provides information about the survey and is therefore to be distinguished from the survey's target data. Paradata provides information about the contact process during the fieldwork: how many contact attempts were made, at what time of the day, by which interviewer, and what the outcomes of these efforts were. Paradata also provides information about the sampling procedures, how the sample units were targeted and so forth. Paradata might also include characteristics of the households or individuals from the population register totals, which are ideal in order to compare respondents and nonrespondents (or the entire population). Throughout this text, these characteristics will be termed '*auxiliary variables*' or '*auxiliary information*'.

The European Social Survey (ESS) is an important data source for assessing the origin and impact of nonresponse. These datasets will act as the first empirical reference in this dissertation. As much attention and effort was given to combating nonresponse in this survey, the ESS can be considered as a high-quality benchmark in the European survey industry. Its high methodological standards with respect to the fieldwork process (e.g. refusal conversion and spreading contact attempts over time and over modes) and the fieldwork targets (e.g. 70% response rate and 3% noncontact rate), and also the central coordination and monitoring, should be a clear indication that the ESS belongs among the best European surveys. Perhaps the strongest of all the quality assets of the ESS is its transparency. All the survey characteristics and decisions are documented and are easily accessible on the Internet. All datasets are freely available, enabling the replication of published analyses. This methodological open-

ness also facilitates the improvement of the survey quality. Hence, even in surveys reaching the highest standards of quality, some shortcomings can still be found.

Apart from the ESS, other secondary surveys will also be used, such as the Flemish Housing Survey 2005 - 2006 (FHS). For a detailed description of these data sources, see the appendices (from page 223 onward). Original analyses will be provided based on these datasets throughout the text, each time introduced by '**Data analysis**'.

Chapter 1

Measuring Nonresponse Error

Ultimately, this chapter seeks to find out whether, and if so to what extent, nonresponse endangers the probabilistic paradigm used to make inferences from survey data. As nonresponse produces gaps in the data, respondents are not identical to nonrespondents. This contrast between the two groups may result in a bias of the respondent group compared with the full sample. As long as nonresponse cannot be accurately measured or controlled, bias can probably not be ruled out completely, leaving traces of uncertainty for which a reasonable but unknown price needs to be paid, at least if survey researchers remain inclined to avoid type I errors. The first step is to find data, by the use of which the unobservable can be observed. Therefore, what are termed 'auxiliary variables' will be used in order to determine the differences between respondents and nonrespondents. Next, a theoretical distinction between the random and fixed response models will be discussed in detail, enabling the construction of many indicators to measure and assess the effects of nonresponse.

1.1 Finding data to observe the unobservable

An essential characteristic of the analysis of nonresponse is that the target problem cannot be observed directly. If that were possible, then the problem would of course completely disappear. Therefore, nonresponse needs

to be observed indirectly, implying only partial information that can be assessed, usually complemented by a set of assumptions.

Research into nonresponse bias seems to be a continually expanding process. Due in particular to the increase in the amount of available paradata over the last ten to fifteen years, the possibilities of assessing the nature and impact of nonresponse continue to grow rapidly. Groves (2006) reviewed five distinct methods for assessing nonresponse bias: (1) response rate comparisons across subgroups; (2) using sampling frame data or supplemental matched data; (3) comparisons with similar estimates from other sources; (4) studying variations within existing surveys: nonresponse follow-up studies; and (5) contrasting alternative post-survey adjustments for nonresponse. Today, these methods are still the basis for nonresponse bias assessment, but because of the growth of supplementary data as a by-product of data collection, these methods have tended to become more powerful and refined.

Two main groups of methods with which to investigate nonresponse are distinguished for further elaboration in this dissertation. The first group uses what are termed 'auxiliary variables' or background information that is available for both respondents and nonrespondents. The second group instead looks for traces of nonresponse bias by monitoring the data flow or by tracking changes in the target variables as a function of increased fieldwork efforts.

It appears as though the strength of the first group of methods is the weakness of the second group. The first starts from the availability of information that covers both respondents and nonrespondents (or the entire population). With this information (hereafter termed '*auxiliary information*'), it is relatively easy to compare respondents with the full sample in order to assess the effect of nonresponse. However, the auxiliary variables are usually not the variables that are of interest, or the variables for which the survey was originally designed.

The second group of methods therefore focuses on the target variables by monitoring them during the fieldwork or through weighting them by a set of auxiliary variables. However, this group of methods cannot fully

assess the impact of nonresponse, as the true target variables are never completely revealed.

In this regard, completeness and relevance are two sides of the same coin. Each method used to investigate nonresponse seems to be obliged to choose one of the two strengths, leaving the other as a weakness. Auxiliary variables are complete, but not necessarily relevant, whilst relevant target data can monitored throughout the data collection process, but will never be complete.

With regard to the first set of methods, Lynn (2008) distinguishes four different ways of collecting such auxiliary data:

1. *Sample frame information.* The list of units from which the sample is eventually drawn may contain useful information about the individuals. In a general population survey, the sample may be drawn from an official database that also contains information about the age and gender of the individuals, as well their marital status and residence information. A survey among the employees of an organization can use information from the personnel department, including salary, years of employment, etc. Groves (2006) advises comparing the response rates of different groups or profiles as defined by the background register data. Instead of comparing the response of different socio-demographic groups, issues caused by nonresponse can be measured by comparing the means (or other statistics of interest) between the recorded data of respondents and nonrespondents. If no differences in response rates or means can be found, the survey is probably better protected against the unfavorable effects of nonresponse, at least if the nonresponse structure is believed to be MAR. However, the absence of evidence for nonresponse bias does of course not imply evidence of the absence of bias.

2. *Linked data.* For address samples (from which the target person still need to be selected), the usefulness of frame information is less obvious. Linking other sources of data might be more advisable in that case. A frequently-used technique is to merge sample with

municipality-level data, using the postal code as the merging key (Kalton & Flores-Cervantes, 2003; Johnson, Young, Campbell, & Holbrook, 2006). This allows enrichment of the data with area-level information such as population density, crime rates, etc.

3. *Interviewer observations.* When trying to make contact with the target persons or household, interviewers can also observe and record aspects of the housing conditions or appearance of the target (Groves & Couper, 1998; Copas & Farewell, 1998; Groves, Wagner, & Peytcheva, 2007). Specifically, in an in-home face-to-face survey, interviewers can be asked to observe and record information about the dwelling and the neighborhood of the household that needs to be contacted. It is even possible to find details of the family composition (children's bikes or child seats in the car), smoking (cigarette butts), etc. Some new information technologies such as *Google Street View* may open up new opportunities in this regard. Interviewer observations are obviously subject to the interpretation of the particular observer. In addition, some interviewers diligently fill in the observable information on their contact sheets, while others systematically skip this part of their task or only sloppily fill in the required area information. Therefore, nonresponse researchers should be careful when using such data (Kaminska & Lynn, 2011; Sinibaldi, Kreuter, & Durrant, 2011; West, 2013).

4. *External data sources.* The respondents in a sample can be compared with the characteristics that are also known about the general population. Sometimes, comparison is also possible with a generally accepted gold-standard survey (e.g. a mandatory labor force survey). The comparison with an external source usually does not only include the measurement of nonresponse error, but also coverage error and sample error, because the respondent set is not compared to the full sample, but to the entire population, including other selection processes than just nonresponse. Even so, measurement error can occur. For example, the distribution of the level of education as

measured from the questionnaires may be compared with an official data source on the overall population. Occasional differences may be (partly) due to the inaccuracy of the answers given by respondents. Another problem with the use of external data is the comparability of population data with survey data, because of the time lag between the two measurements or because of differences in the classification of the sample and the population variables. There are, for example, many ways to classify someone's level of education. If the survey question and the population information use different classifications, comparison becomes rather pointless. This may be even worse in an international context, where different countries use different ways to classify educational achievement.

The particular advantage of these four ways to explore nonresponse damage is the completeness of such auxiliary information. In the first three methods at least, there is a clear distinction between respondents and nonrespondents; for both groups the auxiliary variables provide complete information. Unfortunately, such auxiliary information does not necessarily reflect the error that applies to the target variables of the survey. Auxiliary variables are usually somewhat more factual or administrative compared with survey target data, which sometimes also reflects reported behavior or attitudes. Therefore, the methods in the second group are also interesting to consider, as they monitor target variables throughout the course of the fieldwork or during extended fieldwork efforts.

Usually this second group of methods uses paradata to monitor the target statistics of a survey as a function of the efforts that have been made to get individuals to participate, or the time between the request and the return of the completed questionnaire. Instead of auxiliary data, process data is more important for the methods in this group, which are as follows:

1. *Nonresponse conversion techniques.* There are many reasons why a survey request fails to make an individual or household participate: noncontact, refusal, language barrier, mental or physical disabilities

or illness, bad timing, etc. In many instances, the interviewer or the fieldwork management may decide to re-issue the case and possibly enhance the chances of success by choosing a better time or better incentive for the target person. Participating countries in the ESS are advised to consider refusal conversion attempts in order to improve response. Not only will this increase response rates, but refusal conversion is also considered to be a tool with which to facilitate the inclusion of individuals who are less inclined to participate in a survey. As a consequence, occasional bias because of the non-participation of substantially different profiles may be (partially) anticipated. Survey refusal is probably not a permanent status. As will be argued later on, survey participation can be seen as the realization of a latent individual trait. This latent propensity has evidently no fixed outcome, but can vary, depending on a chance coincidence. Groves and Couper (1998) consider the decision to participate in a survey to be a process that is not well considered, and on which most respondents do not expend a great deal of cognitive effort. A refusal at the first request does not necessarily imply a rejection at the second attempt. When the elapsed time between the two requests is long enough to 'cool down', there is an increased probability of conversion (Groves & Couper, 1998; Triplett, Scheib, & Blair, 2001; Triplett, 2002; Edwards, Martin, DiSogra, & Grant, 2004; Beullens, Billiet, & Loosveldt, 2010). In order to measure (initial) nonresponse effects, the inevitable assumption for conversion activities is that converted individuals have some similarities with final nonrespondents. However, without substantively altering the contact strategy, converted individuals might only be delayed respondents, not (or only slightly) differing from initially cooperative respondents. Therefore, the differences between initial and converted respondents are usually accepted as traces of nonresponse bias, and the absence of such differences is not necessarily considered as evidence of the absence of bias. In addition, the possibility must be considered that converted initial nonrespondents may be more inclined to produce measurement error

such as item nonresponse, hindering a straightforward assessment of unit nonresponse effects (Olson, 2013a).

2. *Survey of nonrespondents.* Follow-up surveys among nonrespondents are similar to conversion efforts. However, the questionnaire or the contact procedure is usually substantively changed. A shorter questionnaire or a more convenient contact mode can facilitate the responsiveness of initially nonresponding cases. Hansen and Hurwitz (1946) proposed that nonrespondents from an initial mail survey should be followed-up by means of face-to-face interviews, whereas Lynn (2003) advised addressing a short questionnaire to nonrespondents in order to partially recover information that would otherwise be missing. This PEDASKI-method (pre-emptive doorstep administration of key survey items) is similar to the basic-question approach as proposed by Kersten and Bethlehem (1984), the difference between the two methods is in the length of the shortened questionnaire. The basic-question approach is usually restricted to only a few items, while PEDASKI method includes more survey questions. Most of these methods target nonrespondents who initially refused to cooperate. Other profiles of nonrespondents, such as non-natives or disabled people, may need an appropriate treatment in order to participate. Therefore, follow-up designs become more sophisticated. For example, the University of Liége is currently developing methods specifically designed to facilitate that participation in the ESS of sensory disabled people by means of web-based questionnaires and face-to-face interviews with specially trained interviewers (Fontaine, 2012). Matsuo, Billiet, Loosveldt, Berglund, and Kleven (2010) also used a short questionnaire survey among nonrespondents in order to inform weighting adjustments.

3. *Early and late responders.* Particularly in self-administered surveys, comparing early and late respondents is relatively easy (Dillman, 1978, 2000). It is assumed that late respondents are informative with regard to estimating the answers of final nonrespondents. Par-

ticularly due to the increase in process data, this kind of nonresponse research has gained attention. During the last decade or so, fieldwork operations have been monitored more frequently in order to manipulate them and optimize the quality of the obtained sample, a process often referred to as adaptive or responsive survey design (Groves & Heeringa, 2006; Schouten, Calinescu, & Luiten, 2011; Couper & Wagner, 2011).

This second group of methods to assess the impact of nonresponse clearly lacks complete target information. Even after considerable conversion attempts, there will still be nonrespondents, implying that some uncertainty about how nonresponse affects the data is still left to consider. Lin and Schaeffer (1995) mention two different assumptions in this respect. First, the continuum of resistance model assumes a (strong) correlation between the effort during the fieldwork and the characteristics of the respondents. This means that late responders are more like the final nonrespondents or that converted refusals are also more like final refusers (Smith, 1984). Under this assumption, the composition of the sample is expected to improve as more fieldwork efforts are made. Second, the classes of nonparticipants model rejects the starting point that there is only one single mechanism that explains survey participation (namely effort), and advances the idea that there are a variety of factors that contribute to the decision of whether to participate. This latter perspective is therefore obviously more skeptical about the usefulness of additional survey efforts in order to assess the effects of nonresponse. In addition, consideration should be given to the fact that additional fieldwork efforts may be directed toward the cases that are deemed more responsive, leaving a potentially different group of persistent nonrespondents out of the picture. Therefore, the classes-of-participants model anticipates an improvement of the sample composition only if the relevant classes of nonrespondents are identified and additional efforts are made to pursue the equality of participation.

Next to the two groups of methods, a third method to address nonresponse also makes use of auxiliary variables, but still assesses the effect of nonresponse with regard to the target variables. Target survey statistics can be compared before and after weighting corrections based on auxiliary variables. It is thereby assumed that statistics which remain stable under different adjustment procedures are less biased than those statistics that are more subject to shifts (Groves, 2006). Nevertheless, this particular method may also combine the disadvantages of the two former groups of methods: (1) because the target variables are still incomplete, it is difficult, if not impossible, to assess the efficacy of the adjustment procedures. (2) The auxiliary variables that are used to inform the adjustment procedure may not be relevant enough to completely remove the problem of nonresponse error.

This overview of methods to assess the impact of nonresponse makes clear that there is no perfect way of looking at the issue. In every case, the information used to assist the process of investigating nonresponse is only partial: it is either complete with regard to irrelevant auxiliary variables, or incomplete with regard to relevant variables. It might even be impossible to aspire to a perfect situation: if all the relevant variables were complete, there would be no need to assess nonresponse, because the problem would have completely disappeared. On the other hand, if the auxiliary variables became relevant, the effort of carrying out the survey would have been in vain, because all the relevant information would have already been available beforehand.

At this point, it may be worthwhile considering the concept of paradata, as such data will be the major source of information used to assess the nonresponse problem. Paradata is usually considered as a by-product of the data collection process (Couper, 1998; Kreuter & Casas-Cordero, 2010; Durrant & Kreuter, 2013; Olson, 2013b), as opposed to the target data that a survey is designed to collect. Paradata tells us something about the target data and can therefore be termed 'second order data'. Paradata might not only relate to the contact process or the interviewer's impression of the neighborhood of the target household. It might also in-

dicate the length of an interview, the speed with which the interviewer asks questions, linked administrative data sources, the recording of mouse movements across the screen during a web survey, and so forth. Given the variety of paradata, there are many opportunities to use such data to assess and augment the quality of survey data. This dissertation does not aspire to give a comprehensive overview of paradata, but will instead use paradata to assess the effects of nonresponse. Therefore, two main categories of paradata that are used in this dissertation will be referred to as 'auxiliary variables' and 'process data'. 'Auxiliary variables' are defined as variables that are available on the individual level, for both respondents and nonrespondents. Usually, this data describes the socio-demographic background of individuals and may comprise information from records, such as age, gender, or region. Interviewer observations such as the assessment of the dwelling and neighborhood of the sample cases can also be considered as auxiliary variables. This first chapter will primarily use such auxiliary variables. 'Process data' describes the contact sequence or the properties of the contact attempts of an interviewer in order to get individuals or households to participate. This kind of data will predominantly be utilized in the second chapter. Although paradata can be used to inform the quality of the target data (Carton, 2008; Kreuter & Casas-Cordero, 2010; Sinibaldi et al., 2011; Matsuo & Loosveldt, 2012), paradata itself is also prone to imperfections. For example, sample frame data may not be up-to-date, interviewer observations may be biased, the contact sequence may be groundlessly upgraded by the interviewers, etc.

It also appears that nonresponse research is impossible without making assumptions about the problem itself (e.g. the continuum of resistance model versus the classes of nonparticipants model) or about the assisting variables that facilitate the assessment of nonresponse. Even with the latest generation of nonresponse models, where process data is combined with auxiliary variables so that the use of circumstantial data is optimized (see, e.g., Dahlhamer & Jans, 2011), uncertainties about the nature of nonresponse and the ability to measure it still remain.

On the one hand, these ever-present uncertainties challenge nonresponse researchers to continually seek alternative and creative ways to measure and combat nonresponse error. For example, Kaminska, Billiet, and McCutcheon (2010) examined the relationship between satisficing in the responses to survey questions and the reluctance to participate in the survey. Kohler (2007) examined internal survey criteria to assess nonresponse. This kind of research presumes that, for example, for every 100 married women in the population, there should also be 100 married men. These proportions should evidently be reflected in the survey sample, unless it has been affected by nonresponse bias. To test this internal criterion, a survey should comprise as many men as women among the group of married people. On the other hand, measurements of nonresponse error can easily be contested or rejected because of these uncertainties. As a result, a wide range of attitudes can be taken toward nonresponse error, going from simply ignoring the consequences of nonresponse to the complete rejection of the use of survey data. Positions in between these opposite attitudes may recognize and mention the threat of nonresponse to some degree, without taking measurements, while others may be more strict and advise the use of surveys only for exploration, rejecting their use for inferential purposes. Currently, the American Association of Public Opinion Research (AAPOR, 2010) recommends providing transparency about how a sample was obtained. Essential elements in this regard are the definition of the population, a description and origin of the sample frame, the sampling design, how respondents are selected and recruited, information about quota and additional sample unit selection, the eventual sample size, estimates of the sampling error, and information about the possible weighting method. When making inferences from survey data, AAPOR advises that

> 'We shall not knowingly imply that interpretations should be accorded greater confidence than the data actually warrant. When we use samples to make statements about populations, we shall only make claims of precision that are warranted by the sampling frames and methods employed. For example, the

> reporting of a margin of sampling error based on an opt-in or self-selected volunteer sample is misleading'.

Such a point of view is relatively strict and encourages a more skeptical attitude toward survey data that is used to deal with nonresponse.

This chapter will deal with the first group of methods to assess the impact of nonresponse (methods based on auxiliary variables), leaving the process-oriented approach for Chapter 2, where (non)response will be considered as the result of a (production) process. This current chapter will instead consider nonresponse from an output-oriented perspective. It will become clear that even within the first group of nonresponse assessment methods, a variety of different assumptions or perspectives can be deployed, potentially leading to different interpretations of the seriousness of the nonresponse problem.

The chapter starts with a theoretical outline of nonresponse. Two models, the random response model and the fixed response model, will be presented to explore the interpretative scope of the nonresponse problem. As already mentioned, by definition nonresponse cannot be observed. This means that trying to examine the problem by only one method or procedure may be problematical. Therefore, a variety of possible viewpoints should be offered. The distinction between the random and fixed models is an interesting starting point, as it offers a wide range of (opposite) angles from which to observe the unobservable. Gradually, this chapter will also present some quantifiers that express or indicate the consequences of nonresponse.

1.2 The fixed and the random nonresponse model

1.2.1 Basic concepts

Nonresponse can be seen either as a cause or as a consequence. In the first interpretation, there is a decision (cause) as to why an individual i

participates ($r_i = 1$) in a survey. As a result (consequence), there are two groups (respondents and nonrespondents) that may differ in respect of some statistics of interest. Considering nonresponse as a cause or as a consequence basically refers to the difference between respectively the random and the fixed nonresponse model (Kalsbeek, 1979; Lessler & Kalsbeek, 1992; Bethlehem, 2009; Bethlehem, Cobben, & Schouten, 2011).

The fixed or deterministic interpretation assumes that the population consists of a fixed group of respondents and a fixed group of nonrespondents. This implies that the response behavior is predetermined for each member of the population: sample elements in the response stratum participate in a survey with certainty; the nonrespondents will never participate.

In the random or stochastic view, however, nonresponse is assumed to be the outcome of a random process. This perspective does not restrict the expected response to either 0 or 1, but allows a response probability or propensity to be $0 < \rho_i < 1$. The response propensity ρ_i can be seen as a latent trait of a person or a sample element that is relatively stable and has a manifest 0-1 outcome each time a survey request is addressed to the individual. Similarly, Schouten, Cobben, and Bethlehem (2009) consider a response propensity as 'a biased coin that a unit carries in a pocket'. This means that an individual may be a respondent on the first occasion, whereas he or she will not participate on the next occasion, even when there are no substantial differences between the two situations.

Although the interpretation of each model is different, many (non)response measurements will be similar. For example, the response rates in both models have a different formal notation, though the result is the same: $E\left(\frac{1}{n}\sum_{i=1}^{n} r_i\right) = E\left(\frac{1}{n}\sum_{i=1}^{n} \rho_i\right) = E\left(\bar{r}\right) = E\left(\bar{\rho}\right)$. The fixed model counts the integers 0 and 1, the random model counts the underlying propensities that result in response or nonresponse, and n refers to the total sample, including both respondents and nonrespondents.

Response propensities can obviously change. Both views on nonresponse acknowledge the fact that survey participation is not only the result of a decision-making process exclusively by the individual sample unit. The circumstances in which the request is presented also determine survey

participation. In this regard, the same individual may be a respondent in a first survey, but may not cooperate on a subsequent occasion, due to changing survey conditions such as the mode of data collection, a different timing of the request, less interesting survey topic, etc. In this regard, Groves and Couper (1998) have conceived a model outlining the possible factors that contribute to the likelihood of participating in a survey. Figure 1.1 shows that survey participation depends on many factors that can be related to the individuals or sample cases, the interviewers (at least in a face-to-face survey), the survey design, and environmental factors such as the survey climate. If, for example, the fieldwork management decides to incentivize prospective respondents by offering lottery tickets, the response propensity may increase (random model), resulting in more respondents in the eventual obtained sample (fixed model). The deterministic approach considers design variables as additional conditions that determine the eventual decision to participate. However, the response propensities in the stochastic approach are also prone to the surrounding characteristics of the survey request. These considerations make it hard to accept that a fixed response propensity under any conditions is realistic (Dalenius, 1983).

Notwithstanding the rather subtle theoretical differences between the two approaches, the fixed and random response models lead to different ways of measuring nonresponse error.

1.2.2 Opposite causal order

The random response model predominantly focuses on the reason why an individual participates in a survey. If an auxiliary variable aux is available for both respondents and nonrespondents, propensities can be determined by $\rho_i = P(r_i|aux_i)$ (Rosenbaum & Rubin, 1983), or the probability of an individual responding positively to a survey request, conditional on the available auxiliary variable(s). Usually, models that estimate response propensities take the response outcome r as the dependent variables that

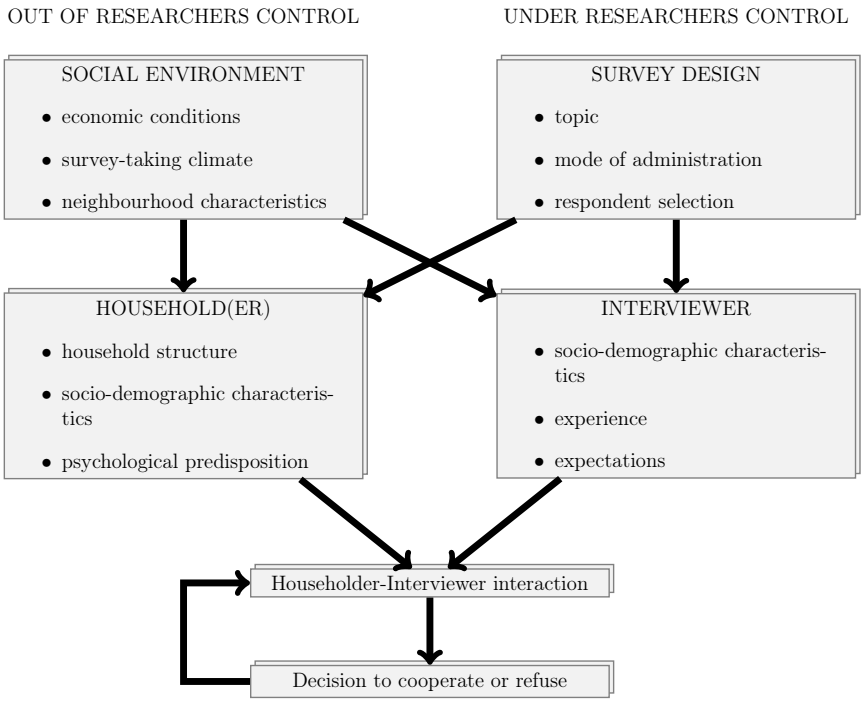


Figure 1.1: A conceptual framework for survey cooperation.
Source: Groves and Couper (1998)

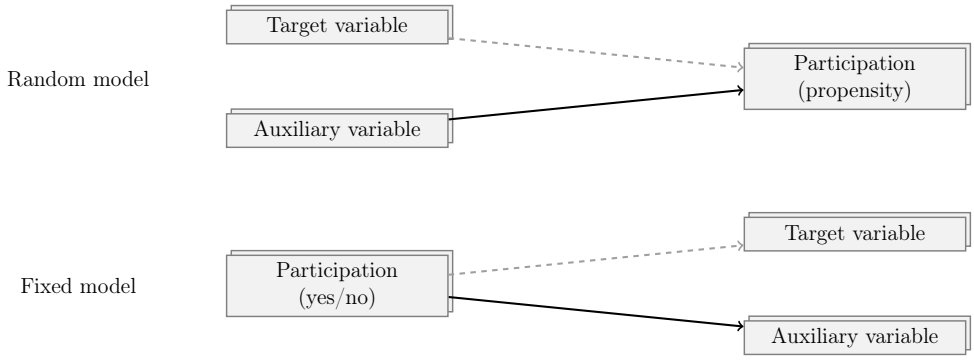


Figure 1.2: Random and fixed response model: opposite causal orders

is explained by independent variables $aux_1, aux_2, \ldots, aux_p$, such that

$$\hat{\rho}_i | aux_i = g^{-1}\left(aux_i'\beta\right) \tag{1.1}$$

The link function $g\,(.)$ can be specified as the logit or the identity (linear) link function (Schouten et al., 2009). Other link functions such as probit have also been taken into account (Laaksonen, 2006). The result is a vector of n elements, where n expresses the total sample, including n_r respondents and n_{nr} nonrespondents. This vector of propensities can subsequently be used for a variety of nonresponse assessment operations such as measurement of the representativeness of the sample, bias estimation, or propensity score weighting. All these will be discussed later on. Theoretically, target variables are also deemed to have some effect on response behavior. However, since target information is not available among nonrespondents, these effects are not reflected in the response propensities. This is the reason why Figure 1.2 only shows a broken line, suggesting an existing but unobservable effect of the target variables.

The fixed response model takes the opposite perspective to its random counterpart. Usually, the auxiliary variables are considered to be the dependent variables, the response indicator r being the independent variable. In the case where the auxiliary variable is continuous, such a nonresponse bias assessment simply reduces to the same setting as a *t-test*. The differences between respondents and nonrespondents can be denoted

as $a\bar{u}x|(r = 1) - a\bar{u}x|(r = 0)$. Similar to the limitations of the random response model, the difference between respondents and nonrespondents can also exist for the target variables, but is again unobservable. This is why only a broken line is shown from the response outcome to the target variables.

1.2.3 Expressions for bias and contrast

Given that an auxiliary variable aux is available for both respondents and nonrespondents, and the target variable y is only available for respondents, the fixed and the random response model provide expressions for bias and contrasts. In both cases, nonresponse bias is defined as the difference between the respondent-only sample and the full sample (respondents and nonrespondents).

The fixed response approach defines bias with regard to the auxiliary variable as

$$bias_{a\bar{u}x_r} = a\bar{u}x_r - a\bar{u}x_f \tag{1.2a}$$

$$= \left(\frac{n_{nr}}{n}\right)(a\bar{u}x_r - a\bar{u}x_{nr}) \tag{1.2b}$$

where $a\bar{u}x_f$ is the full-sample mean of aux, $a\bar{u}x_r$ is the respondent mean for aux, $a\bar{u}x_{nr}$ represents the nonrespondent mean for aux, n is the total sample size, and n_{nr} is the number of nonrespondents. In fact, nonresponse analysis under the fixed interpretation uses auxiliary variables as interim target variables. The same bias estimate for aux can be obtained by

$$bias_{a\bar{u}x_r} \approx \frac{corr_{r,aux}\sigma_r\sigma_{aux}}{\bar{r}} \tag{1.3}$$

where σ_r is the standard deviation of r $(=\sqrt{r\,(1-r)})$, σ_{aux} is the standard deviations of aux, and $\bar{r}$ is the mean response propensity or the response rate.

In the random response approach, bias with regard to a target variable is defined as

$$bias_{\bar{y}_r} \approx \frac{corr_{\rho y}\sigma_\rho\sigma_y}{\bar{\rho}} \tag{1.4}$$

where σ_ρ and σ_y are the standard deviations of ρ and y, and $\bar{\rho}$ is the mean response propensity or the response rate (Bethlehem, 1988). The difference between equations 1.3 and 1.4 concerns the use of the latent propensities ρ_i on the one hand (equation 1.4) and the manifest response indicator r_i on the other hand (equation 1.3). Usually, the propensities are estimated using auxiliary variables as covariates in a (logistic) regression model, predicting the probability that an individual will participate in a survey. This implies that the biases as determined by the fixed and the random models are the same if aux and y are identical.

From equations 1.2a, 1.2b, 1.3, and 1.4 , it is clear that nonresponse bias is higher where the response rate is lower. The bias will also increase if the variance of the response propensities increases or if the target variable (auxiliary variable) is more correlated with the propensities ρ (or response outcome r).

The difference between the respondent mean and the nonrespondent mean can be termed the contrast, and equals the nonresponse bias divided by the nonresponse rate. The contrast in terms of the fixed response model is

$$K_{a\bar{u}x_r} = \frac{bias_{a\bar{u}x_r}}{\left(\frac{n_{nr}}{n}\right)} = a\bar{u}x_r - a\bar{u}x_{nr} \tag{1.5}$$

When elaborating equation 1.3, the contrast can also be expressed as

$$K_{a\bar{u}x_r} \approx \frac{bias_{a\bar{u}x_r}}{1-\bar{r}} \approx \frac{corr_{r,aux}\sigma_r\sigma_{aux}}{\bar{r}\left(1-\bar{r}\right)}$$

Under the random response approach, and provided that the response propensities have been estimated by some set of auxiliary variables aux,

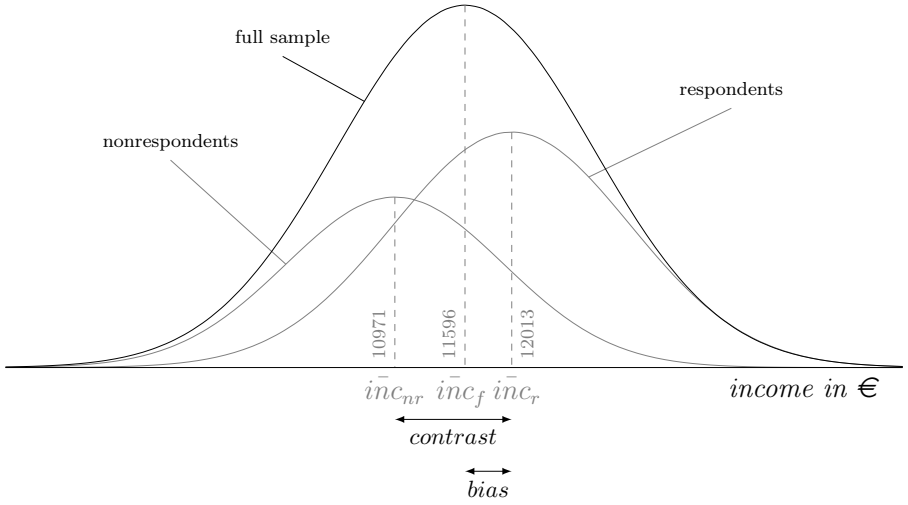


Figure 1.3: Full-sample, respondent-only and nonrespondent-only distributions for *income*, fictitious data

the contrast with regard to target variable y is

$$K_{\bar{y}_r} \approx \frac{bias_{\bar{y}_r}}{(1-\bar{\rho})} \approx \frac{corr_{\rho y}\sigma_\rho\sigma_y}{\bar{\rho}\,(1-\bar{\rho})} \tag{1.6}$$

It is thereby assumed that $corr_{\rho y}$ is the same for respondents and nonrespondents. It is also assumed that that the within group variance is constant among respondents and nonrespondents. If this latter assumption is not fulfilled, σ_ρ cannot be estimated properly, invalidating the contrast estimate. This problem also holds for the bias estimate expressed in equation 1.4.

As a final note, it is worthwhile standardizing the bias and contrast indicators with regard to the scale of y and aux. This allows comparisons to be readily made between (interim) target variables.

In order to illustrate the concepts that have been presented so far, consider a fictitious sample of $n = 1000$ individuals, of whom 600 (n_r) participate in the survey and 400 (n_{nr}) do not participate. As a consequence, $\bar{r} = \bar{\rho} = 0.60$. There is only one auxiliary variable, *income*, which expresses the average income of the municipality an individual lives in. Figure 1.3

illustrates the distribution of *income* for the complete sample, as well as the groups of respondents and nonrespondents separately. As can be seen, the respondents seem to live in municipalities with higher average incomes. The *contrast* $K_{\bar{inc}}$ equals $12013 - 10971 = 1042$. Multiplying the contrast by the nonresponse rate equals the bias ($1042 \times 0.40 = 417$), which is exactly the same as the difference between the respondent-only mean and the mean of the full sample ($12013 - 11596 = 417$). This illustrates equations 1.2a, 1.2b and 1.5 on page 25.

Provided that the correlation between the response indicator r and the auxiliary variable *income* is 0.4473, $\sigma_r = \sqrt{r\,(1-r)} = 0.4901$ and $\sigma_{inc} = 1142$, the bias can also be determined by

$$\begin{aligned} bias_{\bar{inc}_r} &= \frac{corr_{r,inc}\sigma_r\sigma_{inc}}{\bar{r}} \\ &= \frac{0.4473 \times 0.4901 \times 1142}{0.60} \\ &= 417 \end{aligned}$$

This concludes the fixed model application. For the demonstration of the random model, the *income* variable takes the position of the auxiliary variable in order to be able to estimate the individual response propensities $\rho_i = P\,(r_i = 1|inc_i)$. A logistic regression model provides the propensities that are illustrated in Figure 1.4.

Suppose now that the survey seeks to estimate the average annual clothing expenditure per person (*clothes*). For respondents, the correlation between *clothes* and $\hat{\rho}_i|income$ is 0.21. Other information in order to estimate the degree of nonresponse bias with regard to *clothes* includes the standard deviation of the response propensities σ_ρ, and the standard deviation of the target variable $\sigma_{clothes}$. The value of σ_ρ can easily be obtained using the predicted values of the logistic regression (Figure 1.4) and equals 0.2231 in this particular case. However, $\sigma_{clothes}$ is much harder to estimate, because information is only available for respondents. Therefore, $\sigma_{clothes}$ should be replaced by $\sigma_{clothes}|(r = 1)$, although it can be expected that $\sigma_{clothes}|(r = 1)$

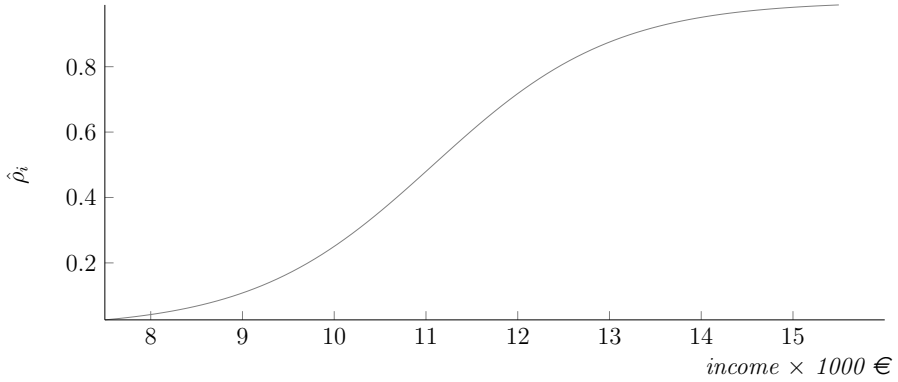


Figure 1.4: Estimated response propensities $\hat{\rho}_i$ as a function of *income*, fictitious data

slightly underestimates $\sigma_{clothes}$. This is due to the fact that, under the assumption of homoscedasticity, the within-group variance is smaller than the total variance of a variable. Hence, $\sigma_{clothes}|(r=1) < \sigma_{clothes}$.

Provided that $\sigma_{clothes}|(r=1) = 1200$€, the estimated bias with regard to *clothes* is

$$\begin{aligned} bias_{\bar{clothes}_r} &\approx \frac{corr_{\rho,clothes}\sigma_\rho\sigma_{clothes}|(r=1)}{\bar{\rho}} \\ &\approx \frac{0.21 \times 0.2231 \times 1200\text{€}}{0.60} \\ &\approx 93\text{€} \end{aligned}$$

Of course, this bias assessment using the random response model only holds true if the estimated propensities based on the income information reasonably reflect the actual response propensities. However, it is fairly obvious that the response propensities are also related to many other variables. Therefore, applying equation 1.4, the bias assessment using the random response model, is relatively speculative, as it probably underestimates the variability of the response propensities. It is also speculative as $corr_{\rho,clothes}$ in the expression determines ρ only conditional on the income variables instead of using the true ρ's, which are evidently unknown.

In survey literature, there are plenty of reports, articles, and book chapters discussing the differences between respondents and nonrespondents (contrasts) or respondents and the full sample (bias), assessed using the fixed model. Assael and Keon (1982) discussed the differences between respondents and nonrespondents in a survey among small businesses and found that responding businesses had higher telephone bills, more telephone lines, and more telephone stations than nonresponding businesses did. They also reported that this kind of nonsampling error far outweighed the error expected from random sampling alone. Bolstein (1991) found that in an analysis based on respondents alone, an African-American Democratic candidate was predicted to win the 1989 Virginia election by 10%, although he only realized a 0.4% victory. The author argues that nonrespondents in particular cast their vote for the white republican candidate. Many other surveys nowadays try to gather information about the entire sample in order to obtain a basic nonresponse analysis. By means of meta analysis, Groves (2006) synthesized 30 studies that examined the differences between respondents and nonrespondents. In total, 235 of such mean contrasts were available. On average, the distance between the respondent mean and the nonrespondent mean is about 0.054 standard deviations.

Using the random model interpretation, Merkle and Edelman (2002) examined the differences in response rates in election exit polls, conditional on available background information for voters. They found that older voters (+60) were usually less inclined to participate. Race and gender, as they found, had less impact on the response rates. In an experimental study to assess the effect of interviewer incentives on the bias of target variables, Peytchev, Riley, Rosen, Murphy, and Lindblad (2010) first estimated the response propensities of individuals based on demographic characteristics and some available housing and mortgage data, available from a prior wave to the actual data collection. With these individual response propensities, the relationship with target variables could be estimated in order to determine the bias of these target variables. The propensities were fairly predictive of the later response outcome, but the estimated biases were not consistent between the control group (no specific incentives) and

the experimental group (interviewer incentives to prioritize low propensity individuals). Another study (Kreuter et al., 2010) comprised five large surveys and considered the relationship between auxiliary variables and the response outcome on the one hand, and the relationship between the same auxiliary variables and the target variables on the other hand. The random response model in particular was used in order to evaluate shifts between the unadjusted and adjusted (weighted) estimates of the target variables. Weights are usually obtained by inverting the response propensities.

Data analysis Instead of extensively discussing examples from literature concerning nonresponse, original analyses are presented here. Three different surveys are used throughout this dissertation to compare theory and practice: the Flemish Housing Survey, the Dutch General Population Survey, and the European Social Survey. The last of these three actually comprises multiple surveys over different countries and rounds.

Because all three surveys provide auxiliary variables, available for both respondents and nonrespondents, the assessment under the fixed response model is a convenient way to become acquainted with nonresponse bias.

Table 1.1 gives an overview of the full-sample distribution of nine auxiliary variables in the FHS, compared with the distributions for respondents only. With these full-sample and respondent-only parameters, biases and contrasts can be easily obtained. For example, the bias with regard to the percentage of age category <30 is $8.50\% - 9.57\% = -1.07\%$. Provided that the response rate is 67.17%, the contrast is the bias divided by the nonresponse rate or $-1.07\%/32.83\% = -3.26\%$, implying that the percentage of the category <30 years old among the nonrespondents is $8.50\% - (-3.26\%) = 11.76\%$. Contrasts or nonrespondent-only parameters are not reported in Table 1.1. The table shows that some variables are severely affected due to nonresponse. Multi-unit dwellings in particular are strongly under-represented in the FHS. Families with a female head are also under-represented. Some other categories show moderate traces of under-representation: young and old family heads, poor-quality buildings, buildings with small frontage, unoccupied and/or neglected neighbor-

Table 1.1: Overview of nonresponse bias for some auxiliary variables of the FHS ($n = 7770, \bar{r} = 67.17\%$)

	Full sample	Respondents only		Full sample	Respondents only
*age class family head*****			*unoccupied or neglected houses in neighbourhood*****		
< 30	9.57%	8.50%	none	68.99%	71.03%
31-40	18.19%	18.12%	unoccupied or neglected	21.60%	20.22%
41-50	20.63%	20.80%	unoccupied and neglected	9.41%	8.75%
51-60	18.15%	19.66%			
61-70	14.54%	15.46%	*type of area*****		
71-80	13.23%	12.96%	exclusively houses	42.09%	43.95%
> 80	5.69%	4.50%	mixed use of buildings	54.14%	51.94%
			rural area	3.78%	4.11%
*male family head*****	74.39%	76.68%	*no green elements in area*	37.88%	37.58%
*overall building quality according to expert report*****			*construction year of building*****		
in good repair	64.83%	65.83%	<1919	5.78%	5.59%
minor repair needed	27.12%	27.14%	1919-1945	15.31%	14.65%
out of repair	8.05%	7.02%	1946-1960	19.48%	18.59%
			1961-1970	18.23%	17.97%
*front width building**			>1970	41.20%	43.21%
<4 m	1.35%	1.20%			
4-6 m	17.12%	16.41%	*Multi-unit building*****	19.43%	15.19%
>6 m	81.53%	82.38%			

χ^2-test(H_0:respondents=nonrespondents) $^{\dagger}: p < 0.1; ^{*}: p < 0.05; ^{**}: p < 0.01; ^{***}: p < 0.001; ^{****}: p < 0.0001$

hoods, older buildings, and areas with mixed purposes (e.g. both living and working or shopping).

Similar to the FHS, manifest effects due to nonresponse can be found in Table 1.2, where the Dutch GPS is documented. The percentage of non-natives is underestimated in the respondent sample by 2.6% and the percentage of listed phone numbers in the full sample is overestimated by more than 4%. Further, males and non-working people or those receiving social allowances are slightly under-represented, as well as single people and single-parent households. Densely populated areas and neighborhoods with lower house values seem to be rather under-represented.

Interviewers fielding the European Social Survey are asked to score their observations about some visual aspects of the neighborhood and the buildings in which their target households or individuals live. Recording information during data collection is a practice that has attracted more attention recently and has provided many nonresponse researchers with paradata that can be used to assess the fieldwork design and monitor the

Table 1.2: Overview of nonresponse bias for the auxiliary variables of the GPS ($n = 32019, \bar{r} = 58.69\%$)

	Full sample	Respondents only		Full sample	Respondents only
Allowance received			*Average house value in neighbourhood (in 1000€)*****		
Disability***	6.65%	6.21%	< 75	6.23%	4.70%
Social****	4.21%	2.91%	75-150	35.91%	33.59%
Unemployment	2.53%	2.53%	150-300	49.11%	52.53%
			>300	8.76%	9.17%
*Type of household*****			*Urbanisation*****		
single	16.97%	14.34%	very strong	16.73%	12.37%
couple without children	33.01%	33.58%	strong	24.11%	23.46%
couple with children	44.12%	47.09%	fairly	20.78%	21.79%
single parent	5.05%	4.41%	little	21.94%	23.74%
other	0.85%	0.58%	not	16.44%	18.63%
*age class*****					
< 30	19.80%	19.06%	*Male**	49.16%	48.68%
31-40	21.72%	22.10%	*has job?*****	48.97%	47.39%
41-50	19.31%	20.26%	*Non-native*****	14.69%	12.09%
51-60	15.93%	15.70%	*listed phone number*****	77.04%	82.32%
61-70	11.48%	11.75%			
> 70	11.76%	11.14%			

Significance test χ^2-test(H_0:respondents=nonrespondents) $^{\dagger}: p < 0.1; ^{*}: p < 0.05; ^{**}: p < 0.01; ^{***}: p < 0.001; ^{****}: p < 0.0001$

ongoing operations. In the European Social Survey, attention has been paid to training interviewers to take greater care with these observations. In the fifth round of the ESS, the interviewers were asked to record the following information about the dwellings and the neighborhood:

- `FLAT`: The sample unit is believed to live in an apartment. (0 = no; 1 = yes)
- `VANDA`: 'How common is vandalism, graffiti or deliberate damage to property?' (1 = Very common, ..., 4 = not at all common)
- `LITTER`: 'In the immediate area, how common is litter or rubbish lying around?' (1 = Very common, ..., 4 = not at all common)
- `PHYS`: 'In what physical state are the buildings or dwellings in this area?' (1 = In a very good state, ..., 5 = Very bad state)
- `ENTRY_PHONE`: 'Before reaching the (target) respondent's individual door, is there an entry phone?' (0 = no; 1 = yes)

- `GATE`: 'Before reaching the (target) respondent's individual door, is there a locked gate or door'? (0 = no; 1 = yes)

The differences between the respondent-only percentage (or mean) and the full-sample counterpart for these six variables are shown in Table 1.3. Only countries for which at least 95% of the observable data was recorded are shown in the table. This means that the reported gross sample size is somewhat lower than reported in the ESS documentation. Notwithstanding a few exceptions, households living in flats or dwellings where the interviewer had to cope with obstacles such as locked gates and doors or entry phones seem to be less inclined to participate in the ESS. With regard to the indicators for the observed quality of the housing unit and the neighborhood, there is no clear overall nonresponse effect for all the countries. In countries such as Belgium and Denmark, nonresponse seems to be related to poor neighborhoods and housing quality. In other countries, these relationships are less explicit or even in the opposite direction. These latter three variables may be deemed more prone to interviewer subjectivity, whereas the first three variables (`FLAT,ENTRY_PHONE and GATE`) instead reflect a factual position.

The possibility cannot be completely excluded that interviewer observations depend on the response outcome, rather than the other way round. In particular, if interviewers do not run the risk of their observations being verified by other interviewers or another person, they may falsely use obstacles such as entry phones, locked gates, or poor-quality neighborhoods in order to justify unsuccessful contact attempts. Using such interviewer observations is therefore not always appropriate in order to assess nonresponse effects, and may even be hazardous to use for correction purposes such as weighting. In this respect, it is advisable therefore to also collect neighborhood and dwelling information that is observed independently of the operating interviewer.

With regard to the ESS3-BE, a severe deviation between the full sample and respondent-only sample can be found with regard to the type of dwelling: apartment dwellers are strongly under-represented (see Table

Table 1.3: Overview of nonresponse bias for interviewer observations, ESS5

	respondent only %		full sample %		respondent only (mean)		full sample (mean)
Belgium (BE) ($\bar{r}$: 52%,n: 3199)							
FLAT	13.81	****	18.74	PHYS	1.87	****	1.96
ENTRY_PHONE	20.89	****	26.48	LITTER	3.76	***	3.72
GATE	16.14	**	18.32	VANDA	3.84	****	3.81
Bulgaria (BG) ($\bar{r}$: 76%, n: 2997)							
FLAT	44.75		45.18	PHYS	2.27		2.27
ENTRY_PHONE	19.23		19.59	LITTER	3.20		3.19
GATE	56.53	****	58.89	VANDA	3.56	**	3.54
Cyprus (CY) ($\bar{r}$: 68%, n: 1564)							
FLAT	21.54	****	27.02	PHYS	2.01	****	2.08
ENTRY_PHONE	27.13	****	30.35	LITTER	3.74	*	3.73
GATE	38.98		38.48	VANDA	3.83	†	3.82
Denmark (DK) ($\bar{r}$: 54%, n: 2856)							
FLAT	21.47	****	26.37	PHYS	1.76	****	1.90
ENTRY_PHONE	15.92	****	19.38	LITTER	3.90	****	3.86
GATE	9.26	****	12.07	VANDA	3.95	***	3.93
Spain (ES) ($\bar{r}$: 66%, n: 2727)							
FLAT	62.39	***	64.42	PHYS	2.11	**	2.14
ENTRY_PHONE	73.47	****	74.73	LITTER	3.78	*	3.77
GATE	28.81	****	28.13	VANDA	3.79		3.78
Greece (GR) ($\bar{r}$: 64%, n: 4230)							
FLAT	50.20	*	50.53	PHYS	2.02	****	2.08
ENTRY_PHONE	46.23	**	47.54	LITTER	3.77	**	3.75
GATE	33.22	****	35.32	VANDA	3.88		3.88
Hungary (HU) ($\bar{r}$: 59%, n: 2612)							
FLAT	27.97		29.12	PHYS	2.21		2.23
ENTRY_PHONE	30.69	**	32.81	LITTER	3.58	*	3.56
GATE	72.26	**	74.46	VANDA	3.69	*	3.67
Israel (IL) ($\bar{r}$: 71%, n: 3230)							
FLAT	64.58	**	62.65	PHYS	2.12	*	2.09
ENTRY_PHONE	30.95	†	31.49	LITTER	2.94		2.96
GATE	21.23	***	22.94	VANDA	2.99	*	3.01
Portugal (PT) ($\bar{r}$: 73%, n: 3264)							
FLAT	43.23	***	45.57	PHYS	2.37		2.36
ENTRY_PHONE	51.62	****	55.67	LITTER	3.78	****	3.80
GATE	59.76	****	62.62	VANDA	3.82	***	3.84
Russian Federation (RU) ($\bar{r}$: 65%, n: 3982)							
FLAT	69.40	****	76.02	PHYS	2.59		2.58
ENTRY_PHONE	55.18	****	63.39	LITTER	2.95	****	2.92
GATE	31.72	****	28.73	VANDA	3.56	****	3.53

Significance test χ^2-test or $t-test$ (H_0: respondents=nonrespondents)

$^{\dagger}: p < 0.1; ^{*}: p < 0.05; ^{**}: p < 0.01; ^{***}: p < 0.001; ^{****}: p < 0.0001$

Table 1.4: Overview of nonresponse bias for the auxiliary variables of the ESS3-BE ($n = 2927, \bar{r} = 61.34\%$)

	Full sample	Respondents only		Full sample	Respondents only
*Age class***			*Percentage of non-Belgians in municipality (in %)****		
< 21	8.37%	9.52%	< 2	19.67%	22.88%
21-40	31.18%	29.15%	2-5	30.19%	32.73%
41-60	35.00%	36.47%	5-15	32.69%	31.02%
> 60	25.45%	24.86%	> 15	17.45%	13.37%
*Population density in municipality (in inh./km^2)****			*Average annual per capita income in municipality (in €)****		
< 200	11.63%	12.43%	< 12000	17.73%	14.85%
201-400	26.04%	29.04%	12000-14000	37.95%	37.90%
401-700	18.01%	19.97%	14000-16000	32.69%	35.42%
701-2500	32.41%	30.80%	> 16000	11.63%	11.83%
>2500	11.91%	7.76%			
Male	48.72%	46.64%	*Multi-unit building****	18.07%	11.61%
*Region****			*Neighbourhood quality****		
Flanders	58.73%	62.87%	poor	21.73%	17.55%
Brussels	9.42%	5.50%	good	30.69%	33.39%
Wallonia	31.86%	31.63%	excellent	47.58%	49.06%

Significance test χ^2-test(H_0:respondents=nonrespondents) $^{\dagger}: p < 0.1; ^{*}: p < 0.05; ^{**}: p < 0.01; ^{***}: p < 0.001; ^{****}: p < 0.0001$

1.4). Furthermore, there seems to be some nonresponse bias at the municipality level: there is an underestimation of high population density municipalities or municipalities counting many immigrants. Lower-income areas are also underepresented. This may also be the reason why the proportions for Brussels or the quality of the neighborhoods (assessed by the interviewers) in particular are biased downwards.

In order to completely fit within the framework of equations as presented above, the auxiliary variables of the ESS3-BE will be interpreted as continuous variables, for which the correlation with the response outcome r can be obtained.

$$\begin{aligned} bias_{\bar{y}_r} &= \frac{corr_{ry}\sigma_r\sigma_y}{\bar{r}} \\ &= \frac{corr_{ry}\sqrt{\bar{r}\,(1-\bar{r})}\sigma_y}{\bar{r}} \end{aligned}$$

Table 1.5: Overview of the nonresponse bias and standardized nonresponse bias for the auxiliary variables of the ESS3-BE ($n = 2927, \bar{r} = 61.34\%$)

Auxiliary variable	$corr_{ry}$	$\hat{S}_r$	$\hat{S}_y$	$\bar{r}$	$bias_{\bar{y}_r}$	$st.bias_{\bar{y}_r}$
Age	-0.0675***	0.4851	18.5017	0.6134	-0.9750	-0.0527
Gender(♂=1;♀=0)	-0.0269	0.4851	0.4996	0.6134	-0.0105	-0.0210
Housing type(flat=1;no flat=0)	-0.1616****	0.4851	0.3691	0.6134	-0.0466	-0.1262
Population density (÷1000)	-0.1532****	0.4851	2.8635	0.6134	-0.3427	-0.1197
Average income (÷1000)	0.0712***	0.4851	0.1841	0.6134	0.0102	0.0556
% non-Belgians	-0.1479****	0.4851	0.0821	0.6134	-0.0095	-0.1155
Flanders(yes=1;no=0)	0.0832****	0.4851	0.4906	0.6134	0.0319	0.0649
Brussels(yes=1;no=0)	-0.1471****	0.4851	0.2826	0.6134	-0.0325	-0.1149
Wallonia(yes=1;no=0)	0.0017	0.4851	0.4649	0.6134	0.0006	0.0013

$corr_{ry}$ is significant at $^{\dagger}: p < 0.1; ^{*}: p < 0.05; ^{**}: p < 0.01; ^{***}: p < 0.001; ^{****}: p < 0.0001$

The resulting biases can also be standardized, allowing for a comparison among all auxiliary variables. Therefore, making $\sigma_y = 1$ reduces the bias expression to

$$\begin{aligned} st.bias_{\bar{y}_r} &= \frac{corr_{ry}\sigma_r}{\bar{r}} \\ &= \frac{corr_{ry}\sqrt{\bar{r}\,(1-\bar{r})}}{\bar{r}} \end{aligned}$$

Table 1.5 provides estimates of bias and standardized bias for all seven auxiliary variables of the ESS3-BE. It seems that the type of housing, the percentage of non-Belgians in the municipality and the Brussels region are the most severely biased variables. For example, the estimate for the percentage of non-Belgians among respondents only is 0.1155 standard deviations smaller than the same parameter estimate for the full sample. Only gender and the Wallonia region appear not to be substantially or significantly biased due to nonresponse.

What is quite remarkable in this set of examples is the ease with which traces of bias can be found in the different surveys. For some variables, such as the type of housing (apartment dwellers being the strongly under-represented category) in Belgium and Flanders, or the fixed telephone lines in the Netherlands (being strongly over-represented), nonresponse bias may even be considered a severe threat to survey quality. Although these auxiliary variables are not really the information that a survey seeks

Table 1.6: Explaining the response outcome r by a set of auxiliary variables, logistic regression parameters, ESS3-BE ($n = 2927, \bar{r} = 61.34\%$)

	Parameter Estimate		Parameter Estimate
*Age class***		*Percentage of non-Belgians in municipality (in %)*	
< 21	0.2643**	< 2	0.1360
21-40	-0.1536*	2-5	0.0693
41-60	0.0361	5-15	-0.0730
> 60	-0.1468*	> 15	-0.1322
Population density in municipality (in inh./km²)		*Average annual per capita income in municipality (in €)*	
< 200	-0.0387	< 12000	0.1509
201-400	0.0733	12000-14000	-0.0147
401-700	0.1242	14000-16000	0.0259
701-2500	-0.1215	> 16000	-0.1620†
>2500	-0.0373		
Male	-0.0967**	*Multi-unit building*	-0.3612****
Region†		*Neighbourhood quality**	
Flanders	0.2550*	poor	-0.1568*
Brussels	-0.3514*	good	0.1647**
Wallonia	0.0964	excellent	-0.0078

† : $p < 0.1$;* : $p < 0.05$;** : $p < 0.01$;*** : $p < 0.001$;**** : $p < 0.0001$

Effect coding is applied: $\sum \beta_j = 0$. Parameter estimates for redundant categories are also provided

to target (except perhaps the Flemish Housing Survey, for which the type of housing is an important variable), their biases should at least be considered as a strong indication of how target statistics may be affected.

Leaving the fixed response model behind, the random response model framework will now be used to explore how target variables may occasionally be affected by nonresponse. Therefore, a vector of response propensities ρ_i has to be determined based on auxiliary variables. Then, the correlation between the propensities and the target variables needs to be estimated. Eventually, the estimates of bias are obtained by equation 1.4.

$$bias_{\bar{y}_r} = \frac{corr_{\rho y}\sigma_\rho\sigma_y}{\bar{\rho}}$$

For the third round of the ESS in Belgium, a response rate $\bar{\rho}$ of 0.6134 has been realized, while σ_ρ is 0.1071. The logistic model that is used

to obtain the propensities is shown in Table 1.6. In particular, the area level variables and 'region' seem to lose some of their explanatory power in the multiple logistic regression context as compared with the bivariate analyses. This may be due to their relatively strong mutual relationships and their strong association with the type of dwelling.

Table 1.7 gives an overview of how the target variables relate to the estimated response propensities. The table can be read in the following way: for example, for the item 'Feeling of safety of walking alone in the local area after dark', the term between brackets *'unsafe'* indicates that higher values on the scale relate to feelings of being unsafe. This item is negatively correlated to the estimated propensities, suggesting that individuals who are more inclined to participate in the survey report lower feelings of being unsafe when walking alone in a local area after dark. This implies that the level of feeling unsafe is slightly underestimated, as indicated by the standardized bias $'st.bias_{\bar{y}_r} = -0.0102'$.

When interpreting all the items in Table 1.7, survey response seems to be (strongly) related to living in the countryside, feelings of happiness, good health, good feelings about the household income, intensive Internet use, high levels of trust in others, democracy, economy, politics, and personal life.

A weakness of this method for assessing nonresponse bias for target variables is probably the poor link between the estimated propensities based on auxiliary information and the presumed true propensities. Indeed, the propensities are determined as conditional on a very selective set of auxiliary variables, implying that these propensities only partially reflect all the possible dimensions of nonresponse mechanisms. This may explain why the standardized biases under the random response model, provided in Table 1.7, are much smaller than their counterparts under the fixed model (see Table 1.5).

Furthermore, the estimated response propensities are linear combinations of auxiliary variables. This means that the possibility cannot be ruled out that $corr_{\rho y}$ is a spurious effect, attributable to the linear combinations of auxiliary variables, rather than to the supposed nonresponse effects. As

an example, one may think of a sample for which the response propensities have been measured, conditional on the age of the (non)respondents. A target variable that is not biased at all may still correlate with the propensities, because of its relationship with age, rather than with (non)response. ■

Before moving on to the next section, it might be worth considering that measurements such as the variance of response propensities, as well as estimates of bias under the fixed model, are never exactly equal to zero. Even in the absence of selective effects (MCAR), there will still be differences between respondents and nonrespondents, implying that σ_ρ and $E(bias)$ are always somewhat biased themselves, simply because of random fluctuations. This might make the damage due to nonresponse somewhat worse than it really is, particularly if the sample size is small and many auxiliary variables are used to determine the propensities. Larger sample sizes or fewer auxiliary variables used in the random and fixed response models will mitigate the overestimation of σ_ρ. In regression analysis, this problem is also known as the difference between the coefficient of determination R^2 and its adjusted counterpart R^2_{adj}.

$$R^2_{adj} = 1 - \left(1 - R^2\right) \frac{n-1}{n-p-1} = 1 - \frac{SS_{err}}{SS_{tot}} \frac{df_{tot}}{df_{err}} \tag{1.9}$$

where p is the number of parameters in the regression model and n is the total sample size. Under the fixed model, R^2 can be considered as $corr^2_{ry}$, that is expected to be overestimated. In this particular case, p=1, as only one of the parameters needs to be estimated in order to distinguish between respondents and nonrespondents. As a result, 1.3 can be rewritten as

$$|bias_{a\bar{u}x_r}| = \frac{\sqrt{1 - \left(1 - corr^2_{r,aux}\right) \frac{n-1}{n-2}} \sigma_r \sigma_{aux}}{\bar{r}} \tag{1.10}$$

Table 1.7: Overview of nonresponse bias of target variables under the random response model, ESS3-BE ($n = 1798$)

Target variable	$corr_{\rho y}$	σ_y	$st.bias_{\bar{y}_r}$
Feeling of safety of walking alone in local area after dark (unsafe)	-0.06*	0.70	-0.0102
Borrow money to make ends meet, difficult or easy (easy)	0.06*	1.26	0.0104
Domicile, respondent's description (countryside)	0.51****	1.08	0.0882
European Union: European unification go further or gone too far (go further)	0.07**	2.60	0.0119
Gays and lesbians free to live life as they wish (disagree)	-0.08***	1.05	-0.0144
Government should reduce differences in income levels (disagree)	0.02	1.09	0.0042
How happy are you (happy)	0.14****	1.58	0.0240
Subjective general health (bad)	-0.10****	0.80	-0.0179
Feeling about household's income nowadays (difficult)	-0.22****	0.84	-0.0391
Immigration bad or good for country's economy (good)	0.00	2.24	0.0002
Important to care for nature and environment (unimportant)	0.07**	0.89	0.0122
Important to make own decisions and be free (unimportant)	0.05*	0.98	0.0092
Important to be rich, have money and expensive things (unimportant)	0.01	1.03	0.0011
Important to live in secure and safe surroundings (unimportant)	0.02	1.10	0.0037
Important to follow traditions and customs (unimportant)	0.06**	1.14	0.0111
Country's cultural life undermined or enriched by immigrants (enriched)	0.02	2.23	0.0044
Immigrants make country worse or better place to live (better)	0.08***	2.07	0.0144
Important to think new ideas and being creative (unimportant)	-0.00	1.13	-0.0008
Important that people are treated equally and have equal opportunities (unimportant)	0.03	0.84	0.0058
Important to help people and care for others well-being (unimportant)	-0.00	0.81	-0.0004
Important to be humble and modest, not draw attention (unimportant)	0.06*	1.01	0.0103
Important to show abilities and be admired (unimportant)	-0.03	1.14	-0.0051
Placement on left right scale (right)	0.03	2.01	0.0047
Personal use of internet/e-mail/www (often)	0.14****	3.04	0.0243
Newspaper reading, total time on average weekday (often)	-0.09***	1.22	-0.0151
Politics too complicated to understand (complicated)	0.07**	1.10	0.0118
Making mind up about political issues (easy)	-0.02	0.94	-0.0033
Most people try to take advantage of you, or try to be fair (fair)	0.06*	2.07	0.0099
Most of the time people helpful or mostly looking out for themselves (helpful)	0.15****	2.14	0.0267
Most people can be trusted or you can't be too careful (trust)	0.08***	2.28	0.0140
How often pray apart from at religious services (never)	0.08***	2.32	0.0138
Ban political parties that wish overthrow democracy (no ban)	0.11****	1.14	0.0185
Radio listening, total time on average weekday (often)	0.06**	2.74	0.0108
How often attend religious services apart from special occasions (never)	0.04	1.39	0.0067
How religious are you (very religious)	0.01	2.95	0.0023
Take part in social activities compared to others of same age (often)	0.02	1.05	0.0031
How often socially meet with friends, relatives or colleagues (often)	0.05†	1.45	0.0080
Modern science can be relied on to solve environmental problems (no)	0.02	0.96	0.0036
How satisfied with the way democracy works in country (satisfied)	0.06*	2.08	0.0100
How satisfied with present state of economy in country (satisfied)	0.11****	1.97	0.0197
State of education in country nowadays (satisfied)	0.21****	2.04	0.0372
How satisfied with the national government (satisfied)	0.07**	1.99	0.0119
State of health services in country nowadays (satisfied)	0.05*	1.67	0.0089
How satisfied with life as a whole (satisfied)	0.15****	1.87	0.0269
Terrorist suspect in prison until police satisfied (no)	-0.03	0.92	-0.0048
Torture in country never justified even to prevent terrorist attack (justified)	0.04†	1.28	0.0070
Trust in the European Parliament (trust)	0.05*	2.17	0.0093
Trust in the legal system (trust)	0.09****	2.38	0.0152
Trust in the police (trust)	0.09****	2.17	0.0156
Trust in politicians (trust)	0.11****	2.11	0.0186
Trust in country's parliament (trust)	0.07**	2.13	0.0122
Trust in political parties (trust)	0.12****	2.17	0.0217
Trust in the United Nations (trust)	0.04	2.28	0.0066
TV watching, total time on average weekday (often)	-0.10****	1.99	-0.0166

$corr_{\rho y}$ is significant at $^{\dagger}: p < 0.1; ^{*}: p < 0.05; ^{**}: p < 0.01; ^{***}: p < 0.001; ^{****}: p < 0.0001$

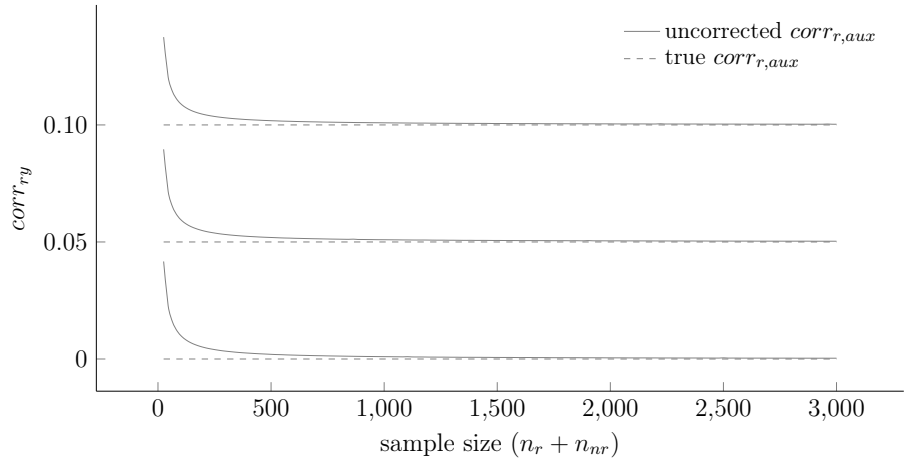


Figure 1.5: Overestimation of $corr_{ry}$ as a function of sample size

Figure 1.5 shows how an increase in sample size gradually removes overestimation of the bias. For different levels of $corr_{ry} = 0$, 0.05 and 0.010, the overestimation seems to become a minor problem for large sample sizes (for example, > 1000).

Similarly, Shlomo, Skinner, Schouten, Carolina, and Morren (2009) have developed a procedure to correct the bias surplus for the random response model. Even when the response outcome and the auxiliary variables are independent, modeling response propensities through a (logistic) regression model still generates variance in the predicted response outcomes. This variance will be overestimated to an even greater degree as the number of parameters in the (logistic) regression model increases or as the sample size decreases.

1.2.4 Bias assessment for location parameters and other types of parameters

The indicators of bias and contrast as discussed so far, refer particularly to location parameters such as the mean of a variable. Peytchev, Carley-Baxter, and Black (2011) correctly note that other types of parameters can also be subjected to nonresponse bias assessment. These may include variances, correlations, or even multivariate distributions. In fact, many

statistics of interest go beyond descriptive measures as means and proportions, and instead focus on the relationships between the target variables.

As an example, consider a case where the bias with regard to a correlation coefficient between variables a and b needs to be assessed. In the context of the fixed response model, the bias assessment is relatively straightforward:

$$\begin{aligned} bias_{corr_{ab}} &= = corr_{ab,r} - corr_{ab,f} && \text{(1.11a)} \\ &= \left(\frac{n_{nr}}{n}\right)(corr_{ab,r} - corr_{ab,nr}) && \text{(1.11b)} \end{aligned}$$

This bias assessment might also be translated to the random interpretation of the response model, by including the inverse response propensities as individual weights, so the bias with regard to $corr_{ab}$ can now be expressed as the difference between the unweighted and weighted correlation coefficient:

$$bias_{corr_{ab}} = corr_{ab,r} - corr_{ab,w}, \tag{1.12}$$

where $w = 1/\rho$ (sometimes weights are also determined by $w = \bar{\rho}/\rho$).

In fact, comparing the weighted and unweighted parameters is always a valuable alternative with which to assess nonresponse bias. Hence, bias assessment with regard to the mean of a variable as discussed in section 1.2.3, can also be carried out by comparing the unweighted respondent-only mean and its weighted counterpart. If the inclusion probabilities (=response propensities in this context) are correctly specified, the *Horvitz-Thompson* estimator (1952) is unbiased and a bias assessment is consequently possible.

The advantage of propensity weighting is the relative ease with which it can be obtained, at least when auxiliary information is readily available. Propensity weighting tends to operate locally: each responding sample is individually weighted, that is, one parameter per individual. As a result, all distributional aspects (mean, variance, etc.) can be properly reconstructed. Conversely, bias equation 1.4 as presented in section 1.2.3 only

uses one linear parameter: $corr_{\rho y}$ and therefore instead operates globally. This implies less flexibility as compared with local propensity weighting and will therefore not properly affect all distributional aspects of the target variables or statistics. Accordingly, local propensity weighting also corrects the variance (and not only the mean) of the target variable, which may be particularly interesting in the case of heteroscedasticity or $\sigma^2_{y,r} \neq \sigma^2_{y,nr}$. When using equation 1.4 under the random model, using only one parameter $corr_{\rho y}$, usually implies the correction of only the mean, leaving the variance (and other distributional aspects) unaltered.

Later on, adjustment techniques will be discussed in more detail. This paragraph instead focuses on the effects of nonresponse on parameters other than location parameters. Relevant literature about surveys has not paid as much attention to variances or correlations as compared with nonresponse effects on means and proportions (Peytchev, 2013). Only a few studies deal with the effects of nonresponse on measurements of association. Lepkowski and Couper (2002) found small traces of bias in associations in a mail survey Martikainen, Laaksonen, Piha, and Lallukka (2007) report that although two variables were biased with regard to their means, the association between occupational social class and sickness absence was not significantly biased.

Data analysis The lack of empirical evidence about the nonresponse effect on measurements of association will be somewhat remedied in this data analysis section. The fixed response model will be used to examine correlations in the full sample and compare these with the respondent-only sample. Of course, this can only be achieved when applying the comparison to auxiliary variables that are available for both respondents and nonrespondents. The random response model will therefore be used to look at the correlations between target variables (only available for respondents) and consider their differences before and after weighting corrections informed by a set of auxiliary variables.

Data from the third round of the Belgian ESS as illustrated in Table 1.5 will be used again for the assessment under the fixed response model.

Table 1.8: Overview of nonresponse bias for correlations of target variables under the fixed response model, ESS3-BE ($n = 3249$)
Upper-right triangle: $corr_{ab,f}$; Lower-left triangle: $corr_{ab,r} - corr_{ab,f}$

	Age	Gender	Flat	Density	non-Belgian	Income	Flanders	Brussels	Wallonia
Age		-0.0478	-0.0043	-0.0526	-0.0524	0.0564	0.0407	-0.0507	-0.0112
Gender	0.0277		0.0112	0.0193	0.0250	-0.0181	0.0367	0.0103	-0.0453
Flat	0.0095	-0.0195		0.3883	0.3672	-0.1796	-0.1100	0.3936	-0.1305
Density	-0.0120	-0.0075	-0.0566		0.7723	-0.4467	-0.3259	0.8593	-0.1943
non-Belgians	0.0068	-0.0326	-0.0806	-0.0546		-0.5534	-0.4744	0.7233	0.0478
Income	0.0319	0.0270	0.0715	0.0845	0.0425		0.4890	-0.3951	-0.2690
Flanders	0.0059	0.0360	0.0543	0.0553	0.0263	-0.0460		-0.3846	-0.8156
Brussels	-0.0131	-0.0134	-0.0893	0.0022	-0.0797	0.0837	0.0707		-0.2205
Wallonia	-0.0060	-0.0287	0.0392	0.0530	0.1022	-0.0386	-0.0695	0.0564	

The triangle to the upper-right in Table 1.8 represents the full-sample pairwise correlation between any two variables. The lower-left triangle shows how strongly the respondent-only correlations deviate from their full-sample counterparts. For example, the correlation between age and gender is -0.0478 in the entire sample; among the respondents only, this correlation is -0.0478 + 0.0277 = -0.0201. Most correlations only seem to be mildly biased. However, for some particular estimates, the difference between the full-sample and the respondent-only estimate becomes more critical, for example some combinations of *'income'* and *'flat'*. The correlation between flat and income in the full sample is -0.1796, suggesting that individuals living in apartments usually reside in municipalities with lower average incomes. However, when only looking at the respondents, there is an upward bias of 0.0715, since the correlation among respondents is only -0.1081. This means that the respondent-only sample underestimates the association between average municipality income and type of dwelling. Indeed, in the full sample the difference with regard to average income between apartment dwellers (€13,022) and non-apartment dwellers (€13,889) is €867. In the respondent-only subset, the difference between apartment dwellers (€13,336) and non-apartment dwellers (€13,931) is only €595.

Table 1.9: Overview of nonresponse bias for correlations of target variables under the random response model, ESS3-BE ($n_r = 1798$)
Upper-right triangle: $corr_{ab,f}$; Lower-left triangle: $corr_{ab,r} - corr_{ab,w}$

	AESFDRK	DOMICIL	HAPPY	HEALTH	HINCFEL	IMUECLT	NETUSE	PRAY	SCLMEET
AESFDRK		-0.1188	-0.0959	0.1334	0.1109	-0.1503	-0.1633	-0.0722	-0.0850
DOMICIL	0.0089		0.0801	-0.0080	-0.1304	-0.0866	-0.0486	0.0199	-0.0216
HAPPY	-0.0125	0.0202		-0.2701	-0.2805	0.1431	0.0724	0.0091	0.1034
HEALTH	0.0090	0.0035	-0.0043		0.2204	-0.2070	-0.3031	-0.1389	-0.1372
HINCFEL	0.0123	-0.0108	-0.0150	0.0111		-0.1474	-0.2160	-0.0945	-0.0768
IMUECLT	-0.0076	-0.0102	0.0082	-0.0075	-0.0004		0.2903	0.0525	0.1468
NETUSE	-0.0102	-0.0143	0.0038	-0.0072	0.0003	-0.0000		0.2502	0.2304
PRAY	-0.0042	0.0022	0.0080	0.0019	0.0029	-0.0158	-0.0082		0.0078
SCLMEET	-0.0052	-0.0048	0.0061	-0.0065	0.0016	-0.0061	-0.0019	-0.0025	

AESFDRK	Feeling of safety of walking alone in local area after dark
DOMICIL	Domicile, respondent's description
HAPPY	How happy are you
HEALTH	Subjective general health
HINCFEL	Feeling about household's income nowadays
IMUECLT	Country's cultural life undermined or enriched by immigrants
NETUSE	Personal use of internet/e-mail/www
PRAY	How often pray apart from at religious services
SCLMEET	How often socially meet with friends, relatives or colleagues

For the random response model assessment, a subset of the variables used in Table 1.3 will be recovered. Correlations between target variables are measured before and after weighting, based on a set of auxiliary variables that are used to determine the response propensities, from which in turn the weight score can be derived. The results are presented in a similar way to the analysis under the fixed response model. The upper-right triangle represents the weighted correlations and the lower-left triangle shows to what extent the respondent-only (unweighted) correlations differ from their weighted counterparts. As has already been found for the location parameters, the biases seem to be much milder than the biases measured under the fixed response model. Again, the incompleteness of the set of auxiliary variables in order to predict the true propensities is probably the underlying reason.

When comparing the bias among correlations (Tables 1.8 and 1.9) and the tables showing biases for means and proportions (see section 1.2.3), it is hard to assess whether or not correlations are more prone to the effects of nonresponse than location parameters, because correlations are

measured on a different scale (from -1 to 1) than the location parameters (interpreted on a standard normal scale). The next sections will therefore further quantify nonresponse damage in order to assess the sensitivity to nonresponse for the different kinds of parameters. ■

1.2.5 From a specific parameter to the level of the survey as a whole

So far, biases and contrasts have only been determined for individual parameters, for example, the difference between the age means of the respondents and nonrespondents, or the bias of a correlation between two target variables. Such indications of the response quality of a particular parameter may be valuable, but may also have their limitations.

Under the random response model, nonresponse bias is the result of the inequality of response propensities. Therefore, the underlying reason for a particular bias is usually to be found in the sample as a whole. Measuring the variance of response propensities therefore becomes a crucial element in the quality assessment of a complete realized sample. Observing the sample level may be particularly interesting for monitoring the fieldwork process, as the survey management may aim for the equality of response propensities at the close of the fieldwork, pursuing the absence of bias for any parameter to be estimated from the realized respondent set.

Under the fixed perspective, bias is measured for the auxiliary variables only, ignoring the bias with regard to target variables. Given the assumption that the biases for auxiliary variables are somehow indicative or representative of the target variables, one is able to determine the expected bias with regard to target survey statistics. Observing the sample as a whole under the fixed response model introduces the idea of a bias distribution that applies to particular target variables.

This shift from the perspective of one particular parameter to the obtained sample as a whole will also be an important bridge when discussing statistical inference from survey data in the presence nonresponse. This subject is discussed in section 1.2.6.

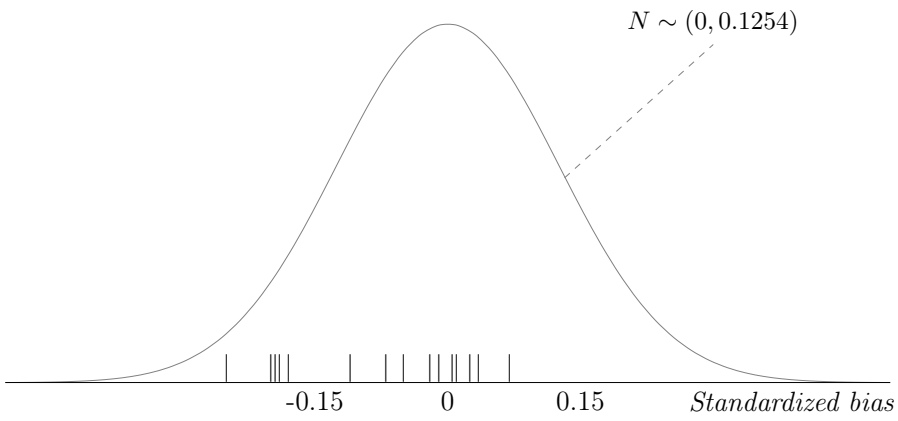


Figure 1.6: 15 nonresponse biases constitute a nonresponse bias distribution, fictitious data

More importantly, this jump will demonstrate that nonresponse bias is not only a fact, but instead is also a risk. As nonresponse is unobservable by definition, the actual nonresponse bias with regard to target variables remains concealed. Nonetheless, nonresponse is a threat, the risk of which needs to be taken into account. Circumstantial information about auxiliary variables may serve to determine that risk.

Fixed response model: a bias distribution

Suppose a survey sample has been completed where $n = 1000$ and $n_r = 500$. Fortunately, about 15 auxiliary variables are available from records or interviewer observations in the field. For each of the variables, the respondent-only mean and the full-sample mean are determined in order to obtain 15 biases. Provided that all 15 variables have first been standardized based on the full-sample mean and standard deviation, the resulting biases can be plotted as shown in Figure 1.6 (short vertical lines).

From the 15 empirically-measured biases, a latent distribution might be constructed. In this particular case, a normal distribution is assumed

with

$$\bar{bias} = 0 \tag{1.13}$$

$$S^2_{bias} = \frac{1}{p}\sum_{j=1}^{p}\left(\frac{a\bar{u}x_{p,r} - a\bar{u}x_{p,f}}{\sigma_{aux_{p,f}}}\right)^2, \tag{1.14}$$

where p is the total number of variables for which the means of respondents-only and the full sample can be compared.

The choice made to consider $\bar{bias} = 0$ is quite evident, as the absence of bias is an obvious reference point. Even more, the bias with regard to, for example, the proportion of males is the exact opposite of the bias with regard to the proportion of females in the obtained respondent sample, perfectly placing zero in the middle of all the biases. As a result, the denominator to obtain the variance has as many degrees of freedom as there are parameters, instead of the number of parameters minus one.

When shifting the perspective from an individual survey statistic to all statistics from the same survey, bias is no longer a single point, but instead becomes a distribution. The distribution expresses the nonresponse quality of the obtained sample as a whole and may be assumed to be indicative for the nonresponse effects that need to be taken into account when considering target variables. Very similar to the example of Figure 1.6, Peytchev and Biemer (2011), building on the work of Groves (2006) and Groves and Peytcheva (2008), have proposed the *Bias effect Size (BES_k)*, or an expectation of the bias with regard to some variable k, relative to its (population) standard deviations. That expectation can be obtained by simply determining the mean and variance of the biases of a set of auxiliary variables. The motivation is to fulfill the need to provide a 'concise, convenient and readily interpretable measure of the overall level of nonresponse bias for a survey' (Peytchev & Biemer, 2011). In their view, the bias effect size is the average of the absolute standardized biases or

$$BES = \frac{1}{p}\sum_{j=1}^{P}\left|\frac{a\bar{u}x_{p,r} - a\bar{u}x_{p,f}}{\sigma_{aux_{p,f}}}\right| \tag{1.15}$$

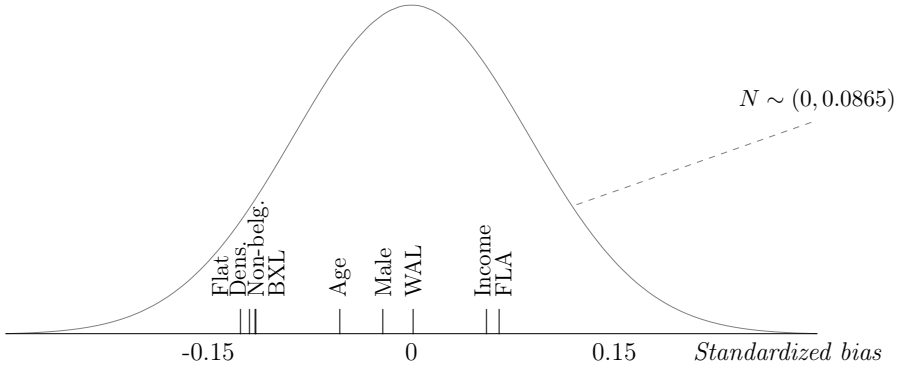


Figure 1.7: 9 nonresponse biases constitute a nonresponse bias distribution, ESS3-BE ($n = 2927$)

and is therefore a small variation on S_{bias} as shown in expression 1.14.

If the auxiliary variables represent the target variables reasonably well, the aspects of the bias distribution can be projected onto the target variables, allowing a realistic estimate of their unknown biases that should be taken into account.

It is obvious that equation 1.14 to obtain S^2_{bias} can also be extended to statistics other than means. Indeed, as suggested in section 1.2.4, separate estimates of S^2_{bias} can be calculated for correlations, regression parameters and so forth.

Data analysis To illustrate the relatively new idea of using a bias distribution in order to assess the quality of an obtained sample with regard to nonresponse, the results from Table 1.5 on page 37 are used. The last column of the table contains standardized bias estimates $\left(\left(a\bar{u}x_{p,r} - a\bar{u}x_{p,f}\right)/\sigma_{aux_{p,f}}\right)$. The average absolute bias (expression 1.15) among these nine biases is 0.0746. This means that, on average and assuming these auxiliary variables are representative of the target variables, it is expected that the target variables will also be subjected to this level of nonresponse bias. Suppose that a target variable indicates whether a respondent currently has a paid job or not and that in the full sample this percentage is 50%, the bias is then expected to be $0.0715 \times \sqrt{50\% \times 50\%} = 3.56\%$. The standard deviation of the biases (expression 1.14) is 0.0865.

Table 1.10: Overview of the unadjusted and adjusted standardized nonresponse bias, ESS3-BE ($n = 2927$)

Auxiliary variable	$corr_{r,aux}$	R^2	R^2_{adj}	$st.bias^2$	$st.bias^2_{adj}$
Age	-0.0675	0.0046	0.0042	0.0028	0.0026
Gender(♂=1;♀=0)	-0.0269	0.0007	0.0004	0.0004	0.0003
Housing type(flat=1;no flat=0)	-0.1616	0.0261	0.0258	0.0159	0.0157
Population density (÷1000)	-0.1532	0.0235	0.0232	0.0143	0.0141
Average income (÷1000)	0.0712	0.0051	0.0048	0.0031	0.0029
% non-Belgians	-0.1479	0.0219	0.0216	0.0133	0.0132
Flanders(yes=1;no=0)	0.0832	0.0069	0.0066	0.0042	0.0040
Brussels(yes=1;no=0)	-0.1471	0.0216	0.0213	0.0132	0.0130
Wallonia(yes=1;no=0)	0.0017	0.0000	-0.0003	0.0000	-0.0002
			$\sum$	0.0673	0.0657

It should be noted that these estimates of the standard deviations of biases may be slightly overestimated because of random sampling fluctuations. Even if nonresponse is completely random, there will still be small differences between respondents and the full sample. A possible way to deal with this problem is not to determine the biases by subtracting the full-sample mean from the respondent-only mean, but instead to employ expression 1.3 that uses the correlation between the variable and the 0-1 response outcome $corr_{r,aux}$. If it holds that

$$\begin{aligned} bias_{a\bar{u}x_r} &\approx \frac{corr_{r,aux}\sigma_r\sigma_{aux}}{\bar{r}} \\ bias^2_{a\bar{u}x_r} &\approx \frac{corr^2_{r,aux}\sigma^2_r\sigma^2_{aux}}{\bar{r}^2} \\ st.bias^2_{a\bar{u}x_r} &\approx \frac{corr^2_{r,aux}r(1-r)}{\bar{r}^2} \end{aligned}$$

$corr^2_{r,aux}$ can be replaced by its adjusted equivalent (R^2_{adj}) as already indicated in expressions 1.9 and 1.10 on page 40. Making use of this adjusted coefficient of determination allows the removal of the upward bias that is due to sampling variance. In this regard, Table 1.10 compares unadjusted and adjusted versions of the computation of $st.bias^2$.

Then, taking the square root of the average of all squared individual standardized biases results in S^2_{bias} and $S^2_{bias_{adj}}$. The unadjusted standard

deviation is $\sqrt{0.0673/9} = 0.0865$ as already shown. The adjusted equivalent equals $\sqrt{0.0657/9} = 0.0854$, which indicates a small downward correction. ■

Random response model: representative response

A bias or a contrast only refers to one single parameter of interest. However, the reason for possible damage to the parameters caused by nonresponse needs to be found in the sample as a whole. This becomes clear when looking the bias equation under the random interpretation.

$$bias_{\bar{y}_r} = \frac{corr_{\rho y}\sigma_\rho\sigma_y}{\bar{\rho}}$$

There are only two ways to ensure that no survey estimates are biased due to nonresponse: (1) obtain a response rate of 100%, or (2) make all response propensities equal, implying perfect representativeness.

Nonresponse avoidance (so that $\bar{\rho} = 1$), might at least theoretically be a secure method to provide immunity from nonresponse bias for all target variables and for all their distributional aspects. Therefore, the response rate might be a valuable tool to assess the potential for bias in surveys. However, because of the declining trend in response rates, the representativeness of the sample is attracting more attention.

The concept of representativeness has a long history in statistical and scientific literature, as well as non-scientific literature, and seems to cover a wide range of (sometimes even contradictory) meanings. Kruskal and Mosteller (1979a, 1979b, 1979c) studied the use of the term *'representative sample'* (or *'representative sampling'*) and found that it had been related to notions such as *'absence of selective forces'*, *'miniature of the population'*, *'typical or ideal case(s)'*, *'sample drawn with mathematical precision'* and *'coverage of the population'*.

Recently, Schouten et al. (2009) dredged up the notion of representativeness and specifically linked it to the variability of the response propensities (σ_ρ) as meant in the bias equation under the random response

model. The focus on representativeness has recently gained much interest, as there is a growing belief that response rates alone are only poor indicators of survey quality. Schouten et al. consider a sample to be representative of the whole population if all units or individuals within the population have an equal possibility of being included in the sample. Applied to nonresponse, representativeness implies that all individuals have an equal response probability $\bar{\rho}$, provided they have been selected from the sample frame or population ($s_i = 1$) (Shlomo, Skinner, Schouten, Bethlehem, & Zhang, 2008). Therefore:

$$\rho_i = P\left(r_i = 1 | s_i = 1\right) = \bar{\rho}, \forall i \tag{1.16}$$

This relatively strong claim for representativeness corresponds with MCAR and can be relaxed to a weaker counterpart, corresponding with MAR, where response propensities are allowed to differ, but only with regard to the categories of an auxiliary variable:

$$\bar{\rho}_h = \frac{1}{N_h} = \sum_{i=1}^{N_h} \rho_{hi} = \rho \quad for \quad h = 1, 2, \ldots, H \tag{1.17}$$

The more the individual ρ_i converge toward the constant $\bar{\rho}$, the more representative the sample will be and the more the risk of bias will be suppressed.

Schouten et al. (2009) define what they term the R-indicator as

$$R_\rho = 1 - 2\sigma_\rho. \tag{1.18}$$

This measures the extent to which an obtained sample deviates from perfect representativeness. The R-indicator does not refer to any specific target variable or estimate, but instead reflects the quality of the obtained sample as a whole. It is clear that the R-indicator $\in [0, 1]$ and that a value closer to one would be preferable. It should be noted that because the maximal possible variance of the propensities is restricted to $\bar{\rho}\left(1 - \bar{\rho}\right)$, the R-indicator depends strongly on the response rate. When the response rate

is close to 1 or 0, the R-indicator will become more favorable, as the potential for variance is restricted in these areas. The R-indicator is usually lower where $\bar{\rho}$ is close to 0.5. At this response rate, the maximal possible propensity variance is 0.5×0.5. Therefore, it is advisable to consider the R-indicator together with the response rate.

An R-indicator does not make any specific or explicit reference to nonresponse bias and therefore has a very distinct meaning. Accordingly (and very similar to the R-indicator), a measurement expressing the maximal absolute bias has been introduced (Schouten et al., 2009). This is the worst-case estimator for $bias_{\bar{y}_r}$ and assumes a perfect correlation between the response propensities and the target variables.

$$\begin{aligned} bias_{\bar{y}_r} &= \frac{corr_{\rho y}\sigma_\rho\sigma_y}{\bar{\rho}} \\ |bias_{\bar{y}_r}| &< \frac{\sigma_\rho\sigma_y}{\bar{\rho}}, \qquad where\ corr_{\rho y} = 1 \end{aligned} \tag{1.19}$$

Additionally, when the target variable is standardized, the expression for the maximal absolute bias reduces to

$$|st.bias_{\bar{y}_r}| < \frac{\sigma_\rho}{\bar{\rho}}, \qquad where\ corr_{\rho y} = 1; \sigma_y = 1 \tag{1.20}$$

Equivalently, the maximal absolute contrast is denoted as

$$|st.K_{\bar{y}_r}| < \frac{\sigma_\rho}{\bar{\rho}\,(1-\bar{\rho})} \qquad where\ corr_{\rho y} = 1; \sigma_y = 1 \tag{1.21}$$

These worst-case indicators clearly suggest the extent to which any estimate in the survey may be biased. For this reason, these indicators of maximal absolute bias and contrast reflect the quality of the respondent sample as a whole rather than an individual parameter.

As an illustration, consider Table 1.11, where a fictitious survey of $n = 1000$ is presented with regard to (non)response for men and women.

For men, the response propensities equal $\frac{200}{500} = 0.40$, whereas for women the response propensities are $\frac{350}{500} = 0.70$, resulting in an overall response

Table 1.11: Illustration of maximal absolute bias and contrast

	respondents	nonrespondents	total
men	200	300	500
women	350	150	500
	550	450	1000

rate of $\frac{0.40\times500+0.70\times500}{1000} = 0.55$. The standard deviation of the propensities can be easily obtained by $\sqrt{\frac{500\times(0.40-0.55)^2+500\times(0.70-0.55)^2}{999}} = 0.1501$. The R-indicator for this survey equals $1-2\sigma_\rho = 0.6998$, while the estimates for the maximal absolute bias and contrast are respectively $\frac{\sigma_\rho}{\bar{\rho}} = 0.2729$ and $\frac{\sigma_\rho}{\bar{\rho}(1-\bar{\rho})} = 0.6064$. This means that the respondent-only mean of a target variable that is maximally correlated to the propensities will be 0.2729 standard deviations away from the full-sample mean (bias) and 0.6064 standard deviations away from the nonrespondent-only mean (contrast).

Data analysis To illustrate the R-indicators, maximal absolute bias, and contrast, the RISQ-project is used. The national statistical institutes of The Netherlands (CBS), Norway, and the Statistical Office of the Republic of Slovenia, as well the Universities of Southampton and Leuven, have formed the RISQ-project (Representativity Indicators for Survey Quality; 2007 - 2010) in order to define and elaborate both theoretically and empirically the concept of representativeness and the associated R-indicator. Table 1.12 gives an overview of the survey-level nonresponse indicators under the random response model. For all surveys, response propensities have been obtained through logistic regression, using the auxiliary variables as explanatory factors. The number of non-redundant categories of the auxiliary variables in the logistic model is provided in the column 'number of parameters'. Only main effects have been used for all the surveys. ■

Schouten et al. (2012) suggest that R-indicators should serve the following purposes:

1. To compare response between different surveys within the same population

Table 1.12: R-indicators, maximal absolute bias and contrast for several European surveys

Survey / auxiliary variables	number of parameters	response rate	σ_ρ	R-indicator	Maximal absolute bias	Maximal absolute contrast
Norway: European Social Survey 2006 (Individuals), n = 2673						
age, gender, urbanicity	44	0.66	0.0583	0.8834	0.0883	0.2598
Norway: Survey on Level of Living 2004 (individuals), n = 4837						
age, gender, urbanicity	47	0.69	0.0640	0.8719	0.0928	0.2994
Slovenia: Labour Force Survey 2007 (Q3), n = 2219						
age, gender, region	9	0.70	0.0728	0.8544	0.1040	0.3467
Slovenia: ICT Survey 2007 (enterprises). n = 1998						
business type, business size	16	0.88	0.0728	0.8544	0.0827	0.6894
Netherlands: Short Term Statistic on IB Industry 2007 (enterprises), n = 64413						
business type, business size, vat	49	0.79	0.0412	0.9175	0.0522	0.2485
Netherlands: Short Term Statistic on retail 2007 (enterprises), n = 93799						
business type, business size, vat	47	0.78	0.0608	0.8783	0.0780	0.3545
Netherlands: Health Survey 2005 (Individuals), n = 15411						
age, gender, marital status, urbanicity, house value, ethnicity, type household, job	34	0.67	0.0959	0.8082	0.1432	0.4338
Netherlands: Consumer Satisfaction Survey 2005 (households), n = 17908						
age, gender, marital status, urbanicity, house value, ethnicity, type household, job	38	0.67	0.0283	0.9434	0.0422	0.1279
Netherlands: General Population Survey, n = 32019						
age, gender, household type, urbanization, house value, ethnicity, job, listed phone number, allowance entitlement	27	0.59	0.1058	0.7883	0.1794	0.4375
Belgium: European Social Survey 2006 (individuals), n = 2927						
age, gender, region, type dwelling, urbanicity, foreigners in area	25	0.61	0.1010	0.7980	0.1656	0.4245
Belgium: Flemish Housing Survey 2005-06 (dwellings), n = 7629						
age, gender, type dwelling, quality score dwelling, front width, green in area, number of houses in area	27	0.67	0.1025	0.7951	0.1529	0.4635
UK: LFS May-June 2001 (households), n = 7830						
age, gender, region	27	0.81	0.0374	0.9252	0.0462	0.2431

Source: Shlomo et al. (2008) and Shlomo et al. (2009), except Flemish Housing Survey and General Population Survey

2. To compare response longitudinally
3. To monitor response during data collection
4. To adapt the data collection process by tailoring based on historic data, frame data, and paradata.

Comparisons between different surveys are evidently only relevant when they use the same auxiliary variables to estimate the response propensities. In a situation where two surveys need to be compared, the first having many and strongly predictive auxiliary variables, and the second having only a few weak auxiliary variables available to determine response propensities, the comparison of the two with regard to the R-indicator, maximal absolute bias, and contrast will not be completely appropriate.

Differences with regard to response rates may also hinder the comparison of R-indicators for different surveys and impede the interpretation of R-indicators during the course of the fieldwork. As the response rate increases over time, the restricting maximal possible variance of response propensities also changes. In fact, during the time the response rate increases from 0% to 50%, the R-indicator may have a tendency to deteriorate. From 50% upwards, increasing response rates usually coincide with improving R-indicators. This is why paradoxical situations may sometimes occur when monitoring fieldwork operations. As an illustration, consider the fictitious example of Table 1.13. The total sample comprises 500 men and 500 women. After some fieldwork efforts at t_1, 16% of the men and 24% of the women have become respondents. The corresponding response rate at t_1 is 0.20. After having determined the variance of the response propensities, the R-indicator is 0.9157, the maximal absolute bias 0.2108 and the maximal absolute contrast 0.2635. Additional fieldwork efforts lead to an overall response rate of 40%, 35% among the men, and 45% among the women. The effects of the additional efforts on the R-indicator, maximal absolute bias, and contrast appear contradictory. On the one hand, the R-indicator drops slightly, indicating that the representativeness of the sample has become worse. On the other hand, the bias and contrast indications seem to have improved. The reason for these seemingly diverging

Table 1.13: Fictitious fieldwork monitoring example, $n_{men} = 500; n_{women} = 500$

	t_1	t_2
ρ_{men}	0.16	0.35
ρ_{women}	0.24	0.45
$\bar{\rho} = \bar{r}$	0.20	0.40
S_ρ	0.0422	0.0527
R-indicator	0.9157	0.8946
Max. bias	0.2108	0.1318
Max. contrast	0.2635	0.2196

conclusions is whether or not the response rate information is used when calculating the indicators. All three indicators use the same propensity variance, but the R-indicator does not take the response rate into account.

These subtle differences between the indicators based on the random response model suggest that the nonresponse quality of an obtained sample may have different interpretations. Indeed, the R-indicator measures the representativeness of the sample, whereas the indication of absolute maximal bias and contrast measure a different aspect of the same problem. Nevertheless, when evaluating the effects of nonresponse, it is necessary to be aware of these different angles from which the nonresponse problem can be assessed.

However, when comparing Table 1.12, (illustrating maximal absolute biases under the random model ranging from 0.05 to 0.18) with Table 1.5, (illustrating observed biases with regard to auxiliary variables under the fixed model; on average 0.0746), it is quite remarkable that what is termed 'maximal' possible bias is similar to what the fixed model generates as 'expected' bias.

It seems advisable always to examine thoroughly the assumptions and the data that are used for a nonresponse bias assessment. The examples presented here so far seem to suggest that bias assessments based on the random model using auxiliary variables to construct propensities, are more positive than straightforward comparisons between respondent-only and full-sample estimates.

1.2.6 Nonresponse and inference

Strictly speaking, whenever nonresponse occurs, it is very likely that the means of the full sample and the respondent-only will differ, even if the MCAR mechanism is operating. This elementary statistical principle should also be integrated into this discussion, not only for the sake of fine tuning the theoretical aspects, but because the combination of the bias and the variance of an estimate may have far-reaching consequences with regard to statistical inferences.

The relevance of distinguishing between bias and variance can be explained by statistical concepts *'type I error'* and *'type II error'*. Type I error occurs when the null hypothesis is true, although the statistician rejects it. This can also be termed a false positive event and can be compared with an innocent person being convicted of a crime. In the context of nonresponse, such an error occurs when survey data shows evidence that a certain correlation is significant or a proportion has changed over time, although the observed structure is in fact the result of a data imperfection such as nonresponse. In other words, type I error can be related to a researcher being too eager to find revelatory results from a survey, and ignoring the possibility of an error-generating structure that biases the data. Errors of this first type are usually more troublesome than errors of the second type. Particularly in large-sample surveys, small standard errors are obtained, reducing the risk of having to deal with type II errors. These type II errors apply when some structure needs to be revealed from the data, although the researcher cannot find it. This is also termed a false negative inference, and can be compared with the acquittal of a guilty offender. In this respect, survey researchers might be too careful or too skeptical of survey error, and thereby incapable of finding the correct population structure.

This section will use the fixed and random response models to explore the conditions under which valid inferences can be made in the presence of survey nonresponse. Nonresponse, particularly when it biases survey estimates, may have a substantive effect on type I error. Therefore, this

type of error will be of primary interest when discussing nonresponse and inference.

Nonresponse and inference: The fixed response model

Under the random response model and in particular under the fixed response model, making a distinction between structural bias and variance is appropriate, because both errors may interact when making inferences from survey data. Both sources of error can be brought together in the following expression:

$$mse_{\bar{y}_r} = bias^2_{\bar{y}_r} + \sigma^2_{\bar{y}_r} \tag{1.22}$$

Because nonresponse is by definition a problem that cannot be observed directly, the assessment of the bias component of equation 1.22 is particularly problematic. As a consequence, safely determining the *mse* may also be considered troublesome. Estimating the variance component on the other hand is usually not compromised by unknown factors, but can still be very difficult, particularly when the sampling design is complex and possibly coping with (known) disproportionality or clustering.

Making inference about a particular parameter

Instead of simply adding both components together to provide *mse*, the interplay between bias and variance can also be taken into account. When bias is likely to occur, but is not taken into consideration, the confidence interval will be dislocated and will be less likely to cover the interval compared with a situation where there is no bias. In such a case, a researcher runs a higher risk of producing a type I error. Therefore, small confidence intervals only seem to be interesting in the absence of bias, while large confidence intervals seem to be a necessity where the likelihood of bias is considerable.

Figure 1.8 provides an interesting two-dimensional alternative to the unidimensional interpretation of *mse*. Four different parameters (respon-

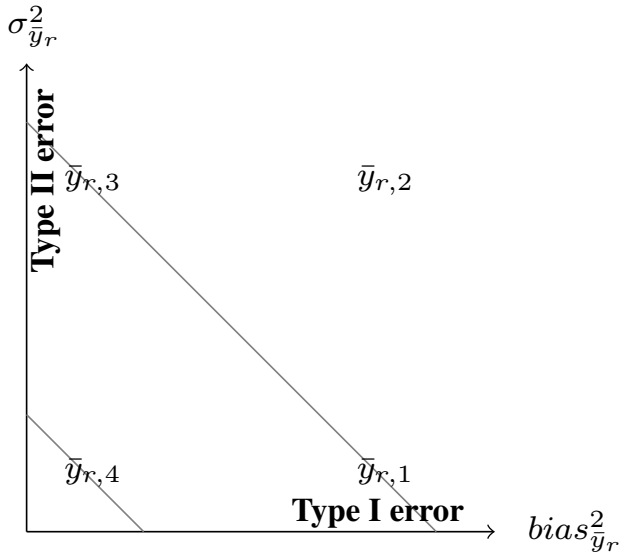


Figure 1.8: Trade-off between bias and variance

dent means) are plotted as a function of their bias (x-axis) and variance (y-axis). The bias is the result of the difference between the full sample and the respondent-only sample, the variance is determined by the respondent-only sample size (based on a simple random sampling survey design). The grey lines in the plot connect the points where *mse* is equal.

It is clear that $\bar{y}_{r,4}$ has the lowest *mse* as it combines low nonresponse bias and high precision, resulting in fairly accurate estimates. Situation $\bar{y}_{r,1}$ is probably the worst. It also has a small confidence interval, but may nonetheless provide a poor estimate because of the substantive bias. As a result, a researcher could become very sure about something that is actually wrong: clearly a situation where a type I error occurs. Although $\bar{y}_{r,3}$ is on the same *mse* level as $\bar{y}_{r,1}$, $\bar{y}_{r,3}$ is probably more safe than $\bar{y}_{r,1}$ because the probability of making a type I error is relatively low. However, the confidence interval of $\bar{y}_{r,3}$ may be large, implying an unfavorable possibility of type II error. Even though it is on a higher *mse* curve, $\bar{y}_{r,2}$ may run a lower risk of type I error than $\bar{y}_{r,1}$.

Instead of measuring the unidimensional *mse*, the two-dimensional alternative seems to offer a more insightful indication of how nonresponse

affects the quality of the estimates. Therefore, consider the two following indicators, based on the earlier work of Bethlehem and Kersten (1985), Bethlehem (2009), Bethlehem et al. (2011) Skinner (1996) and Eltinge (2002).

$$\begin{aligned} safety_{\theta_r} =& \Phi\left(\frac{(\theta_r + Z_{\alpha/2}\sigma_{\theta_r}) - \theta_f}{\sigma_{\theta_f}}\right) - \\ & \Phi\left(\frac{(\theta_r - Z_{\alpha/2}\sigma_{\theta_r}) - \theta_f}{\sigma_{\theta_f}}\right) \end{aligned} \tag{1.23}$$

$$\begin{aligned} waste_{\theta_r} =& 1 - \Phi\left(\frac{(\theta_f + Z_{\alpha/2}\sigma_{\theta_f}) - \theta_r}{\sigma_{\theta_r}}\right) \\ & + \Phi\left(\frac{(\theta_f - Z_{\alpha/2}\sigma_{\theta_f}) - \theta_r}{\sigma_{\theta_r}}\right) \end{aligned} \tag{1.24}$$

θ_f and θ_r represent the full-sample and the respondent-only parameter estimates and σ_{θ_f} and σ_{θ_r} refer to their respective standard errors. The *safety*-equation (also termed 'confidence level' (Bethlehem & Kersten, 1985) or 'interval coverage rate' (Skinner, 1996; Eltinge, 2002)) starts from the full-sample distribution of the mean that is defined by $N\left(\theta_f, \sigma^2_{\theta_f}\right)$ and measures the proportion of its density that is covered by the respondent-only interval that is informed by θ_r and σ_{θ_r}. The safety indicator therefore expresses the probability of making a type I error due to nonresponse and it is preferable for this to be close to 1. In fact, when $\alpha = 0.05$, the safety is expected to be 95%. Complementary to the safety indicator, but probably not as important, is the *waste* indicator, which expresses the proportion of the respondent-only density that is not covered by the full-sample interval. This proportion can be considered as not useful or redundant. Waste values close to 0 are preferable.

Figure 1.9 rearranges the four parameters of Figure 1.8 into the framework of the safety and waste indicators. Suppose that the full-sample interval is the reference benchmark to which each of the four respondent-only parameter intervals can be compared. In Figure 1.8, $\bar{y}_{r,4}$ was found to

be the best performing parameter estimate as it showed virtually no bias combined with relatively low variance. Therefore, it seems to completely cover the full-sample interval (safety = 97.80%), where only 13.48% of its interval is redundant (waste). This loss of precision is usually due to a smaller sample size compared with the size of the full sample. Again, $\bar{y}_{r,1}$ is found to perform very poorly, despite having an mse similar to that of $\bar{y}_{r,3}$. $\bar{y}_{r,1}$ does not cover the full-sample interval at all because a combination of considerable bias and relatively low variance. Apart from $\bar{y}_{r,4}$, $\bar{y}_{r,3}$ is the most preferable parameter estimate, despite the rather high variance. $\bar{y}_{r,2}$ performs worse than θ_3 because of the bias, but is still better than $\bar{y}_{r,1}$, as it still covers a small part of the full-sample interval.

Data analysis The concepts of safety and waste can be applied to real datasets. As an example, reconsider the auxiliary variables of the third round of the ESS for Belgium. Table 1.4 on page 36 includes 25 parameters that are expressed as a percentage. Some variables have a continuous counterpart (age, percentage of foreigners, population density, and average municipality income), resulting in four additional location parameters.

Apart from the 29 location parameters, 42 measurements of association are prepared. These are a combination of all the auxiliary variables where only one (logistic) regression parameter is needed in order to express the relationship between the two variables (see Table 1.14).

For all the 29 location parameters and 42 measurements of associations, both the full-sample and respondents-only estimates and their respective standard errors are known, with which the safety and waste indicators can be calculated. It should be noted that location parameters expressing a proportion have first been transformed into logits ($ln\left(\pi/(1-\pi)\right)$), to provide the convenience of symmetric parameter density. The results are plotted in Figure 1.10.

As can be seen, almost all of the points are above the line connecting the best possible situation (where safety = 1 and waste = 0) and the worst possible situation (where safety = 0 and waste = 1). A plausible explanation for this may be that because of the smaller respondents-only sample

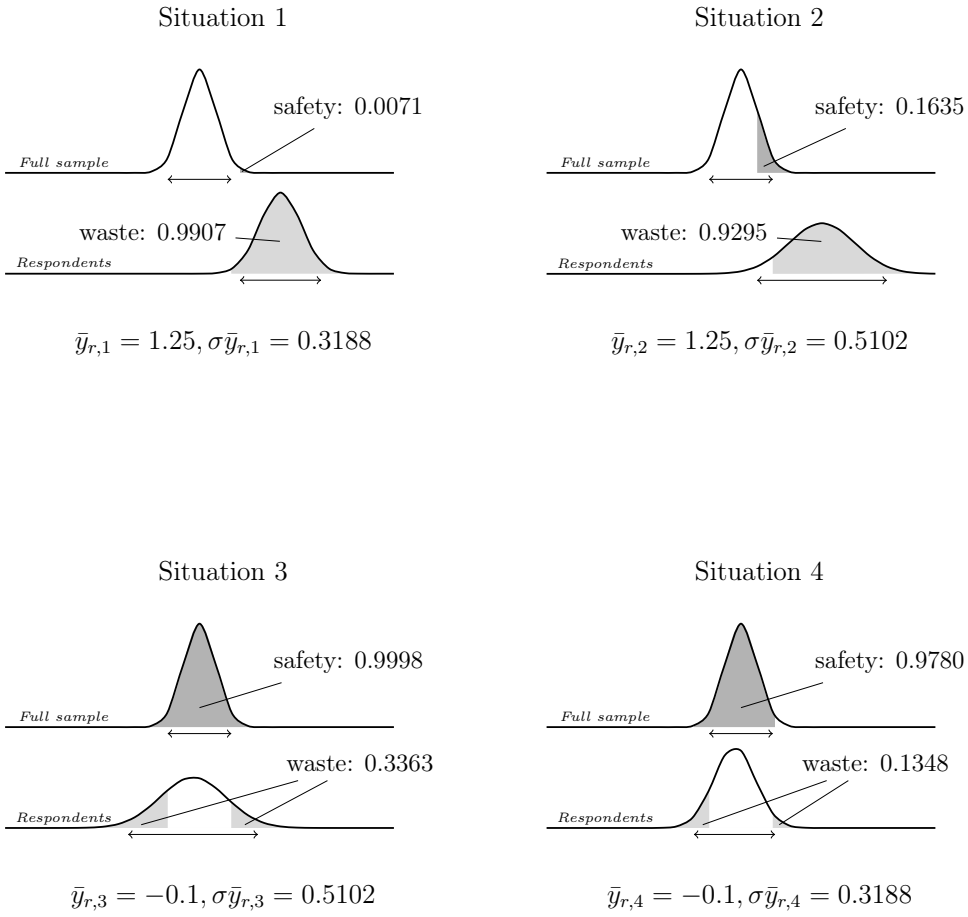


Figure 1.9: Parameter densities and confidence intervals for the full-sample mean and respondent-only means of four variables

Table 1.14: Construction of association parameters, ESS3-BE

Dependent	Link	Independent	Dependent	Link	Independent
Male	Logit	Foreigners (cont.)	Flat	Logit	(Reg. = Flanders)
Flat	Logit	Foreigners (cont.)	Flat	Logit	(Reg. = Brussels)
Foreigners (cont.)	Linear	Neighb. qual. (cont.)	Flat	Logit	(Reg. = Wallonia)
Foreigners (cont.)	Linear	Pop. dens. (cont.)	Neighb. qual. (cont.)	Linear	Pop. dens. (cont.)
Foreigners (cont.)	Linear	Income (cont.)	Neighb. qual. (cont.)	Linear	Income (cont.)
Age class (cont.)	Linear	Foreigners (cont.)	Age class (cont.)	Linear	Neighb. qual. (cont.)
Foreigners (cont.)	Linear	(Reg. = Flanders)	Neighb. qual. (cont.)	Linear	(Reg. = Flanders)
Foreigners (cont.)	Linear	(Reg. = Brussels)	Neighb. qual. (cont.)	Linear	(Reg. = Brussels)
Foreigners (cont.)	Linear	(Reg. = Wallonia)	Neighb. qual. (cont.)	Linear	(Reg. = Wallonia)
Flat	Logit	Male	Pop. dens. (cont.)	Linear	Income (cont.)
Male	Logit	Neighb. qual. (cont.)	Age class (cont.)	Linear	Pop. dens. (cont.)
Pop. dens. (cont.)	Linear	Male	Pop. dens. (cont.)	Linear	(Reg. = Flanders)
Income (cont.)	Linear	Male	Pop. dens. (cont.)	Linear	(Reg. = Brussels)
Age class (cont.)	Linear	Male	Pop. dens. (cont.)	Linear	(Reg. = Wallonia)
Male	Logit	(Reg. = Flanders)	Age class (cont.)	Linear	Income (cont.)
Male	Logit	(Reg. = Brussels)	Income (cont.)	Linear	(Reg. = Flanders)
Male	Logit	(Reg. = Wallonia)	Income (cont.)	Linear	(Reg. = Brussels)
Flat	Logit	Neighb. qual. (cont.)	Income (cont.)	Linear	(Reg. = Wallonia)
Flat	Logit	Pop. dens. (cont.)	Age class (cont.)	Linear	(Reg. = Flanders)
Flat	Logit	Income (cont.)	Age class (cont.)	Linear	(Reg. = Brussels)
Flat	Logit	Age class (cont.)	Age class (cont.)	Linear	(Reg. = Wallonia)

size compared with the full sample, the standard errors will increase, automatically leading to more waste even when there is no bias. Occasional cases below the line may indicate a situation where the variance of a target variable among respondents only is smaller than in the full sample. This suggests that not only the points estimates may be biased, but that nonresponse can also affect other distributional aspects, interfering with the estimation. In this particular case, heteroscedasticity or the inequality of the variances of respondents and nonrespondents cannot be ignored when making inferences.

Examining all the plotted safety/waste combinations in Figure 1.10, it appears that the impact of nonresponse affects a wide range of possible outcomes with respect to statistical inference. In the lower-right corner, some estimates can be found that are not affected by nonresponse at all. However, the upper-left corner illustrates that many other respondent-only estimates completely fail to cover their true full-sample counterparts. Some respondents-only estimates are in the middle, only partially covering the full-sample estimates.

Furthermore, it seems that association parameters are somewhat safer than location parameters. On average, full-sample location parameters are only covered for about 42% by their respective respondent-only intervals, while full-sample association parameters are covered for about 74%.

In the Dutch General Population Survey, the average safety for location parameters (0.27; waste = 0.78; 69 parameters) is also worse compared with the average safety for association parameters (0.69; waste = 0.50; 77 parameters). Similar results apply to the Flemish Housing Survey: the average safety of location parameters is 0.39 (waste = 0.69; 53 parameters), the average safety of association parameters is 0.92 (waste = 0.24; 72 parameters).

These empirical results all suggest that nonresponse may have a detrimental effect on the validity of the conclusions drawn from survey data. It is agreed that ideally, the risk of a type I error is limited to only 5%. However, the empirical findings suggest that this risk can be as high as 73% (GPS, location parameters), and seems to be at least greater than 8% (FHS, association parameters). ■

Making inferences about target variables

Under the fixed response model, normally only the auxiliary variables are used for bias assessment. Apart from individually assessing the safety and waste for these variables, it might also be interesting to formulate an expectation of what might happen when projecting the safety and waste structure onto the auxiliary variables toward the target variables.

Two methods will be further developed under the fixed response approach, which integrate the presence of bias into the estimation of target statistics. The first method starts from the assumption that bias is a distribution that can be measured among the auxiliary variables and that can be projected onto the target variables. It will typically assess the degree of bias in '$mse_{\bar{y}_r} = bias^2_{\bar{y}_r} + \sigma^2_{\bar{y}_r}$', in order to obtain a realistic estimate of the total nonresponse error that should be taken into account when constructing confidence intervals for target statistics. The second method

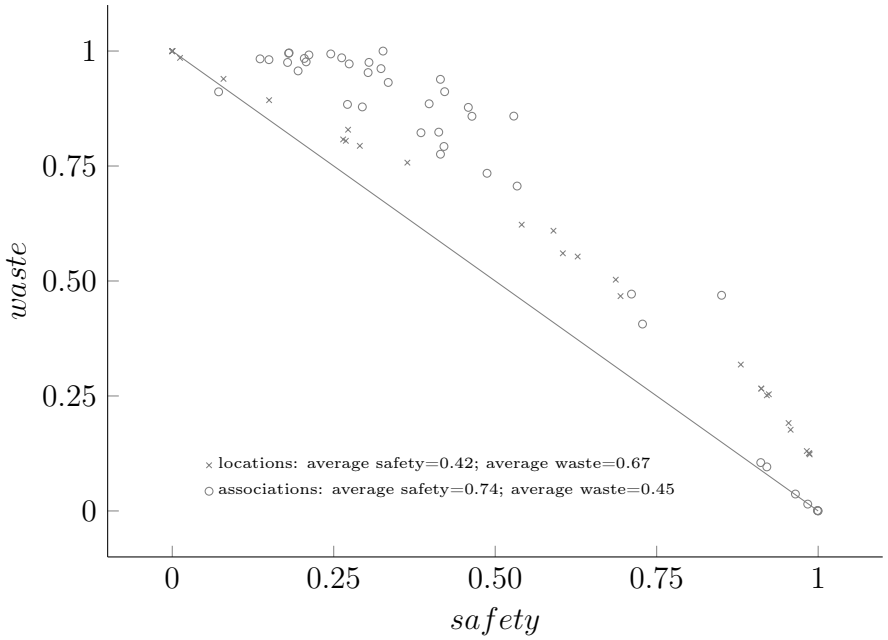


Figure 1.10: Safety-waste plot for location and association parameters, ESS3-BE ($n = 2927$)

determines the necessary degree of variance inflation that should be applied to all auxiliary variables in order to obtain a pre-specified aggregate safety level (for example 95%). It is then assumed that the same amount of variance inflation is also applicable to the target statistics.

1. *Assuming a bias-distribution*
   Suppose a set of respondents is randomly selected from the full sample under the assumption of MCAR. The mse of the mean of a target variable can be determined by $mse_{\bar{y}_r} = \sigma^2_{y_r}/n_r$, assuming there are no sources of error other than sampling error (based on simple random sampling). If y has first been standardized, the expression can be further reduced to $mse_{\bar{y}_r} = 1/n_r$. However, it is clear that nonresponse also introduces a sort of uncertainty that is rarely reflected in the mean squared errors. Further, informed by former sections, nonresponse bias is a source of error that is considered very likely to occur. Therefore, it may be wiser to introduce a level of

uncertainty that should be taken into account in order to protect estimates against nonresponse bias. Starting from the equation that $mse_{\bar{y}_r} = bias^2_{\bar{y}_r} + \sigma^2_{\bar{y}_r}$, the terms can be rewritten as

$$bias_{\bar{y}_r} = \frac{corr_{ry}\sigma_r\sigma_y}{\bar{r}}$$

and

$$\sigma^2_{\bar{y}_r} = \frac{\sigma^2_y(1 - corr^2_{ry})}{n_r}$$

The bias-expression has been encountered earlier (see equation 1.3 on page 25). The variance-expression, referring to the sampling error, starts from the full-sample variance of y, that is corrected downwards, because the within-group variance of the respondents only is usually somewhat smaller than the full-sample variance. Particularly because of nonresponse bias, respondents and nonrespondents differ, so that the full-sample variance of y can be separated into '*between*' and '*within*' variances. Assuming homoscedasticity, the full-sample variance can be easily determined by introducing the factor $(1 - corr^2_{ry})$. The mse can now be further elaborated as

$$\begin{aligned}
mse_{\bar{y}_r} &= bias^2_{\bar{y}_r} + \sigma^2_{\bar{y}_r} \\
&= \left(\frac{corr_{ry}\sigma_r\sigma_y}{\bar{r}}\right)^2 + \frac{\sigma^2_y(1 - corr^2_{ry})}{n_r} \\
&= \left(\frac{corr^2_{ry}\left(\sqrt{\bar{r}(1-\bar{r})}\right)^2\sigma^2_y}{\bar{r}^2}\right) + \frac{\sigma^2_y(1 - corr^2_{ry})}{n_r} \\
&= \sigma^2_y\left(\frac{corr^2_{ry}\,(1-\bar{r})}{\bar{r}} + \frac{(1 - corr^2_{ry})}{n_r}\right) \\
&= \sigma^2_y\left(\frac{corr^2_{ry}\,(1-\bar{r})\,n}{n_r} + \frac{(1 - corr^2_{ry})}{n_r}\right) \\
&= \sigma^2_y\left(\frac{(1-\bar{r})\,corr^2_{ry}n + 1 - corr^2_{ry}}{n_r}\right)
\end{aligned}$$

$$= \sigma_y^2 \left(\frac{(n(1-\bar{r})-1)\, corr_{ry}^2 + 1}{n_r} \right)$$
$$= \left(\frac{(n(1-\bar{r})-1)\, corr_{ry}^2 + 1}{n_r} \right),\ where\ \sigma_y^2 = 1 \qquad (1.25)$$

σ_y^2 can be isolated and when standardizing y, the expression reduces to equation 1.25.

Now, $corr_{ry}$ is not supposed to be a single value, but a distribution. Therefore, $E\left(mse\left(\bar{y}_r\right)\right)$ has to be determined. Using Taylor expansions, the expected mse can be approximated by $E\left[g\left(X\right)\right] \approx g\left(\mu_X\right) + \frac{g''(\mu_X)}{2}\sigma_X^2$. This means that

$$\begin{aligned} E\left(mse_{\bar{y}_r}\right) &\approx \left(\frac{(n(1-\bar{r})-1)\, c\bar{o}\bar{r}r_{ry}^2 + 1}{n_r} \right) \\ &+ \frac{1}{2}\frac{(-2(n(\bar{r}-1)+1))\sigma_{corr_{ry}}^2}{n_r} \end{aligned} \qquad (1.26)$$

As discussed earlier, each variable y has a complement $-y$, compensating their respective correlations with r. It can therefore be argued that $c\bar{o}\bar{r}r_{ry}$ is supposed to be zero, reducing the first term of equation 1.26 to $1/(n\bar{r})$. Furthermore, fixing $c\bar{o}\bar{r}r_{ry}$ at zero implies that the computation of the variance does not need the additional degree of freedom. It is now expected that

$$\begin{aligned} E\left(mse_{\bar{y}_r}\right) &\approx \left(\frac{1}{n\bar{r}}\right) + \frac{(-(n(\bar{r}-1)+1))\sigma_{corr_{ry}}^2}{n_r} \\ &\approx \left(\frac{1}{n\bar{r}}\right) + \frac{(n(1-\bar{r})-1)\,\sigma_{corr_{ry}}^2}{n_r} \\ &\approx \left(\frac{1 + (n(1-\bar{r})-1)\,\sigma_{corr_{ry}}^2}{n_r} \right) \end{aligned} \qquad (1.27)$$

This means that the effect of nonresponse on the expected mse can be assessed whenever the response, the sample size and variance of the $corr_{ry}$'s are known.

Evidently, $\sigma_{corr_{ry}}$ is unknown by definition, but can be replaced by $\sigma_{corr_{r,aux}}$, assuming that the nonresponse mechanism has a similar effect on both the auxiliary and target variables.

Once $E\left(mse_{\bar{y}_r}\right)$ has been determined, it can be compared with the naive mse, which is the mean square error where only sampling error and not nonresponse error has been taken into account. In this case, the naive mse simply equals $1/n_r$. Both quantities can be used to determine what is termed the *variance inflation factor* (VIF), or the factor by which the naive mse should be multiplied in order to obtain $E\left(mse_{\bar{y}_r}\right)$. Dividing the respondent-only sample size by the variance inflation factor gives what is termed the *effective sample size*. The effective sample size (n_{eff}) expresses the statistical power of a (usually complexly designed) sample that is similar to the statistical power of a simple random sample of n_{eff} elements. In this case, it can be expressed as:

$$\begin{aligned} n_{eff} &= \frac{n_r}{VIF} \\ &= \frac{n_r}{\frac{E(mse_{\bar{y}_r})}{\frac{1}{n_r}}} \\ &= \frac{1}{E\left(mse_{\bar{y}_r}\right)} \\ &= \frac{n_r}{1+\hat{S}^2_{corr}\left(n\left(1-\bar{r}\right)-1\right)}. \end{aligned} \tag{1.28}$$

When reconsidering equation 1.27, expressing the expected mse, the total sample size n may have an upper bound: the effect of increasing the sample size on the mse gradually becomes smaller. As $n \to +\infty$, the marginal effect of one additional sample unit will eventually

converge to zero.

$$\lim_{n\to+\infty} \frac{n_r}{1+\hat{S}^2_{corr}\left(n\left(1-\bar{r}\right)-1\right)} = \frac{\bar{r}}{\hat{S}^2_{corr}\left(1-\bar{r}\right)}. \tag{1.29}$$

This equation expresses the point where $\sigma^2_{\bar{y}_r}$ or $1/n_r$ becomes practically zero.

2. *Avoiding type I error by deliberately increasing the variance*

   An alternative way to determine the effective power of a survey sample that is affected by nonresponse bias, is to deliberately increase the variance of all estimates by a constant factor ψ, until the average safety indicator over all auxiliary variables meets a pre-specified level (for example 95%) or

$$\frac{\sum_{j=1}^{p} safety_{\theta_{r,inf,j}}}{p} = 0.95 \tag{1.30}$$

   where p is the number of auxiliary variables and $safety_{\theta_{r,inf,j}}$ is the safety with regard to the parameter estimate of auxiliary variable j after its variance is inflated by a constant factor ψ. Equation 1.30 can be rewritten as

$$\frac{1}{p}\sum_{j=1}^{p}\left[\Phi\left(\frac{(\theta_{r,j}+Z_{\alpha/2}\sqrt{\psi\sigma^2_{\theta_{r,j}}})-\theta_{f,j}}{\sigma_{\theta_{f,j}}}\right) - \Phi\left(\frac{(\theta_{r,j}-Z_{\alpha/2}\sqrt{\psi\sigma^2_{\theta_{r,j}}})-\theta_{f,j}}{\sigma_{\theta_{f,j}}}\right)\right] = 0.95 \tag{1.31}$$

   In Figure 1.11, the full-sample (black) and the respondent-only (grey) densities are illustrated for 10 fictitious parameters. Usually, the sample parameter densities are somewhat smaller because $n > n_r$. In some cases, biases can easily be observed (parameters θ_1 and θ_5), suggesting that making inferences may become risky without allow-

ing space for additional uncertainty due to nonresponse. A constant variance inflation factor ψ has therefore been determined, by which each parameter variance has to be multiplied. As a result, the average safety after variance inflation is now 0.95 (type I error is 0.05), whereas the average safety before inflation is only 0.8036, indicating the risk of a type I error of about 20%. The inflated densities are indicated by the broken lines. Nevertheless, individual safety indications may deviate from the desired 0.95. Indeed, parameter estimates that show hardly any bias are actually overprotected, whereas severely biased estimates do not reach the prescribed level of safety = 0.95.

Nevertheless, as it is assumed that any new estimate for which no full-sample information is available is affected by nonresponse similarly then the 10 known parameters, it is expected that ψ should also be applied to this new respondent-only estimate (see Figure 1.12). The value for ψ can easily be found numerically. After having determined ψ, the effective sample size is also easily derived.

In order to illustrate the two methods as presented above, suppose a full sample of $n = 2000$ has $n_r = 1000$ respondents. There are 20 auxiliary variables that have first been standardized, so that $a\bar{u}x_j = 0$ and $\sigma^2_{aux_j} = 1$, implying that $a\bar{u}x_r = bias_{aux_r}$. Since $\sigma^2_{aux_j} = 1$ and $\bar{r} = 0.5$, the bias expression reduces to

$$\begin{aligned} bias_{aux_r} &= \frac{corr_{r,aux}\sigma_r\sigma_{aux_j}}{\bar{r}} \\ &= \frac{corr_{r,aux}\sqrt{\bar{r}\,(1-\bar{r})}\sigma_{aux_p}}{\bar{r}} \\ &= \frac{corr_{r,aux}\sqrt{0.5 \times 0.5} \times 1}{0.5} \\ bias_{aux_r} &= corr_{r,aux} \end{aligned}$$

Therefore, the first two columns of Table 1.15 are identical, given that all correlations between the response outcome r and the auxiliary variables, $\sigma_{corr_{r,aux}}$ can be determined. However, since $corr_{r,aux}$ also contains some sampling fluctuations, $\sigma_{corr_{r,aux}}$ may be somewhat overestimated. The ad-

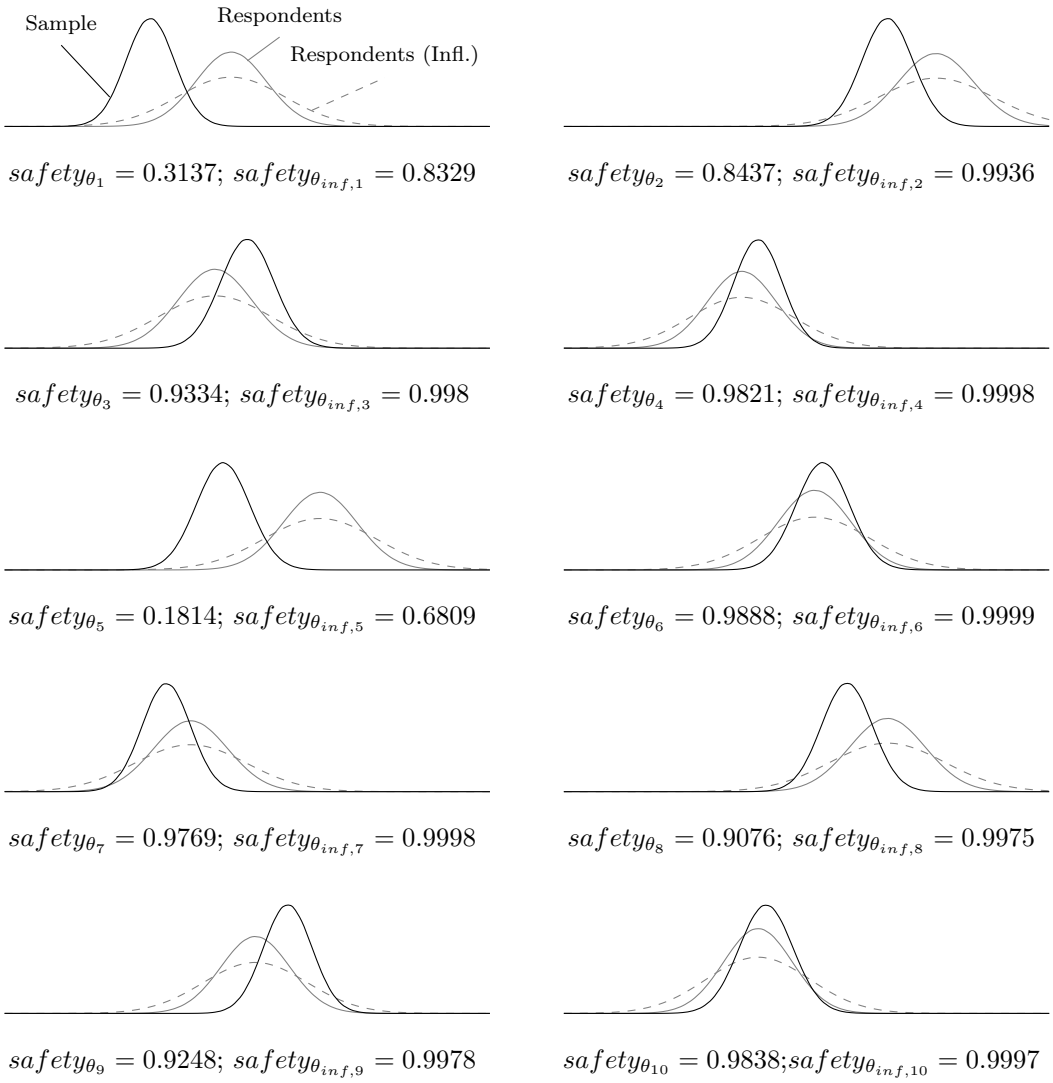


Figure 1.11: Applying a constant variance inflation factor $\psi = 2.28$ to all respondent-only parameter densities in order to cover 95% of the full-sample parameter densities

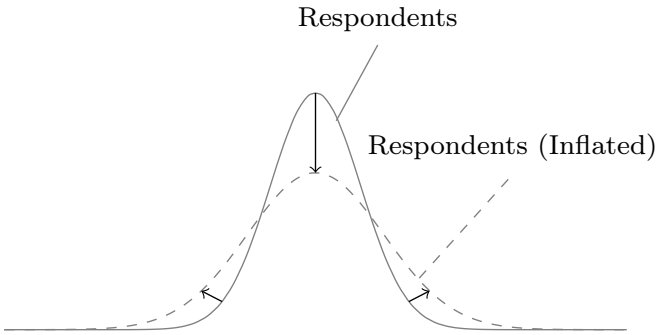


Figure 1.12: Applying a constant variance inflation factor $\psi = 2.28$ to a target parameter estimate in order to obtain type I of 0.05

justed $corr_{r,aux}$ or R^2_{Adj} is used instead. Therefore, the average bias is not $\sqrt{0.0035} = 0.0594$, but instead $\sqrt{0.0030} = 0.0548$. Filling out equation 1.27, produces:

$$\begin{aligned} E\left(mse_{\bar{y}_r}\right) &\approx \left(\frac{1 + \left(n(1-\bar{r}) - 1\right)\sigma^2_{corr_{ry}}}{n_r}\right) \\ &\approx \left(\frac{1 + \left(2000(1-0.5) - 1\right)0.0030}{1000}\right) \\ &\approx 0.004033 \end{aligned}$$

As compared with the naive mse of $1/\left(n_r\right) = 1/1000 = 0.0010$, the actual mse needs to be inflated by a factor ψ of 4.0331, because of the additional uncertainty due to nonresponse bias. This also means that instead of the 1000 respondents, the effective sample size should be downgraded to a level of $\frac{1000}{4.0331} = 248$ units. Since nonresponse bias is considered to be independent of the sample size, increasing the sample size may only affect the variance term in the mse, implying that the marginal effect on the effective sample from adding one case to the full sample, decreases as the full-sample size grows. Eventually, this marginal effects converges to zero as the full sample size approaches infinity. In this particular case, applying equation 1.29, the effective sample size will never exceed $\frac{\bar{r}}{\hat{S}^2_{corr}\bar{r}} = \frac{0.5}{0.0030 \times 0.5} = 329.36$,

Table 1.15: 20 variables to determine a distribution of bias: fictitious data

# aux	$bias_{y_r}$	$corr_{r,aux}$	$corr^2_{r,aux}$	R^2_{Adj}	$safety$	$safety_{inf}$
1	-0.1285	-0.1285	0.0165	0.0160	0.0011	0.3729
2	-0.0873	-0.0873	0.0076	0.0071	0.1315	0.9560
3	-0.0854	-0.0854	0.0073	0.0068	0.1501	0.9630
4	-0.0706	-0.0706	0.0050	0.0045	0.3270	0.9893
5	-0.0608	-0.0608	0.0037	0.0032	0.5246	0.9980
6	-0.0078	-0.0078	0.0001	-0.0004	0.9912	1.0000
7	-0.0103	-0.0103	0.0001	-0.0004	0.9899	1.0000
8	-0.0008	-0.0008	0.0000	-0.0005	0.9946	1.0000
9	-0.0239	-0.0239	0.0006	0.0001	0.9538	1.0000
10	0.0127	0.0127	0.0002	-0.0003	0.9905	1.0000
11	0.0008	0.0008	0.0000	-0.0005	0.9946	1.0000
12	0.0248	0.0248	0.0006	0.0001	0.9430	1.0000
13	0.0335	0.0335	0.0011	0.0006	0.8957	1.0000
14	0.0186	0.0186	0.0003	-0.0002	0.9738	1.0000
15	0.0271	0.0271	0.0007	0.0002	0.9468	1.0000
16	0.0290	0.0290	0.0008	0.0003	0.9282	1.0000
17	0.0503	0.0503	0.0025	0.0020	0.7038	0.9996
18	0.1034	0.1034	0.0107	0.0102	0.0307	0.8224
19	0.0609	0.0609	0.0037	0.0032	0.5161	0.9978
20	0.0953	0.0953	0.0091	0.0086	0.0659	0.9009
Average			0.0035	0.0030	0.6527	0.9500

provided that the nonresponse mechanism operates irrespective of the full-sample size. This concludes the illustration of the first method.

The second method seeks to find a constant factor ψ by which to multiply the variances of the estimates in order to obtain an average safety of 95%. Applied to the situation as presented in Table 1.15, the variance inflation factor ψ equals 4.0584. This variance inflation factor shifts an initial safety of 0.6527 toward 0.95. However, some of the estimates may still be biased (for example the first estimate).

The variance inflation factor ψ that is found through the second method is quite comparable with the variance inflation of 4.0331 from the first method, assuming a distribution of biases. The additional advantage of this second method is its flexibility to incorporate parameters of all kinds

(means, logits, regression parameters, etc.), as it requires only point estimates and associated variances. The first method is constrained to means only. Therefore, the next presentation of empirical evidence will employ the second method, using the variance inflation factor ψ, particularly because in addition to location measures, association measures will also be assessed in this nonresponse analysis.

Data analysis For all the datasets in which some auxiliary information is available for both respondents and nonrespondents, the average safety and waste values will be determined, as well as the variance inflation necessary in order to obtain an average safety of 95%. For the first three datasets in Table 1.16, the ESS3-BE, the FHS, and the GPS, the safety and waste indicators have already been discussed on page 66. For example, for the ESS3-BE, 29 location parameters have been identified, of which 25 are expressed as a proportion and 5 as an average, together with 42 parameters that describe a relationship between any two auxiliary variables. For each of the variables, the point estimate and its respective variance is available for the full sample as well as the respondent-only sample. For each parameter, the safety and waste can be determined, the averages for which are presented in Table 1.16. In the ESS3-BE, ψ is 11.74, meaning that all variances of the respondent-only location parameters should be multiplied by a factor 11.74 in order to attain an overall safety level of 95%. As a consequence, to retain statistical power the sample reduces to an effective sample size of only 153 units. This means that a simple random sample of 153 units is as powerful as the entire ESS3-BE when the nonresponse deficit is taken into account. This effective sample size refers to location parameters. Measures of association offer a more positive impression, the variance inflation factor ψ is 6.28, resulting in an effective sample size of 286.

For the FHS and the GPS, where the average safety and waste indicators are also presented on page 66, the effective sample size seems to be somewhat higher compared with the ESS3-BE. However, since their full

samples are greater than that of the ESS3-BE, ψ is also greater (particularly for the GPS).

This is also the case for the ESS5 where interviewer observations are used as auxiliary variables, as already presented in Table 1.3 on page 35. In Table 1.16, all possible associations between the auxiliary variables have also been integrated in the analysis. In all countries, the safety indicator is lower (sometimes much lower) than 95%, suggesting that a substantive amount of variance inflation is needed in order to hedge against type I errors.

All the surveys in the different countries seem to have lower safety indicators, higher indications of *waste* and higher ψ-values for location parameters than for parameters of association. As a result, the effective statistical power, as expressed by $n_{eff}(= n_r/\psi)$, of location parameters is lower than that of association parameters.

Irrespective of the type of parameter, the indicators of safety, waste, ψ, and the effective sample size vary considerably between the surveys. Average safety indicators range from 0.03 to 0.94, variance inflations are no smaller than 1.07 and can even rise to 81. Of course, some of the auxiliary variables have a limited range in covering all the different nonresponse dimensions. They may also have been measured poorly (for example, the interviewer observations). However, it is clear that the possibility cannot be ruled out that nonresponse may have a severe impact on survey quality and that far-reaching precautionary measures should sometimes be taken in order to prevent type I errors. ■

Nonresponse and inference: The random response model

The random response model starts from the assumption that different sample units may have different propensities to respond to a survey request. If target statistics need to be obtained, it is necessary to take into account this variability of response probabilities. Usually, auxiliary variables are deployed in order to inform a weighting procedure, hoping to bridge possible gaps between respondents and nonrespondents. If the correct in-

Table 1.16: Average safety, waste, and variance inflation factor ψ for various surveys

Survey	n	n_r	$\bar{r}$	kind of parameter	$\bar{safety}$	$\bar{waste}$	ψ	n_{eff}
ESS3-BE	2927	1798	61.43%	location (29)	0.42	0.67	11.74	153
				association (42)	0.74	0.45	6.28	286
GPS	32019	18792	58.69%	location (69)	0.27	0.79	81.22	231
				association (77)	0.69	0.50	7.33	2564
FHS	7770	5216	67.13%	location (53)	0.39	0.69	14.17	368
				association (72)	0.92	0.24	1.29	4043
ESS5-BE	3199	1669	52.16%	location (6)	0.08	0.95	9.09	183
				association (15)	0.89	0.37	1.49	1118
ESS5-BG	2997	2280	76.06%	location (6)	0.82	0.24	2.29	996
				association (15)	0.90	0.18	1.38	1647
ESS5-CY	1564	1059	67.69%	location (6)	0.53	0.55	5.17	205
				association (15)	0.83	0.33	1.95	544
ESS5-DK	2856	1554	54.41%	location (6)	0.03	0.97	14.82	105
				association (15)	0.89	0.36	1.49	1042
ESS5-ES	2727	1794	65.79%	location (6)	0.78	0.35	2.01	892
				association (15)	0.87	0.23	1.74	1030
ESS5-GR	4230	2715	64.18%	location (6)	0.63	0.49	3.81	712
				association (15)	0.94	0.20	1.07	2533
ESS5-HU	2612	1547	59.24%	location (6)	0.75	0.46	2.08	744
				association (15)	0.87	0.30	1.56	994
ESS5-IL	3230	2294	71.02%	location (6)	0.77	0.36	2.13	1076
				association (15)	0.81	0.31	2.79	821
ESS5-PT	3264	2382	72.99%	location (6)	0.40	0.68	5.32	448
				association (15)	0.86	0.24	1.66	1432
ESS5-RU	3982	2595	65.17%	location (6)	0.24	0.81	40.42	64
				association (15)	0.81	0.30	2.10	1236

clusion probabilities are known, the eventual weighted estimate (using the inverse of the inclusion probabilities) will be unbiased (Horvitz & Thompson, 1952).

There are many ways to adjust sample estimates (see, among others, Kalton & Flores-Cervantes, 2003; Bethlehem et al., 2011). They all basically start from the availability of unbiased auxiliary information that can be used as points of support to correct possibly biased target survey statistics. Adjustment methods seek to accomplish correspondence between the respondent sample and the whole population (or full sample) with respect to the auxiliary information, in the expectation that all the respondent-only target statistics will become similar to the full sample or population. These unbiased reference variables usually come from population totals, official records, sample frame data, complete sample information from interviewer observations, etc. Strictly speaking, when investigating the effects of nonresponse it is preferable to have auxiliary variables that refer only to the full sample. Population data is less adequate as it cannot distinguish between the diverse sources of survey error such as nonresponse, coverage, and sampling error.

A general framework for weighting is presented by Deville and Särndal (1992) and is termed *calibration weighting*, where weights should be conceived so that the weighted sample distribution with respect to some auxiliary variable(s) corresponds with the full sample or population distribution. Additionally, the weights should be as close to 1 as possible. When two or more auxiliary variables are used and all weighted sample cells in the cross tabulation are supposed to match the population (or full sample), this situation coincides with *cell weighting* or *post-stratification* (Holt & Smith, 1979). Here, the sizes of the cells or strata $h = 1, 2, \ldots, H$ have to be known so that $w_h = (N_h/N)/(n_h/n)$. Relaxing the model requirements so that only the marginal or univariate distributions of the auxiliary variables need to achieve correspondence is also known as *raking* (Skinner, 1991).

Generalized regression estimation (GREG) can also be used to adjust nonresponse bias (Bethlehem, 1988). This method starts from the re-

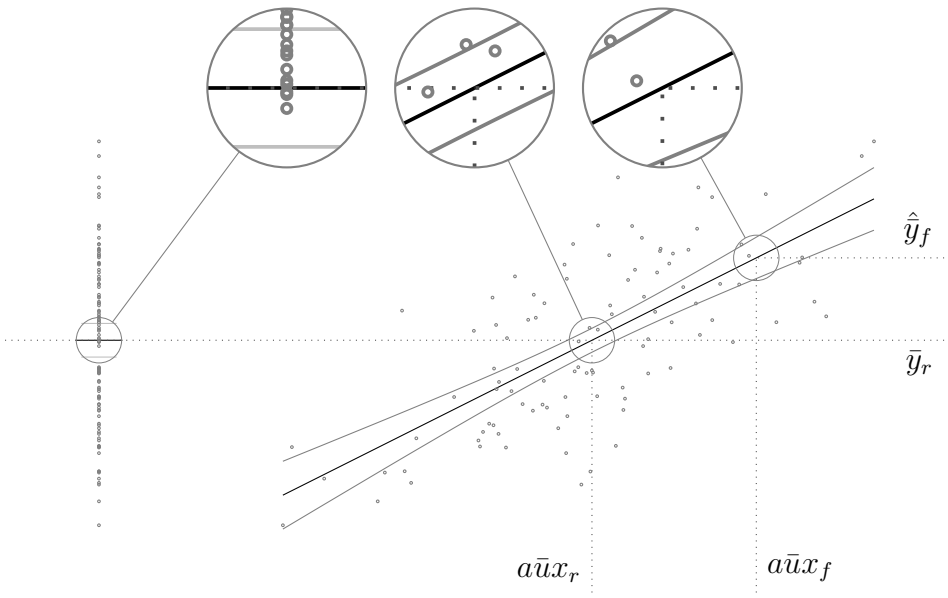


Figure 1.13: Regression estimation: effects on point estimates (——) and confidence intervals (——)

lationship between the auxiliary variable(s) and the target variable, as expressed by regression parameters $\hat{\beta}_r = (b_0, b_1, \ldots, b_p)'$, and where the regression parameters can be obtained from the respondents only as $\hat{\beta}_r = (aux'aux)^{-1} aux'y$. The adjusted mean of the target variables y can then be estimated by $\bar{y}_r^{GR} = a\bar{u}x\hat{\beta}_r$. When using only one auxiliary variables, *GREG*-estimation can be illustrated by Figure 1.13. The dots in the plot represent the respondents. They have a mean for the target variable $\bar{y}_r$ and for the auxiliary variable $a\bar{u}x_r$. Because of the relationship between the variables (known only for respondents), the full-sample mean of the auxiliary variable $a\bar{u}x_f$ can be projected onto the regression line, resulting in the adjusted mean of the target variable $\hat{\bar{y}}_f$. Although regression estimation does not explicitly uses weights, the method produces comparable adjustments to methods that deploy weight scores (Bethlehem et al., 2011).

A more flexible way of nonresponse adjustment is offered by *propensity weighting* in which a response propensity $\hat{\rho}_i$ is assigned for each responding unit. Inverting the propensities into a weight vector, this can then be used for the adjustment. Theoretically, propensity weighting could use as many correction parameters as there are individuals in the respondent sample (one unique weight score per case), however, regression estimation is instead restricted to the number of parameters in the model. Therefore, propensity weighting operates locally (on the individual level), while regression estimation instead operates globally (on the respondent sample level). In practice, the propensities are estimated by the same set of auxiliary variables as entered in the regression model, so that both methods use the same set of linear combinations of auxiliary variables in order to correct for nonresponse.

All the diverse methods usually produce very similar estimates (Kalton & Flores-Cervantes, 2003). The occasional differences are often due to the choice of auxiliary variables and the way they are used to inform the weighting. Sometimes, only the marginal distributions of the auxiliary variables are taken into account, whereas interaction terms can also be considered (as already mentioned when discussing post-stratification and raking).

Further, it is important not only to take the shifting parameter estimates into consideration, but also the effects of weighting adjustments on the their standard errors (Kish, 1965). Determining weighted point estimates is relatively straightforward; however, assessing the effect of weighting on the standard errors is comparatively complicated and requires advanced mathematical procedures and software tools (Kish, 1992). Generally, nonresponse correction techniques lead to variance inflation. A simple explanation for this phenomenon relates to the under-representation of particular groups or profiles in a complete sample. Profiles that are relatively scarce are assigned higher weight scores, resulting in them having a relatively strong influence on the eventual parameter estimates. Because of the dependence on these higher-weighted cases, relevant survey estimates should allow for a greater degree of uncertainty compared with a sample

where all units have equal weights. (Kish, 1992) expresses the variance inflation factor as

$$F = 1 + CV_w^2 \tag{1.32}$$

where CV_w is the coefficient of variance $(S_w/\bar{w})$ with respect to the individual weight scores among the respondents. The equation considers each respondent to have a weight score w_i, where $\bar{w}$ is the average weight score and S_w is the associated standard deviation. The symbol F is used, in accordance with Kalton and Flores-Cervantes (2003). It should be noted that the symbol F will be used from here onwards to indicate variance inflation in the random response model, while ψ indicates the variance inflation under the fixed response interpretation.

Little and Vartivarian (2005) have argued that equation 1.32 is only applicable if the target variable is not correlated with the weights. Otherwise, the standard error tends to decrease because the target variable is enriched by the auxiliary variables constituting the weight vector. Many authors (see, among others, Kalton & Flores-Cervantes, 2003; Little & Vartivarian, 2003, 2005; Groves, 2006; Kreuter et al., 2010) therefore agree that applying weighting will only be effective if

1. The weight variables are related to survey participation or (non)response
2. The weight variables are related to target survey variables.

Särndal and Lundström (2006) also add that the most substantial or important domains or subpopulations within the survey should be identified.

Indeed, when considering Figure 1.13 again, it seems that the confidence interval is smaller at $a\bar{u}x_r$ as compared with the situation where no relevant information is provided in order to facilitate the estimation. In this respect, consider the confidence bound at $a\bar{u}x_r$ and the situation where no auxiliary information is provided (the univariate and uninformed representation of y on the left-hand side of the figure). This last situation has a larger confidence interval. However, it seems that the standard error of the estimate increases as the mean of the auxiliary

variables ($a\bar{u}x_f$) moves away from the center of the cloud of the auxiliary data (respondents only). This means that the first condition (the relationship between r and aux) has an increasing effect on the standard error, whereas a substantive relationship between the target variable and the auxiliary variable (second condition) tends to reduce the standard error. This is because the proportion of unexplained variability with regard to the target variable will also decrease, leaving less uncertainty when estimating the mean. Formally, the variance of the conditional mean $\hat{\bar{y}}_f$ is expressed as $S^2_{\hat{\bar{y}}_f} = a\bar{u}x'_f mse\left(aux'_r aux_r\right)^{-1} a\bar{u}x_f$. When only one auxiliary variables is involved, the equation can be simplified as $S^2_{\hat{\bar{y}}_f} = mse\left(\frac{1}{n_r} + \frac{\left(a\bar{u}x_f - a\bar{u}x_r\right)^2}{\sum_{i=1}^{n_r}(aux_i - a\bar{u}x_r)^2}\right)$. Clearly, the smaller the mse (second condition), the smaller the variance of $\hat{\bar{y}}_f$. This variance will increase as the mean of the full sample ($a\bar{u}x_f$) becomes more distant to the mean of the respondent-only ($a\bar{u}x_r$) sample (first condition).

When weight scores are used instead of the regression framework, standard errors can be determined by using Taylor-series approximation or other advanced techniques such as jackknife or replication methods (Groves et al., 2009; Lohr, 1999). They all follow the principle that an increasing distance between the full sample and the respondent-only sample leads to variance inflation, whereas stronger correlations between the weight scores and the target variable have a decreasing effect on the variance of $\hat{\bar{y}}_f$. This has led to the conclusion that nonresponse does not necessarily imply less precision (=variability). Little and Vartivarian (2005) even suggest that '[...] the most important feature of variables for inclusion in weighting adjustments is that they are predictive of survey outcomes; prediction of the propensity to respond is a secondary, though useful, goal.'

This assertion is quite remarkable in two respects. First, the second condition (correlation between weight scores and target variables) is an attractive statistical property, even when going beyond the scope of nonresponse. Simply adding supplementary or relevant information during the estimation process will usually lead to a decrease in the standard error. In other words, there is no real need to use nonresponse as a reason or an ex-

cuse to deploy auxiliary variables in order to enhance the efficiency of the estimation. Therefore, the net effect of nonresponse on standard errors can only be disadvantageous. Consider in this regard Figure 1.13 again and suppose that $a\bar{u}x_f$ would be on the same location as $a\bar{u}x_r$. In that situation, taking advantage of the auxiliary information leads to more efficient estimates compared with the uninformed situation (left-hand side of Figure 1.13). Now, as the distance grows between $a\bar{u}x_f$ and $a\bar{u}x_r$, the corrected mean needs to be estimated outside the center of the data, leading to larger confidence intervals.

Second, if the second condition is to be considered more important than the first, survey researchers may be tempted to only include auxiliary variables that decrease the variance, without really eliminating the bias. As a result, it might be possible to become very confident about an estimate that is still potentially wrong. Therefore, it might be wiser not to aspire to the smallest possible standard error, but instead to strive for correct standard errors. Such standard errors allow the full-sample estimate or true parameter to belong to the respondent-only confidence interval. If this means that standard errors need to be somewhat larger, survey researchers have no other choice than to accept. Unfortunately, such standard errors are very hard to obtain, as the full-sample target estimates are unknown, as is the mechanism that leads to nonresponse.

For this reason, survey researchers tend to make assumptions about the nonresponse mechanism, such as assuming MAR instead of MNAR. However, it is striking that the additional uncertainty induced by the assumptions is rarely reflected in the standard errors. As a result, these standard errors may be (severely) underestimated. This leads to the hypothesis that variance inflation factors F under the random response model are (much) smaller than their fixed response model counterparts.

For the ESS2, Vehovar (2007) determined weight scores to correct for nonresponse for all the participating countries, using gender (male, female), age (15 to 34,35 to 54 and 55+) and educational level (lower secondary or less, higher secondary and post-secondary). For some countries (Belgium, the Czech Republic, Estonia, Hungary, Greece, Ireland, the Nether-

lands, Norway, Poland and Portugal) the raking method was used, for all other countries, post-stratification was used. In some countries, the design weights were also included in order to correct for unequal selection probabilities, due to the fact that the sampling was based on addresses rather than a list of individuals (giving relatively higher selection probabilities to members of small families). The variance inflation factors of all the 24 ESS countries in round 2 ranged from 1.02 (Finland and Poland) to 3.31 (Ukraine) and even 4.02 (France). Compared with the variance inflation factor (ψ) under the fixed response model (see Table 1.16), these variance inflation factors F seem to be substantively lower.

The fact should be noted that the variance inflation factor in the random model, as determined by Vehovar (2007), is obtained for some countries by comparing the obtained respondent sample and the total population, instead of the full sample. This means that weights do not only accommodate nonresponse imperfections, but also take frame error and sampling error (and even measurement error) into account. Therefore, the actual net effects of nonresponse on these variance inflation factors F might be different to those presented by Vehovar (2007).

Data analysis For all the surveys that were presented in Table 1.16 where the variance inflation factors ψ under the fixed model were determined can now also be used to determine their counterpart F under the random response model. Therefore, individual propensities ρ_i are estimated by logistic regression, using auxiliary variables as independent variables. In order to assess the extent to which standard errors change under the random response model, variance inflation factors F will be estimated based on equation 1.32, informed by the coefficient of variation among the weight scores $w_i = \bar{\rho}/\rho_i$.

Table 1.17 shows the previous results under the fixed model in grey (see also Table 1.16). The last two columns in Table 1.17 show the variance inflation factor F under the random model, one when only main effects are allowed in the logistic model (F_{main}), the other where all pairwise interactions between all auxiliary variables are added to the model ($F_{interact}$).

Table 1.17: Comparison of survey-level nonresponse analysis between the random and fixed response model for various surveys

				Fixed		Random	
Survey	n	n_r	$\bar{r}$	ψ_{loc}	ψ_{ass}	F_{main}	$F_{interact}$
ESS3-BE	2927	1798	0.6143	11.74	6.28	1.04	1.09
GPS	32019	18792	0.5869	81.22	7.33	1.05	1.13
FHS	7770	5216	0.6713	14.17	1.29	1.03	1.10
ESS5-BE	3199	1669	0.5216	9.09	1.49	1.05	1.06
ESS5-BG	2997	2280	0.7606	2.29	1.38	1.01	1.02
ESS5-CY	1564	1059	0.6769	5.17	1.95	1.03	1.14
ESS5-DK	2856	1554	0.5441	14.82	1.49	1.05	1.06
ESS5-ES	2727	1794	0.6579	2.01	1.74	1.01	1.02
ESS5-GR	4230	2715	0.6418	3.81	1.07	1.01	1.02
ESS5-HU	2612	1547	0.5924	2.08	1.56	1.01	1.03
ESS5-IL	3230	2294	0.7102	2.13	2.79	1.01	1.03
ESS5-PT	3264	2382	0.7299	5.32	1.66	1.01	1.02
ESS5-RU	3982	2595	0.6517	40.42	2.10	1.04	1.05

Under both the main effects model and the extended logistic model with interactions, the variance inflation factor remains relatively low. For each of the individual surveys, the estimated variance inflation factors F are far below the estimated variance inflation factors ψ_{loc} for location parameters and ψ_{ass} for association parameters that are obtained under the fixed response model, supporting the hypothesis stated earlier: that variance inflation factors F under the random response model will be (much) smaller than their fixed response model counterparts ψ.

A possible explanation for the strong differences between the two nonresponse model interpretations might be the fact the estimated response propensities under the random model reflect the true propensities fairly well, correcting occasional bias to such an extent that additional variance inflation is no longer a real need. However, such speculation leads to the paradox that the fewer (relevant) auxiliary variables that are used to model the response propensities, the less the variance among propensities that is generated, which in turn leads to less variance inflation in the eventual estimates. Conversely, one would instead expect that a more extensive

model (comprising more auxiliary variables and interaction terms) would more adequately reflect the underlying nonresponse mechanism, somewhat increasing the variance inflation factors F. A more adequate explanation is that the auxiliary variables in the random model are only capable of installing a weak form of representativeness, and only protecting the auxiliary variables themselves against bias. As a consequence, target variables are only partially protected, and still in need of additional uncertainty bounds. ■

1.2.7 Revisiting the fixed response model and the random response model

The distinction between the fixed and random response models has been chosen as the *leitmotiv* for structuring this chapter, because the two models look at the problem of nonresponse from completely different angles, although in practice they use the same information. The random response model considers nonresponse as the probability (or differences between probabilities) of reacting positively to a survey request, or as the possible result of weighing up the pros and cons of survey participation by the individual (non)respondents, given the survey design. Therefore, the random model is somewhat process oriented. It uses auxiliary variables to roughcast the nonresponse mechanism and to determine response propensities that, in turn, can be used to determine the damage due to nonresponse caused to the representativeness of the survey, or to particular statistics in the survey.

The fixed response model instead considers nonresponse as the reason why some auxiliary variables (or at least variables available for both respondents and nonrespondents) are biased. This model is therefore more output oriented.

However, it is quite striking that evaluation of the possibly detrimental effects of nonresponse on the quality of the survey leads to divergent conclusions: empirical findings suggest that the random model provides a more positive impression than the fixed model does.

It is very likely that the only limited set of auxiliary variables under the random response model is not fully capable of completely explaining the underlying (non)response mechanism. As a result, the variance of response propensities is too small, underestimating the damage to the representativeness of the respondent-only sample as well as its associated risk of bias. This can easily be observed when considering the expression for bias and representativeness (R-indicator) under the random response model:

$$\begin{aligned} R_\rho &= 1 - 2\sigma_\rho \\ bias_{\bar{y}_r} &= \frac{corr_{\rho y}\sigma_\rho\sigma_y}{\bar{\rho}} \end{aligned}$$

In particular, the variance of the estimated response propensities may in many cases be too small, as it is only determined by a limited set of auxiliary variables. Usually, only some socio-demographic variables, such as age and gender are available with which to determine the propensities. This means that other dimensions that have an effect on (non)response are not reflected in the variability of response propensities.

Minimal variance is also an explicit objective in the context of survey adjustment through weighting or calibration. The adjustment method is driven by the need to find weight scores such that the distributions of a set of auxiliary variables are identical between the respondent set and the full sample (or population). There may be a countless number of possible weight vectors that satisfy the condition, however, there is only one vector that also has the lowest possible variance of weight scores. This latter vector is the one that is usually chosen. In this regard, consider Table 1.18, representing fictitious data.

The full sample comprises six men and six women. Only two men and four women responded positively to the survey request. There may be an infinite number of possible weight vectors that establish the correspondence between the obtained respondent set and full sample with regard to the gender distribution. Only four of the possibilities are presented in the table (w_{min}, w_{alt1}, w_{alt2} and w_{alt3}). Usually, only w_{min} would be applied

Table 1.18: Alternative weight vectors, obtaining the same point estimate (% males) but with different variance inflation factors F, fictitious example

aux	r	w_{min}	w_{alt1}	w_{alt2}	w_{alt3}	y
male	1	1.50	0.98	0.45	2.22	1.50
male	1	1.50	0.26	0.40	0.05	2.50
male	0					?
male	0					?
male	0					?
male	0					?
female	1	0.75	2.48	3.24	0.27	1.75
female	1	0.75	0.54	0.85	0.23	2.25
female	1	0.75	2.83	2.98	0.63	2.00
female	1	0.75	0.66	0.23	3.33	3.00
female	0					?
female	0					?
% males		0.50	0.50	0.50	0.50	
F		1.15	1.71	2.02	2.43	

as it (1), sufficiently corrects the point estimate with regard to gender distribution and (2), provides the smallest variance inflation (Deville & Särndal, 1992; Deville, Särndal, & Sautory, 1992; Bethlehem et al., 2011). The first argument ($\theta_{r,gender} = \theta_{f,gender}$) is straightforward and does not require further discussion. However, the second argument (bringing all w_i as close as possible to 1, resulting in the minimization of the variance inflation factor F) is much harder to justify. Because the mean of the target variable is not known, the probability of the vector w_i leading to invalid inference is quite substantial. Alternative weight vectors might be more adequate, but are usually not applied. This point is particularly relevant as gender is probably not the only variable that is related to survey participation. This means that within the categories of men and women unknown variability of response propensities should be accounted for. Using w_{min} clearly fails to reflect such within-group variability.

Data analysis These considerations may lead to the conclusion that response propensities, estimated conditional on auxiliary variables, are very

Table 1.19: Evolution of the R-indicator per added auxiliary variable - GPS ($n = 32019$)

Added variable	**R-indicator**
Age classes	0.9434
Disability allowance received	0.9390
Ethnic background	0.8738
Gender	0.8734
Has job	0.8707
Household type	0.8424
House value in neighborhood	0.8317
Marital status	0.8207
Phone (listed)	0.7941
Social allowance received	0.7936
Unemployment allowance received	0.7935
Urbanization	0.7772

likely to overestimate the representativeness of a survey. As an illustration, consider Table 1.19, where auxiliary variables from the Dutch GPS are used. First, all auxiliary variables have been sorted alphabetically. Next, propensities have been determined only conditional on the first auxiliary variable (age), with which the R-indicator is calculated. Next, the propensities are estimated conditional on the first two auxiliary variables from which the R-indicator is recalculated, and so forth.

The crucial idea behind Table 1.19 is that adding new auxiliary information usually decreases the R-indicator. Suppose now that instead of the twelve auxiliary variables, only the first six had been available. In this case, the R-indicator of 0.8424 is obviously an overestimation of the true representativeness indicator. This suggests that even when considering the full set of twelve auxiliary variables, it is very likely that the R-indicator still has the potential to decrease further. ■

Considering the expression of bias under the random response model again, not only is the variance of response propensities very likely to be underestimated, but the correlation between response propensities and the target variable may also be a problem when determining the bias with

regard to the target variables. As the response propensities are informed only by a limited and selective set of auxiliary variables, the correlations between propensities and target variables will probably also reflect that selectivity. This implies that a target variable that is strongly related to the auxiliary variables is more likely to be found biased than a target variable that is not related to the auxiliary variables, even if both target variables have, in reality, the same level of bias. This further implies that adjustment procedures will not correct all target variables equally. In this regard Särndal and Lundström (2008) comment that 'in practice, it is impossible to designate a vector that will completely eliminate the bias. Even the best of auxiliary vectors leave some bias remaining ...'. In section 1.3.2, this problem will be further elaborated.

The implications of working with the random response model are to accept that the data is missing at random (MAR), and to assume that the nonresponse mechanism can be completely explained by a set of available auxiliary variables. Andridge and Little (2011) have recently developed a method that relaxes the MAR assumption toward MNAR, by allowing an additional parameter to determine how strongly the auxiliary variable(s) and the target statistic are related. This parameter λ takes the value 0 when assuming MAR and can move toward infinity where it accommodates the most extreme position of MNAR. When moving between 0 and infinity, the estimated target statistic can take several outcomes, supporting the idea that sensitivity analysis is most appropriate in the context of nonresponse. The continuum between MAR and MNAR can also be expressed as the fraction of missing data that is lost due to nonresponse (Wagner, 2010).

Is it quite clear that the random response model, based on response propensities that are estimated conditional on auxiliary variables, is very likely to underestimate the impact of nonresponse. The fixed model, on the other hand, is much more difficult to position on the continuum of under or overestimating the nonresponse effects. This latter model only uses auxiliary variables to provide an idea of how nonresponse affects the precision of survey estimates. The critical question is whether these auxiliary vari-

ables are representative of the surveys' target variables. This is clearly an untestable problem. Nevertheless, taking auxiliary variables as (imperfect) representatives or substitutes for the target variable will provide a more realistic idea of nonresponse damage in a survey. It is therefore advisable not simply to collect every possible auxiliary variable, but to look for a more balanced set of variables that covers as many dimensions as possible with regard to the nonresponse mechanism.

Stating that auxiliary variables are good representatives of target variables is of course an assumption, for which additional uncertainty should be provided. This is an argument to support the idea that even the fixed model underestimates the problem of nonresponse and that the estimates of ψ as presented in the empirical section should be somewhat higher.

Nonetheless, studying nonresponse will always be hindered by the fact that the central point of interest remains concealed. Therefore, nonresponse researchers need to find ways to observe the unobservable, which inevitably implies making assumptions or even speculations about the research subject. It is therefore better not to rely on only one perspective or method. The distinction between the random and the fixed response models can be considered as a good starting point to assess nonresponse error. Going further in the development of advanced and innovative methods in nonresponse research should be encouraged. The next section discusses some of these new ideas.

1.3 Advanced methods to assess the effects of nonresponse

This section does not seek to give a complete overview of recent research activities concerned with assessing the impact of nonresponse, but instead highlights some of the possibilities of combining some of the elements presented so far into new assessment methods. In addition, integrating new sources (paradata or process data in particular) may throw a promising and inspiring light on our understanding of nonresponse in surveys.

The first methods described in this section use fieldwork process data to estimate the variance of response propensities in an alternative way. More specifically, measuring the progress of response rates during the fieldwork in order to derive the variance of the response propensities. This method can obviously be classified under the random response model. The second method will combine the random and the fixed response models by iteratively regrouping a set of auxiliary variables into one group of weight variables, leaving the remaining group as interim target variables in order to assess the effects of propensity weighting.

1.3.1 Monitoring the progress of response rates to determine the variance of response propensities

In practice, response propensities are determined by regressing the 0-1 response outcome by a set of auxiliary variables. In previous sections it has been argued that this may lead to severe underestimation of the variance of such obtained propensities, consequently also underestimating the damage resulting from nonresponse. Therefore, it may be worthwhile looking for methods of measuring the effects of (non)response that go beyond the use of auxiliary variables. Recently, some survey researchers have started exploring the possibilities of using what is termed paradata (Couper, 1998). Paradata is not necessarily the target data of a survey, but rather comes as a by-product and includes information such as interviewer observations, the number of call attempts, the time of day of the visit, the contact methods, etc. For an overview of the use and definitions of paradata, see e.g. Kreuter and Casas-Cordero (2010) or Olson (2013b). In the second part of this dissertation, paradata will be used more when discussing field operations and strategies. Here, paradata is used within a very specific context in order to estimate the variance of response propensities.

A specific field in the use of paradata consists of estimating response propensities. Biemer and Link (2007) and Biemer, Chen, and Wang (2012), elaborating on the ideas of Drew and Fuller (1980) propose modeling response propensities as a function of the level of effort (LOE) needed to

turn a sample case into a respondent. This approach assumes a 'continuum of resistance model', where individuals who need more effort in order to become respondents are deemed to have lower response propensities than individuals who respond immediately to the survey request.

This track will be further investigated. The idea is to ignore auxiliary variables and determine the 'true' or at least a more realistic estimate of the variance of response propensities in an alternative way. A simple illustrative introduction reads as follows.

Suppose in sample A 100 individuals all have a response propensity of $\rho_i = \bar{\rho} = 0.5$, implying that the propensity variance equals 0. After the first contact attempt, the expected response rate will evidently reach 50%, increasing to 75% when subsequently re-contacting all initial nonrespondents during the follow-up. In sample B 50% of the sample cases have a propensity of 100% and the other 50% have a 0% response propensity, implying the maximal possible variance of propensities $\sigma^2_\rho = 0.25$. Whereas in sample A, the response rate evolves from 50% to 75%, the response rate in sample B will not increase after a follow-up attempt among the initial nonrespondents. This means that the progress of the response rate can be used as a means to determine the variance of response propensities:

$$\sigma^2_{\rho_1} = 1 - (1 - \bar{\rho}_1)^2 - \bar{\rho}_2 \tag{1.33}$$

where $\sigma^2_{\rho_1}$ is the propensity variance at the first contact attempt and $\bar{\rho}_1$ and $\bar{\rho}_2$ are respectively the cumulative response rates after the first and the second contact attempts. Starting from the cumulative distribution function of a geometric distribution $1 - (1 - \rho)^k$, the expected probability after two trials is expressed as $\rho_2 = 1 - (1 - \rho_1)^2$. If $\bar{\rho}_1$ and $\sigma^2_{\rho_1}$ were known, $\bar{\rho}_2$ could be obtained by using Taylor expansions for the moments of functions of random variables: $E[g(X)] \approx g(\mu_X) + \frac{g''(\mu_X)}{2}\sigma^2_X$. Applied to the progress of the first and the second attempt response rates, this means that:

$$\bar{\rho}_2 = 1 - (1 - \bar{\rho}_1)^2 + \frac{g" \left(1 - (1 - \bar{\rho})^2\right)}{2} \sigma^2_{\rho_1}.$$

The second derivative of the equation $1 - (1 - \rho_1)^2$ equals -2, such that

$$\begin{aligned} \bar{\rho}_2 &= 1 - (1 - \bar{\rho}_1)^2 + \frac{-2}{2} \sigma^2_{\rho_1} \\ &= 1 - (1 - \bar{\rho}_1)^2 - \sigma^2_{\rho_1} \end{aligned}$$

Rearranging the terms implies that

$$\sigma^2_{\rho_1} = 1 - (1 - \bar{\rho}_1)^2 - \bar{\rho}_2$$

In sample A, $\bar{\rho}_1 = 0.50$ and $\bar{\rho}_2 = 0.75$, implying that $\sigma^2_{\rho_1} = 1 - (1 - 0.50)^2 - 0.75 = 0$. In sample B, $\bar{\rho}_1 = 0.50$ and $\bar{\rho}_2 = 0.50$, implying that $\sigma^2_{\rho_1} = 1 - (1 - 0.50)^2 - 0.50 = 0.25$.

A more intuitive way to understand the progress of response rates as an indication of the variance of response propensities is the following: provided that not all individuals have the same propensity to respond positively to the survey request, the first contact will probably skim off the high propensity cases. The group of remaining nonrespondents will therefore contain individuals who are not so inclined to participate. This means that a second attempt to convert these initially nonresponding cases will yield less cases compared with a situation where all the cases had equal response propensities.

Though the model to obtain the propensity variance is fairly simple, the estimation of the two response rates can be somewhat complicated. Many practical obstacles, such as interviewer memory, under-reporting of visits, and the definition of response rates hinder an easy computation.

1. The cumulative geometric distribution function does not take into account memory effects that inevitably operate during the fieldwork. On the one hand, serial correlation can be expected with regard to

the respondents, e.g. noncontact or refusal may occur when a unit is contacted twice on the same day. Such effects probably decline as the elapsed time increases. Such a memory effect suppresses $\bar{\rho}_2$ and may lead to an overestimation of $\sigma^2_{\rho_1}$. Conversely, and probably the strongest memory effect, the tactics of the interviewer (or fieldwork management) such as sending another interviewer, alternative timing, other doorstep arguments, etc., will probably enhance $\bar{\rho}_2$. Moreover, fieldwork tactics may anticipate the unfavorable effect of the memory of initially nonresponding sample cases (e.g. postponing the second attempt or purposely choosing a different time of the day or week). Therefore, it is expected that $\bar{\rho}_2$ will be overestimated rather than underestimated, leading to the underestimation of $\sigma^2_{\rho_1}$.

2. Not all initial nonrespondents are given a second opportunity. Chapter 2 will discuss that high propensity cases usually have a greater probability of being re-selected for conversion activities. As a result, it can be expected that generalizing the response among the attempted-only cases toward all initial nonrespondents implies an overestimation of $\bar{\rho}_2$, leading to an underestimation of $\sigma^2_{\rho_1}$. It can therefore be recommended to predict the response probabilities of unattempted cases, conditional on their initial reason for non-participation (e.g. refusal, illness, noncontact, language barrier, etc.).

3. A third issue relates to the quality of the contact history data. For a proper estimation of $\bar{\rho}_1$ and $\bar{\rho}_2$, all first and second contact attempt should be reported. Under-reporting of attempts by the interviewer is expected to occur more frequently than over-reporting (Bates, Dahlhamer, Phipps, Safir, & Tan, 2010; Wang & Biemer, 2010). If the amount of under-reporting is relatively stable at all stages of the fieldwork, than the estimation of the propensity variance will not be influenced substantively. However, as it is unclear whether the first or the second attempts are more under-reported, the potential influence on the estimation of $\sigma^2_{\rho_1}$ is very difficult to assess.

Other drawbacks when using expression 1.33 relate to the fact that the variance is only available on the sample level; no individual propensities can be estimated. In addition, the variance can only be determined for the first contact attempt, while the variance of the propensities at the end of the fieldwork is usually much more important. Chapter 2, however, will deal with the way the sample cases are prioritized during the fieldwork, suggesting an increase or decrease of the variance toward the closure of the data collection. Finally, the estimated propensity variance also has a standard error, which can be considerable high, particularly if a second contact attempt is assigned to a selected subset of nonrespondents.

This method is interesting, as it confronts more traditional methods of response propensity modeling, particularly those based on auxiliary variables. Nevertheless, it is highly speculative as it builds on many unknowns and assumptions. Therefore, this method should be considered as an additional instrument to assess nonresponse, in addition to the more traditional methods presented earlier.

Data analysis In the FHS, after one contact attempt the initial response rate $\bar{\rho}_1$ equals 35.44%. Of the remaining nonrespondents, only 81.44% were approached again, implying that $\bar{\rho}_2$ cannot be determined without obtaining an expectation of the success probability of these censored cases. Fortunately, information about the reason for non-participation is available for all initial nonrespondents (noncontact, soft refusal, hard refusal, illness, language barrier, etc.), therefore for all these nonresponding profiles, a response probability is available after the second contact attempt. Imputing the success probabilities among non-issued nonrespondents for this second attempt, conditional on the information of the first attempts, $\hat{\bar{\rho}}_2$ equals 53.91%. As a result, the variance of the response propensities after the first contact attempt is $1-(1-\bar{\rho}_1)^2-\hat{\bar{\rho}}_2 = 1-(1-0.3544)^2-0.5391 = 0.0441$. The resulting $R-indicator$ is 0.58, the maximal absolute standardized bias is 0.59 (contrast=0.92).

Unfortunately, the expression to estimate the variance assumes that the two attempts are carried out independently. Given the unavoidable mem-

ory effects operating during the fieldwork, the assumption of independence seems to be too strong. However, some considerations can be taken into account, suggesting whether the obtained result is an over or underestimation of the propensity variance. If it is assumed that the initial response behavior is perpetuated on a following occasion as a result of the memory of the nonrespondents, then memory will have a decreasing effect on $\hat{\tilde{\rho}}_2$, suggesting that the propensity variance is rather overestimated under such circumstances. Conversely, there are many more reasons to suggest that $\hat{\tilde{\rho}}_2$ will increase because of knowledge about previous fieldwork events, particularly by the interviewers and/or fieldwork management. These latter survey agents have a particular interest in enhancing response rates and can therefore use previous information in order to improve their fieldwork successes. As an illustration, consider the relationship between the probability of an initially nonresponding sample case being converted and the same case being selected for a conversion attempt. As interviewers are usually paid per completed interview, their interest is predominantly in reissuing cases for which a relatively high success rate is expected. Hence, the correlation between the re-selection probability and the conversion probability among initially nonresponding cases is strictly positive: 0.46. This means that initial nonrespondents who are deemed to be more responsive on the next occasion will have a higher probability of being revisited. Initial nonrespondents that interviewers think will not cooperate at renewed contact attempts, are usually less likely to be re-contacted. As such, fieldwork interventions can be considered to follow the path of least resistance (see Chapter 2 for a more detailed discussion). This also underscores the importance of estimating the response probabilities of non-issued cases conditional on their specific nonresponse profile.

The variance obtained by comparing the response rates after the first two attempts can now be compared with the variance determined by auxiliary variables (an overview of the variables is given in Table 1.1). Accordingly, a logistic model is built, regressing the response outcomes after the first contact attempt. The model produces propensities of which the variance equals only 0.0076 (as compared with 0.0441 provided by the

alternative method). Consequently, it seems that the method using the progress of the response rates generates about six times the variance that is produced by the method based on auxiliary variables. This means that the overall capacity of the auxiliary variables to restore the representativeness of the respondent sample is relatively low and this also supports the argument already presented throughout this current chapter: that the use of auxiliary variables in the random response model is very likely to underestimate the impact of nonresponse. ■

1.3.2 Combining the fixed and the random model: an assessment of the efficacy of weighting adjustments

Weighting techniques are generally considered as means to combat the unfavorable effects of nonresponse. Although these techniques may build on theoretically and mathematically convincing arguments, they still rely on assumptions that are inevitably made about the data. As already mentioned on page 82, many authors (see, among others, Kalton & Flores-Cervantes, 2003; Little & Vartivarian, 2003, 2005; Groves, 2006; Kreuter et al., 2010) argue that two condition should be met in order to apply these correction techniques:

1. The weight variables are related to survey participation or (non)response
2. The weight variables are related to target survey variables.

A crucial and virtually untestable assumption in this regard is the relationship between the response outcome and the target variable, as indicated by Figure 1.14.

The relationship between the response outcome r and the target variable y is unknown. If both factors are substantially related, bias is very likely to occur. The weight vector w that is informed by the auxiliary variable(s) can only remove occasional bias under very specific conditions. In fact, the relationship between the corners of the triangle in Figure 1.14

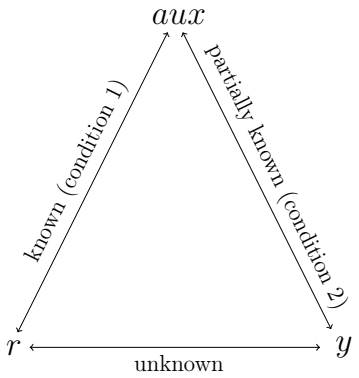


Figure 1.14: Relationships between auxiliary variable(s) aux, response outcome r and target variable y

should satisfy the condition that the partial correlation $corr_{ry.w}$ is acceptably close to zero. Otherwise, some unfavorable situations may emerge as illustrated in Figure 1.15.

In any of the six possible configurations shown in Figure 1.15, it is assumed that the auxiliary variable(s) aux explains some proportion of the response behavior r, therefore $corr_{r,aux}$ is expected to be substantial. It is also assumed, mostly for reasons of convenience, that the relationship between the auxiliary variable(s) and the target variable is either zero or positive.

The different scenarios also indicate the distance between the full-sample mean $\bar{y}_f$, the unweighted respondent-only mean $\bar{y}_{unw}$, and the weighted mean $\bar{y}_w$. For example, in the (*'no problem'*) scenario, the three means do not differ (no dash in-between the different $\bar{y}$'s) because y is not related to either (non)response or the auxiliary variable(s). This first scenario (*'no problem'*) is probably the most preferable, as it does not change an already unbiased estimate. The second scenario is also acceptable, as is corrects the initial bias completely. The remaining four configurations are obviously less favorable. Some initial (unweighted) estimates may not be shifted at all (*'no correction'*), or only partially (*'partial correction'*) toward the full-sample mean, other estimates may be shifted even further away (*'even worse'*). It is also possible that an initial unbiased parameter becomes biased after an adjustment procedure (*'bias creation'*).

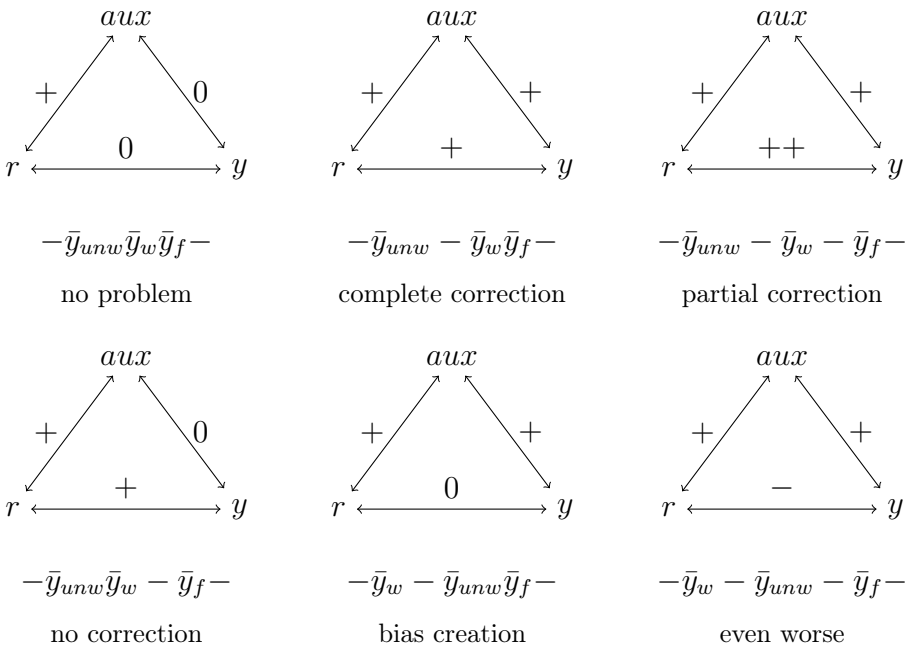


Figure 1.15: Six possible triangular relationships between auxiliary variable(s) aux, response outcome r and target variable y

Since the relationship between the target variable(s) and the response outcome is unknown, it is only possible to judge the impact of weighting by using the other two relationships. From that perspective, scenario 1.15a (*'no problem'*) and 1.15d (*'no correction'*) are similar, in that the unweighted mean will not be shifted. The remaining scenarios indicate a shift from unweighted to weighted means, in which the last two situations (*'bias creation'* and *'even worse'*) are clearly undesirable false positive reactions to the weighting procedure. Consequently, because $corr_{ry}$ is unknown, assessment of the efficacy of post-survey weighting procedures is practically impossible, leaving the adjuster with only a taste of hope rather than indisputable certainty.

Indeed, empirical findings in relevant literature seem to endorse the difficulty of finding a set of auxiliary variables that eliminate non-response bias, without possibly even making the position worse (Brick, Lê, & West, 2003; Little, , Heeringa, Lepkowski, & Kessler, 1997; Little & Vartivarian, 2003; Peytcheva & Groves, 2009). Dey (1997) reports that weighting for

non-response bias in a student survey seemed to be highly effective with regard to univariate distributions, whereas the effectiveness of weighting procedures on correlations and regression estimates was less clear. In a longitudinal setting, Vandecasteele and Debels (2007) found that weighting provided a solution for bias in some cases. However, weighting could even adversely affect the estimated parameters, although only slightly in some cases. They also mention that the inclusion of survey-related design information into the weighting model may improve the effectiveness. Peytchev et al. (2011) report that survey estimates may improve somewhat after weighting adjustments. As demographic background variables are normally used to construct the weight vector and are usually not strongly associated with the target statistics, some bias will nevertheless still remain. Other researchers also encourage the collection of more and better auxiliary information, such as interviewer observations, in order to correct estimates that are potentially biased due to non-response (Kreuter et al., 2010).

Although the fixed model and the random model can be considered as theoretical opposites, they can be integrated into one method in order to assess the efficacy of weighting. Suppose that only two auxiliary variables are available for both respondents and nonrespondents. The first variable can then be used as a temporary or interim target variable, while the second can be used to determine the individual weight scores. Next, the second variable will act as a target variable, while the first is considered as a weight variable. The underlying assumption here is that the way this casual or interim target variable reacts to the weighting adjustment is indicative for a real target variable. The greater the number of auxiliary variables that are available, the greater the number of combinations of auxiliary sets that can be constructed and the greater the number of interim target parameters that can be monitored. It should also be noted that not only location parameters such as means and proportions can be assessed. Associations such as correlations between the target variables can also be evaluated. This is an opportunity, as survey research often exceeds the level of descriptive statistics. Research on (nonresponse) bias nevertheless

most frequently focuses only on means and proportions (Peytchev et al., 2011).

As an illustration of the method, suppose that three auxiliary variables A, B, and C are available for all respondents and non-respondents, and are used to build eight different weighting models (see Table 1.20). It is clear that the first set coincides with the unweighted statistics and that the eighth set is no longer relevant, as no target information is left to assess. In any of the other sets, variables that are not included in the weighting model can be used as interim target variables, either in the form of location parameters or combined as association statistics. If a variable is continuous, its mean will be monitored. In the categorical case, the proportions of the categories will be monitored. For association measures, a regression model can be run where the response variable is continuous and the independent variable is either binary or continuous. Alternatively, a binary variable can be taken as the response variable in a logistic regression. In both situations (linear or logistic regression), the slope parameters can be monitored through the different weighting configurations. Association parameters are less likely to act as a target statistic, compared with location statistics, since both their constituent variables should be excluded from the auxiliary set. As a result, any location parameter can be monitored $\binom{p-1}{0}+\binom{p-1}{1}+\ldots+\binom{p-1}{p-1}$ times as an interim target parameter, and association parameters can be monitored $\binom{p-2}{0}+\binom{p-2}{1}+\ldots+\binom{p-2}{p-2}$ times. Here, p represents the number of variables available for both respondents and non-respondents. Applied to the situation in Table 1.20, the location parameter for, for example variable A, can be monitored $\binom{3-1}{0}+\binom{3-1}{1}+\ldots+\binom{3-1}{3-1}=4$ times. An association parameter such as A$\times$B only appears twice as a target parameter.

Considering Table 1.20, it becomes clear how the fixed model and random model are integrated. The first column of the table (weighting variables) clearly takes the position of the random response model as it assumes that all individual sample cases have a propensity to respond to the survey request. Regressing (e.g. logistic regression) the response outcome (response versus non-response) by the particular set of auxiliary variables

Table 1.20: Mutually exclusive sets of auxiliary variables and interim target variables, illustration

	Set of weighting variables	Target location parameters	Target associations
1	∅	A B C	A×B A×C B×C
2	A	B C	B×C
3	B	A C	A×C
4	C	A B	A×B
5	A B	C	∅
6	A C	B	∅
7	B C	A	∅
8	A B C	∅	∅

easily generates a vector containing all the individual response propensities ρ_i, from which the weight scores $w_i = \bar{\rho}/\rho_i$ can be derived. The second and the third columns represent the target variables (fixed model interpretation) from which the impact of the adjustment procedures can be assessed. In the empirical example, the aforementioned safety indicator and waste indicator (see equation 1.23 and 1.24 on page 62) will be used to evaluate how well the (weighted) respondent-only confidence interval spans the density of the full-sample parameter estimate and vice versa. When applying equations 1.31 on page 71, the variance inflation factor ψ is determined in order to have an average safety indicator of 95%.

Ultimately, this method is indicative of the extent to which propensity weighting (random response model) is successful in eliminating occurring nonresponse bias, and to what extent the variance inflation factor ψ (fixed response model) should compensate in order to avoid type I errors. Given the considerable number of auxiliary variables in the datasets, the method presented above also has the advantage that it provides a large number of estimates under different adjustment combinations, such as unweighted estimates, weighted by only one variable, weighted by two variables, and so forth. The ability to combine the interim target variables into association statistics is a further advantage of this technique.

However, a possible disadvantage is the dependence on a particular set of auxiliary variables. Usually, auxiliary variables provide residence-related information about the area in which the sampled household or individual lives, interviewer observations about the type of housing, the presence of green spaces in the neighborhood, etc. Such information may only cover one or a few non-response dimensions, whereas other reasons for non-participation such as attitudes or behavior will remain concealed. If the available auxiliary variables are mutually strongly correlated, and are not so strongly correlated with the real target variables, the efficacy of weighting adjustments may be overestimated.

Data analysis The same data from the ESS3-BE as presented on page 36 will be used in order to present these theoretical considerations using empirical evidence. The dataset includes 8 auxiliary variables, representing 25 parameters that are expressed as percentages (4 age classes, 5 population density classes, 3 regions, 4 classes of the percentages of non-Belgians in the municipalities, 4 municipality income classes, 3 neighbourhood quality classes and 2 parameters indicating type of dwelling and gender). Some variables also have a continuous counterpart (age, percentage of foreigners, population density and average municipality income), resulting in four additional location parameters. It should be noted that the analyses below will be weighted so that each of the eight variables contributes equally. For example, this means that the single location parameter associated with 'type of housing' will be assigned more importance than each of the 6 (5 categories and 1 continuous variable) location parameters associated with 'population density'.

In addition to the 29 location parameters, 42 measurements of association are prepared. These are combinations of all the auxiliary variables where only one (logistic) regression parameter is needed in order to express the relationship between two variables (see Table 1.14). It is clear that these parameters are only taken into account where the contributing variables do not take part in the weighting model to obtain the response propensities ρ_i. For each location parameter, $\left[\binom{7}{0} + \binom{7}{1} + \ldots + \binom{7}{7}\right] = 128$

different weight models will be run. Association parameters will each have $\left[\binom{6}{0} + \binom{6}{1} + \ldots + \binom{6}{6}\right] = 64$ weight configurations. For example, the percentage of apartment dwellers can be weighted consecutively by 7 weight vectors, with only one auxiliary variable determining the response propensities. Given that there are 29 location parameters to be monitored, there are 203 (7×29) location parameters to be monitored that are weighted by only one auxiliary variable. For each location parameter, 21 weight vectors can be determined combining any two remaining auxiliary variables, resulting in 609 (21×29) weighted parameter estimates, etc. These figures can also be found in Table 1.21, where the number of monitored target parameters are given as a function of the number of auxiliary variables used for weighting, both for location parameters and association parameters.

All the weight vectors are obtained by modeling the 0-1 response outcome through logistic regression, taking the auxiliary variable(s) as explanatory variables and only allowing main effects. Only the categorical versions of the auxiliary variables are considered as explanatory variables, not their continuous counterparts. This means that, for example, the continuous variable 'age' will only be used as an interim target variable, never as an explanatory variable in order to determine the response propensities $\hat{\rho}_i$.

In total, 3712 location parameters and 2688 association parameters can be monitored. Not only can the point estimates of these parameters be obtained, but their standard errors can also be determined. In this analysis, the standard errors are determined through the Taylor Linearisation Method, using `PROC SURVEYMEANS, SURVEYREG and SURVEYLOGISTIC` in SAS. Since the full-sample counterparts for each of these estimates are also available, the safety and waste indicators can be determined.

In Table 1.21, with regard to the line where no auxiliary variables have been used, the same results can be found as in Table 1.16 on page 78. These results suggest that for location parameters, the unweighted intervals only cover 42% of the total full-sample intervals, so that a variance inflation factor ψ of 11.74 is needed in order to obtain an average safety of 95% (type I error of 0.05%). Under the same unweighted conditions, association

Table 1.21: Confidence interval coverage under different weighting configurations and the variance inflation factors ψ necessary to avoid type I error, ESS3-BE ($n = 2927$)

# auxiliary variables used	# target parameters monitored	$\bar{safety}$	$\bar{waste}$	ψ	# auxiliary variables used	# target parameters monitored	$\bar{safety}$	$\bar{waste}$	ψ
0	29	0.42	0.67	11.74	0	42	0.74	0.45	6.28
1	203	0.54	0.57	8.44	1	252	0.80	0.41	4.24
2	609	0.64	0.50	5.77	2	630	0.84	0.37	2.86
3	1015	0.70	0.44	4.20	3	840	0.87	0.35	2.05
4	1015	0.75	0.41	3.42	4	630	0.89	0.33	1.63
5	609	0.77	0.38	3.03	5	252	0.91	0.32	1.40
6	203	0.79	0.36	2.81	6	42	0.92	0.31	1.26
7	29	0.81	0.35	2.70					
location parameters					association parameters				

parameters seem to be covered for about 74% (safety), where the variance inflation factor ψ needs to be 6.29.

Among the location parameters, weighting using only one auxiliary variable improves the average safety indicator from 42% to 54%. It even improves to a level of 81% where the data is weighted by seven auxiliary variables. Inversely, the proportion of redundant interval space (waste) decreases from 67% (unweighted) to 35% (seven auxiliary variables). The need for additional variance inflation is still substantive, as ψ is 2.70.

For association parameters, the effects are similar but less impressive. This is probably due to the fact that location parameters are initially more subject to a lack of safety, leaving greater potential for improvement. At first glance, weighting procedures seem to improve parameter estimates considerably. However, some considerations should be made.

As not all bias can be removed by weighting, some false negative cases can be found. These are situations in which there is initial bias, but the weighting procedures cannot, or can only partially, bridge the gap (*'no correction'* and *'partial correction'* in Figure 1.15). For all the location parameters, about 87% is initially biased ($safety < 0.95$). Out of the total of the biased parameters, 12% is sufficiently corrected ($safety \geq 0.95$) when weighted by only one auxiliary variable and about 54% is sufficiently corrected when deploying seven auxiliary variables, leaving 46% in the bias

zone. Initially, association parameters tend to be less prone to bias (only 62% is initially biased). Of the initially biased association parameters, 15% can be corrected when using only one auxiliary variable to inform the weighting vector, rising to 46% when using all six auxiliary variables, leaving 54% in the bias zone. This suggests that false negative effect might still frequently occur.

On average, about 31% of all the weighted parameters are worse than their unweighted counterparts, with no notable differences between location and association parameters. It is consequently necessary to consider and accept the possibility that weighting procedures can occasionally produce false positive results as well. This is the situation in which a shift from unweighted to weighted parameter estimate can be observed, but it shifts in the wrong direction, further away from the full-sample estimate. Fortunately, as the number of auxiliary variables increases, the percentage of deteriorated parameters seems to decrease, from 39% when only a single variable is used to determine the weight scores, to 27% when seven auxiliary variables are deployed. It should also be noted that parameter improvement is usually more significant than parameter deterioration (see Table 1.22). For all the improved location parameters, the safety increases impressively from 0.29 (unweighted) to 0.73 (fully weighted), whereas only a mild safety decrease can be observed among the deteriorated parameter estimates.

Table 1.22 can be read as follows. Among all the weighted location parameters, 72% improve because of weighting as their weighted point estimate is closer to the full-sample estimate compared with their unweighted counterpart ($\theta_{unw} > \theta_w > \theta_f < \theta_w < \theta_{unw}$). Out of all the weighted parameters, 28% deteriorate ($\theta_w > \theta_{unw} > \theta_f < \theta_{unw} < \theta_w$). Among all the improved parameters, the initial average safety grows from 0.29 (unweighted) to 0.73 (fully weighted), indicating a strong improvement. For the deteriorated parameters, the average safety decrease is clearly not as great as the safety increase among improved parameters.

It should further be noted that in this particular dataset, the improvements are predominantly due to shifts in the point estimates, whereas

Table 1.22: Positive and false positive effects of weighting, ESS3-BE (n = 2927)

	location parameters		
	% of parameters	initial $\bar{safety}$ no auxiliary variables	fully weighted $\bar{safety}$ 7 auxiliary variables
improved parameters	72	0.29	0.73
deteriorated parameters	28	0.72	0.65
	association parameters		
	% of parameters	initial $\bar{safety}$ no auxiliary variables	fully weighted $\bar{safety}$ 6 auxiliary variables
improved parameters	68	0.65	0.85
deteriorated parameters	32	0.91	0.89

the standard errors remain relatively stable or even slightly increase. Unweighted location parameters have a mean relative bias of 0.0510 (bias relative to the standard deviation of the target variables), decreasing to a level of 0.0195 when weighted for all seven auxiliary variables. The relative standard error (standard error relative to the standard deviation of the target variable) slightly increases in this regard, from 0.0183 to 0.0193. For association parameters, this average of relative biases reduces from 0.0267 (unweighted) to 0.0153 (six auxiliary variables), while the relative standard errors increase from 0.0200 to 0.0212. Therefore, the reduction of type I error is predominantly due to shifts of the point estimates, instead of increases of standard errors.

Weighting seems to be a favorable method with which to (partially) adjust for non-response bias. However, the rather frequent occurrence of false positive and false negative cases should warn researchers not to rely too heavily on weighted survey outcomes.

On average, parameter estimates seem to improve after weighting adjustments. However, a particular parameter may be differently changed, depending on the particular set of auxiliary variables used to build the weight vector. In this regard, a weighted parameter cannot to be considered as a final destination or endpoint, but should instead be seen as a realization of the distribution of possible weighted outcomes. This is

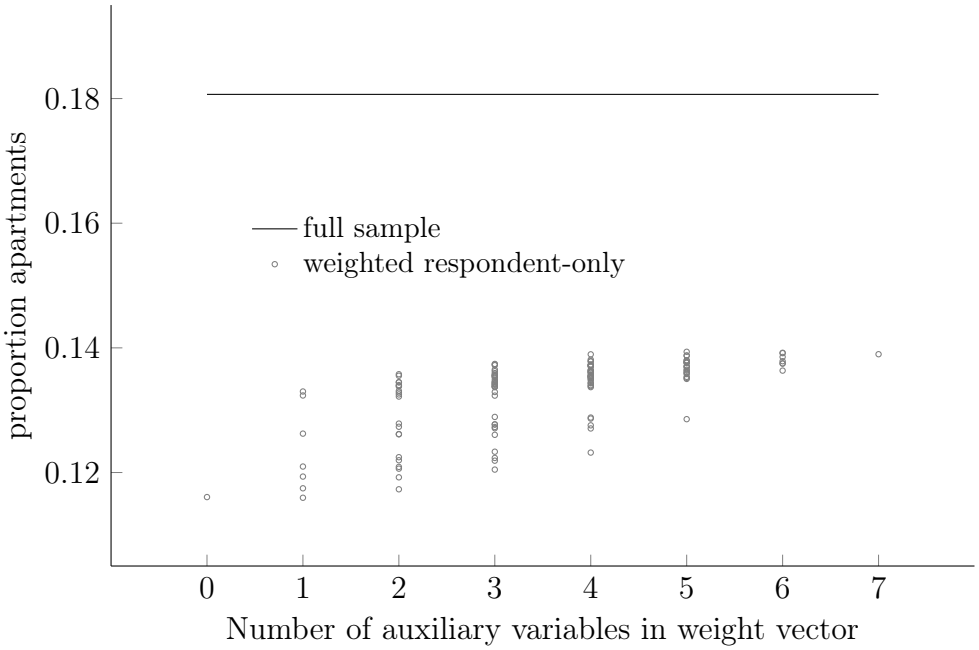


Figure 1.16: Distribution of weighted parameter estimates for apartment dwellers as a function of the number of weight variables, ESS3-BE ($n = 2927$)

particularly the case when the assisting weight model is believed to be underspecified. Such a point of view implies a double source of uncertainty. First, there is the usual standard error that depends on the sampling mechanism (simple random sampling or more complex), and that can usually be obtained theoretically (by analytic or numerical methods). The second source of uncertainty is empirical and depends on the remaining bias in the weighted point estimate. Figure 1.16 illustrates this kind of uncertainty. All 128 alternative estimates for the proportion of apartment dwellers are portrayed as a function of the number of auxiliary variables used to establish the weight vector, ranging from 0 to 7 auxiliary variables. As can be expected, not all weight configurations lead to the same parameter estimate. Therefore, the quality of weighting strongly depends on the set of auxiliary variables.

Usually, only the first source of uncertainty (theoretical standard error) is taken into account when estimating a parameter. However, Table

1.21 indicates that these routines are unsatisfactory, as they are insufficient when trying to cover the full-sample confidence interval. The ψ values in the table are determined by applying equation 1.30 and are read as follows: In order to cover 95% of the total length of all full-sample confidence intervals, the variances of the 29 unweighted location parameters need to be inflated by a factor of 11.74 (see Table 1.21). In order to cover 95% of the total length of all full-sample confidence intervals, the variances of the 203 location parameters, weighted by only auxiliary variables, need to be inflated by a factor of 8.44, and so forth. Even when the parameters are weighted by a set of seven auxiliary variables, the ψ is still considerable. To the extent that these interim target variables are representative of the real target location parameters, a deliberate ψ of about 2.70 needs to be accepted in order to restore the possibility of making a type I error to a level of $\alpha = 0.05$. With regard to association parameters, the problem is less burdensome, although variance inflation is also an adequate method with which to guard against type I errors. It is also clear that the effective sample size should be lower, compared with the nominal 1798 interviewed cases in ESS3-BE. Depending on the type of variable or the number of auxiliary variables, the effective sample size ranges from 213 ($\frac{1798}{8.44}$, location parameters, only one auxiliary variable) to 1427 ($\frac{1798}{1.26}$, association parameters, six auxiliary variables).

As a final comment, the method of iteratively replacing the set of auxiliary variables and subsequently monitoring the remaining variables that take the role of interim target variables may be somewhat problematic. In particular, it is debatable whether these interim target variables are truly representative of the real target variables. All eight auxiliary variables, except for *age* and *gender*, correlate relatively strongly with one another. This means that they each act as strong levers to adjust the biases of the others. On average, the absolute correlation coefficient between any two auxiliary variables is 0.22. The real target variables in the ESS clearly refer to a much wider range of social dimensions than just residence-related information. Indeed, the average absolute correlation between a real target variable (attitudes toward politics, feelings of social trust, media use, etc.)

and one of the auxiliary variables is only 0.07 (measured among respondents only). Unless these real target variables are not as severely biased, compared to the eight auxiliary variables, the real target variable oriented ψ values may be worse than portrayed in Table 1.21. ■

In survey practice, data is often weighted by a set (or multiple sets) of auxiliary variables, hoping to bridge the gap between the respondent-only and the full-sample data. However, since nonresponse can be regarded as a complex configuration that is not simply explained by a set of background variables, the assisting weight model is very likely to be underspecified. Therefore, a particular outcome of a weighted target statistic can be considered as a realization of a distribution of possible and imperfect outcomes, depending on the auxiliary variables that are available. Consequently, awareness may arise that additional uncertainty should be accepted in the presence of survey nonresponse, even when weighting using a set of (relevant) auxiliary variables. Empirical results indicate that weighted respondent-only intervals better cover the full-sample intervals, as more auxiliary variables are taken into account. However, additional variance inflation is in any case still needed.

The analysis also draws attention to the two-dimensionality of the concept of mean squared error. As the mse is expressed as the sum of the squared bias and the variance of the parameter, it may be tempting to minimize the known part of the mse: the variance. It is clear that this only makes sense in the absence of bias, which is usually unknown. Particularly in the presence of bias, small standard errors only aggravate the possibility of a type I error. This should prompt survey researchers to prioritize the attention to bias, rather than standard errors, and to be more careful with regard to statistical inference, even if the data is weighted by powerful auxiliary variables.

However, it is unclear how strong the precautionary measures should be in this regard. The greater the number of relevant auxiliary variables that are available, the fewer the precautions that need to be taken. Empirical arguments also suggest that measures of association are safer than

location parameters. Nevertheless, as nonresponse bias is hard to measure and may strongly depend on the particular survey or survey variable, this kind of uncertainty is very hard to control for or remedy. Unlike sampling error, which produces 'certain uncertainties', systematic survey error such as nonresponse produces uncertain or unknown uncertainties that go beyond probability theory. Consequently, the possibility of making inferences based on survey data may be threatened.

The idea of deliberately increasing standard errors, as reflected by the variance inflation factor ψ, opposes the views of many survey researchers who have developed methods and procedures to intentionally make confidence intervals smaller. Once again, it is not the smallest standard errors that should be aimed for, but instead the correct ones. In this respect, some weighting techniques that propose censoring or trimming of large weight scores to reduce them to a pre-specified maximum (Kalton & Flores-Cervantes, 2003; Bethlehem et al., 2011) should be considered with caution. Large weights are considered as undesirable as they inflate the standard errors. When specific groups are extremely underrepresented, their relative absence is compensated for by increased weight scores. Extreme weight scores (for example $w > 10$) are then deliberately truncated to that maximum. Potter (1993) argues that 'the ultimate goal of weight trimming is to reduce the sampling variance more than enough to compensate for the possible increase in bias and, thereby, reduce the mean square error'. Not only is the one-sided focus on the mean square error questionable in this situation; what is even worse is the deliberate manipulation of survey data in order to provide an artificial overvaluation of the quality of a survey. Similar techniques consist of predominantly selecting auxiliary variables that correlate more with the target variables than with the response outcome (Little & Vartivarian, 2005). Such interventions also lower the standard errors, while the bias is not necessarily reduced, particularly because this latter is usually unknown.

It may appear conservative, but interventions such as weighting techniques that intentionally generate groundlessly smaller confidence intervals, or weight trimming, may lead to inferences being conferred with

greater confidence than the data actually warrants. In this respect, bias due to nonresponse should be acknowledged and accepted, instead of being kicked into the long grass.

1.4 Discussion: How strongly are surveys affected by nonresponse

This chapter strongly supports diversifying the methods that are used to estimate the impact of nonresponse on data quality. As the effect of nonresponse on survey estimates cannot be measured directly, it is necessary to take some detours and approach the problem indirectly. Each alternative path toward the measurement of nonresponse damage has its own pros and cons: some will lead to overly-optimistic results, others rely on too many assumptions or require too much data. It seems therefore essential to use a mixture of approaches.

When a new car rolls off the production line, it is relatively easy to instantly check its quality. Yet after a while, the driver might find some hidden defects. The 'proof of the pudding' is an advantageous property of many industrial products. However, in survey research, the quality of the results is much harder to assess, as there is usually no external reference with which target estimates can be compared. This is both comforting and threatening. From the viewpoint of the data supplier, it is comforting that the quality of survey data is hard to evaluate. Estimates may be wrong, but nobody will notice. There are, however, some historical examples of how survey estimates can be contradicted by reality. The U.S. presidential elections of 1936 and 1948 were predicted to be won by a landslide by respectively Landon and Dewey, whereas their opponents Roosevelt and Truman were actually elected as presidents. *'Dewey defeats Truman'*, was the headline in *The Chicago Tribune* on November 3, 1948, the day after the American voters had elected Truman as their new president. The fatal flaw was reliance on public opinion polls. In this regard, Squire (1988) commented on the Landon-Roosevelt election of 1936 that

in particular a low nonresponse rate and the difference between respondents and nonrespondents contributed to the failure to predict the correct winner. Apparently, the *Literary Digest* poll of 1936 counted many more nonrespondents among Roosevelt voters than Landon voters.

Nevertheless, in most cases the true parameters remain concealed. Unverifiable outcomes of a survey may be a blessing to data suppliers, allowing them to enter the market or public media, communicating their results at relatively low cost. However, from the perspective of data users, the inability to compare survey estimates with true parameters is threatening. In any of the surveys examined in this chapter, it seems relatively easy to find variables that substantively and significantly differ between respondents and nonrespondents. Other ways of estimating damage due to nonresponse also urge that conscious and careful efforts should be made to deal with this issue.

The findings related to the European Social Survey and presented in this chapter oppose the views of Stoop et al. (2010): Looking back over four successive rounds of the ESS, they concluded that (p. 302) '[...] using all the information that is available and based on different approaches - we have no evidence of serious nonresponse bias in the ESS' or (p. 294) 'Somewhat reassuringly, there are as yet virtually no indications that nonresponse bias is extensive in the ESS data, [...]'.

Not only is nonresponse bias as such a threat to survey research, but it is also very hard to measure and to provide adequate protection against. Indeed, it is very difficult to determine how strong the protective measures against nonresponse bias should be, particularly because the many different methods used to assess nonresponse bias or error may lead to divergent conclusions. Therefore, unlike sampling error, nonresponse produces uncertainties that are almost impossible to control for. As a result, the capacity of a survey to enable inferences to be drawn is strongly undermined. Simply building on the principles of probability theory and ignoring systematic error such as nonresponse may be considered as naive (Särndal, 2010). This should warn survey researchers and urge them to assess critically the survey analyses on which they build their scientific claims. Most

software packages provide survey estimates, together with standard errors and p-values, that are only valid under very strong statistical assumptions. Correct inferences can only be made if all population members have an equal probability of being included in the sample (MCAR) or if the (different) probabilities of inclusion are known (MAR). Given that MNAR is likely to be the most realistic scenario, the validity of p-values read from software output is highly questionable.

The question is how survey researchers should act, and how to direct their attitudes toward the potential of a survey being affected by nonresponse. In order to provide some provocative reflections on this subject, consider the following list of attitudes or ethical positions toward surveys, ordered from a position of simply ignoring the problem of nonresponse, to one of extreme skepticism.

1. *Ignore nonresponse*

   In the era when response rates were high, nonresponse was simply not an important issue. Following the decline in response rates, researchers started worrying about the validity of surveys being affected by nonresponse. It is clear that simply ignoring nonresponse or assuming it has no noteworthy impact can be considered as too naive. However, many polls that appear in newspapers or other media do not seem to take into account any survey error at all.

2. *Transparency with regard to response rates*

   Many international survey authorities are highly committed to survey transparency. Organizations such as the American Association for Public Opinion Research (AAPOR), the Council of American Survey Research Organizations (CASRO), the European Society for Opinion and Marketing Research (ESOMAR), the International Statistical Institute (ISI), and the American Statistical Association (ASA) mention the importance of providing the necessary information about the sample design or the methods of data collection. AAPOR (2011) has even developed a set of definitions and standards for the calculation of response rates. At the same time, many survey researchers

have become aware that the response rate alone is a very poor indicator of survey quality, particularly when response rates are low and leave more potential for bias to occur.

3. *Provide a nonresponse analysis*

   Comparisons between respondents and nonrespondents with regard to some auxiliary variables often suggest a substantial danger that survey statistics are biased. Other methods such as the comparison between weighted and unweighted estimates also seem to indicate that nonresponse unfavourably affects the quality of surveys.

4. *Take weak protective measures against nonresponse (weighting adjustments)*

   Many survey researchers have resorted to corrective measures such as post-survey weighting adjustments. These are believed to (partially) improve the quality of survey statistics. However, their effectiveness is hard to prove because the full-sample target statistics are unknown by definition.

5. *Take strong protective measures against nonresponse (deliberately enlarge the standard errors)*

   In order to guard against type I errors, margins of uncertainty may be deliberately enlarged (by variance inflation factor ψ). However, it is unclear how far such measures should go in order to completely safeguard the survey statistics. On the other hand, overprotection or variance inflation may also lead to the dilution of survey research as a contributor to our understanding of the economy or society as a whole.

6. *Abandon any inferential claim based on survey data (use surveys only for exploratory purposes)*

   Survey research seems to be forced to leave the paradigm that starts from the assets of probability sampling. In this regard, Groves (2006)

points out that 'Because of falling response rates, legitimate questions are arising anew about the relative advantages of probability sample surveys. Probability sampling offers measurable sampling errors and unbiased estimates when 100% response rates are obtained. There is no such guarantee with low response rates. Thus, within the probability sampling paradigm, high response rates are valued.' Similarly, Särndal (2010) suggests that 'The probability sampling (scientific sampling) tradition is a reflection of an idyllic past. We are now in 2010, not 1950. On what grounds is it still defendable, in our time?'

7. *Reject any kind of survey research that is subject to nonresponse*

   This last option on the scale of attitudes toward nonresponse is probably a step too far. Surveys still deliver data that is supposed to be more informative than an uninformed guess about the poverty rate, the average health, political trust, etc. Survey data at least provides a reflection of an underlying, but hard to measure, reality. However, it is hard to evaluate what the added value of (biased) survey data is, compared with having no data at all.

Currently, most survey researchers seem to endorse the idea that nonresponse poses a threat to the reliability of survey statistics. The dominant practice to deal with the problem of nonresponse seems to include the observation of a large-sized sample, which is weighted with regard to a set of available auxiliary variables, accompanied by a set of assumptions, and of which the response rate is maximized. These assumptions usually refer to the expectation that these adjustments eventually control and remedy nonresponse bias, so that inferences can still be valid. Making such statistical assumptions is a widespread practice. However, because assumptions always imply an uncertainty, additional provisions should be made, accommodating the risks and perils of such assumptions. This means that the survey community should move toward the more skeptical end of the nonresponse attitude scale as presented above. The former AAPOR president Martin (2004) discussed the problem of nonresponse in her presidential

address and noted that 'we need to acknowledge and address the gaps in our knowledge, and explain the limitations of the data. I don't think we are doing a good job of that. If we can't or won't do it, then we ought to abandon claims that our work is scientific'.

Not only should survey researchers lower their expectations from survey data, but survey response is also an outcome of an underlying production process. The next chapter therefore will examine ongoing and desired practices in survey fieldwork in order to reveal some imperfections in the production process of a survey sample.

Chapter 2

Do Current Fieldwork Procedures Actually Reduce Nonresponse Bias?

2.1 Introduction: from Total Survey Error to Total Quality Management

There is a general belief that process quality is the key to final product quality. This belief is perhaps strongly institutionalized in many manufacturing disciplines; however, survey researchers have only recently acknowledged the usefulness of process data with regard to improving survey quality. Process quality is different to output quality. In the context of nonresponse, in the output approach one typically assesses the differences between respondents and nonrespondents (after refusal conversions, etc.) or the differences between respondents and the total population with regard to one or more variables. By comparison, the focal point of the process approach is the construction or realization of a sample, including the preparation of the sample frame, the sampling procedure, and the fieldwork. The process approach focuses on the selection of sample units, the timing and sequence of contacts, the efforts made to make people participate, etc. Moreover, the process precedes the output and therefore the

process quality determines the output quality, or, 'If the process of gathering data is good, there is no need to worry about the quality of the final product' (Lyberg & Biemer, 2008). In addition, whereas the assessment of output is usually respondent-oriented, the process approach focuses on a broader range of survey agents, such as the interviewers or the fieldwork management, as these have an important impact on the selection and treatment of sample units. Further, the assessment of process quality requires more data and documentation about the selection, timing, contact attempts, etc. That is why the availability of good paradata is critical for fieldwork monitoring and fieldwork improvement.

Process monitoring is relatively new in survey practice. The term paradata (and the ambition to use it) first appeared in 1998 (Couper). Dippo (1997) recognizes that the integration of continuous quality improvement into statistical services requires a broader approach than in manufacturing. The processes that need to be addressed are typically not physical by nature, but instead human actions, decisions, and the path those decisions take. Aitken, Hörngren, Jones, Lewis, and Zilhão (2004) also argue that literature on quality improvement and process monitoring is relatively scarce in survey research, notwithstanding the unmistakable benefits suggested by their examples. A general theoretical framework to inspire fieldwork monitoring with respect to representativeness or nonresponse bias is adopted from Morganstein and Marker (1997) and is presented in Figure 2.1. Their framework builds on the achievements of the Total Quality Management (TQM) paradigm.

The first step in Morganstein and Marker's framework is to identify the critical quality characteristics. Applied to the problem of nonresponse, these may include all the different nonresponse measures as presented in Chapter 1. It should be noted that the critical characteristics not only address the final response, but also specific parts of the process that involve the (non)selection of sample units (e.g. noncontacts, refusals, and other nonresponses).

The next step in improving statistical products is to develop a flow chart, yielding a better understanding of the related (sub)processes. Figure

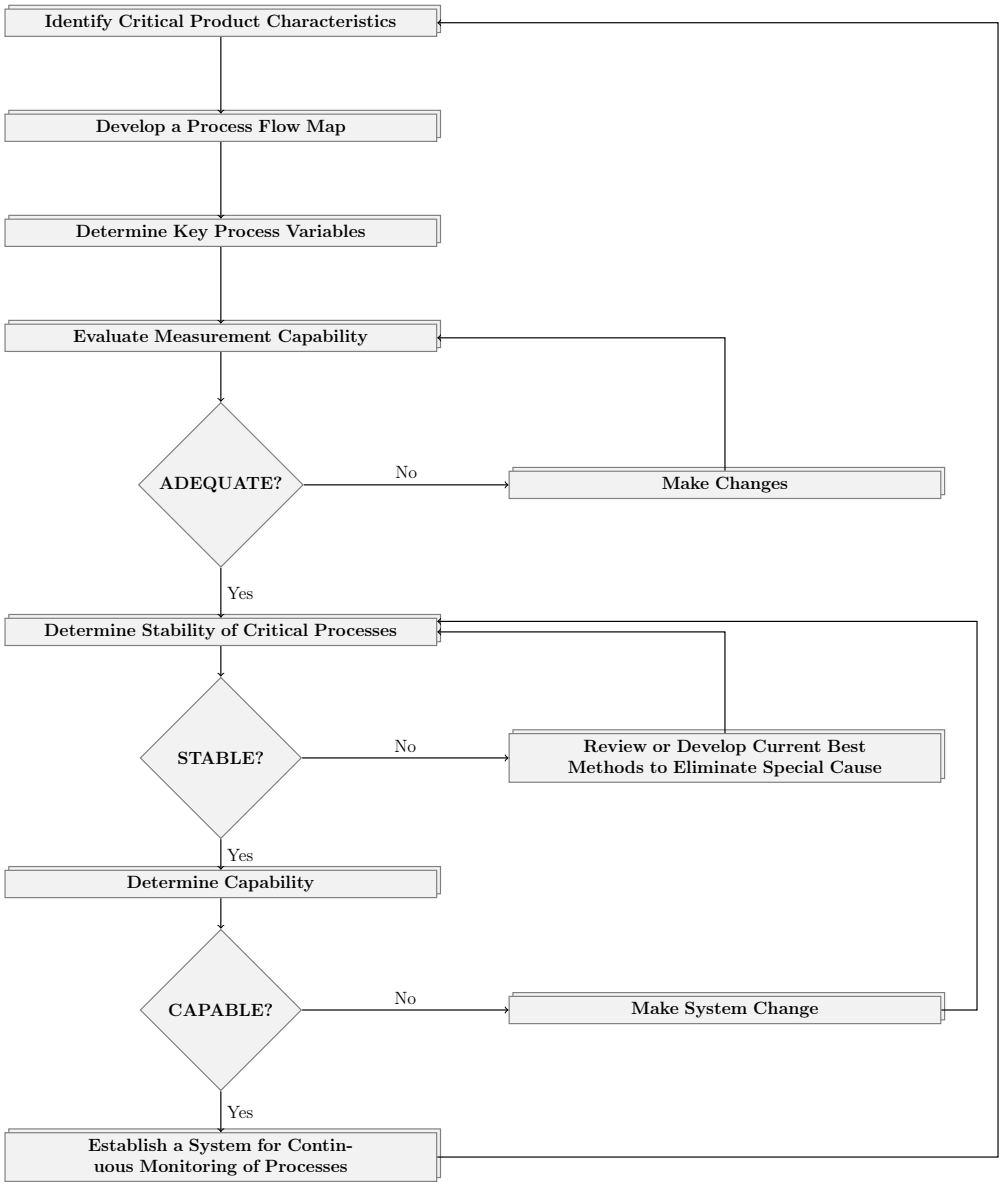
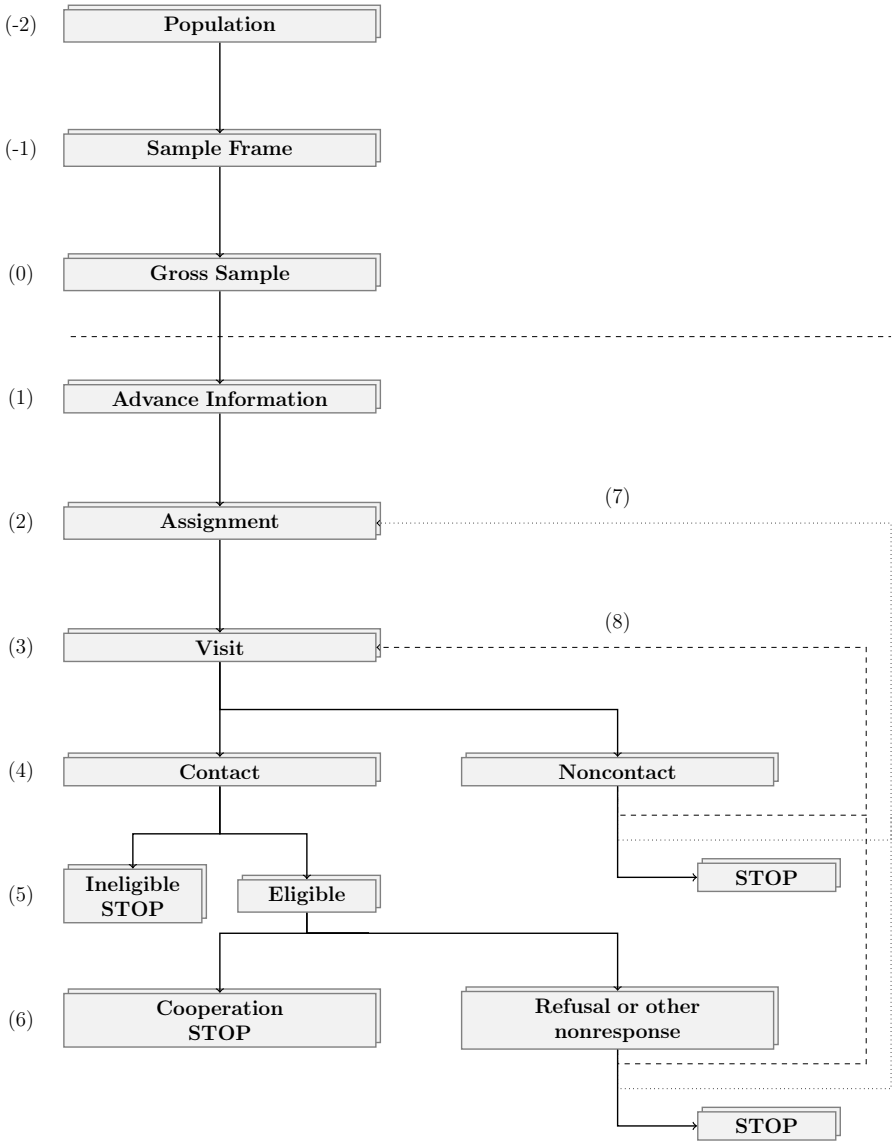


Figure 2.1: A Plan for continuous quality improvement
Source: Morganstein and Marker, 1997

2.2 shows an example of the process flow related to the sample construction activities for the third round of the Belgian ESS. Three components should be in the flow chart. First, the sequence of processes is delineated, indicating all the decisive points. Second, the owners (agents or stakeholders) are identified. Interviewers, (non)respondents, and fieldwork management take the most prominent positions in the flow. Third, the key process variables are listed. These factors can vary with every repetition of the process and affect critical product characteristics. The key process variables represent the quality of the sub-processes, such as the assignment of interviewers, waiting periods between contact attempts, etc. Key process variables may also be termed 'treatment variables', as they can be considered controllable variables and serve as an input to improve the survey process and consequently also the survey quality.

The evaluation of measurement capability entails the quality of the process data and refers to a wide range of information about the process. As argued by Morganstein and Marker (1997), the measurement error with respect to process data is often one of the least appreciated aspects of quality improvement procedures: 'Researchers often select a process because it is easy to measure, rather than choosing a more important but harder-to-measure process'. Such a claim can be endorsed, as many fieldwork organizations and interviewers do not have a long tradition of documenting or recording their contact activities, nor of organizing or archiving such data.

The next step involves the actual monitoring of fieldwork. Here, the sources of process and output variations are identified. Nonresponse measurements such as the R-indicator or the maximal absolute bias may show sudden jumps, or a gradual deterioration or improvement of nonresponse error. This may be accomplished by using control charts and other statistical tools or methodologies. The empirical parts throughout this chapter will deal with some of these tools.



	Process Owner
(-1)	government, statistical agency, private company, ...
(0)	government, statistical agency, private company, fieldwork management, ...
(1)	fieldwork management, interviewer
(2)	fieldwork management
(3)	interviewer
(4)	interviewer, (non)respondent, family member, neighbour, ...
(5)	interviewer, (non)respondent, family member, neighbour, ...
(6)	interviewer, (non)respondent, family member, neighbour, ...
(7)	fieldwork management
(8)	fieldwork management, interviewer

Figure 2.2: Flow diagram of the contact process - ESS3-BE

2.2 Setting the fieldwork objectives

This section will explore how fieldwork objectives may affect sample representativeness or nonresponse bias. First, a brief overview will be provided about the current attitudes toward fieldwork objectives. Then, a simulation study will be set up, assessing a variety of fieldwork objectives. The simulation study discusses earlier work (Beullens & Loosveldt, 2012) in more detail.

2.2.1 Attitudes toward fieldwork objectives

In most businesses, targets are usually set aimed at maximizing sales or manufacturing, and minimizing costs. Commercial companies are not alone in following this practice, government agencies and social profit organizations may also abide by these rules of efficiency: saving as many people from poverty as possible at minimal cost, vaccinating as many babies as possible against poliomyelitis at maximal efficiency, etc. This objective has also been applied to surveys: minimal costs or effort for maximal response.

However, recent evolutions in the survey climate now urge researchers to reconsider their fieldwork objectives. As response rates have declined, nonresponse bias has become a more important threat. Therefore, survey researchers and fieldwork managers may be more inclined to pursue representative samples, rather than merely large samples. Indeed, it seems that the 'highest response at lowest cost' objective is steadily changing into 'high but equal inclusion probabilities at lowest possible costs'.

Nevertheless, the objective of maximizing response rates still holds a dominant position. Many survey agencies have the tendency to measure and communicate the quality of their respondent samples in terms of response rates. In addition, AAPOR mentions in its best practices to 'maximize cooperation or response rates [...]' (`www.aapor.org`). Low nonresponse rates are believed to reduce the potential for nonresponse bias. Indeed, as bias can be determined by the response rate multiplied by the contrast between respondents and nonrespondents with regard to the mean

of a variable, response rate maximization seems to be a reasonable strategy to reduce nonresponse bias. High response rates are also easy to calculate, and since anyone is expected to interpret their meaning and relevance, they are easy to communicate. Moreover, a response rate is a clear strategic objective to focus on during fieldwork activities.

Response rates hold a prominent position throughout survey literature as well as in survey practice. The main reason for the attention to response rates is the conviction that high response rates protect surveys from nonresponse bias: the smaller the proportion of nonrespondents, the less damage they can cause. Recent handbooks on survey research pay much attention to the contrast between respondents and nonrespondents as an important part of the quality assessment of a survey. In social research handbooks written before around 2000, authors have suggested response rates above which bias is unlikely to occur. In this regard, Babbie's 'The practice of social research' is a good illustration of how attention has shifted from a response rate centered orientation to a bias oriented perspective. One of the first editions of the handbook (1975) mentions that '[...] a response rate of at least 50% is *adequate* for analysis and reporting. A response rate of at least 60% is *good*. And a response rate of 70% or more is *very good*.' The seventh edition of the handbook (1995) repeated the same recommendation, while the twelfth edition (2009) states that 'there is no absolutely acceptable level of response [...], except for 100%'. Whereas the first editions of the handbook proposed acceptable levels of response, they also warned the reader 'to bear in mind, however, that these are only rough guides, they have no statistical basis, and a demonstrated lack of response bias is far more important than a high response rate'. Bailey (1987) proposes a minimal acceptable response rate of 75%. For Fowler (1985, 1993, 2002, 2009), there is no agreed-upon standard for a minimal acceptable response rate, but it is nevertheless observed that important agencies ask for high response. The Office of Management and Budget of the American federal government (OMB, 2006) requests that survey procedures are generally designed to yield an 80% response rate. The European Social Survey requires response rates of 70% or more from all participating countries.

There seems to be an increasing awareness amongst survey researchers that a single-minded focus on response rates alone is not advisable. As Groves (2006) finds, high response rates do not necessarily imply low bias in the eventual survey estimates, more refined strategies may more appropriate. Instead of the blind pursuit of high response rates, an informed pursuit of high response rates is wiser. This is why many survey researchers have started to collect auxiliary variables for both respondents and nonrespondents, or other relevant paradata in order to guide the fieldwork activities, hopefully resulting in samples that are more representative.

Krosnick (1999) also claims that representativeness does not necessarily improves by simply increasing the response rate. Langer (2003) states that '[...] comparisons of response rates are exceedingly difficult, and their value unclear; [...] a higher response rate is not automatically indicative of better data'. Peytchev et al. (2010) even hypothesize that response rate maximization and bias reduction are strategically incompatible objectives, particularly in face-to-face surveys. Interviewers, often paid per completed interview, are usually evaluated on their individual response rates, probably prioritizing the cases they estimate to be more responsive and leaving the low propensity cases unattended. Such prioritization regimes may only enlarge the gap between high and low propensity cases, endangering the representativeness of the survey and facilitating survey estimates being affected by nonresponse bias. This consideration makes it clear that the survey industry does not necessarily need to follow the strategic principles of any other line of business. Sales managers are usually inspired by the idea of selling as many products at the lowest possible cost, thereby encouraging their staff to first concentrate on the 'low hanging fruit'. Hence, many businesses follow the line of least resistance. However, survey research pursuing representative samples should instead instruct their interviewer force in such a way that any member of the population or the gross sample has an equal probability of being included in the respondent set. This is clearly a much more complicated objective than following the line of least resistance that predominantly targets the low hanging fruit in order to maximize the response rate at the lowest possible cost.

2.2.2 A simulation study

This section will explore how fieldwork objectives may affect the quality of the obtained respondent set. Setting an objective affects the process of data collection, which in turn has an impact on the output quality.

Suppose a gross sample of $n = 3000$ individuals needs to be fielded and the survey budget allows for $E = 10,000$ efforts or contact attempts. The average success probability for one contact attempt in the sample is $\bar{\rho}_1 = 0.25$ and the variance of these probabilities or propensities about $\bar{\rho}_1$ is $\sigma^2_{\rho_1} = 0.03$. These parameters are chosen as they closely resemble the situation in the ESS3-BE. Indeed, this survey also started from about 3000 cases, out of which about 25% were interviewed after the first contact attempt. The entire fieldwork comprised about 10,000 contact attempts. The variance of the response propensities is of course much harder to determine. The value of $\sigma^2_{\rho_1} = 0.03$ is chosen because it is higher than the variance when it is estimated only based on auxiliary variables (see Table 1.6 on page 38) where the propensity variance is 0.01, but still somewhat lower than the 0.0441 found for the FHS based on the progress of the response rates (see the real data example on page 97). Figure 2.3 illustrates what the distribution of propensities about $\bar{\rho}_1$ may look like.

The shape of the distribution of the propensities is assumed to have an underlying normal distribution: the propensities themselves are not normally distributed, but their logits are. So, given that $x\beta_i = ln(\frac{\rho_i}{1-\rho_i})$, it holds that $x\beta \sim N(\mu, \sigma^2)$.

Under different optimization criteria, the 10,000 contact attempts or efforts (E) need to be distributed over the 3000 individuals. Each individual i will be assigned k_i contact attempts, such that $\rho_{final,i} = 1-(1-\rho_{1,i})^{k_i}$ and $\sum_{i=1}^{n} k_i = E = 10,000$. It should be noted that the connection between the 'one-shot'-propensity ρ_1 and the final propensity ρ_{final} is accommodated by the geometric distribution function. Such a function assumes no memory between the contact attempts, which is probably a simplification of reality. It is expected that individuals in the sample may recall having been contacted earlier in the fieldwork, or that interviewers may learn from

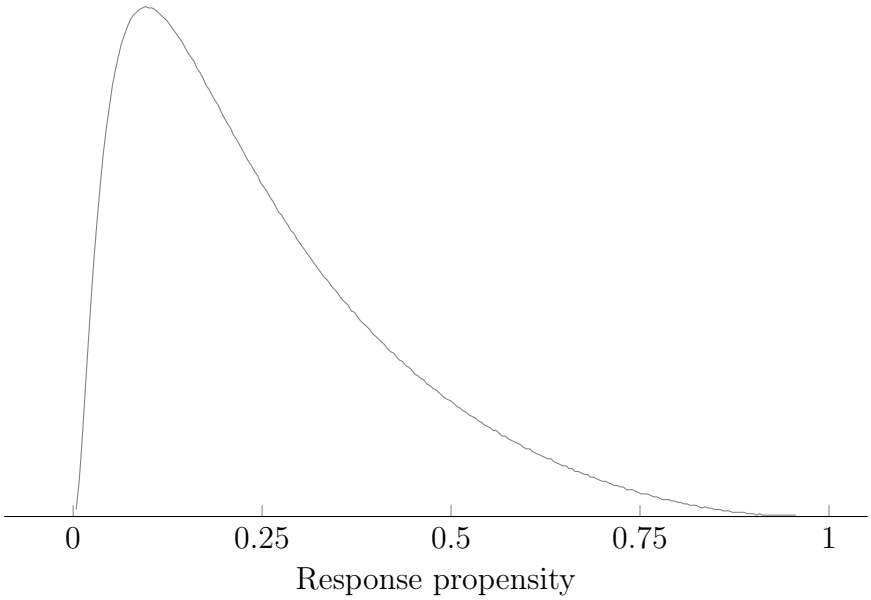


Figure 2.3: Distribution of response propensities where $\bar{\rho}_1 = 0.25$, $\sigma^2_\rho = 0.03$ and $ln(\frac{\rho_i}{1-\rho_i}) \sim N(\mu, \sigma^2)$

earlier contact attempts and adapt their subsequent call attempts accordingly. In particular, consideration of the interviewers may suggest that the success probability may slightly increase with each renewed contact attempt. However, for reasons of simplicity, such memory effects will not be dealt with.

Under the random response model, the first chapter of this dissertation provides some quality measures that can be optimized for the sample as presented above. For each of the optimization criteria, a separate effort assignment strategy will be applied, resulting in a vector of final response propensities ρ_{final} based on which the quality measures can be obtained: R-indicator, maximal absolute bias, and maximal absolute contrast. Five effort assignment strategies will be explored.

1. *Response rate maximization.* (Strategy A1) It is expected that this strategy will prioritize the high propensity cases or follow the line of least resistance. The possibility can not be ruled out, despite the high response rate, that the contrast between respondents and

nonrespondents will be forced upward, probably leading to a higher risk of nonresponse bias. This strategy requires that a vector K needs to be found, containing all individually assigned contact attempts k_i for all 3000 sample members, so that $\sum_{i=1}^{n} 1-(1-\rho_{1,i})^{k_i}$ is maximized, where $k_i \geq 1$ and $\sum_{i=1}^{n} k_i = E = 10,000$.

2. *R-indicator maximization.* (Strategy B) The maximization of the R-indicator coincides with the minimization of the variance of the final propensities. Such a strategy is expected to prioritize the low propensity cases. As a result, much of the fieldwork efforts will be devoted to these low propensity cases leaving less attempts for the more high propensity cases, so that the response rate will be somewhat lower compared with following the line of least resistance in the response rate maximization strategy. As the contrast between respondents and nonrespondents is expected to be rather small, the risk of nonresponse bias will also be small. This strategy requires a vector K that minimizes $\frac{\sum_{i=1}^{n}\left(1-(1-\rho_{1,i})^{k_i}\right)^2}{n} - \left(\frac{\sum_{i=1}^{n} 1-(1-\rho_{1,i})^{k_i}}{n}\right)^2$, where $k_i \geq 1$ and $\sum_{i=1}^{n} k_i = E = 10,000$. It should be noted that this optimization makes of use of the property that $Var(X) = E(X^2) - E(X)^2$.

3. *Bias minimization.* (Strategy C) This strategy does not focus exclusively on the response rate or the propensity variance alone, but instead on the quotient $\sigma_\rho/\bar{\rho}$, expressing the maximal possible bias for the mean of a standardized variable. A vector K is required that minimizes
$$\frac{\sqrt{\frac{\sum_{i=1}^{n}\left(1-(1-\rho_{1,i})^{k_i}\right)^2}{n} - \left(\frac{\sum_{i=1}^{n} 1-(1-\rho_{1,i})^{k_i}}{n}\right)^2}}{\frac{1}{n}\sum_{i=1}^{n} 1-(1-\rho_{1,i})^{k_i}},$$
provided that $k_i \geq 1$ and $\sum_{i=1}^{n} k_i = E = 10,000$.

4. *Random allocation strategy.* (Strategy D1) This is of course not a strategy in the strict sense of the word. Nevertheless, this assignment regime is reasonably feasible as does not use any information whatsoever concerning the responsiveness of the individuals in the

sample: all remaining nonrespondents need to have an equal probability to a renewed contact attempt.

Without any restrictions on the total number of attempts E, the expected number of contact attempts needed to convert an individual is $1/\rho_{1,i}$. However, since the total number of effort E is restricted and all individuals should have the same re-issue probability, the expected number of contact attempts needs to be written as $k_i = \frac{1}{1-\lambda(1-\rho_{1,i})}$, where λ is a constant controlling the overall re-issue probability so that $\sum_{i=1}^{n} k_i = \sum_{i=1}^{n} \frac{1}{1-\lambda(1-\rho_{1,i})} = E$, provided that $k_i \geq 1$.

In order to illustrate the relevance of the parameter λ, consider an individual with a response propensity of $\rho_1 = 0.75$ (see Table 2.1). Without any restrictions on re-selection ($\lambda = 1$), the probability of a renewed attempt after the first attempt is $1-0.75 = 0.25$. Success at the second attempt equals $0.25 \times 0.75 = 0.1875$, leaving a probability of $0.25 - 0.1875 = 0.0625$ that the individual needs to be re-issued at the third attempt, etc. The sum of all re-selection probabilities equals $1/\rho_1$ or $1/0.75 = 1.3333$. This is exactly the expected number of contact attempts $E(k)$.

Now, in the case where the re-selection is restricted (e.g. $\lambda = 0.8$), P(re-select) = P(failure) $\times\lambda$, implying that the invested effort at the next contact attempt is also decreased. Eventually, the sum of all efforts is $\frac{1}{1-\lambda(1-\rho_1)}$, in this case $\frac{1}{1-0.8(1-0.75)} = 1.25$.

Table 2.1: How to determine the re-selection probability in a geometric distribution, illustration

attempt	effort	P(success)	P(failure)	P(re-select)
1	1.0000	0.7500	0.2500	0.2500
2	0.2500	0.1875	0.0625	0.0625
3	0.0625	0.0469	0.0156	0.0156
4	0.0156	0.0117	0.0039	0.0039
5	0.0039	0.0029	0.0010	0.0010
6	0.0010	0.0007	0.0002	0.0002
7	0.0002	0.0002	0.0001	0.0001
8	0.0001	0.0000	0.0000	0.0000
9	0.0000	0.0000	0.0000	0.0000
10	0.0000	0.0000	0.0000	0.0000
$\sum$	1.3333			

$\lambda = 1$, $\rho_1 = 0.75$

attempt	effort	p(success)	p(failure)	p(re-select)
1	1.0000	0.7500	0.2500	0.2000
2	0.2000	0.1500	0.0500	0.0400
3	0.0400	0.0300	0.0100	0.0080
4	0.0080	0.0060	0.0020	0.0016
5	0.0016	0.0012	0.0004	0.0003
6	0.0003	0.0002	0.0001	0.0001
7	0.0001	0.0000	0.0000	0.0000
8	0.0000	0.0000	0.0000	0.0000
9	0.0000	0.0000	0.0000	0.0000
10	0.0000	0.0000	0.0000	0.0000
$\sum$	1.2500			

$\lambda = 0.8$, $\rho_1 = 0.75$

Thus, when applying a constant re-selection probability λ to all sample members, the fieldwork is not driven by any knowledge about the responsiveness of the individual sample members. For this optimization problem, λ is the only parameter that needs to be chosen in $\sum_{i=1}^{n} k_i = \sum_{i=1}^{n} \frac{1}{1-\lambda(1-\rho_{1,i})} = E$. In other words, λ needs to be chosen such that the contact attempts are kept within the total number of contact attempts E, provided that all units have an equal re-selection probability λ after an unsuccessful contact attempt. This equal re-selection probability is a crucial element in this strategy as it makes sure that the fieldwork decisions are blind or uninformed.

5. *One-shot-only strategy.* (Strategy E1) For this strategy, the gross sample size is equal to the total number of contact attempts or $n = E$. This means that any nonrespondent is replaced by a substitute. This also means that the final response propensities $\rho_{final,i}$ are identical to the one-shot propensities $\rho_{1,i}$. This strategy is particularly relevant as it provides the basic principle of quota sampling, notably the substitution of unsuccessful cases. Later on, this strategy will be restricted, satisfying the requirements for a predefined quota to be obtained with regard to auxiliary variables such as age and gender.

Each of these strategies will be applied to the sample as presented above and the quality of the resulting samples are assessed by the final response rate $\bar{\rho}_{final}$, the R-indicator, the maximal possible bias $\sigma_{\rho_{final}}/\bar{\rho_{final}}$ and the maximal possible contrast $\sigma_{\rho_{final}}/(\bar{\rho}_{final}(1-\bar{\rho}_{final}))$.

The vector containing the number of expected contact attempts for all individuals is obtained through numerical optimization, using the `OPTMODEL` procedure in SAS. This means that no formal solution will be presented. An important restriction regarding all the strategies is that every individual has to be attempted to be contacted at least once ($k_i \geq 1$).

Numerical optimizations can sometimes provide irregular or invalid results as they risk being locked up in a local instead of the global optimum. However, repeating each optimization problem with different starting values for k_i and obtaining identical solutions, provides a strong guarantee that the global optimum has been found, instead of a suboptimal local solution.

Figure 2.4 shows how the contact attempts (first panel) are distributed according to the one-shot response propensities $\rho_{1,i}$. For the one-shot-only strategy, a straight horizontal at $k = 1$ may be added. However, the way the four other strategies distribute their efforts is more important. The bias minimization strategy is very similar (or even almost identical) to the strategy that maximizes the representativeness of the sample. Both strategies seem to expend most of the available contact attempts on the low propensity cases (the high hanging fruit), while the response rate maximization strategies seems to direct more effort to the high propensity cases (the low hanging fruit). In fact, the correlation between the probability of having a renewed contact attempt and the initial response propensity ρ_i under the response rate maximization regime is 0.88, suggesting that this strategy follows the line of least resistance. This correlation under the bias minimization strategy is -0.97 and -0.96 for the strategy maximizing representativeness. This means that response rate maximization is strategically incompatible with, or even opposite to, strategies that seek to guard against nonresponse damage. The blind assignment strategy seems to be located in between the two opposite fieldwork strategies. This strat-

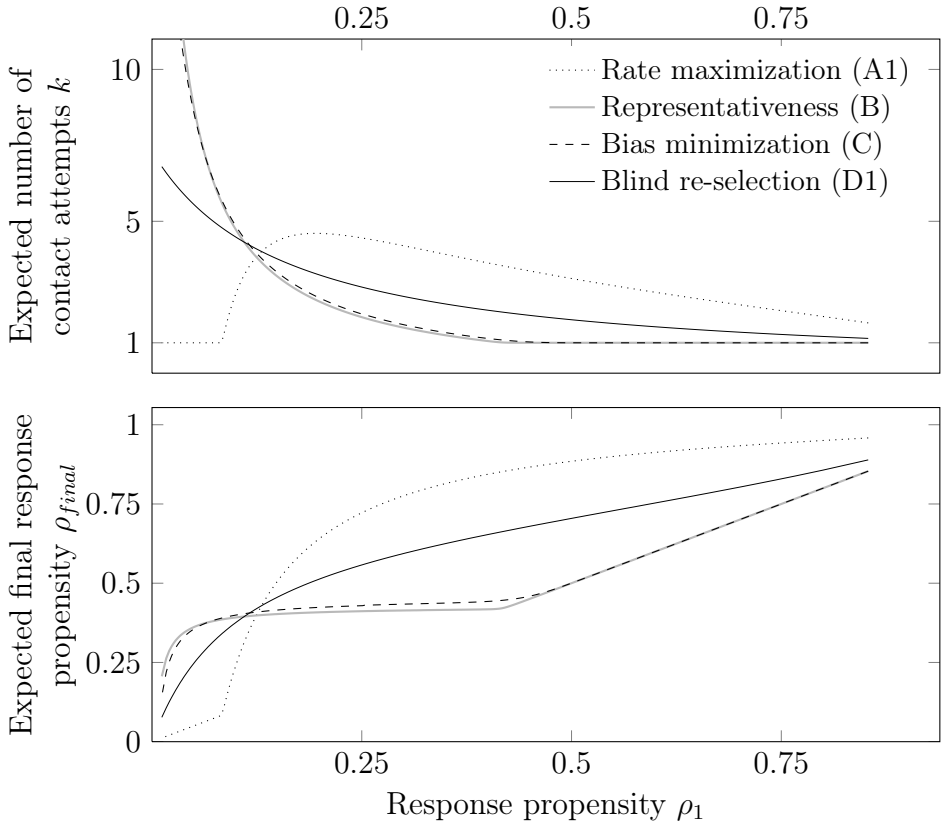


Figure 2.4: Distribution of contact attempts and final propensities over response propensities according to different fieldwork strategies, simulations

Table 2.2: Obtained sample quality indicators for five different fieldwork strategies, simulations

		Response rate	R-indicator	Max. bias	Max. contrast
Basic strategies					
A1	Response rate maximization	0.5647	0.4068	0.5253	1.2067
B	R-indicator maximization	0.4237	0.8457	0.1821	0.3160
C	Bias minimization	0.4335	0.8435	0.1806	0.3187
D1	Blind assignment strategy	0.5094	0.6828	0.3113	0.6346
E1	One-shot only ($n = E = 10,000$)	0.2500	0.6536	0.6928	0.9238
Strategies restricted by stratification quota					
A2	Response rate maximization	0.5138	0.4647	0.5209	1.0714
D2	Blind strategy (restricted)	0.4667	0.7091	0.3117	0.5844
E2	Quota strategy ($n = E = 10,000$)	0.2097	0.7432	0.6124	0.7749

egy gives all sample units, after an unsuccessful contact attempt, an equal re-issue probability (λ) of about 86%.

The second panel in Figure 2.4 shows how the distribution of contact attempts results in the final response propensities. Clearly, the response rate maximization strategy makes the low hanging fruit even more responsive, leaving the high hanging fruit rather underrepresented. As a result, the spread of final response propensities is therefore highest in this strategic scenario. The two opposite strategies (maximization of representativeness and minimization of bias) both seem to reduce the final response propensities among the low hanging fruit and improve the propensities among the high hanging fruit, leading to a relatively low spread of final propensities.

Table 2.2 shows the results of the fieldwork strategies with regard to the quality indicators of their respective obtained respondent samples. As can be seen, the first three strategies seem to have succeeded in the optimization of their objectives: the first strategy indeed has the highest response rate (A1), the second has the highest R-indicator (B) and the third has the lowest bias (C). The most striking finding from this table is that the strategy that maximizes the response rate is apparently confronted with a high level of bias. Only the one-shot strategy (E1) seems to perform even worse. Guided by the findings in Table 2.2, the strategy of response rate maximization, although designed to protect a sample from bias, cannot be considered as a useful way in which to provide a good quality sample. The

essential problem with response rate maximization lies in the prioritization of the 'low hanging fruit'. In this regard, no prioritization at all (as in the blind fieldwork operation; D1) seems to provide a better respondent set.

The fieldwork simulations presented so far assume that the propensities are perfectly known. However, in real fieldwork operations, only some information with regard to auxiliary variables is known, for example when response rates are different between some identified subgroups in the sample. Trying to equalize the response rates in this regard is a recommended objective (Groves, 2006). Nevertheless, if only the between-group differences are known, there may still be a considerable amount of unknown within-group differences that cannot be used strategically. Therefore, a simulation will be set up that starts from a blind assignment regime, which is restricted by the objective of attaining equal response rates for some identified subgroups in the sample. In this simulation strategy, the total variance of response propensities is still 0.03, of which one third (0.01) can be explained by the use of one auxiliary variable counting five different categories (see Table 2.3). It should be noted that the five groups presented in Table 2.3 have the same size in the full sample, although they differ with regard to their average response propensities. Group one has the highest average response propensity, group five has the lowest average response propensity. The simulation will try to assign the available contact attempts $E = 10,000$ in such a way that the eventual response rate among the five groups is equal. In other words, the between group propensity variance will be minimized. The within-group propensity structure will not pursue any kind of prioritization. Formally, within each class h a constant λ_h needs to be found that assures an equal re-issue probability for nonrespondents within that class, aiming at eventually equal response rates for each class, provided that $k_i \geq 1$ and $\sum_{i=1}^{n} k_i = E = 10,000$.

Table 2.2 provides the results of this simulation. Compared with the blind assignment strategy (D1), the response rate has decreased to 0.4667 (from 0.5049) because of the restrictions (D2), forcing more efforts to be assigned to low response groups 4 and 5, at the expense of groups 1 and 2. Conversely, the R-indicator, as well as the maximal possible contrast

Table 2.3: Five categories of response propensities

Group	rel. size	$\bar{\rho}_1$	σ^2_ρ	$F^{-1}_{0.05}$	$F^{-1}_{0.95}$
Overall		0.2500	0.0300	0.5979	0.0442
1	20%	0.4095	0.0322	0.7256	0.1396
2	20%	0.3021	0.0257	0.6064	0.0863
3	20%	0.2372	0.0200	0.5144	0.0610
4	20%	0.1819	0.0144	0.4215	0.0428
5	20%	0.1194	0.0079	0.2980	0.0254

between respondents and nonrespondents, has improved because of the fieldwork restrictions. In this particular case, it is hard to say whether the maximal possible bias has improved because of the restriction. There seems to be a compensating effect between the response rate and the contrast, resulting in status quo with regard to the bias. However, a favorable effect on the bias is probably strongly dependent on the capacity of the auxiliary (class) variable to explain a substantial part of the real response propensities. In the case where 0.018 (instead of 0.01) of the 0.030 is explained by the class variables, the bias improves to 0.2797 (compared with 0.3117), where the response rate is only 0.4341 and the contrast 0.4941.

Imposing fieldwork restrictions in order to achieve equal response rates in combination with response rate maximization with regard to the within-group structure (A2) does not seem to produce less bias than its unrestricted counterpart (A1). This might be due to the restrictions aimed at achieving equal response rates among subgroups resulting in a lower overall response rate. Again, this might also depend on the capacity of the auxiliary variable(s) to capture the variability of the real response propensities.

An interesting variation on this strategy is the one-shot strategy (E1) as presented above, where it is now imposed that the five subgroups should be represented according to the population parameters. Table 2.3 indicates that the five groups comprise equal proportions of the population. This implies that fieldwork efforts should be directed toward obtaining equal group sizes in the obtained sample as well. This strategy is particularly interesting as it is very similar to what is known as *quota sampling*. The process

tries to make the obtained sample correspond with the population with regard to a predefined set of auxiliary variables (for example age and gender classes), allowing nonrespondents to be replaced by substitutes of the same class. This simulation will push the substitution of nonrespondents toward an extreme position: only one attempt per individual is permitted and nonrespondents are immediately replaced. Formally, the strategy is forced to have that $\sum_{c=1}^{5} n_h = E$ and $n_{r,1} = n_{r,2} = n_{r,3} = n_{r,4} = n_{r,5}$. Thus, in this simulation, the unknowns $n_1, n_2, \ldots, n_5$ need to be found.

For the quota strategy, the 10,000 contact attempts are divided among the five classes, resulting in equal group sizes in the obtained sample. This results in $n_1 = 1024, n_2 = 1389, n_3 = 1768, n_4 = 2306, n_5 = 3512$. Given their class-specific response rates, the full-sample response rate is 2097/10,000 = 0.2097. The only remaining variance in the sample is the within class variance, which can be determined by $\frac{1}{n}\sum_{c=1}^{5} n_h \sigma^2_{h,\rho}$. With the final response rate and propensity variance, the different sample quality indicators can constructed again, as shown in Table 2.2.

The quality of this quota sample does not seem to be very attractive (E2). It combines a relatively high contrast between respondents and nonrespondents and a very low response rate. This results in the highest bias of all the restricted fieldwork strategies. The ease with which nonresponding individuals (mostly low propensity cases) are replaced by new units is probably the cause for this high level of nonresponse bias.

When reconsidering the response rates of the simulations in Table 2.2, it seems that they are all somewhat below the response rate of the real survey (about 56%, 61% ineligibles excluded) that was used as the empirical reference for the simulations. Only the unrestricted and perfectly informed response rate maximization strategy (A1) has succeeded in obtaining such a response rate. A few hypotheses can be formulated as possible explanations for this problem.

First, the variance of response propensities $\rho_{1,i}$ may be slightly overestimated. In the case where $\sigma^2_{\rho_1} = 0.02$ instead of $\sigma^2_{\rho_1} = 0.03$, the response rate of the bias minimization strategy will increase to 47.51% instead of 43.35%. This is probably due to the fact that the wider the range of re-

sponse propensities, the greater the number of cases that belong to the low propensity tail of the distribution, probably absorbing a great deal of the efforts. Only when $\sigma^2_{\rho_1} = 0.01$, the response rates of the different strategies are relatively close to the real response rate, as they range from 54.49% (bias minimization) to 59.56% (response rate maximization). A propensity variance of 0.01 is however not realistic as it is close to the propensity variance that can be generated by only using auxiliary variables to explain differences in response behavior. It is clear that the real propensity variance should much larger than the propensity variance that is induced by such auxiliary variables.

A second possible hypothesis concerns the number of budgeted contact attempts. This was fixed at $E = 10,000$ and equals the total number of reported contact attempts in the contact sheet dataset. As it can be expected that contact attempts are more frequently under-reported than over-reported (Bates et al., 2010; Wang & Biemer, 2010), the real number of contact attempts may be slightly higher than the reported $E = 10,000$.

A third plausible hypothesis to explain why the simulations have lower response rates than the real survey lies in the nature of the geometric distribution. The geometric distribution, and the exponential distribution to which it is closely related, make no allowance for memory. This means that the fieldwork operations in the simulation do not account for learning from previous attempts, implying that one-shot response propensities are assumed not to change during the course of the fieldwork. However, in a real fieldwork setting, interviewers might learn that Monday mornings are not convenient for individual A and that individual B had a flu at the first attempt, which is informative enough for the interviewer to wait another week before initiating the next attempt. These learning curves may contribute to higher success rates as more information is gathered about the sample cases.

Concluding this section, it should be noted that the only difference between the simulations is prioritization, except for strategies E1 and E2 which start from a larger gross sample. Spending more effort on the low hanging fruit generally leads to higher response rates, but also generates

a higher risk of bias, as they enlarge the potential distance between respondents and nonrespondents. From these simulations, it becomes clear that following the line of least resistance is not the ideal way to conduct survey fieldwork. The strategies similar to quota sampling seem to be the worst: they combine a relatively large contrast between respondents and nonrespondent on the one hand, and a low response rate on the other hand, resulting in the most severe risk of nonresponse bias. Hence, the rejection of quota sampling in most survey handbooks seems to be perfectly justified (see, among others, Biemer & Lyberg, 2003; Babbie, 2009). In particular, the ease with which nonresponding individuals are replaced by substitutes can be considered as a negative property of quota sampling. Nevertheless, many survey agencies still seem to continue to use such methods. Strategies that reduce the risk of bias or pursue representativeness seem to offer the best results. Their response rates may be somewhat lower, but since the variation in the response propensities that are finally obtained is minimized, the risk of bias is also strongly reduced. Unfortunately, how strongly an individual is inclined to participate in a survey is hard to assess, making it difficult to carry out fieldwork operations under such perfect conditions. Nevertheless, even an uninformed fieldwork process might still be a reasonable option, possibly complemented by balancing the response rates with regard to known auxiliary variables. However, as these auxiliary variables only (very) partially explain the response propensities, this strategy (D2) is still suboptimal and cannot be considered as a method to satisfactorily combat nonresponse bias.

2.3 Evaluating real fieldwork activities under the random response model: following the line of least resistance

Having critically assessed the fieldwork objectives that should and should not be followed, real fieldwork operations can now be monitored in order to evaluate the extent to which survey agencies follow the line of least

resistance. If response rate maximization is still the dominant strategy, fieldwork decisions are expected to be directed toward the low hanging fruit. In the ESS, not only are all participating countries required to attain a response rate of 70%, but other elements of the fieldwork protocol may also advance the prioritization of high propensity cases.

In this regard, consider the strategic choice to pay interviewers per completed interview, which is a fieldwork practice adhered to by most participating ESS countries. Paying interviewers per completed interview could prompt them to maximize their income by rationalizing their efforts; that is by giving priority to the low hanging fruit (Peytchev et al., 2010). For example, in the fifth round of the ESS, only Finland and Norway had an hourly rate arrangement to pay their interviewers. The Netherlands, Switzerland, and Spain combined an hourly rate and payment per completed interview. Sweden paid its interviewers a fixed regular salary. All the other countries paid per completed interview. In some countries such as Belgium, interviewers could earn an additional payment for a successful refusal conversion. Other countries compensated for specific or difficult locations (France).

In the next two paragraphs, the fieldwork operations of some surveys (predominantly the ESS) will be analyzed with the explicit objective of assessing the way cases are prioritized. The first paragraph will use auxiliary variables to determine the likelihood of the individuals to respond positively, the next paragraph will use process data to assess the prioritization mechanism. In fact, previous outcomes of contact attempts may be predictive of the extent to which an individual is likely to participate on a subsequent occasion. It is hypothesized that this process information is more powerful than traditional auxiliary variables in distinguishing between the low and the high hanging fruit. Furthermore, the prioritization may also depend more strongly on this kind of information compared with auxiliary variables.

This is a particularly appealing element in survey research, as auxiliary variables are usually employed to equalize response rates between different subgroups (such as age classes or geographic regions). Indeed, as already

suggested in Chapter 1, there might be greater response propensity variance within the classes of auxiliary variables than between the classes. This means that fieldwork operations that are only guided by auxiliary variables may obtain only a very partial picture of the response propensities.

2.3.1 Monitoring fieldwork operations based on auxiliary variables

Groves (2006) states that 'blind pursuit of high response rates in probability samples is unwise; informed pursuit of high response rates is wise'. Indeed, as the simulations in section 2.2 suggest, the maximization of response rates may induce a fieldwork mechanism that follows the line of least resistance. Therefore, many survey researchers endorse the principles of informed response maximization.

In addition, the ESS does not encourage the pursuit of high response rates as such, but instead advises participating countries to reduce the negative effects of nonresponse in an informed way (Stoop et al., 2010). The Core Scientific Team (CST) of the European Social Survey has therefore proposed a set of response enhancement recommendations to be read by national coordinators when preparing fieldwork in their countries (Koch, Fitzgerald, Stoop, Widdop, & Halbherr, 2012). One of the concerns of the CST is that an 'essential element is the need to achieve high response rates in all participating countries, and to ensure that the people interviewed in each country closely represent the country's total population'. Apart from the minimal target response rate of 70%, participating countries are invited to consider a range of fieldwork techniques including 'advance letters, toll-free telephone numbers for potential respondents to contact, extra training of interviewers in response-maximization techniques and doorstep interactions, implementing refusal avoidance and conversion techniques, re-issuing of refusals and noncontacts' (Koch et al., 2012). In particular, the CST asks the national coordinators to 'be mindful of the need to boost levels of response among all groups of the population and to bring response rates to a more consistent level across subgroups, if possible'.

This balancing of response rates across subgroups closely resembles strategy A2 in Table 2.2: for some pre-defined variables, correspondence needs to be met between the finally obtained respondent set and the full sample (or population). These variables usually only very partially reflect the true propensity variance of the sample, probably only revealing the tip of the iceberg and leaving a great deal of relevant propensity information concealed below the line. Some available auxiliary variables may be used for the fieldwork modification in order to achieve balanced response rates. Other variables that are not pre-defined for obtaining representativeness, however, may be seen as traces of evidence that the fieldwork indeed follows the line of least resistance. Anyway, balancing response rates does not seem to be a guarantee to reduce nonresponse bias as both the propensity variance, but also the overall response rate seems to decrease.

Data analysis There may be many ways to assess how the allocation or prioritization of fieldwork efforts influences quality indicators such as the R-indicator, or maximal absolute bias and contrast. In this data analysis section, two perspectives will be explored. This first will look at the fieldwork process, taking the ranking of consecutive contact attempts within each individual as the main perspective. This type of perspective cannot be used for real-time fieldwork observation, but is interesting as it considers prioritization at the individual level. Second, the fieldwork process and its resulting quality indicators may also be seen through the perspective of the progressing fieldwork period. This approach can be used for real-time fieldwork monitoring and is more the point of view taken by the survey management, who need to decide how prioritization should be altered as the fieldwork progresses. Particularly in the context of adaptive survey designs, this is an interesting approach.

Assessing fieldwork operations as a function of the rank of the contact attempts

In the ESS3-BE, the gross sample contains 3249 individuals, all of whom have been visited at least once. Table 1.4 on page 36 also indicates the

Table 2.4: Univariate success probabilities after one contact attempt($^\diamond$), ESS3-BE ($n = 3249$)

	Success Probability		Success Probability
Age class		*Percentage of non-Belgians in municipality (in %)*	
< 21	26.47%	< 2	29.73%
21-40	15.10%	2-5	23.34%
41-60	22.87%	5-15	21.19%
> 60	26.96%	> 15	11.29%
Population density in municipality (in inh./km^2)		*Average annual per capita income in municipality (in €)*	
< 200	23.02%	< 12000	14.93%
201-400	28.13%	12000-14000	22.63%
401-700	23.76%	14000-16000	24.95%
701-2500	20.80%	> 16000	20.63%
>2500	6.46 %		
Gender		*Type of housing*	
Male	20.09%	Multi-unit	7.48%
Female	23.41%	Single-unit	24.95%
Region		*Neighbourhood quality*	
Flanders	25.68%	poor	14.45%
Brussels	6.54%	good	25.58%
Wallonia	19.13%	excellent	22.70%

($^\diamond$) An appointment at the first contact attempt, immediately followed by an interview at the second contact, is also considered a successful first contact attempt

auxiliary variables that are available for both respondents and nonrespondents, such age, gender, region, or the type of housing. It is possible to estimate the propensities for responding positively after only one contact attempt. These estimated propensities could serve to inform the success probability of an individual, conditional on his or her auxiliary information. Table 2.4 provides per auxiliary variable the conditional probability of a successful first contact attempt.

In a multivariate setting, the success at the first contact attempt can be modeled using logistic regression, where the auxiliary variables are covariates. This models can estimate the individual first contact response propensities, of which the distribution is shown in Figure 2.5.

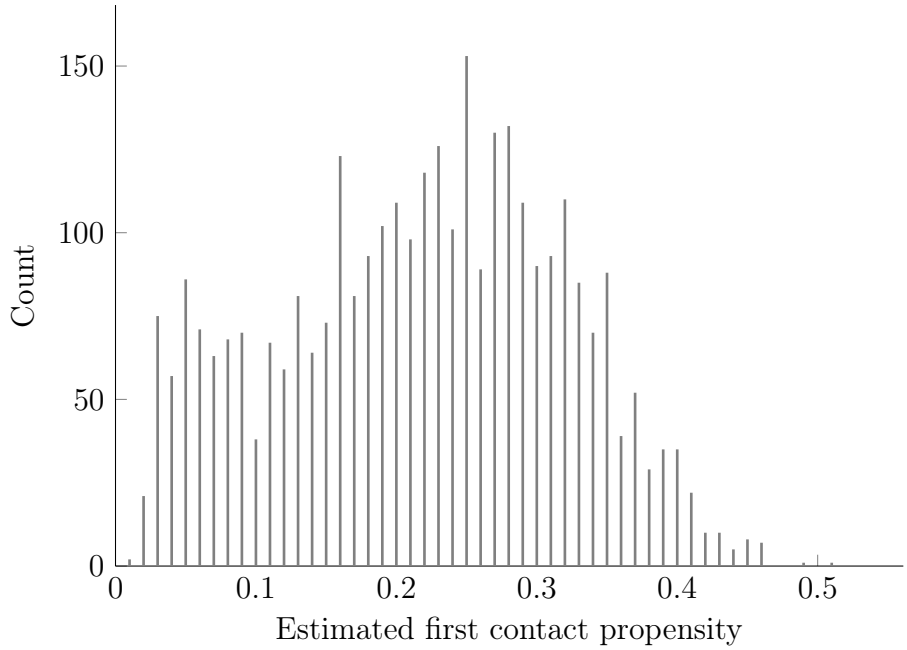


Figure 2.5: Distribution of first contact response propensities, $\bar{\rho} = 0.2179$, $\sigma^2_\rho = 0.0103$

It is assumed that these first contact propensities are indicative of the probability of successful subsequent contact attempts. For example, individuals living in Brussels, or individuals in the age class 21-40 are expected to be harder to convert after the first contact attempt than profiles with relatively high first contact propensities. To some extent, these first contact response propensities reflect the responsiveness of the diverse profiles in the full sample, and which may be used to direct the prioritization of cases during the further course of the fieldwork. However, it may be possible that the first contact propensities differ from the second (or subsequent) contact propensities. In this context, consider the hypothetical situation where men tend to be more inclined to participate at t_1, whereas women are more likely to participate when more contact efforts have been made. It may be possible to test this hypothesis by estimating the response at all contact attempts and allowing interaction terms between the auxiliary variables and the rank of the contact attempts in the logistic regression models. As substantive effects emerge from the interaction terms,

the assumption may be violated that the first attempt propensities are good predictors for the further course of the fieldwork. Specifically, in the ESS3-BE the success probability among apartment dwellers on the first contact attempt is 7.48%, on subsequent occasions, the probability seems to have increased to 15%. Such interactions could not be found for other auxiliary variables. If this interaction between contact rank and type of housing truly reflects an increase in response propensities among apartment dwellers, then considering the first contact response propensity may not be the best way to assess the attractiveness of pending nonrespondents.

However, this increase may also be due to the accumulation of information gathered by interviewers during previous contact attempts. For example, an interviewer may have had difficulty gaining access to the building in which the apartment unit was located. Overcoming this obstacle is more likely during subsequent contact attempts. The selection mechanism may also be considered in explaining possible changes in propensities. Suppose that during subsequent contact attempts only the high propensity apartments have been prioritized, the apparent increase of response propensities for this particular group may only be a by-product of fieldwork selectiveness.

Nevertheless, the first contact attempt can be seen as an unprejudiced visit, where no memory effect or selectivity is operating. These relatively unbiased first contact propensities can be used later to estimate how the fieldwork decisions prioritize certain profiles in the group of pending nonrespondents.

First, consider Table 2.5 where the fieldwork can be monitored as a function of the efforts or contact attempts. At the first contact attempt, the sample comprises 3249 individuals to start the fieldwork operations with. Contact with all of them needs to be attempted at least once, implying that no cases can be dropped in this early stage of the fieldwork process. Out of all the attempted cases, 708 are interviewed, leaving 2541 cases as pending nonrespondents for the second contact attempt. Subsequently, 451 cases are interviewed at t_2, but 359 cases are dropped and considered as final nonrespondents. These individuals will never be visited again. After the

second contact attempts, there are still 1731 pending nonrespondents, and so forth.

After each contact attempt, some respondent-set quality indicators are determined. The response rate measurement is the cumulative number of respondents divided by the gross sample size of 3249. After each contact attempt, a renewed response propensity is determined for each individual by regressing the respondent set membership by the set of auxiliary variables. The aggregated mean and variance can be used to calculate the maximal absolute bias (see equation 1.20 on page 54), the maximal absolute contrast (see equation 1.21) and the R-indicator (see equation 1.18 on page 53). In addition, the cumulative number of contact attempts (efforts) has been recorded.

Examining these quality indicators, it can be observed that the maximal contrast only starts to decrease after the second contact attempt. The maximal bias seems to benefit from a substantial response rate increase after the first contact attempt and a reduction to the maximal contrast after the second attempt. Thereafter, the changes in response rate and contrast are only modest, so the bias can only improve slightly. The R-indicator is somewhat harder to interpret, as the variance of the response propensities strongly depends on the response rate. Response rates closer to 50% usually generate larger variance compared with response rates closer to 0% or 100% (for a more detailed discussion of this problem, see page 58).

The last four columns in Table 2.5 indicate the prioritization during fieldwork. After each contact attempt, the means of the individual first attempt response propensities are determined within the groups of dropped individuals (final nonrespondents), sample units selected for conversion activities, and converted cases. It should be noted that the dropped and the selected cases are mutually exclusive and the respondent set is a subset of the selected set. For example, at t_2, 708+451 cases have been interviewed, their average response propensity as determined at t_1 is 0.2548. Among the 359 cases that have been dropped at t_2, the average propensity at t_1 is 0.2031, among the selected nonrespondents for the follow-up attempts this average propensity is 0.2052. The last column (prioritization) mea-

Table 2.5: Sample quality indicators for subsequent contact attempts, ESS3-BE ($n = 3249$)

contact attempt	number started	number dropped	number selected	number responded	cumulative response rate	maximal bias	maximal contrast	R-indicator	cumulative effort	$\bar{\rho}_1$ dropped	$\bar{\rho}_1$ selected	$\bar{\rho}_1$ responded	prioritization$^{\diamond}$
1	3249	0	3249	708	0.2179	0.4411	0.5639	0.8078	3249		0.2179	0.2651	
2	2541	359	2182	451	0.3567	0.3627	0.5639	0.7412	5437	0.2031	0.2052	0.2548	0.0616
3	1731	234	1497	272	0.4404	0.2781	0.4970	0.7550	6940	0.2049	0.1950	0.2458	-0.2266
4	1225	181	1044	145	0.4851	0.2569	0.4988	0.7508	7990	0.2037	0.1903	0.2428	-0.2578
5	899	154	745	87	0.5119	0.2431	0.4980	0.7511	8740	0.1729	0.1887	0.2410	0.2695
6	659	126	533	56	0.5291	0.2325	0.4937	0.7540	9276	0.1830	0.1868	0.2397	0.0730
7	477	95	382	33	0.5392	0.2284	0.4958	0.7536	9661	0.1944	0.1830	0.2393	-0.2272
8	350	73	277	27	0.5476	0.2221	0.4908	0.7568	9941	0.1928	0.1766	0.2383	-0.3699
9	250	68	182	18	0.5531	0.2140	0.4788	0.7633	10125	0.1831	0.1740	0.2375	-0.1015
10	165	50	115	7	0.5553	0.2120	0.4766	0.7646	10241	0.1658	0.1792	0.2372	0.1622
11	108	44	64	5	0.5568	0.2095	0.4727	0.7667	10306	0.1640	0.1918	0.2370	0.2611
12	60	20	40	4	0.5580	0.2081	0.4708	0.7678	10346	0.2057	0.1889	0.2368	-0.1813
13	36	9	27	2	0.5586	0.2063	0.4675	0.7695	10373	0.2361	0.1814	0.2366	-0.2236
14	25	11	14	2	0.5593	0.2059	0.4671	0.7697	10387	0.2064	0.1701	0.2365	-0.1547
15	12	6	6	1	0.5596	0.2048	0.4649	0.7709	10393	0.1848	0.1667	0.2364	-0.0841
16	5	1	4	0	0.5596	0.2048	0.4649	0.7709	10397	0.2832	0.1645	0.2364	-0.5642
17	4	1	3	0	0.5596	0.2048	0.4649	0.7709	10400	0.2409	0.1391	0.2364	-0.5675
18	3	1	2	0	0.5596	0.2048	0.4649	0.7709	10402	0.0576	0.1799	0.2364	0.7809
19	2	2	0	0	0.5596	0.2048	0.4649	0.7709	10404	0.1799	0.0000	0.2364	
20	2	1	1	0	0.5596	0.2048	0.4649	0.7709	10405	0.2363	0.1234	0.2364	-1.0000

$^{\diamond}$ prioritization is measured as $corr_{\rho_1, p(reselection)}$

sures the correlation between the first contact response propensities and re-selection probability, conditional on the auxiliary variables. If, for example, low propensity cases such as apartment dwellers or individuals in the age class 21-40 have a relatively high re-selection probability compared with high propensity profiles such as the Flemish or single-unit dwellers, the correlation is negative and suggests fieldwork operations that prioritize low propensity cases. Positive correlations suggests that the fieldwork follows the line of least resistance, prioritizing the low hanging fruit.

The results in Table 2.5 do not suggest a clear prioritization mechanism. The mean of the first attempt response propensities among the selected and non-selected (dropped) cases are usually not significantly different. The correlations indicating prioritization also do not show a stable pattern. Measured over all contact attempts, this correlation is -0.01, suggesting practically no prioritization at all. This clearly does not support the expectation that fieldwork activities tend to follow the line of least resistance.

Nevertheless, some considerations should be taken into account.

1. As was already mentioned, the response propensity of an individual sample unit is measured at the first contact attempt, assuming that this is indicative of the success probability of subsequent attempts. However, the possibility cannot be ruled out that some profiles in the sample may initially be reluctant, but tend to be more cooperative on the next occasion. As an illustration, consider the first contact success probabilities related to the type of housing in Table 2.5: apartment dwellers react positively to the survey request in only 7.48% of the attempts, compared with 24.95% for single-unit dwellers. On the second occasion, among nonrespondents single-unit dwellers have a 22.84% success rate, which is a small decrease compared with the first contact attempts, whereas the apartment dwellers now have a success rate of 15%, which is a strong increase.

   There are many (untestable) reasons why success rates might shift between subsequent contact attempts. First, an undocumented selec-

tion mechanism might have caused only the higher propensity apartment dwellers to be given a second chance. Second, interviewers may have learnt what the impediments to contact are from the first attempt and consequently persuaded the individual to participate, for example asking the janitor or the neighbors when would be a more convenient time to revisit the household or individual. These two elements may even be closely related. Interviewers collect some information about the individual's or household's situation. They may estimate the subsequent success rate and work out an appropriate strategy to successfully revisit the case based on observations such as a barking dog, a neighbor stating the individual is at work, a very mildly or alternatively very strongly reluctant attitude of the target person at the first occasion, etc. Third, a real increase in response propensity might have occurred. For example, some individuals may become aware that their participation is more important than they initially thought.

These elements indicate that assessing the prioritization during fieldwork operations is difficult, as there are many unknowns in the data about the fieldwork process.

2. With regard to the quality of the sample frame from which the gross fieldwork sample is selected, a number of ineligible addresses may emerge, such as deceased people, addresses of schools or private companies, untraceable addresses, premises not yet built or demolished, and so forth. Such cases obviously do not belong to the research population and need to be omitted from the sample. It is relatively easy to delete these cases from the sample records once the fieldwork has ended. The gross sample will therefore be somewhat smaller than the initial sample and the response rate needs to be determined only from the set of eligible cases. However, during the course of the fieldwork, eligibility needs to be assessed by the interviewer and/or fieldwork management. In practice, this assessment may sometimes be rather

ambiguous. This may explain why cases that are initially coded as ineligible may be issued again and even converted into respondents.

The analysis shown in Table 2.5 can be carried out again, omitting the final ineligibles in order to assess the sensitivity of the prioritization pattern with and without the ineligible cases. Some 322 cases are finally coded as ineligible in the ESS3-BE, representing about 10% of the initial gross sample, which is a considerable amount. This is probably due to flaws in the sample framework, which is built using the commercial database of 'Orgassim'. Using the national population register, Orgassim has developed a database with statistics of 'inhabitants per building'. With this database, it is possible to make an individual database including age, gender, and address for each person. Names are not available in this database, therefore this individual database is linked with another commercial database and enriched with names (65% matches). A person is identified by his or her name or the combination of gender and age.

The analysis based on the eligible cases alone does not show different patterns in the evolution of the quality indicators, such as the maximal possible bias or contrast. Without the ineligible cases, the maximal absolute contrast seems to be somewhat lower during the entire course of the fieldwork. The maximal contrast after the first contact attempt is 0.5050 compared with 0.5639 for the situation with ineligibles. At the end of the fieldwork, the different in maximal contrasts is comparable: 0.4125 (without ineligible cases) and 0.4649 (ineligible cases included). Further, the prioritization pattern is not substantially altered by the deletion of ineligible cases. The overall correlation between the selection probability and the first contact response propensity is -0.01 (ineligible cases included) and -0.07 (ineligible cases excluded).

Although no overall prioritization mechanism can be observed when monitoring the fieldwork, some auxiliary variables are indisputably used to inform the selection of cases for additional contact efforts. In this particular

survey, the age class information substantively determined the probability of renewed contact attempts. In this respect, examine Table 2.6. For each contact attempt, the four columns '% selected' report the selection probability of a renewed attempt for each of the four age classes. The effects of this prioritization can be seen in the subsequent four columns, indicating the cumulative progress of the response rates for each of these four age classes. The last column, 'maximal contrast', is built on a set of response propensities where only the age class information is used in the propensity model, or where age is the only auxiliary variable explaining the response. This means that the column 'maximal contrast' gives a reasonably good summary of the four columns reflecting the progress in cumulative response rates. It can be seen that the last age class (>60) is systematically less prioritized. The age class 21-40 is seen to be more prioritized. As the class >60 has a relatively high success rate and the age class 21-40 is less inclined to participate, the class-specific response rates converge slightly, leading to smaller indications of maximal contrast.

In fact, Table 2.6 supports the idea that fieldwork operations are doing the exact opposite of following the line of least resistance, in their obvious attempt to equalize the response rates among different age classes.

However, the age class variable seems to be the only variable for which the prioritization has a positive effect on the composition of the sample. For the other auxiliary variables, traces of prioritization are hard to find, or may even have detrimental consequences. Table 2.7 shows that Brussels was given somewhat less attention at contact attempts 3, 4, and 5. Given that individuals are harder to convert in the Brussels area, the indication of maximal contrast also increases during this sequence of fieldwork efforts. Again, 'maximal contrast' is measured using region as the only auxiliary variable to model the propensities, suggesting that this metric serves as a good summary of the class-specific response rates.

It seems that there is some evidence of prioritizing, though it is unclear what purpose it was intended to actually serve. Some variables may have been targeted at the beginning of the fieldwork as candidates for which correspondence between the gross sample and the final respondent sam-

Table 2.6: Sample quality indicators for subsequent contact attempts for age variable, ESS3-BE ($n = 3249$)

contact attempt	number started	% selected <21	21-40	41-60	>60	cumulative response rate per class <21	21-40	41-60	>60	overall response rate	maximal contrast
1	3249	1.00	1.00	1.00	1.00	0.26	0.15	0.23	0.27	0.2179	0.2814
2	2541	0.90	0.84	0.89	0.83	0.42	0.28	0.37	0.41	0.3567	0.2394
3	1731	0.90	0.89	0.87	0.80	0.53	0.38	0.46	0.46	0.4404	0.1870
4	1225	0.86	0.87	0.88	0.79	0.56	0.43	0.50	0.50	0.4851	0.1593
5	899	0.82	0.87	0.83	0.77	0.58	0.46	0.53	0.52	0.5118	0.1425
6	659	0.83	0.81	0.81	0.79	0.61	0.49	0.55	0.53	0.5291	0.1427
7	477	0.81	0.84	0.81	0.70	0.61	0.49	0.56	0.54	0.5392	0.1457
8	350	0.83	0.82	0.82	0.62	0.62	0.50	0.57	0.54	0.5476	0.1463
9	250	0.67	0.76	0.74	0.65	0.63	0.51	0.58	0.55	0.5531	0.1496
10	165	0.83	0.69	0.67	0.75	0.64	0.52	0.58	0.55	0.5552	0.1438
11	108	0.78	0.58	0.53	0.75	0.64	0.52	0.58	0.55	0.5568	0.1424
12	60	0.71	0.67	0.67	0.62	0.64	0.52	0.58	0.55	0.5580	0.1414
13	36	1.00	0.86	0.50	0.80	0.64	0.52	0.58	0.55	0.5586	0.1410
14	25	0.60	0.64	0.40	0.50	0.64	0.52	0.58	0.55	0.5592	0.1383
15	12	0.67	0.60	0.00	0.50	0.64	0.52	0.58	0.55	0.5596	0.1369
16	5	0.50	1.00	0.00	1.00	0.64	0.52	0.58	0.55	0.5596	0.1369
17	4	0.00	1.00	0.00	1.00	0.64	0.52	0.58	0.55	0.5596	0.1369
18	3	0.00	0.50	0.00	1.00	0.64	0.52	0.58	0.55	0.5596	0.1369
19	2	0.00	1.00	0.00	1.00	0.64	0.52	0.58	0.55	0.5596	0.1369
20	2	0.00	1.00	0.00	0.00	0.64	0.52	0.58	0.55	0.5596	0.1369

Table 2.7: Sample quality indicators for subsequent contact attempts for region variable, ESS3-BE ($n = 3249$)

contact attempt	number started	% selected VLA	BXL	WAL	cumulative response rate per class VLA	BXL	WAL	overall response rate	maximal contrast
1	3249	1.00	1.00	1.00	0.26	0.07	0.19	0.2179	0.3374
2	2541	0.85	0.84	0.88	0.41	0.16	0.31	0.3567	0.3353
3	1731	0.85	0.87	0.89	0.49	0.24	0.40	0.4404	0.3081
4	1225	0.83	0.85	0.88	0.54	0.27	0.45	0.4851	0.3192
5	899	0.83	0.68	0.88	0.56	0.28	0.49	0.5118	0.3291
6	659	0.79	0.64	0.88	0.58	0.29	0.51	0.5291	0.3343
7	477	0.76	0.81	0.85	0.59	0.30	0.53	0.5392	0.3324
8	350	0.77	0.85	0.80	0.59	0.31	0.54	0.5476	0.3233
9	250	0.74	0.66	0.74	0.59	0.32	0.54	0.5531	0.3173
10	165	0.70	0.62	0.71	0.60	0.32	0.55	0.5552	0.3191
11	108	0.65	0.30	0.60	0.60	0.32	0.55	0.5568	0.3160
12	60	0.55	0.50	0.79	0.60	0.32	0.55	0.5580	0.3166
13	36	0.73	1.00	0.75	0.60	0.33	0.56	0.5586	0.3129
14	25	0.64	0.00	0.50	0.60	0.33	0.56	0.5592	0.3142
15	12	0.20	0.00	0.71	0.60	0.33	0.56	0.5596	0.3149
16	5	0.00	0.00	0.80	0.60	0.33	0.56	0.5596	0.3149
17	4	0.00	0.00	0.75	0.60	0.33	0.56	0.5596	0.3149
18	3	0.00	0.00	0.67	0.60	0.33	0.56	0.5596	0.3149
19	2	0.00	0.00	1.00	0.60	0.33	0.56	0.5596	0.3149
20	2	0.00	0.00	0.50	0.60	0.33	0.56	0.5596	0.3149

VLA = Flanders, BXL=Brussels, WAL = Wallonia

ple should be aimed at. Other variables may not have been selected for monitoring purposes, so that the line of least resistance is applied to these variables.

Assessing fieldwork operations as a function of spent fieldwork time

An interesting alternative fieldwork monitoring variant is to screen the evolution of the sample quality indicators during the course of the fieldwork rather than assessing the fieldwork as a function of the rank of the contact attempts.

The top panel of Figure 2.6 shows the quality indicators of the obtained sample under the random response model. After each fieldwork day, a model is run to obtain individually updated response propensities based on the auxiliary variables, with which in turn the response rate, R-indicator, maximal absolute contrast, and bias can be determined. As can be observed, the maximal absolute contrast (expressing the worst-case difference between respondents and nonrespondents) does not notably reduce as the fieldwork progresses, meaning that the reduction of maximal absolute bias is instead due the increase in the response rate. As a reminder, the maximal absolute bias expresses the worst-case difference between respondents and the full sample. Apart from the first two weeks of fieldwork, after about 30 days the maximal absolute contrast seems to peak for the first time, and peaks again after about 80 days of fieldwork. In between these two peaks, the lowest point is after about 65 days. Not surprisingly, this profile is also reflected, although somewhat rotated, in the maximal absolute bias and to a lesser extent the R-indicator curve. It should be noted that the R-indicator mirrors the two other indicators. Indeed, the R-indicator should be as high as possible, whereas the maximal absolute bias and contrast should be as low as possible.

The second panel of Figure 2.6 shows the prioritization during the fieldwork. At each fieldwork day, the probability of a renewed contact attempt is estimated for each pending nonrespondent, conditional on the auxiliary variables. These selection probabilities are then correlated with the

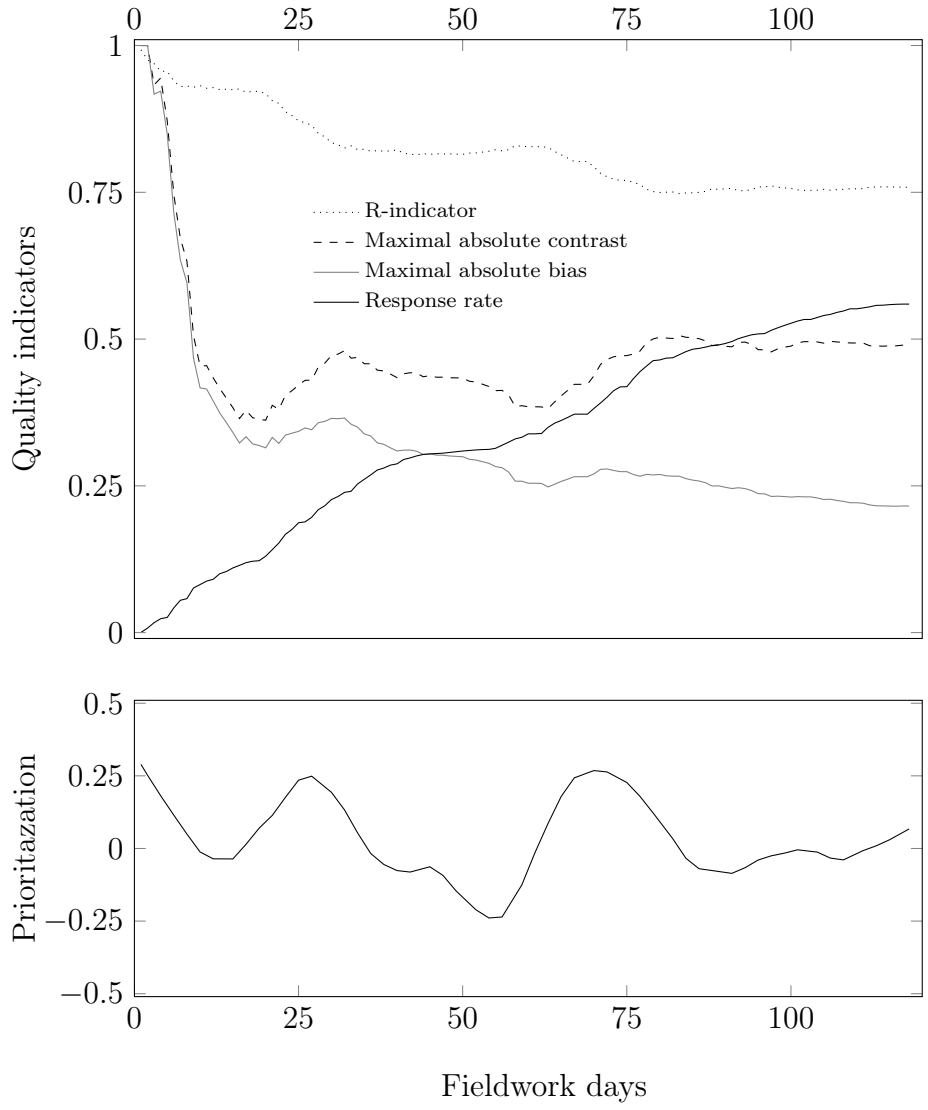


Figure 2.6: Assessing the evolution of the sample quality indicators and prioritization in the course of the fieldwork, ESS3-BE ($n = 3249$)

response propensities determined for the previous day. Positive correlations express the tendency to select high propensity cases rather than low propensity cases. The underlying reasoning is that an adaptive fieldwork strategy may regularly update the composition of the sample by determining the response rate or propensity of individuals or classes of respondents as identified by a set of auxiliary variables. If the fieldwork management finds that particular profiles are under-represented, survey efforts may be allocated to these profiles.

On average, the fieldwork may prioritize high propensity cases slightly more than low propensity cases, as the prioritization line is on average slightly above zero (the average correlation is 0.07). More important, is that the curve describing the maximal absolute contrast tangibly coincides with this prioritization curve. This means that improvements to the obtained sample quality may be influenced by fieldwork efforts. Moreover, the peaks in the maximal contrast curve that coincide with the peaks in the prioritization curve are also reflected in relatively strong increases in the response rate. For example, between fieldwork days 20 and 45, and again between fieldwork days 70 and 80, relatively strong accelerations can be observed in the response rate curve, as well as increases in the prioritization of high propensity groups and an increase in the maximal absolute contrast and bias. This may illustrate that increases in response rates do not necessarily imply the reduction of nonresponse bias and seems to confirm the findings from the simulations (see section 2.2.2), where it was found that response rate maximization may have detrimental effects on the nonresponse bias because of the prioritization of high propensity cases. ■

2.3.2 Slipping through the nets of auxiliary variables

As already mentioned in Chapter 1, auxiliary variables such as age, gender, or residence information may be too weak to substantially explain response behavior. Such variables may indicate that survey estimates are prone to be biased, or even indicate the direction of the expected bias, but it can be expected that many of the reasons why individuals or households

participate in a survey remain below the line. Therefore, only taking the above the line information into account may be quite hazardous. Further, the categories of these above the line variables may be deliberately used to pursue equal response rates between categories. This implies that representativeness can only be accomplished in the weak sense, thus only partially. Survey researchers may find some comfort when the age and gender class distribution of the obtained sample corresponds to the full-sample or the population distribution. However, this may only provide a false sense of security if the propensity variance within such classes exceeds the differences between the classes of auxiliary variables.

With the rise of paradata in the survey industry, a considerable part of the production process of the sample is documented by information that may go far beyond the relatively vague description of the properties of individuals, such as age, gender, type of dwelling, and so forth. The knowledge that someone who initially refused or could not participate because of language barriers may be much more informative and relevant for assessing the problem of nonresponse as compared with socio-demographic variables. In recent years, many nonresponse researchers have become aware of the opportunities provided by contact data in order to better understand the underlying process that leads to (non)response (see, among others, Kreuter & Kohler, 2009; Bates et al., 2010; Biemer et al., 2012).

First, it can be hypothesized that initial or preceding response behavior is a fairly good predictor (at least better than auxiliary variables) of the success of subsequent attempts. This has already been mentioned in the somewhat speculative analysis presented in Chapter 1, where the progress of response rates was used to estimate the variance of response propensities. The second research question of this section is to find out whether initial or preceding (non)response behavior also predicts the success probability of a renewed contact attempt. Inspired by the hypothesis that fieldwork operations follow the line of least resistance, it is expected that low propensity nonresponse profiles such as hard refusals, non-natives, or mentally or physically disabled people will also have a low re-selection probability. Profiles that are deemed to be converted more easily (soft refusals, non-

contacts, broken appointments, etc.) are expected to be more frequently given a second chance.

In order to explain the method by which the hypotheses will be tested, consider the fictitious sample as shown in Table 2.8. The sample consists of six individuals for whom the contact history is presented. Three individuals (1, 3, and 5) eventually participate, while the other three will be considered as final nonrespondents after respectively 1, 3, and 3 unsuccessful contact attempts. In fact, the non-selection of an individual for a renewed attempt is the reason the case is considered as a final nonrespondent. In this particular example, the last outcomes of the contact sequence are refusals or a language barrier; nonresponding profiles that can probably be expected to have a low follow-up success rate. Further, other nonresponse outcome codes that relate to relatively high follow-up success probabilities, such as noncontact or a broken appointment, seem to be consistently re-selected. Therefore, this extract seems to strongly reflect the idea that fieldwork decisions follow the line of least resistance.

Using a much larger dataset, the probability of a renewed contact can be modeled after each contact attempt, using the preceding contact history as covariates or

$$\begin{aligned} ln\left(\frac{p_{selection,it}}{1-p_{selection,it}}\right) &= \#noncontacts_{i(t-1)}\beta_1 \\ &+ \#refusal.by.target_{i(t-1)}\beta_2 \\ &+ \#refusals.by.proxy_{i(t-1)}\beta_3 \\ &+ \ldots + \#language.barrier_{i(t-1)}\beta_C \end{aligned} \tag{2.1}$$

where C is the number of possible nonresponse outcome categories such as refusals or noncontacts, i is the index for the individual and t represents the contact number (chronological rank order in the contact attempt sequence within individual i). The variables $\#noncontacts$ $\#soft.refusals$, ... $\#language$ count the number of previous contact outcomes of that particular nonresponse code within the sequence of the call records of an

Table 2.8: Fictitious extract of contact process data. Strong evidence for the line of least resistance

ID	attempt number	outcome	response?	re-selection?
1	1	noncontact	0	1
1	2	noncontact	0	1
1	3	refusal by proxy	0	1
1	4	interview	1	n.a.
2	1	refusal by target	0	0
3	1	appointment	0	1
3	2	noncontact	0	1
3	3	interview	1	n.a.
4	1	moved	0	1
4	2	noncontact	0	1
4	3	refusal by target	0	0
5	1	refusal by proxy	0	1
5	2	noncontact	0	1
5	3	language barrier	0	0
6	1	just away	0	1
6	2	interview	1	n.a.

individual. For example, ID 1 in table 2.8 will obtain $\#noncontacts = 2$ and $\#soft.refusals = 1$ at $t = 4$. Regressing the selection for a renewed contact attempt (*re-selection?*) by this history of nonresponse events enables modeling a selection probability for each individual at each available contact attempt. This resulting $p_{selection,it}$ can be used as an indication of the likelihood that an individual will be re-selected for nonresponse follow-up and is therefore informative for the applied prioritization.

Similarly, before each contact attempt, the probability of a successful follow-up can be modeled, conditional on the contact history or

$$\begin{aligned} ln\left(\frac{p_{success,it}}{1-p_{success,it}}\right) &= \#noncontacts_{i(t-1)}\beta_1 \\ &+ \#refusal.by.target_{i(t-1)}\beta_2 \\ &+ \#refusals.by.proxy_{i(t-1)}\beta_3 \\ &+ \ldots + \#language.barrier_{i(t-1)}\beta_C \end{aligned} \tag{2.2}$$

It should be noted that the contact history only includes the events until the previous unsuccessful attempt $(t-1)$. Therefore, it is not possible to estimate the success probabilities for the first contact attempt, unless a constant $\bar{\rho}_1$ is imputed for all individuals. Specifically, at $t = 1$, there is no contact history available, implying that all individuals should be considered as equally responsive.

It is expected that the correlation between the selection probabilities $p_{selection,it}$ and the follow-up success probabilities $p_{success,it}$ is strictly positive, consistent with the hypothesis that the fieldwork follows the line of least resistance. This is the first hypothesis to be tested. A second hypothesis is that the contact history has additional predictive power for the subsequent response success, in addition to the explanatory power of auxiliary variables such as age, gender, type of dwelling, area information, and so forth. This latter hypothesis is particularly relevant, as it might suggest that an obtained respondent sample that corresponds to the full sample or population with regard to a set of auxiliary variables, can only claim

representativeness in a weak sense. This means that even though the response rates among classes of auxiliary variables may have been equalized by deliberately targeting the fieldwork efforts, there still is a considerable amount of propensity variance within these classes, threatening the representativeness of the obtained sample.

Data analysis The ESS3-BE, FHS, and ESS5 will be used for this analysis, as they all provide contact history data and auxiliary variables. The ESS3-BE is first shown in more detail in order to demonstrate the method. Table 2.9 shows the predicted selection probabilities (applying equation 2.1) for the second contact attempt, conditional on the nonresponse code of the first attempt, and the follow-up success probability (applying equation 2.2) at the second contact attempt, also conditional on the result of the first visit. It can be easily observed that nonresponse profiles that have relatively high follow-up rates (for example 'moved to other address') also have a higher probability of being re-selected. Low propensity profiles such non-natives or disabled (mentally or physically) are less likely to be given a second chance. The raw correlation between the two columns at $t = 2$ is 0.39 and the same correlation throughout the entire course of the fieldwork is 0.32. This latter correlation is estimated using the selection and follow-up success probabilities for all contact attempts in the dataset, as prescribed by equations 2.1 and 2.2. When also including the auxiliary variables in both models to estimate the selection and follow-up success probabilities the correlation reduces slightly to 0.26. Compared with the correlation of -0.01, between selection and follow-up success probabilities informed by auxiliary variables only (see section 2.3.1; -0.07 among only eligible; -0.01 for all individuals), contact history data gives more distinct support to the hypothesis of the line of least resistance.

In the FHS, the correlation between the probability of being re-selected for follow-up success and the actual follow-up success probability is -0.13, when both probabilities are determined conditional on auxiliary variables alone (type of dwelling, quality of the buildings, etc. For an overview of the covariates, see Table 1.1). When the probabilities are estimated based

Table 2.9: Re-selection and follow-up success probabilities for the second contact attempt, conditional on the nonresponse code of the first contact attempt, ESS3-BE ($n = 3249$)

nonresponse code at t_1	proportion at t_1	selection probability at t_2	succes probability at t_2
Refusal by target person	0.0853	0.8013	0.1128
Refusal by proxy	0.0080	0.9352	0.1099
Noncontact	0.3927	0.9659	0.1392
Language barrier / non-native	0.0132	0.4616	0.1254
Moved to other address	0.0240	0.9158	0.3468
Unavailable until ...	0.0705	0.9672	0.2762
Mentally or physically unable	0.0191	0.3235	0.1191
Partial/broken interview	0.0015	0.8429	0.7333
Broken appointment	0.0416	0.9770	0.3237
Other	0.0723	0.9711	0.2609
Ineligible	0.0539	0.4013	0.0678
Interview	0.2179		

on contact history data, the correlation is 0.22, reducing to 0.17 if the auxiliary variables are also included in the model used to estimate the selection and follow-up success probabilities. These results appear quite similar to those for the ESS3-BE.

The Central Scientific Team of the ESS has always encouraged refusal conversion activities among the participating countries. In round three, particularly Belgium, but also Switzerland, Spain, and the Netherlands have followed this directive. This may explain why refusals have been re-selected somewhat more than could be expected from their follow-up success. In addition, noncontacts have been re-issued consistently, as it is required in the ESS that at least four noncontacts are needed before considering a case as a final nonrespondent.

Because of the availability of contact history data in the ESS, the tendency to prioritize can be measured for many of the participating countries. For all participating countries in the fifth round, Table 2.10 shows the correlations between the probabilities of being re-selected for a renewed

Table 2.10: Correlations between follow-up selection and follow-up success probabilities, ESS5

Country	average selection probability	average one shot response probability	average # contact attempts $\bar{k}$	final response rate	prioritization$^{\diamond}$
BE	0.8303	0.1757	2.9682	0.5345	0.2392
BG	0.8357	0.4827	1.5759	0.8143	0.0372
CH	0.9043	0.1007	5.2719	0.5331	0.1349
CY	0.7972	0.4094	1.6531	0.6973	0.1633
DK	0.8203	0.2015	2.7050	0.5540	0.2570
EE	0.7385	0.2639	2.0369	0.5622	0.2458
ES	0.8874	0.1937	3.3969	0.6852	0.0167
FI	0.8997	0.1330	4.4141	0.5945	0.3054
FR	0.8098	0.1406	3.0725	0.4705	0.4890
UK	0.8836	0.1202	4.3436	0.5630	0.2599
GR	0.7366	0.4074	1.5754	0.6560	0.0790
HR	0.7286	0.2785	1.9153	0.5449	0.5348
HU	0.8040	0.2732	2.1685	0.4915	0.4790
IL	0.3783	0.5920	1.1997	0.7285	-0.2726
NL	0.8783	0.1510	3.8013	0.6003	0.0743
NO	0.8202	0.2175	2.6091	0.5804	0.6699
PL	0.8122	0.3160	2.0740	0.7026	0.2988
PT	0.8755	0.3233	2.2573	0.6708	0.1464
RU	0.8052	0.3253	2.0033	0.6664	0.4638
SE	0.8683	0.1302	3.8959	0.5099	0.3890
SI	0.8167	0.2698	2.3098	0.6439	0.0754
SK	0.8479	0.3795	1.9564	0.7466	0.3241
UA	0.7764	0.3872	1.6607	0.6441	0.3442

$^{\diamond} corr_{p_{selection_{it}}, p_{conversion_{it}}}$

contact attempt and the expected follow-up success, based on previous contact attempt outcomes.

For most countries, the correlation is substantially positive, suggesting that fieldwork operations seem to prioritize cases that are expected to be relatively easy to convert. However, in some countries, particularly the ones that seem to combine relatively short contact sequences (suggesting a relative ease to make someone participate) with relatively high response rates, the prioritization of low response propensity profiles is not very strong. These countries include Bulgaria, Cyprus, Greece, and Israel, and to a lesser extent Slovakia. It is also remarkable that these countries have a response rate that exceeds the prescribed 70% rule. Possibly, these participating countries do not need to follow the line of least resistance to obtain the desired response rate, as cooperation seems to be easily obtained. All

Table 2.11: Modeling the success probability after the first contact attempt, model comparisons for ESS3-BE ($n = 3249$) and FHS ($n = 7770$)

ESS3-BE		
	log likelihood	df
Auxiliary variables alone	132.3825	19
Auxiliary variables and contact history	390.1084	30
Deviance	257.7259	11
FHS		
	log likelihood	df
Auxiliary variables alone	366.3982	41
Auxiliary variables and contact history	504.4744	56
Deviance	138.0762	15

other countries seem to have greater difficulty in obtaining positive responses and need a much longer average contact sequence per individual to achieve a reasonable response rate. Therefore, such countries may prioritize the easy cases very explicitly, such as France, the United Kingdom, and Finland. For a few countries (for example Spain, Slovenia, Switzerland, and the Netherlands) the prioritization is in the expected direction but still relatively weak. For Spain and Switzerland, this is possibly due a strong refusal conversion program. Almost all initially reluctant cases are re-issued in these countries, usually by a more experienced interviewer. Generally, it seems that despite the CST's directives to advance the use of refusal conversion efforts, most ESS countries seem to follow the line of least resistance.

The second hypothesis related to the use of contact history data is that this has a (strong) added value over traditional auxiliary variables (background socio-demographic information) in predicting the outcome of renewed contact attempts. In this regard, consider Table 2.11, where the fit statistics of two logistic regressions are shown, modeling the success after the first unsuccessful attempt. Auxiliary variables seem to be able to predict some variation in contact successes, but a substantive proportion of response propensities seem to slip through the net of auxiliary variables,

as the contact history of nonresponding individuals explains an additional substantive proportion of the response successes. ■

The awareness may grow that auxiliary variables could fail to completely reveal the true response propensities, as contact history data adds a substantive part of observed propensity variance. Still, one may also expect that this contact history information remains imperfect. Hence, even when using auxiliary information and when contact history data is available to measure propensities, monitoring and adjusting fieldwork operations based on such information will probably not completely satisfy the need to obtain satisfactory levels of representativeness in the respondent sample. Indeed, process data as employed in the ESS uses (non)response categories that may still be too raw to completely grasp the subtlety or nuance of the true response propensities. Further, interviewers may be cutting corners when filling out contact forms, even sometimes biasing the true flow of contact events.

2.4 Monitoring fieldwork activities under the fixed response model: do additional fieldwork efforts pay-off?

In Chapter 1, it is mentioned that there are two overall ways through which the effects of nonresponse can be observed. The first is to collect auxiliary variables, such as socio-demographic background variables, that are available for both respondents and nonrespondents (or available on the aggregated full-sample level). These variables enable the estimation of a statistic without any nonresponse, and thereby provide a true or unbiased reference. Unfortunately, auxiliary variables are usually not of much interest as they are observed outside the focus of the survey questionnaire. Therefore, the effect of nonresponse on the target variables should be measured differently. Hence, a second group of methods advances the use of contact history data with which the evolution of a target statistic can be

monitored as a function of the fieldwork efforts. It is thereby assumed that individuals who need more fieldwork efforts in order to be become respondents are somehow representative of all nonrespondents. It is usually considered as an indication of nonresponse bias, when late responders differ from early responders with regard to some target variables.

Because there is never an objective or true reference of a survey's target statistic, the assessment of nonresponse effects through extended fieldwork efforts is somewhat speculative. Even after a vigorous fieldwork strategy, some uncertainty about the remaining nonrespondents should be intercepted in a set of sometimes hard assumptions. In this regard, the aforementioned distinction between the continuum of resistance model and the classes of nonparticipants model (Lin & Schaeffer, 1995) reflects such sets of assumptions. The first model assumes a rather uni-dimensional scale reflecting the amount of efforts needed to convert an initially nonresponding individual. This approach is therefore more quantitative (Stoop, 2005): the more effort needed to convert a particular individual, the more he/she resembles the final nonrespondents. The second model is more qualitative, as it assumes that different classes of (non)participants also reflect different forms of nonresponse mechanisms and ditto the effects on the bias. Indeed, there are many reasons why an individual might not be willing or able to participate: not at home, bad timing, language barrier, soft/hard refusal, etc. Each of these profiles may coincide with particular answers to the target questions of the survey. Moreover, extended fieldwork efforts may differently affect the likelihood of participation in the case of an initial refusal, noncontact and so forth.

A complicating factor is not only the awareness that individuals may react differently on additional fieldwork efforts, but also that these fieldwork operations are directed by the coordinator or the interviewer, anticipating the expected success of any renewed contact. In fact, the observation that fieldwork activities are driven by the line of least resistance jeopardizes to some extent the ability to measure the effects of nonresponse by monitoring target statistics as a function of extended fieldwork efforts.

Table 2.12: Frequencies of nonresponse categories at t_1 and number of finally converted cases (between brackets), ESS5

Country	refusal by target	refusal by proxy	broken appoint-ment	non-contact	away until ...	Ineli-gible	phys./ ment. unable	moved	lang-uage barrier	other	broken inter-view
BE	362(19)	61(6)	149(67)	1241(619)	194(93)	106(13)	80(1)	15(6)	52(3)	277(148)	2(1)
BG	150(40)	112(37)	103(46)	894(631)	170(145)	203(0)	9(0)	0(0)	4(0)	4(4)	22(0)
CH	317(67)	62(19)	239(144)	1529(719)	25(15)	61(2)	30(10)	18(7)	21(0)	31(6)	2(2)
CY	67(0)	43(0)	11(3)	505(308)	94(76)	95(0)	4(0)	0(0)	92(7)	1(1)	0(0)
DK	345(1)	18(0)	328(193)	1099(554)	169(105)	20(2)	68(3)	34(12)	14(0)	164(65)	0(0)
EE	212(4)	82(13)	185(110)	936(294)	220(121)	142(2)	48(1)	194(77)	3(0)	320(181)	10(6)
ES	105(27)	28(12)	597(394)	1155(728)	14(8)	223(43)	39(1)	32(17)	10(0)	0(0)	26(17)
FI	243(33)	13(1)	112(85)	925(489)	1(0)	8(0)	37(5)	6(3)	8(1)	1426(841)	4(3)
FR	61(1)	387(24)	249(89)	2183(843)	10(0)	235(8)	39(0)	1(1)	29(0)	101(55)	0(0)
UK	202(45)	261(49)	219(119)	2557(1227)	341(196)	177(3)	22(0)	0(0)	9(1)	137(65)	2(2)
GR	89(0)	709(30)	57(38)	943(521)	5(4)	265(0)	4(1)	0(0)	37(0)	0(0)	4(4)
HR	353(13)	331(43)	47(19)	941(369)	67(32)	84(0)	19(3)	9(2)	1(0)	98(34)	25(24)
HU	236(12)	71(8)	175(68)	890(398)	83(44)	76(0)	35(0)	39(34)	4(0)	31(3)	1(0)
IL	160(51)	294(46)	10(6)	619(182)	39(26)	78(0)	21(0)	2(0)	15(0)	10(1)	13(13)
NL	330(110)	202(73)	184(127)	1513(815)	38(18)	135(20)	14(1)	0(0)	20(0)	244(157)	0(0)
NO	450(69)	37(8)	512(320)	231(84)	501(275)	61(1)	34(1)	51(25)	25(1)	78(25)	9(6)
PL	205(44)	60(9)	111(59)	536(270)	340(201)	117(0)	24(0)	174(100)	0(0)	63(35)	12(12)
PT	28(0)	263(0)	77(39)	1790(1181)	48(32)	56(0)	0(0)	1(0)	2(0)	1(0)	0(0)
RU	65(7)	370(21)	156(88)	1633(825)	60(42)	79(4)	12(0)	0(0)	0(0)	0(0)	12(0)
SE	587(74)	14(1)	191(143)	62(11)	493(227)	91(18)	89(4)	18(11)	32(1)	693(319)	0(0)
SI	214(33)	28(4)	175(113)	702(386)	279(178)	85(0)	44(4)	38(12)	2(0)	13(5)	1(0)
SK	106(5)	107(7)	136(85)	819(462)	42(29)	15(0)	7(2)	1(0)	1(1)	41(38)	9(9)
UA	413(1)	76(5)	65(19)	577(273)	218(72)	10(0)	16(0)	21(0)	1(0)	160(109)	23(20)

In this regard, Table 2.12 gives an overview per nonresponse profile (determined at the first unsuccessful contact attempt) for the fifth round of the ESS: how many individuals these profiles include and how many became respondents as a result of deploying conversion programs. It becomes quite clear that for some countries a nonresponse assessment based on fieldwork efforts is not very feasible, as conversion rates are minimal, or even zero, for some of even the large nonresponse profiles. For example, Bulgaria, Cyprus, Denmark, Greece, Portugal, and the Ukraine converted hardly any refusals. As no information is collected within this group, the effects of nonresponse can only be measured based on the remaining converted groups such as noncontacts, broken appointments, or the 'unavailable until ...' group. Such profiles are relatively easily converted and can therefore be expected to generate only 'more of the same' or at least generate a selective proportion of initial nonrespondents. The next cluster seems to include countries that have converted some cases in the difficult profiles (refusals, the disabled, or language barriers). These countries include Belgium, Estonia, Finland, France, the United Kingdom, Croatia, Hungary, the Russian Federation, Sweden, Slovenia, and Slovakia. The few cases in the difficult categories that were converted should be weighted, in order to restore the balance among the nonrespondents. In many instances this means that converted refusals should be given extremely high weight scores (>10) compared with, for example, converted noncontacts. As a result, estimates weighted according to their profile-specific conversion rates may become very unstable, as this leads to an inflation of the variance of the estimate.

For example, in Belgium each nonresponse class has at least one unit that has been converted. In some classes, however, the few cases need to represent numerous others, leading to extremely high weight scores (i.e. w=80 for physically and mentally unable, $w = 19.05$ for refusals by target). Figure 2.7 depicts the confidence intervals of the means for ten randomly chosen target variables. The black intervals represent the early respondents-only situations. Early respondents are those individuals who participated after only one contact attempt. It should be noted that all

variables have first been standardized according to the means and standard deviations of these early respondents only. Therefore, the black intervals have the same location and length over all ten variables.

The grey intervals represent the confidence bounds in the situation where the remaining late respondents are added to the early one, and where all late respondents are given equal weights ($w_{late} = 2.60$), so that the sum of all weights equals the total (gross) sample size of 3267. The early respondents are assigned a weight score of 1, so that $1 \times 728 r_{early} + 2.60 \times 975 r_{late} = 3267$. The intervals with broken lines are similar to the grey intervals. The only difference is that the weights of the initial nonrespondents are determined based on the nonresponse outcome at t_1. In a sense, the nonresponse outcome at t_1 can be considered as an auxiliary variables to determine the response propensities and weights for the initially nonresponding individuals. A quite remarkable finding from Figure 2.7 is that the lengths of the intervals with broken lines are equal to and sometimes larger than the black intervals, even though they are based on a sample ($n = 1703$) that is more than twice as large ($n = 728$). Indeed, when estimating statistics based on extended fieldwork efforts, one strongly depends on the few observations in the sparsely populated classes of nonrespondents.

Strong evidence of bias in the Belgian part of the ESS5 seems to be found with regard to reported subjective health. Somewhat surprisingly, the health situation is expected to be underestimated when only considering the early respondents. Adding the nonrespondents and weighting them as a function of their nonresponse outcome class, the average health status shifts to the positive end of the scale. Surprisingly, the only converted nonrespondent among the initial physically or mentally unable nonrespondents reported an excellent subjective health status ...

Out of the 23 countries in Table 2.12, only a few can be used to reasonably estimate the effects of nonresponse. However, as was already mentioned, it should also be assumed that the converted nonrespondent are to some extent representative of all the other cases in their nonresponse class. Switzerland, Spain, and the Netherlands seem to deliver a substantial amount of data for all nonresponding profiles, supporting a thorough

Figure 2.7: Effects of extended fieldwork efforts on some target variables, European Social Survey, ESS5-BE ($n = 3267$)

nonresponse analysis based on extended fieldwork efforts. Figure 2.8 provides the same estimates as the Figure 2.7, applied to the Netherlands. Because all nonresponse profiles include a substantive amount of converted cases (except for language barriers), the nonresponse analysis based on extended fieldwork efforts is less hindered by the negative impact of variance inflation.

Taking a closer look at the Dutch results for the ESS5, one may be inclined to the finding that the augmented and weighted sample scores higher for 'feeling more safe when walking alone in the dark' compared with the early responders. Such a finding rather contradicts the expectation that anxious people are more like final nonrespondents. This would have instead shifted the estimate to the other end of the scale. Under the same extended effort and weighting conditions, the full sample tends to be more tolerant toward gays and lesbians, feels somewhat more happy and clearly more healthy, and watches somewhat less television.

Obviously, as the true full-sample estimates for the variables are not available, the effectiveness of the extended fieldwork efforts is impossible to assess, unless one is prepared not to consider the evolution of the target variables, but instead the less relevant auxiliary variables. Therefore consider Figure 2.9, illustrating the effect of extended fieldwork efforts on the auxiliary variables of the ESS3-BE. In this figure, the full sample instead of the early responders is chosen as the reference for the confidence interval. This is why the thick black intervals are fixed for all auxiliary variables.

In the ESS3-BE, 708 individuals participated at the first contact attempt (appointments followed by an interview are also considered as first contact successes) and some 1099 individuals only participated after conversion attempts. However, weighting these converted cases according to their initial nonresponse status again induces a considerable increase in estimation variance. This is why the lengths of the thin black lines (early respondents only) and the broken grey lines (all respondents; weighted by (non)response status) are quite comparable, even though the sample size of the second group is more than twice as large as the group of early respondents only. This is probably due to the sparseness of some marginal

Figure 2.8: Effects of extended fieldwork efforts on some target variables, European Social Survey, ESS5-NL ($n = 3186$)

nonresponse groups such as those due to language barriers or illnesses. Indeed, in Table 2.9, it seems that these profiles have both small re-selection and conversion probabilities, implying high weight scores for these profiles.

It appears that biases when only considering the early respondents (e.g. neighborhood conditions, type of dwelling, percentage of non-Belgians, population density, and the Brussels region) are somewhat remedied by extended fieldwork efforts and/or weighing initial nonresponse profiles. Nevertheless, this kind of nonresponse remedy is only partial, as it does not succeed in spanning the full-sample intervals (e.g. type of dwelling or percentage of non-Belgians in the municipality). This means that additional uncertainty or variance should be taken into account in order to avoid type I errors.

The latter analysis can be repeated in a more systematic and refined way. First, only considering the difference between early and late respondents does not take into account that some cases are dropped after the second contact attempt, while other cases receive intensified efforts extending to a third, fourth, or even a tenth contact attempt. Second, the effects of extended fieldwork efforts should not be measured only with regard to location parameters such as means and proportions, but also with regard to association measures such as regression parameters. Third, extended fieldwork efforts are not intended only to enlarge the obtained respondent sample in terms of completed interviewers, the aim is also to gain statistical power. However, as it is possible that extended fieldwork efforts only generate 'more of the same', bias may not be (or only partially) removed, leaving the obtained sample with a relatively low-quality status, implying that additional efforts have not led to the efficient use of the survey resources.

Therefore, reconsider the inferential framework under the fixed response model as presented in section 1.2.6. This perspective considers auxiliary variables as interim target variables, for which the measurable effects of nonresponse are illustrative of what might happen to the real target variables. Whenever a set of auxiliary variables is available for the entire sample, the confidence intervals can be compared between respondents (after

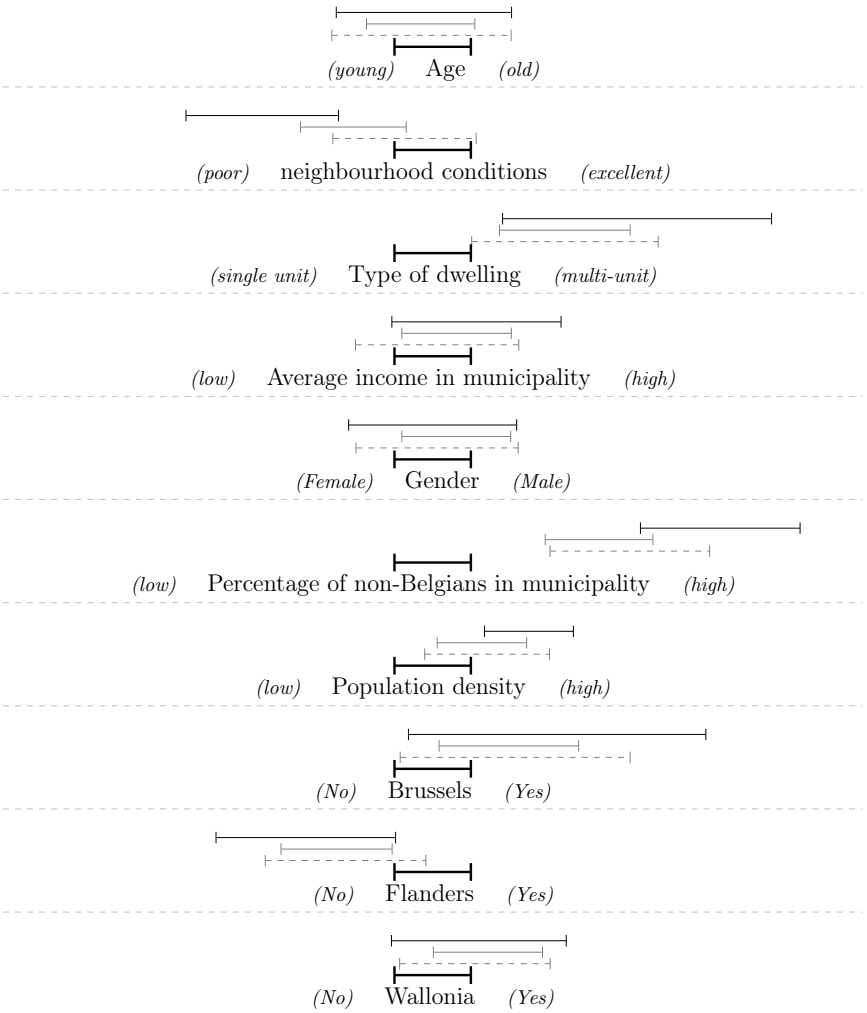


Figure 2.9: Effects of extended fieldwork efforts on some auxiliary variables, European Social Survey, ESS3-BE ($n = 3249$)

the first, second, third , etc.) contact attempt on the one hand and the full sample on the other hand. At each stage of the fieldwork ($t = 1, 2, \ldots$) it is possible to estimate the variance inflation ψ, to be applied to the respondent set, needed to obtain a 95% coverage of full-sample confidence intervals (see equation 1.31 on page 71). Instead of the variance inflation factor ψ, the effective sample size is reported in the monitor dashboards (see Figures 2.10 and 2.11). This can be carried out separately for location and association parameter. In the meantime, the average safety and waste indicators can be measured at each stage of the fieldwork.

Data analysis Figure 2.10 shows eight panels, of which all those on the left refer to location parameters (29) and those on the right refer to association parameters (42). All these parameters are discussed in the real data example referring to the third round of the Belgian ESS (see page 63). In fact, the example will be repeated here, and then during the fieldwork as a function of the number of contact attempts. In each of the panels, the x-axis shows how many contact attempts have been made during the course of the fieldwork, starting from 3249 (applied to all sample members). Adding all the second attempts, the cumulative number of contact attempts is 5387, adding all the third contacts attempts the number of attempts equals 6808, and so on. Finally, about 10,000 attempts will have been made. In the first panel at the top, the y-axis represents the total number of completed interviews. For example, after every individual has been visited once, 728 cases have responded positively, increasing to a number of more than 1200 if all second contact attempts are added, and so forth.

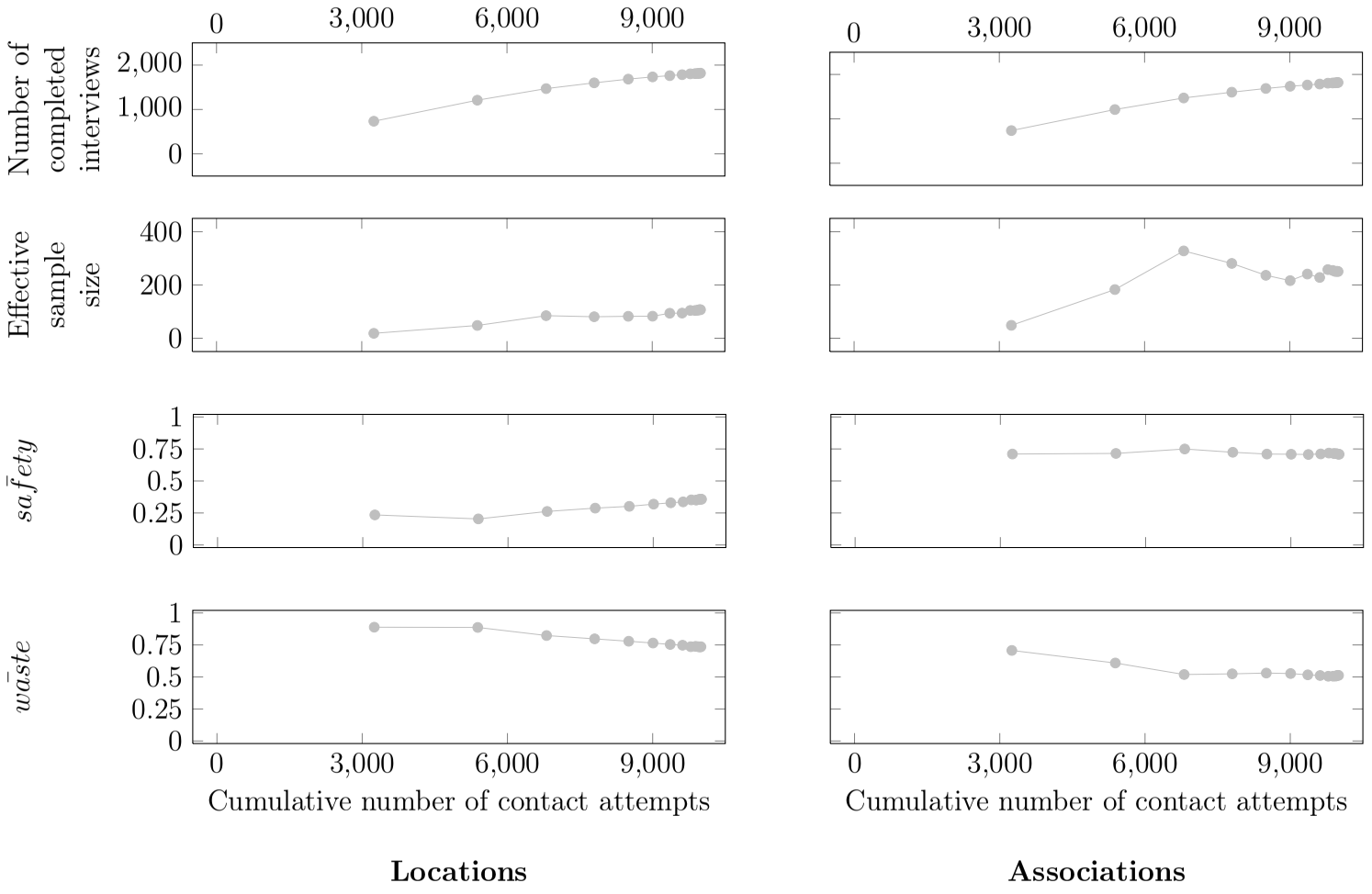


Figure 2.10: Fieldwork monitoring according to cumulative fieldwork efforts, ESS3-BE ($n = 3249$)

The second panel of the monitor dashboard shows the evolution of the effective sample size, or the statistical power of the sample. This effective sample size is the number of completed interviews divided by the variance inflation factor ψ, needed to obtain a 95% coverage of all full-sample intervals. As was already found in Chapter 1, the effective sample size seems to be very low in ESS3-BE, not exceeding 200 effective sample elements for location parameters. Among the association parameters, nonresponse does not have such a drastic effect on the statistical power compared with its effect on location parameters, but it can hardly be considered satisfactory.

The third panel plots the average safety of the parameters. The safety of a parameter indicates the proportion of the density of the full-sample estimate that is covered by the 95% confidence interval determined by the respondents only. The waste indicator, for which the average is plotted in the fourth panel, indicates the proportion of the density of the respondent-only sample estimate that is not covered by the 95% confidence interval determined by the full sample. Here again, the results for the association parameters seem to be slightly more optimistic compared with the location parameters.

Quite striking is the kink in the effective sample size curves after the third contact attempt, appearing in both the location and the association parameter situations, and also reflected to a degree in the average safety and waste curves. Apparently, a large amount of fieldwork effort has not generated additional quality, although the number of completed interviews has increased. A slight improvement can only be observed at the very end of the fieldwork. In fact, if the fieldwork had been discontinued after the third contact attempts, the quality of the respondent set would have been very similar to the real situation, implying that much of the fieldwork effort might have been in vain. Depending on the variable cost of the survey (cost per additional completed interview), a considerable amount of survey budget might have been saved or more efficiently spent elsewhere.

The effect of this leveling off might be a consequence of the fieldwork tactics that are observed in section 2.3.1, Table 2.5: following the line of least resistance. Here, it was also found that from the third contact attempt

onward, the maximal absolute contrast and the R-indicator hardly change, decelerating the decrease of maximal absolute bias. However, strong evidence of this relationship between the random and fixed fieldwork monitoring results is hard to deliver. After all, the fixed models considers auxiliary variables as interim target variables, assuming that these variables represent the real target variables relatively well. Nevertheless, monitoring under the random response model, as well as under the fixed response model seems to produce consistent results.

In the Flemish Housing Survey, the number of completed interviews increases from more than 2000 after the first attempt, to more than 5000 at the end of the fieldwork. However, the implied quality of the obtained respondent set clearly does not improve to the same degree. The average safety for the location parameters even decreases at the beginning of the fieldwork; only at the end of the fieldwork operations is there an improvement in the safety indicator. Because of this, the effective sample size has a major problem in increasing to an acceptable level. This evolution also applies to the association parameters, but is less distinct.

As was also found with regard to the ESS3-BE, much of the fieldwork effort in the FHS could have been used more efficiently, resulting in more safety, less waste, and a greater effective sample size. ■

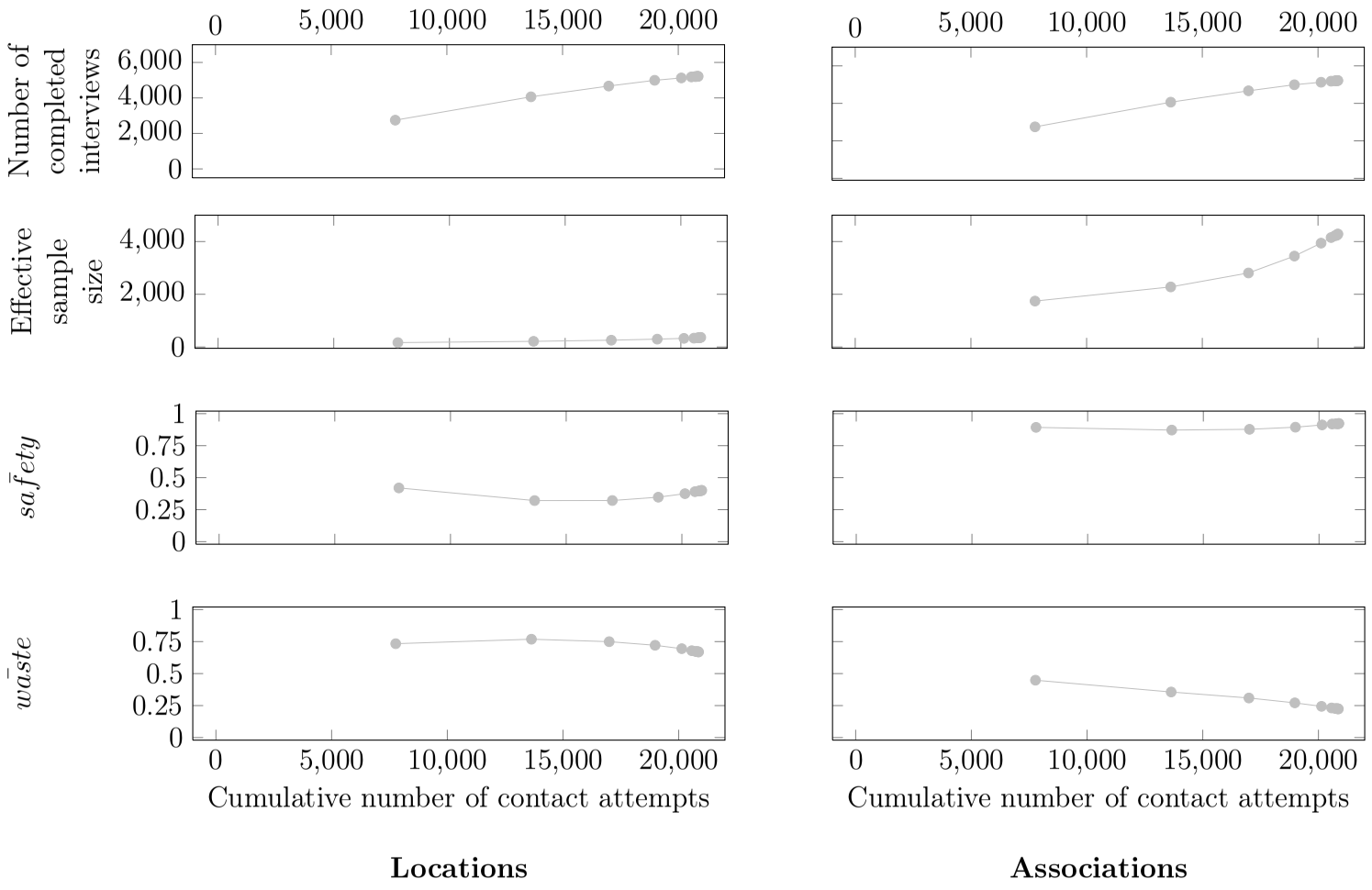


Figure 2.11: Fieldwork monitoring according to cumulative fieldwork efforts, FHS ($n = 7770$)

2.5 Toward smaller sample sizes?

In this last section, two strategies will be presented in order to deal with the nonresponse problem, based on the findings of Chapter 1 and the first sections of Chapter 2.

In Chapter 1 it is discussed that selectiveness in response behavior not only invokes biases with regard to variables that can be observed for both respondents and nonrespondents, but selectiveness also introduces uncertainties with regard to variables that that are observed among respondents only. Although it is very hard to estimate the additional uncertainty that should be taken into account, it is quite reasonable to suggest that it cannot be ignored. In other words, nonresponse induces variance inflation, although it is unclear how much additional variance should be provided. This variance inflation factor ψ may be estimated, based on the known parameter estimates of auxiliary variables projected onto the target variables. However, this also builds on an assumption for which additional uncertainty should be taken into account. This implies that the effective sample size is very likely to be (much) lower than the number of completed interviews or questionnaires. Such a consideration may give rise to the awareness that large samples do not necessarily produce the quality that is initially hoped for, and that the costs of obtaining these large samples are hard to justify. At least two strategies can be considered in this regard.

The first strategy to cope with nonresponse is to substantially reduce the costs of a survey, without significantly reducing the (already low) quality. In fact, decreasing the number of completed interviews does not necessarily imply the same reduction in the effective sample size. Second, instead of following a cost-reduction strategy, survey researchers and agencies may also hope to improve the quality of the realized survey samples by investing in the underlying production process. Instead of following the line of least resistance in order to obtain high response rates, fieldwork operations should be guided by the objective of giving every individual an equal probability of inclusion. In practice, this implies the opposite of

taking the line of least resistance, and prioritizing the low propensity cases such as refusals, language barriers, disabled, etc. instead. The inversion of the line of least resistance supposes that many fieldwork efforts should be directed toward low propensity cases at the expense of the high propensity ones, leading to smaller sample sizes.

Indeed, both strategies will probably result in smaller sample sizes. For the second strategy it holds that 'less is better', whereas the first strategy is inspired by 'less is at least cheaper'.

2.5.1 Reducing costs

In this strategy, the only decision that needs to be taken is to determine the size of the gross sample. Fieldwork agents (interviewers and management) can continue to follow their prevailing fieldwork tactics, implying that the response rate and the level of nonresponse bias are assumed to remain the same, irrespective of the sample size.

When considering $MSE = bias^2 + var$, increasing the sample size will only reduce the second term; the favorable effect of a larger sample will gradually reduce to (practically) zero. Although the number of completed cases may increase, the improvement of statistical quality only increases in a degressive way: the marginal added contribution of each additional completed interview decreases as gross sample size increases. This statistical quality can be expressed as the effective sample size, or the power of a simple random sample that delivers the same margins of error. In Chapter 1, this was determined to be $n_{eff} = \frac{n\bar{r}}{1+\hat{S}^2_{corr}(n(1-\bar{r})-1)}$, with an upper bound of $\frac{\bar{r}}{\hat{S}^2_{corr}(1-\bar{r})}$ (see page 70). $\hat{S}^2_{corr}$ expresses the variability of the correlations between target variables and the response outcome.

Figure 2.12 shows the relationship between the full sample n, the number of respondents n_r, and the effective sample size n_{eff}. In this figure, the average relationship between the target variables and the response outcome is $\hat{S}_{corr}$=0.05, reflecting the fact that, on average, the correlation between any target variable and the response outcome is 0.05. It should be noted that in the ESS3-BE, $\hat{S}_{corr}$ is estimated at 0.11, measured among

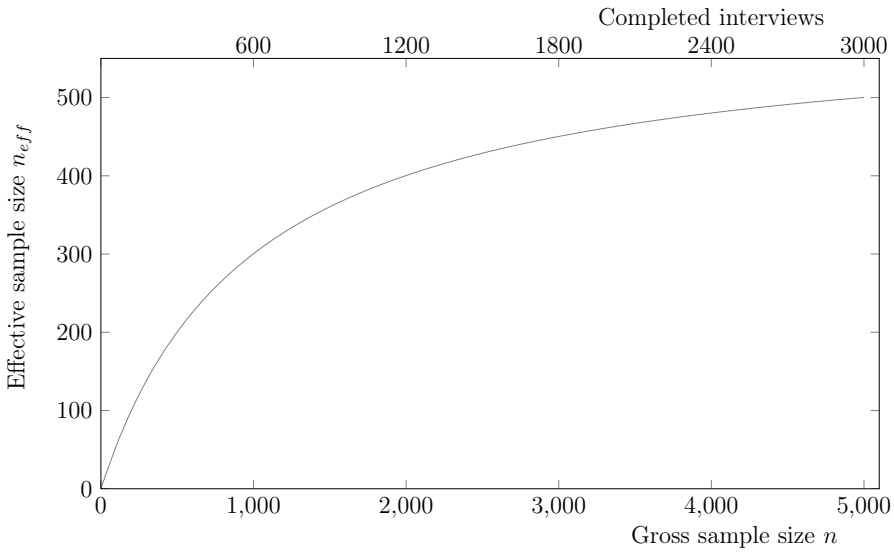


Figure 2.12: Expected effective sample size n_{eff}, conditional on the expected variance of the correlation between r and y ($\hat{S}_{corr}$=0.05), response rate (0.60), and the gross sample

the available auxiliary variables. This average correlation refers to how the mean of a variable is related to the response behavior. In addition to location statistics such as means, other statistics such as regression can be expected to be related somewhat more weakly to the response outcome, implying an expected level of bias that is lower.

Another way to understand how bias affects the effective sample size is to consider the variance inflation needed to make the respondent-only sample cover the full-sample confidence interval. The effective sample size can then be determined by dividing the number of completed interviews by the obtained variance inflation factor (see equation 1.31 on page 71).

Figure 2.12 clearly demonstrates that in the presence of selective non-response, improving the quality becomes increasingly expensive. As the gross sample increases under equal response conditions, the marginal quality increase in terms of effective sample size will be degressive. At a certain point, it might be decided that it is no longer justifiable to make additional efforts for one more unit of sample quality, because the costs to achieve this exceed the benefits.

Suppose that the fixed cost of carrying out a survey is €20,000 (c_{fix}), needed to manage and monitor the fieldwork and for the training of the interviewers. Also suppose that the cost per completed interview is €60 (c_{var}) (interviewer salary, incentive for respondents, etc.). Under these conditions, Figure 2.13 plots the total costs, the mean cost (per unit), the gross and the net sample size as a function of the effective sample size. Again, $\hat{S}_{corr}$ is expected to be 0.05 and the response rate is 0.60.

Because of the degressively increasing effect of efforts on quality, the marginal cost per additional unit of quality (unit effective sample size) increases strongly. This is the reason that large sample sizes are hard to legitimize. On the other hand, the fixed costs (c_{fix}) need to be optimally spread over the realized sample, implying that the sample should not be too small either. This optimization problem is shown in the fourth panel of Figure 2.13. The average cost per quality unit is minimized where the effective sample size is about 256, corresponding to a gross sample size of 744 (a net sample size of 447). The total cost of the survey is estimated at €47.000, about €183 per effective sample unit. Below this optimum, the decrease in sample quality exceeds the cost reduction. For example, decreasing the effective sample size by half to a level of 128 would only result in a total cost reduction of about 36% (from €46,799 to €29,738). In the case where the effective sample size surpasses 256 units, the total cost increase is greater than the quality improvement. For example, doubling the total cost to €93,599 would result in an increase of quality of only 57% (from 256 to 403 effective sample units).

It should be noted that the minimal average cost per unit can be obtained by finding the point where the first derivative of the average cost function is zero. The average cost function is

$$\bar{c}_{eff} = \frac{c_{fix} + c_{var}\bar{r}n}{n_{eff}} \tag{2.3}$$

where $\bar{c}_{eff}$ is the average cost per quality unit, n is the gross sample size, $\bar{r}$ is the response rate and the cost structure is indicated by the fixed costs c_{fix} and the variable cost per unit c_{var}. The average cost is minimized if

$$min\left(\bar{c}_{eff}\right) = \frac{\bar{r}}{\sqrt{\frac{c_{var}\bar{r}(1-\bar{r})S^2_{corr}(1-S^2_{corr})}{c_{fix}}} + S^2_{corr}(1-\bar{r})} \tag{2.4}$$

The ideal gross and net sample sizes seem to be somewhat lower than most surveys currently try to achieve. The ESS3-BE (in which more than 3000 units were selected, of whom about 1800 responded) is located at the very right-hand end of the four curves in Figure 2.13, suggesting that this survey may be too expensive for the obtained quality in terms of effective sample size.

This cost reduction strategy minimizes the average cost per effective sample unit. However, in the case where the fixed cost of a survey is relatively low, the optimal sample size may be very low (e.g. < 10 cases). Indeed, when C_{fix} in equation 2.4 is low, the denominator will increase, making the optimal cost per quality unit low. In the extreme case where $C_{fix} = 0$, the optimal average cost will also be zero, resulting in an empty sample. Of course, this is neither a desirable nor a realistic choice of sample size, suggesting that other considerations should also be taken into account. As a suggestion, the survey agency or sponsor may determine an upper limit of the total survey cost, a lower limit of the effective sample size, or a level of type I error that should not be exceeded.

Indeed, the advantages of this small sample strategy are not only the cost reduction, but also that the risk of making a type I error can be substantially reduced if one is not aware of the risk of nonresponse bias in the estimated statistics. Smaller (nominal) samples provide larger confidence intervals due to sampling variance. Larger samples, however, tend to have smaller confidence intervals, ignoring the uncertainty caused by the risk of bias. In this regard, consider Figure 2.14, expressing the risk of making a type I error as a function of the gross sample size. Again, $\hat{S}_{corr}$ is expected

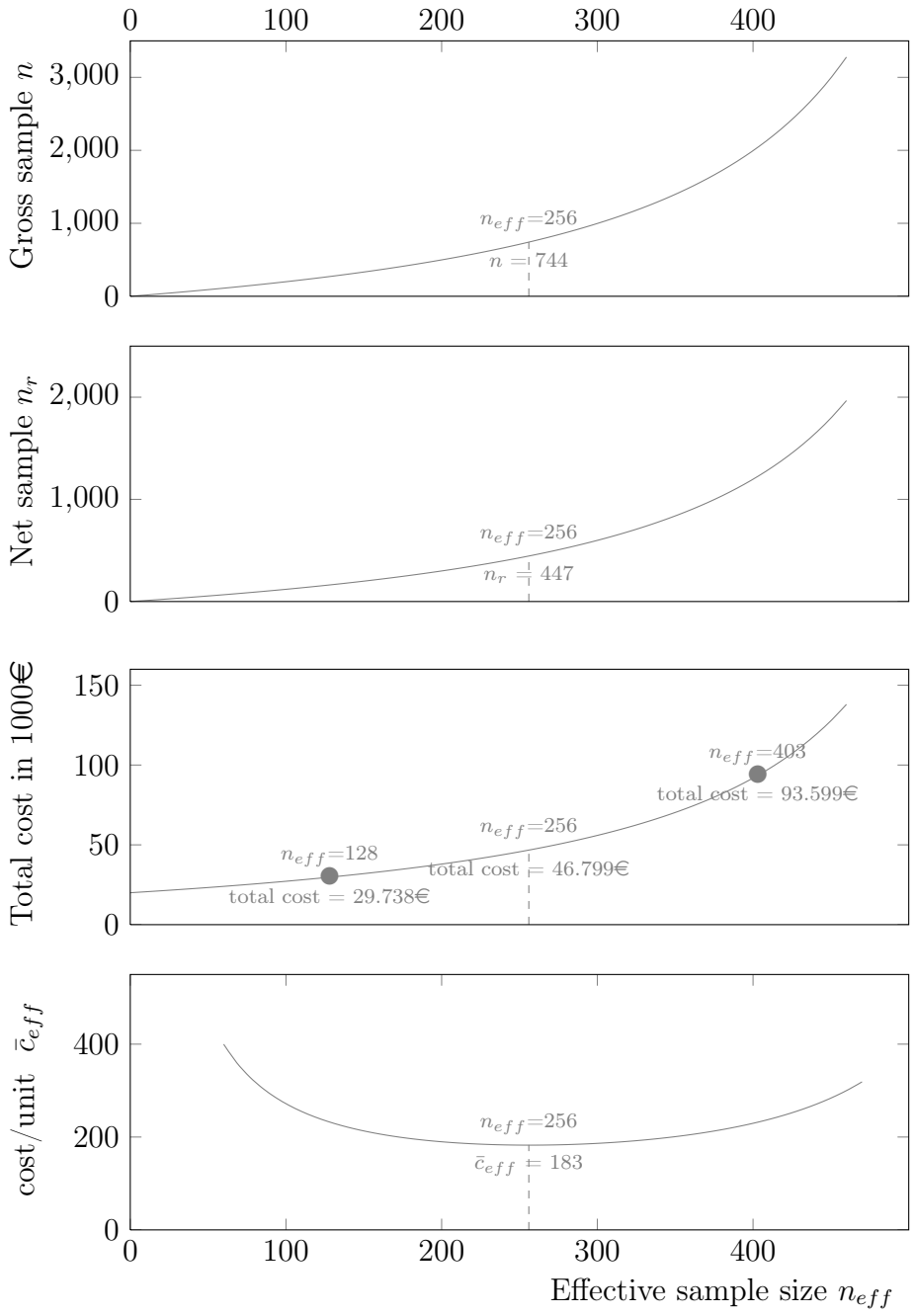


Figure 2.13: Sample size n and n_r, cost $\bar{c}_{eff}$ and total cost as function of effective sample size n_{eff}

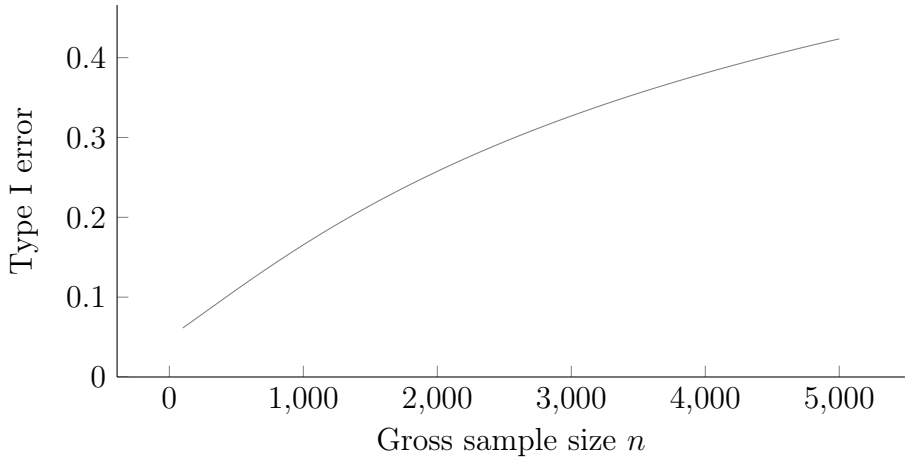


Figure 2.14: Type I error, conditional on the expected variance of the relation between r and y ($\hat{S}_{corr}$=0.05), response rate (0.60) and the gross sample

to be 0.05 and the response rate is 0.60. The Type I error as expressed in Figure 2.14 is the complement of the safety concept as presented in section 1.2.6, equation 1.23. The risk of Type I error can be obtained here by measuring the extent to which the confidence interval with respect to the effective sample size is covered by the number of completed interview. For example, when considering Figure 2.12, a gross sample size of 3000 cases yields 1800 completed interviews, corresponding with an effective sample size of 450. Now, the confidence intervals for the net sample size (1800 cases) are much smaller than the confidence intervals for the effective sample size, implying that the latter intervals can only partially be covered by the first. In this practical example, the type I error is 1-safety or

$$Type_I\left(\theta_r\right) = 1 - \left[\Phi\left(\frac{Z_{\alpha/2}\sqrt{\frac{1}{n_{eff}}}}{\sqrt{\frac{1}{n\bar{r}}}}\right) - \Phi\left(\frac{-Z_{\alpha/2}\sqrt{\frac{1}{n_{eff}}}}{\sqrt{\frac{1}{n\bar{r}}}}\right)\right]$$

and is estimated to be 33%, while usually only 5% is tolerated.

The researcher may decide to allow a type I error of, for example, 10%, corresponding to a gross sample size of about 400 elements (net size: ±204; effective sample size ±170). Again, Figure 2.14 makes it clear that

confidence intervals and inferences drawn only from the naive net sample size face a considerable risk of type I error, only because a certain amount of expected nonresponse bias is not taken into account.

2.5.2 Turning the line of least resistance

Instead of reducing costs (without drastically compromising quality), one may also be inclined to improve the quality of the realized sample by trying to invert the line of least resistance, prioritizing the high hanging fruit rather than the low hanging fruit. Such a drastic reallocation of fieldwork efforts should then aim for representativeness of the obtained sample, instead of maximizing the response rate.

Assuming that the survey budget is fixed, more effort should be allocated to low propensity cases. As a result, fewer interviews will be realized. Nevertheless, as all the sample units should be attempted at least once, many efforts are still directed to the high propensity cases, so that the objective of obtaining a representative respondent set is not optimally achieved.

In order to demonstrate this, the random response model is used again to run some additional simulations allowing the gross sample size to vary and assessing its effect on response rates and the maximal absolute bias. In section 2.2 where the simulations of fieldwork strategies are discussed, a sample is presented with a mean response propensity of $\bar{\rho}_1 = 0.25$ and $S^2_\rho = 0.03$. These properties should somewhat reflect the real propensity structure of the ESS3-BE.

Figure 2.15 represents the maximal absolute bias when (1) maximizing the representativeness (Strategy B: R-indicator maximization) for different gross sample sizes and (2) randomly assigning the conversion attempts (Strategy D1: blind assignment strategy). In both simulations, the total number of efforts (contact attempts) to be distributed over the gross sample of n individuals is $E = 10,000$. Each individual should be visited at least once.

In order to optimize the representativeness, all sample members should have a final response propensity that is very close to the response rate. In the situation with perfect representativeness, $1-(1-\rho_{1,i})^{k_i} = \bar{\rho}_{final}$, which is exactly ρ_1 of the highest propensity case. Due to the fact that contacting each individual is expected to be attempted at least once, the expected final response propensity the highest order propensity case is $1-(1-\rho_{1,i})^1 = \rho_{1,i} = \bar{\rho}_{final}$. All other cases need to be assigned k_i contact attempts to achieve this final propensity. At a certain point, the gross sample size becomes so large that there are not enough contact attempts available to satisfy the ideal of perfect representativeness.

In this particular example, the maximal absolute bias can be perfectly suppressed as long as the gross sample size is lower than about 600-700 units (see reference point A in Figure 2.15).This is exactly the point where there are enough survey efforts available to make all individuals participate. Indeed, the response rate for this scenario is very close to 100%. When selecting more units, the response rate starts dropping, whereas the bias increases. From this point onwards, there are clearly not enough contact attempts available to allocate to all the sample units, the low propensity cases in particular. When the gross sample comprises more than 3000 cases (corresponding to the real ESS3-BE), the risk of bias seems to become practically unavoidable. At this point, the maximal possible bias is 0.23 (0.23 standard deviations between the full sample and the respondent-only sample), implying that bias probably occurs quite regularly in real survey practice.

Because the realization of (perfect) representativeness implies that all individual propensities have to be known, such a strategy is not very likely to apply in a real situation. Therefore, the results of the blind fieldwork strategy are also shown in Figure 2.15. In this strategy, expected to be more feasible in a real survey then pursuing perfect representativeness, all individuals have an equal probability of being re-selected after an unsuccessful attempt. The curve indicating the maximal absolute bias under this strategy is somewhat above the perfectly informed strategy maximizing the representativeness. As long as the gross sample is below a certain

point (in this case about $n = 1250$; see reference point B), the maximal absolute bias remains relatively stable. From that point onward, the sample size seems to become too large to be able to serve all individuals enough to constrain the risk of bias.

The region between the two strategic lines in the graph indicates the bounds of what is realistically feasible (blind strategy; reference point B) and what is optimally desirable (maximize representativeness; reference point A). An optimal gross sample size for this particular situation would consequently be between about 600-700 and 1200-1300 units, and is far below the actual 3249 selected units in the ESS3-BE.

Even repeating the fieldwork simulation in a situation where the variance of the response propensities is lower (for example $S^2_\rho = 0.02$ instead of 0.03), the gross sample size should be limited to about 1100 (new reference point A) and 1600 cases (new reference point B). If $S^2_\rho = 0.01$ (this is comparable to the level of variance explained by a set of socio-demographic background variables), the preferable gross sample size would be about 1900-2100. This latter situation is probably not realistic as has already been argued, since the real propensity variance is expected to be (much) greater than the propensity variances determined by auxiliary variables. A more pessimistic view of the variability of response propensities ($S^2_\rho = 0.04$), leads to an optimal gross sample size between about 400 and 900.

The augmentation of the final response probabilities for the low propensity cases is the basic concern of the quality improvement strategy. Survey researchers have become increasingly interested in finding ways to increase response, such as renewed contact attempts, (monetary) incentives, providing different modes or languages for survey administration, sending a more experienced interviewer, and so forth. Contrary to the somewhat oversimplified settings of the simulations presented above, Peytchev, Baxter, and Carley-Baxter (2009) mention that not all survey effort is equal, suggesting that different recruitment methods should be optimally linked to different profiles of nonrespondents. Using only one survey protocol to convert initial nonrespondents may only yield 'more of the same'. These consid-

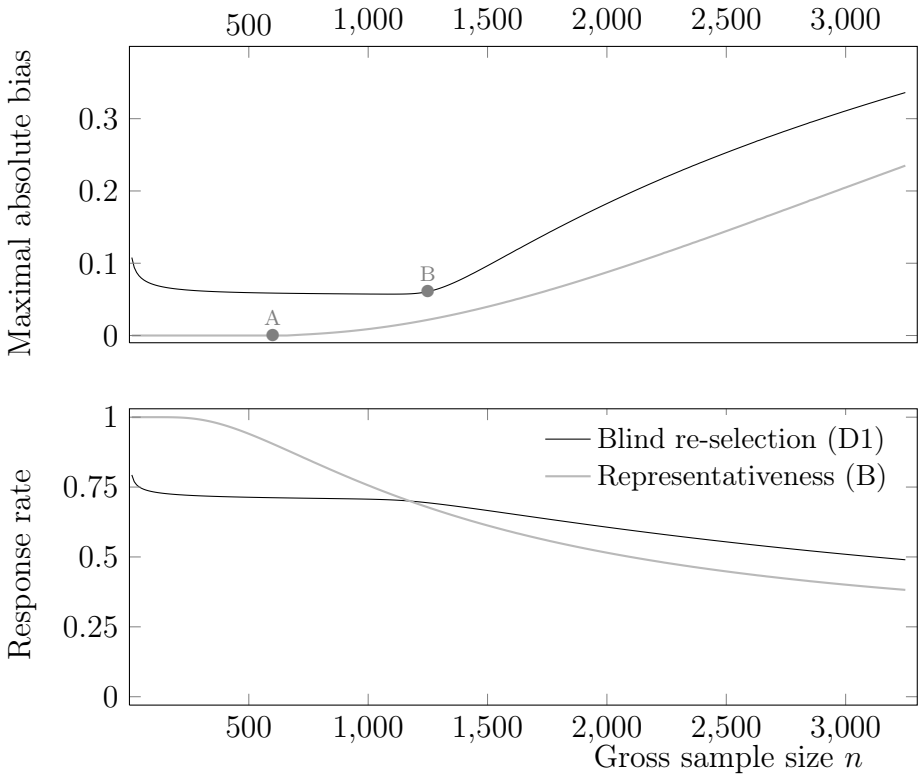


Figure 2.15: Maximal absolute bias and response rates for two fieldwork strategies as a function of the gross sample size

erations are inspired by the Leverage-Salience Theory (Groves, Singer, & Corning, 2000). This theory assumes that individuals value the features of a survey differently and may consequently react differently to a specific mix of design characteristics (Groves, Presser, & Dipko, 2004; Groves et al., 2006). This means that although a response propensity is a unidimensional concept, the implementation to invoke the propensity has many attributes. These may include the name of the sponsor, topic of the survey, (monetary) incentive, interviewer characteristics, contact method and timing, etc. As a result, many survey researchers may advocate the principle of tailoring the recruitment procedure to the specific desires and requirements of the individual.

In light of the above, the concept of the contact attempts needed to convert an individual (k_i) as presented in the simulations is somewhat naive. However, k_i may also be directly related to the cost that is needed to recruit respondents. Low propensity cases such as hard refusals may need more incentives than high propensity cases. Therefore, using the geometric distribution and the number of contact attempts is still realistic, although it should be interpreted as a broader range of fieldwork efforts. Making surveys accessible to different kinds of problematic nonresponse profiles may include many alternative recruitment protocols other than merely renewed contact attempts.

Contrary to the objective of improving the responsiveness among low propensity cases, one may also consider the possibility of mitigating the final inclusion probabilities of high propensity cases. In fact, this means that the $k_i \geq 1$ restriction should be omitted. In other words, instead of only intensifying the efforts toward the difficult cases, the easy cases should receive less attention in order to avoid their prominent presence in the final respondent sample. If the $k_i \geq 1$ restriction is dropped, then a k_i can be found for each individual, satisfying $1 - (1 - \rho_{1,i})^{k_i} = \rho_{final,i} = \bar{\rho}_{final}$, provided that $k_i > 0$ and $\sum_{i=1}^{n} k_i = E$ where E is the total sum of available

contact efforts. Under these conditions, it holds that

$$\begin{aligned}
\bar{\rho}_{final} =& 1-(1-\rho_{1,i})^{k_i} \\
1-\bar{\rho}_{final} =& (1-\rho_{1,i})^{k_i} \\
ln\,(1-\bar{\rho}_{final}) =& k_i ln\,(1-\rho_{1,i}) \\
k_i =& \frac{ln\,(1-\bar{\rho}_{final})}{ln\,(1-\rho_{1,i})}
\end{aligned}$$

Given the fact that all k_i should sum up to E, an expression can be found to determine the final response rate.

$$\begin{aligned}
\sum_{i=1}^{n} \frac{ln\,(1-\bar{\rho}_{final})}{ln\,(1-\rho_{1,i})} =& E \\
ln\,(1-\bar{\rho}_{final}) =& \frac{E}{\sum_{i=1}^{n} \frac{1}{ln(1-\rho_{1,i})}} \\
1-\bar{\rho}_{final} =& e^{\frac{E}{\sum_{i=1}^{n} \frac{1}{ln\left(1-\rho_{1,i}\right)}}} \\
\bar{\rho}_{final} =& 1-e^{\frac{E}{\sum_{i=1}^{n} \frac{1}{ln\left(1-\rho_{1,i}\right)}}}
\end{aligned}$$

Now, replacing the expression for $\bar{\rho}_{final}$ in $k_i = \frac{ln\left(1-\bar{\rho}_{final}\right)}{ln(1-\rho_{1,i})}$ gives

$$\begin{aligned}
k_i =& \frac{ln\left(1-\left(1-e^{\frac{E}{\sum_{i=1}^{n} \frac{1}{ln\left(1-\rho_{1,i}\right)}}}\right)\right)}{ln\,(1-\rho_{1,i})} \\
=& \frac{\frac{E}{\sum_{i=1}^{n} \frac{1}{ln\left(1-\rho_{1,i}\right)}}}{ln\,(1-\rho_{1,i})} \\
=& \frac{E}{ln\,(1-\rho_{1,i}) \sum_{i=1}^{n} \frac{1}{ln(1-\rho_{1,i})}}
\end{aligned} \tag{2.5}$$

Expressed in words, equation 2.5 provides each sample unit with a number of contact attempts k_i in order to achieve perfect representativeness. Applied to a small fictitious sample, it becomes clear that many individuals

should receive less attention than even a single contact attempt (see Table 2.13). The sample consists of 10 individuals with response propensities ranging from 0.05 to 0.50. The total number of dividable efforts E in this small example is 20.

Table 2.13: Example of equalizing final response propensities, $E = 20$, fictitious data

$\rho_{1,i}$	k_i	$\rho_{final,i}$
0.05	7.3161	0.3129
0.10	3.5617	0.3129
0.15	2.3091	0.3129
0.20	1.6817	0.3129
0.25	1.3044	0.3129
0.30	1.0521	0.3129
0.35	0.8711	0.3129
0.40	0.7346	0.3129
0.45	0.6277	0.3129
0.50	0.5414	0.3129

This strategy, which reduces the level of high propensity cases, seems to result in lower response rates and thus also relatively small samples. When applying equation 2.5 to the presumed properties of the ESS3-BE ($\bar{\rho}_1 = 0.25$, $S^2_\rho = 0.03$, $n = 3249$ and $E = 10,000$), the final response rate in order to obtain perfect representativeness is only 35.14% or 1142 completed interviews, compared with 1798 completed interviews in the real survey.

This last strategy opposes common views on fieldwork operations. Usually, prospective respondents should be converted into completed interviews. Here, however, cases that are too easily converted should be avoided. An even more troublesome question is how these easy cases could be identified, and what recruitment procedures are less effective or expensive than one simple contact attempt. Nevertheless, whatever strategy is used to invert the line of least resistance, it always results in smaller sample sizes.

2.5.3 Other pros and cons of small sample sizes

Apart from the fact that small samples may lead to a reduction of survey costs or an increase in quality (under certain conditions, see section 2.5.2), some other advantages may be taken into consideration. As already mentioned, small sample sizes have larger confidence intervals, which are convenient if some amount of unavoidable but uncontrollable nonresponse bias needs to be taken into account. Most statistical software programs assume simple random sampling to build their inferences on. Some advanced procedures also take into account more complex sampling structures as long as these structures are known. Since survey error such as nonresponse is usually (partially) unknown, even after weighting adjustments, the error margins provided by these software programs are probably underestimated, invalidating inferential claims on theoretical statistical distributions such as the t-distribution, the F-distribution or the χ^2-distribution, and their respective p-values. Hence, as smaller sample size provide large confidence intervals, part of the nonresponse bias might also be covered by these additional margins of uncertainty, reducing type I errors. Small samples also reduce the administrative burden, by reducing the organizational complexity. Furthermore, small samples may reduce the social costs of surveying, as fewer people are asked to participate.

Nevertheless, arguments can be found in favor of large samples. First, in a multi-purpose survey such as the European Social Survey, frequently only a subset of the data may be used (e.g. only those in the 65+ age group or only those in employment). These subsets may become too small if the total set is already small. Second, as the distributions of some particular variables such as monthly expenses are very unstable, large sample sizes may be better in order to obtain estimates that are more robust.

2.6 Discussion

Strategically, surveys need to achieve a very uncommon goal: instead of selling as many units of their product as possible, surveyors are expected

to 'sell' equal probabilities. As illustrated in the first chapter, inequality of participation propensities is a prominent threat that fieldwork agents need to combat. The pursuit of participative equality is at least a theoretically justifiable goal; however, survey practice is still predominantly inclined toward the maximization of response rates (or to some extent also the maximization of the sample size).

Then why do survey researchers still insist on maximizing response rates? A first reason is the belief that low nonresponse rates correlate with lower levels of bias and large sample sizes (implying lower standard errors). Indeed, if the proportion of nonresponse is relatively low, the potential for bias to occur is rather limited. However, this does not necessarily mean that response rates should be an end in themselves. In this respect, response rate maximization is probably a good example of what Merton (1940) termed 'displacement of goals'. If avoiding bias is the main objective and low nonresponse rates are believed to restrict the potential for nonresponse bias, the fieldwork objective may shift toward maximization of the response rate, losing sight of the initial objective: bias reduction. A second reason why response rates are so dominant is the ease with which they can be pursued, calculated, and reported. Focusing on the risk of bias is, by comparison, much more problematic.

As the managerial differences between rate maximization and bias minimization strategies seem to be so fundamental, many organizational fieldwork parameters may need to be drastically altered. These include, for example, the training, method of remuneration, and allocation of interviewers, or the incentivizing of nonrespondents. Many of these organizational parameters, however, still seem to be deeply rooted in the tradition of response rate maximization. As an illustration, consider the strategic compatibility between the objective for response maximization and the payment of interviewers per completed interview. Driven by rational choice, interviewers might consider optimizing the trade-off between effort and remuneration, resulting in the prioritization of the low hanging fruit.

However, times may be changing. There seems to be a growing awareness among scholars, survey users, and producers that nonresponse has a

detrimental effect on the quality of survey statistics and that data collection methods should anticipate this threat. Currently, attempts are being made to try to achieve correspondence between the obtained sample on the one hand and the gross sample (or entire population) on the other hand, with regard to a set of variables or statistics. Usually, fieldwork is directed toward the objective of equalizing response rates between classes of sample units, e.g. equalizing the response rate of age $\times$ gender combinations. However, it is clear that such interventions only lead to a partial solution to the problem. As long as such guiding information (auxiliary variables) only represents the tip of the iceberg, leaving most of the response propensity generating information concealed, the strategy of equalizing response rates will only be suboptimal. Strategy D2 in Table 2.2 illustrates this problem. Unless the fieldwork efforts are increased, balancing response rates does not guarantee a reduction of risk of nonresponse bias.

Combating the disadvantages of survey nonresponse therefore requires a more fundamental approach. It is unacceptable to ignore the absence of hard refusals, non-natives, disabled or sick people, or other low propensity profiles when generating statistics from a respondent survey. Fieldwork tactics should therefore be directed toward these groups, making them an offer that cannot be refused. Refusers should be incentivized to cooperate and other investments should be made in order to lower the threshold for other low propensity cases. Survey participation among ethnic minorities, for example, is usually somewhat lower than among the indigenous population, leading to the necessity for adjusted recruitment methods in order to deal with cultural and linguistic barriers (Feskens, Hox, Lensvelt-Mulders, & Smeets, 2006; Deding, Fridberg, & Jacobsen, 2008; Laganá, Elcheroth, Penic, Kleiner, & Fasel, 2011). Physical or mental conditions may also hinder survey participation and require adapted recruitment strategies (Mitchell, Ciemnecki, CyBulski, & Markesich, 2006; Fontaine, 2012), such as adapted questionnaires or third party assistance. Such survey fieldwork innovations should be encouraged. These are particularly relevant when considering the scope of the social aspects that are examined in the European Social Survey. As health issues, ageing,

and forms of discrimination take a prominent place in the ESS questionnaire, fieldwork operations should be more explicitly directed toward these minorities.

An important consideration in this regard is to reconcile survey cost and quality. Because 'more is not necessarily better', survey researchers and fieldwork managers need to reflect on the possibility of using smaller sample sizes. If representativeness or bias reduction really are legitimate fieldwork objectives, the line of least resistance needs to be inverted, probably forcing survey efforts to be directed toward the high hanging fruit, ignoring the low hanging fruit to a degree. This would eventually lead to smaller sample sizes. In addition, the cost argument is an important consideration. Increasing the number of completed interviews is no guarantee that the power of an obtained respondent sample is also linearly improved. This means that smaller samples may be of slightly lower quality, while the costs may be heavily reduced.

Discussion: Cracks in the nonresponse paradigm?

The first chapter of this dissertation focuses on the measurement and assessment of the negative impact of nonresponse on surveys, and includes nonresponse adjustment techniques. Not only might nonresponse result in factual bias with regard to auxiliary variables, nonresponse is also very likely to infect target statistics. However, as such biases cannot be observed, they should not be seen as a fact, but instead as a risk, so that additional uncertainty and thus variance inflation should be allowed.

The awareness that surveys are susceptible to error such as nonresponse gives rise to the belief that the probabilistic starting point is too sterile to support the credibility of surveys as inferential tools for scientific research. Strictly speaking, whenever uncertainties that cannot be dealt with probabilistically enter the data, inferential statistics such as p-values or confidence intervals can no longer be validly warranted. Typically, nonresponse error is one such type of error, the uncertainties of which are unknown and cannot be taken into account in the inferential framework. Evidently, other hard to measure sources of survey error corrode even further the scientific claims for which survey research is conceived. Indeed, not only is nonresponse a prominent source of survey error, but measurement error may also be an even tougher problem than nonresponse, particularly because of the lack of true external references with which survey statistics can be compared. How can the obtained distribution of a survey question such as 'Which political party would you vote for if there were elections today?' be compared with its population counterpart?

In the absence of 'true' references with which survey outcomes can be compared, the only way to assure the quality of a survey is to guarantee an immaculate production process. However, even if all recipes for the construction for a perfect survey sample were known and understood, their practical implementation might still leave much to be desired.

In this respect, the second chapter primarily discusses the detrimental effect of response rate maximization on the quality of the obtained respondent sample. This strategy encourages survey agents such as interviewers to follow the path of least resistance. High response propensity cases will therefore be prioritized, endangering the ideal that all sample cases should be given an equal probability of being included in the respondent set that is eventually obtained. Notwithstanding the growing skepticism about the current practices adhered by survey researchers, surveys are currently still expected to be built on large sample sizes, of which the response rate needs to be maximized.

Despite the widespread agreement that survey data is contaminated by diverse sources of bias or error, survey researchers still seem to utilize the full (naive) power of a survey, as determined by the sample size. This suggests that the survey community is living beyond its means and should, from the perspective of professional ethics, be prepared to pay the price for the additional uncertainties originating from non-sampling survey error. However, it is very difficult, if not impossible, to set a fair price for the insurance of survey statistics.

As a consequence, survey research downgrades from an inferential or confirmatory status to an exploratory or a somewhat illustrative one. It is therefore rather remarkable, or even not understandable, that so much (monetary) effort is invested in a scientific tool that is only able to take a very blurred picture of society. Therefore, surveyors need to decide whether survey costs should be reduced or survey quality should be improved.

An enduring problem when conducting nonresponse research is that the data needed to carry out the analyses (process data and auxiliary variables) may be imperfect or even biased itself. Auxiliary variables that are available for both respondents and nonrespondents (or full-sample or pop-

ulation totals), are by no means able to completely reveal all nonresponse mechanisms. Neither do they undoubtedly represent real target data with which to assess the consequences of nonresponse. Process data, a relatively new and still unfamiliar source of data for monitoring the construction of target survey data, is also not uncontested with regard to completeness, or may even be deliberately manipulated by interviewers or fieldwork administrators in order to meet pre-defined fieldwork targets. Admittedly, such considerations obscure the view of how nonresponse really takes control of survey quality. Consider in this respect a survey for which some auxiliary information is available, informative for the degree of nonresponse bias. However, as there is uncertainty about the auxiliary variables being representative of the target variables, even more uncertainty may eventually be passed on to the target variables. Further, process data is not free from error: (non)response outcome categories may be too raw or may not completely correspond to the actual course of contact events. This hinders the monitoring, evaluation, and improvement of fieldwork procedures. Therefore, it is advisable to consider more extensive investment in the quality measurement capabilities of surveys. These requirements for high-quality paradata may comprise a more diverse set of auxiliary variables in order to cover more dimensions of the nonresponse mechanism and provide more accurate and validated records of the contact and doorstep events.

In sum, how should survey researchers deal with nonresponse?

- Try to scrutinize nonresponse from many different perspectives and indicators. The distinction between the fixed and the random response model is advisable, as both try to perceive nonresponse from completely different theoretical angles and deploy a variety of indicators to express the impact of nonresponse. It is suggested that survey researchers use not only the response rate, R-indicator, maximal absolute bias, and contrast under the random model, but also the possibly more confronting indicators under the fixed response model, such as the average bias, variance inflation factor ψ, or the expected type I and type II error.

- Survey researchers should abandon the *idée fixe* of obtaining as high as possible a response rate. Innovative strategies should be further developed that predominantly focus on the reduction of risk of bias or on the maximization of strong representativeness. This may eventually lead to smaller but better samples.

- Most of all, survey researchers should acknowledge the fragile limits of a survey as a scientific tool. Although it may be very hard to prove survey findings wrong as there simply are no objective references (unless there is another relevant survey), researchers may be tempted to live beyond their means. The survey scene should therefore instead accept the statistical flaws of a survey.

References

AAPOR. (2010). *AAPOR Code of Professional Ethics and Practices.* Available from `http://www.aapor.org/AAPOR_Code_of_Ethics/4249.htm`

AAPOR. (2011, May). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys.* Available from `http://www.aapor.org/Standard_Definitions/1818.htm`

Aitken, A., Hörngren, J., Jones, N., Lewis, D., & Zilhão, M. (2004). *Handbook on Improving Quality by Analysis of Process Variables.* Luxemburg: Eurostat.

Andridge, R. A., & Little, R. J. (2011). Proxy Pattern-Mixture Analysis for Survey Nonresponse. *Journal of Official Statistics*, *27*(2), 153–180.

Assael, H., & Keon, J. (1982). Nonsampling vs. Sampling Errors in Survey Research. *Journal of Marketing*, *46*(2), 114–123.

Atrostic, B., Bates, N., Burt, G., & Silberstein, A. (2001). Nonresponse in U.S. Government Household Surveys: Consistent Measures, Recent Trends, and New Insights. *Journal of Official Statistics*, *17*(2), 209–226.

Babbie, E. (1975). *The Practice of Social Research.* Belmont: Wadsworth.

Babbie, E. (1995). *The Practice of Social Research* (7th ed.). Belmont: Wadsworth.

Babbie, E. (2009). *The Practice of Social Research* (12th ed.). Belmont: Wadsworth.

Bailey, K. (1987). *Methods of Social Research* (3rd ed.). New York: Free Press.

Bates, N., Dahlhamer, J., Phipps, P., Safir, A., & Tan, L. (2010). Assessing Contact History Data Quality and Consistency Across Several Federal Surveys. In *Proceedings of the Survey Research Methods Section.* American Statistical Association.

Bethlehem, J. (1988). Reduction of Nonresponse Bias Through Regression Estimation. *Journal of Official Statistics*, *4*(3), 251–260.

Bethlehem, J. (2009). *Applied Survey Methods. A Statistical Perspective.* Hoboken (N.J.): Wiley.

Bethlehem, J., Cobben, F., & Schouten, B. (2011). *Handbook of Nonresponse in Household Surveys.* Hoboken (N.J.): John Wiley & sons.

Bethlehem, J., & Kersten, H. (1985). On the Treatment of Nonresponse in Sample Surveys. *Journal of Official Statistics*, *1*(3), 287–300.

Beullens, K., Billiet, J., & Loosveldt, G. (2010). The Effect of the Elapsed Time between the Initial Refusal and Conversion Contact on Conversion Success: Evidence from the 2nd Round of the European Social Survey. *Quality & Quantity*, *44*(6), 1053–1065.

Beullens, K., & Loosveldt, G. (2012). Should High Response Rates Really Be a Primary Objective. *Survey Practice*, *5*(3), 1–5.

Biemer, P. (2010). Nonresponse Bias and Measurement Bias in a Comparison of Face to Face and Telephone Interviewing. *Public Opinion Quarterly*, *75*(4), 817–848.

Biemer, P., Chen, P., & Wang, K. (2012). Using Level-of-Effort Paradata in Non-response Adjustments with Application to Field Surveys. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *176*, 147–168.

Biemer, P., & Link, M. W. (2007). Evaluating and Modelling Early Cooperator Effects in RDD Surveys. In J. Lepkowski et al. (Eds.), *Advances in Telephone Survey Methodology* (pp. 587–618). New York (N.Y.): Wiley & Sons.

Biemer, P., & Lyberg, L. (2003). *Introduction to Survey Quality.* Hoboken (N.J.): Wiley.

Bolstein, R. (1991). Comparison of the Likelihood to Vote among Preelection Poll Respondents and Nonrespondents. *Public Opinion Quar-*

terly, *55*(4), 648–650.

Brick, J., Lê, T., & West, J. (2003). Dealing with Movers in a Longitudinal Study of Children. In *Proceedings of Statistics Canada Symposium 2003 Challenges in Survey Taking for the Next Decade.* Statistics Canada.

Brick, J., & Williams, D. (2013). Explaining Rising Nonresponse Rates in Cross-Sectional Surveys. *The ANNALS of the American Academy of Political and Social Science*, *645*(1), 36–59.

Carton, A. (2008). *Are our Paradata Always of Good Quality.* (Paper Presented at 19th International Workshop on Household Survey Nonresponse, Ljubljana, Slovenia, September 15-17)

Copas, A., & Farewell, V. (1998). Dealing with Non-ignorable Nonresponse by Using an 'Enthusiasm-to-Respond' Variable. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *161*(3), 385–397.

Couper, M. (1998). Measuring Survey Quality in a CASIC Environment. In *Proceedings of the Survey Research Methods Section* (pp. 41–49). American Statistical Association.

Couper, M., & Wagner, J. (2011). Using Paradata and Responsive Design to Manage Survey Non-response. In *Invited Paper Presented to the World Statistics Congress of the International Statistical Institute Conference.*

Curtin, R., Presser, S., & Singer, E. (2005). Changes in Telephone Survey Nonresponse over the Past Quarter Century. *Public Opinion Quarterly*, *69*(1), 87–98.

Dahlhamer, J., & Jans, M. (2011). *With an Eye Toward Responsive Design: The Development of Response Propensity Models with the National Health Interview Survey.* (Paper Presented at 22nd International Workshop on Household Survey Nonresponse, Bilbao, Spain, September 5-7)

Dalenius, T. (1983). Some Reflections on the Problem of Missing Data. In I. Olkin & W. Madow (Eds.), *Incomplete Data in Sample Surveys* (Vol. 3, pp. 411–413). New York (N.Y.): Academic Press.

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

Deding, M., Fridberg, T., & Jacobsen, V. (2008). Non-response in a Survey among Immigrants in Denmark. *Survey Research Methods*, *2*(3), 107–121.

Deville, J.-C., & Särndal, C.-E. (1992). Calibration Estimators in Survey Sampling. *Journal of the American Statistical Association*, *87*, 376–382.

Deville, J.-C., Särndal, C.-E., & Sautory, O. (1992). Generalized Raking Procedures in Survey Sampling. *Journal of the American Statistical Association*, *88*, 1013–1020.

Dey, E. (1997). Working with Low Survey Response Rates: The Efficacy of Weighting Adjustments. *Research in Higher Education*, *38*(2), 215–227.

Dillman, D. (1978). *Mail and Telephone Surveys. The Total Design Method*. New York (N.Y.): Wiley.

Dillman, D. (2000). *Mail and Internet Surveys. The Tailored Design Method*. New York (N.Y.): Wiley.

Dippo, C. (1997). Survey Measurement and Process Improvement: Concepts and Integration. In L. Lyberg et al. (Eds.), *Measurement and Process Quality* (pp. 457–474). New York (N.Y.): Wiley.

Drew, J., & Fuller, W. (1980). Modelling Nonresponse in Survey with Callbacks. In *Proceedings of the Survey Research Methods Section* (pp. 639–642). American Statistical Association.

Durrant, G., & Kreuter, F. (2013). Editorial: The Use of Paradata in Social Survey Research. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *176*(1), 1–3.

Edwards, S., Martin, D., DiSogra, C., & Grant, D. (2004). Altering the Hold Period for Refusal Conversion Cases in an RDD Survey. In *Proceedings of the Survey Research Methods Section* (p. 3440-3443). American Statistical Association.

Eltinge, J. (2002). Diagnostics for the Practical Effects of Nonresponse Adjustment Methods. In R. Groves, D. Dillman, J. Eltinge, & R. J. Little (Eds.), *Survey Nonresponse* (pp. 431–443). New York (N.Y.): Wiley.

Feskens, R., Hox, J., Lensvelt-Mulders, G., & Smeets, H. (2006). Collecting Data among Ethnic Minorities in an International Perspective. *Field Methods*, *18*(3), 284–304.

Fontaine, S. (2012). *Specific Mixed-Mode Methodology to Reach Sensory Disabled People in Quantitative Surveys.* (Paper Presented at the International Conference on Methods for Surveying and Enumerating Hard-to-Reach Populations, New Orleans, October 31 - November 3)

Fowler, F. (1985). *Survey Research Methods.* Beverly Hills (Calif.): Sage.

Fowler, F. (1993). *Survey Research Methods* (2nd ed.). Beverly Hills (Calif.): Sage.

Fowler, F. (2002). *Survey Research Methods* (3rd ed.). Thousand Oaks (Calif.): Sage.

Fowler, F. (2009). *Survey Research Methods* (4th ed.). Los Angeles: Sage.

Groves, R. (2004). *Survey Errors and Survey Costs.* Hoboken, (N.J.): Wiley.

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

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

Groves, R., Couper, M., Presser, S., Singer, E., Tourangeau, R., Acosta, G., et al. (2006). Experiments in Producing Nonresponse Bias. *Public Opinion Quarterly*, *70*(5), 720–736.

Groves, R., Fowler, F., Couper, M., Lepkowski, J., Singer, E., & Tourangeau, R. (2009). *Survey Methodology* (Second ed.). Hoboken: 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 (Statistics in Society)*, *169*(3), 439–457.

Groves, R., & Lyberg, L. (2010). Total Survey Error. Past, Present and Future. *Public Opinion Quarterly*, *74*(5), 849–879.

Groves, R., & Peytcheva, E. (2008). The Impact of Non-response Rates on Non-response Bias: a Meta-Analysis. *Public Opinion Quarterly*, *72*(2), 167–189.

Groves, R., Presser, S., & Dipko, S. (2004). The Role of Topic Interest in Survey Participation Decisions. *Public Opinion Quarterly*, *68*(2), 299–308.

Groves, R., Singer, E., & Corning, A. (2000). Leverage-Saliency Theory of Survey Participation: Description and an Illustration. *Public Opinion Quarterly*, *64*(3), 299–308.

Groves, R., Wagner, J., & Peytcheva, E. (2007). Use of Interviewer Judgments About Attributes of Selected Respondents in Post-Survey Adjustment for Unit Nonresponse: An Illustration with the National Survey of Family Growth. In *Proceedings of the Survey Research Methods Section* (p. 3428-3431). American Statistical Association.

Hansen, M., & Hurwitz, W. (1946). The Problem of Nonresponse in Sample Surveys. *Journal of the American Statistical Association*, *41*, 517–529.

Holt, D., & Smith, T. (1979). Post Stratification. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *142*(1), 33–46.

Horvitz, D., & Thompson, D. (1952). A Generalization of Sampling Without Replacement from a Finite Universe. *Journal of the American Statistical Association*, *47*, 663–685.

Johnson, T., Young, I., Campbell, R., & Holbrook, A. (2006). Using Community-level Correlates to Evaluate Nonresponse Effects in a Telephone Survey. *Public Opinion Quarterly*, *70*(5), 704–719.

Kalsbeek, W. (1979). A Conceptual Review of Survey Error Due to Nonresponse. In *Proceedings of the Survey Research Methods Section* (pp. 131–163). American Statistical Association.

Kalton, G., & Flores-Cervantes, I. (2003). Weighting Methods. *Journal of Official Statistics*, *19*(2), 81–97.

Kaminska, O., Billiet, J., & McCutcheon, A. (2010). Satisficing Among Reluctant Respondents in a Cross-National Context. *Public Opinion Quarterly*, *74*(5), 956–984.

Kaminska, O., & Lynn, P. (2011). *Interviewer Observation: Are the Observations What we Want or What Interviewers Think we Want?* (Paper Presented at 22nd International Workshop on Household Survey Nonresponse, Bilbao, Spain, September 5-7)

Kersten, H., & Bethlehem, J. (1984). Exploring and Reducing the Nonresponse Bias by Asking the Basic Question. *Statistical Journal of the U.N. Economic Commission for Europe*, *2*, 369–380.

Kish, L. (1965). *Survey Sampling*. New York: Wiley.

Kish, L. (1992). Weighting for Unequal P_i. *Journal of Official Statistics*, *8*(2), 183–200.

Koch, A., Fitzgerald, R., Stoop, I., Widdop, S., & Halbherr, V. (2012). *Field Procedures in the European Social Survey Round 6: Enhancing Response Rates* (Tech. Rep.). Mannheim: European Social Survey, GESIS.

Kohler, U. (2007). Surveys from Inside: An Assessment of Unit Nonresponse Bias with Internal Criteria. *Survey Research Methods*, *1*(2), 55–67.

Kreuter, F., & Casas-Cordero, C. (2010). *Paradata, Working Paper 136* (Tech. Rep.). German Council for Social and Economic Data Working Paper Series, Berlin, Germany.

Kreuter, F., & Kohler, U. (2009). Analyzing Contact Sequences in Call Record Data: Potential and Limitation of Sequence Indicators for Nonresponse Adjustment in the European Social Survey. *Journal of Official Statistics*, *25*(2), 203–226.

Kreuter, F., Olson, K., Wagner, J., Yan, T., Ezzati-Rice, T., Casas-Cordero, C., et al. (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 (Statistics in Society)*, *173*(2), 389–410.

Krosnick, J. (1999). Survey Research. *Annual Review of Psychology*, *50*, 537–567.

Kruskal, W., & Mosteller, F. (1979a). Representative Sampling, 1: Non-scientific Literature. *International Statistical Review*, *47*(1), 13–24.

Kruskal, W., & Mosteller, F. (1979b). Representative sampling, 2: Scientific Literature, Excluding Statistics. *International Statistical Review*, *47*(2), 111–127.

Kruskal, W., & Mosteller, F. (1979c). Representative Sampling, 3: The Current Statistical Literature. *International Statistical Review*, *47*(3), 245–265.

Laaksonen, S. (2006). Does the Choice of Link Function Matter in Response Propensity Modelling? *Model Assisted Statistics and Applications*, *1*(2), 95–100.

Laganá, F., Elcheroth, G., Penic, S., Kleiner, B., & Fasel, N. (2011). National Minorities and Their Representation in Social Surveys: Which Practices Make a Difference? *Quality & Quantity*, *47*(3), 1287–1314.

Langer, G. (2003). About Response Rates. Some Unresolved Questions. *Public Perspective*, *May/June 2003*, 16–18.

Lepkowski, J., & Couper, M. (2002). Nonresponse in the Second Wave of Longitudinal Household Surveys. In R. Groves, D. Dillman, J. Eltinge, & R. J. Little (Eds.), *Survey Nonresponse* (pp. 259–272). New York (N.Y.): Wiley.

Lessler, J., & Kalsbeek, W. (1992). *Nonsampling Error in Surveys.* New York (N.Y.): Wiley.

Lin, F., & Schaeffer, N. (1995). Using Survey Participation to Estimate the Impact of Nonparticipation. *Public Opinion Quarterly*, *59*(2), 236–258.

Little, R. J., , Heeringa, S., Lepkowski, J., & Kessler, R. (1997). Assessment of Weighting Methodology for the National Comorbidity Survey. *American Journal of Epidemiology*, *146*(5), 439–449.

Little, R. J., & Rubin, D. (1987). *Statistical Analysis with Missing Data.* New York (N.Y.): Wiley.

Little, R. J., & Rubin, D. (2002). *Statistical Analysis with Missing Data* (2nd ed.). London: Wiley.

Little, R. J., & Vartivarian, S. (2003). On Weighting the Rates in Non-response Weights. *Statistics in Medicine*, *22*(9), 1589–1599.

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

Lohr, S. (1999). *Sampling: Design and Analysis*. Duxbury: Pacific Grove.

Lyberg, L., & Biemer, P. (2008). Quality Assurance and Quality Control in Surveys. In E. de Leeuw, J. Hox, & D. Dillman (Eds.), *International Handbook of Survey Methodology* (pp. 421–441). New York (N.Y.): Erlbaum.

Lynn, P. (2003). PEDASKI: Methodology for Collecting Data about Survey Non-respondents. *Quality & Quantity*, *37*, 239–261.

Lynn, P. (2008). The Problem of Nonresponse. In E. de Leeuw, J. Hox, & D. Dillman (Eds.), *International Handbook of Survey Methodology* (pp. 35–55). New York (N.Y.): Erlbaum.

Maitland, A. and, & Bianchi, S. (2006). Nonresponse in the American Time Use Survey: Who is Missing from the Data and How Much Does it Matter? *Public Opinion Quarterly*, *70*(5), 676–703.

Martikainen, P., Laaksonen, M., Piha, K., & Lallukka, T. (2007). Does Survey Non-response Bias the Association between Occupational Social Class and Health? *Scandinavian Journal of Public Health*, *35*(2), 212–215.

Martin, E. (2004). Presidential Address. Unfinished Business. *Public Opinion Quarterly*, *68*(3), 439–450.

Matsuo, H., Billiet, J., Loosveldt, G., Berglund, F., & Kleven, Ø. (2010). Measurement and Adjustment of Non-response Bias Based on Non-response Surveys: The Case of Belgium and Norway in the European Social Survey. *Survey Research Methods*, *4*(3), 165–178.

Matsuo, H., & Loosveldt, G. (2012). *Data Quality in Interviewer Observation Data: Interviewer Burden Perspective from ESS Round 5.* (Paper Presented at the International Conference on European Social

Survey, Nicosia (Cyprus), November 23-25)

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

Merton, R.-K. (1940). Bureaucratic Structure and Personality. *Social Forces*, *18*(4), 560–568.

Mitchell, S., Ciemnecki, A., CyBulski, K., & Markesich, J. (2006). *Removing Barriers to Survey Participation for Persons with Disabilities* (Tech. Rep.). Rehabilitation Research and Training Center on Disability Demographics and Statistics, Cornell University, Ithaca, NY.

Morganstein, D., & Marker, D. (1997). Continuous Quality Improvement in Statistical Agencies. In L. Lyberg et al. (Eds.), *Measurement and Process Quality* (pp. 475–500). New York (N.Y.): Wiley.

Olson, K. (2013a). Do Non-response Follow-ups Improve or Reduce Data Quality?: A Review of the Existing Literature. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *176*(1), 129–145.

Olson, K. (2013b). Where Do We Go from Here? Nonresponse and Social Measurement. *The ANNALS of the American Academy of Political and Social Science*, *645*(1), 222–236.

OMB. (2006). *Standards and Guidelines for Statistical Surveys.* Available from `http://www.whitehouse.gov/omb/inforeg_statpolicy`

Peytchev, A. (2013). Consequences of Survey Nonresponse. *The ANNALS of the American Academy of Political and Social Science*, *645*(1), 88–111.

Peytchev, A., Baxter, R., & Carley-Baxter, L. (2009). Not All Survey Effort is Equal. Reduction of Nonresponse Bias and Nonresponse Error. *Public Opinion Quarterly*, *73*(3), 785–806.

Peytchev, A., & Biemer, P. (2011). *A Standardized Indicator on Survey Nonresponse Bias Based on Effect Size.* (Paper Presented at 22nd International Workshop on Household Survey Nonresponse, Bilbao, Spain, September 5-7)

Peytchev, A., Carley-Baxter, L., & Black, M. (2011). Multiple Sources of Nonobservation Error in Telephone Surveys: Coverage and Nonres-

ponse. *Sociological Methods and Research*, *40*(1), 138–168.

Peytchev, A., Riley, S., Rosen, J., Murphy, J., & Lindblad, M. (2010). Reduction of Nonresponse Bias in Surveys Through Case Prioritization. *Survey Research Methods*, *4*(1), 21–29.

Peytcheva, E., & Groves, R. (2009). Using Variation in Response Rates of Demographic Subgroups as Evidence of Nonresponse Bias in Survey Estimates. *Journal of Official Statistics*, *25*(2), 193–201.

Potter, F. (1993). The Effect of Weight Trimming on Nonlinear Survey Estimates. In *Proceedings of the Survey Research Methods Section* (p. 758-763). American Statistical Association.

Rogers, A., Murtaugh, M., Edwards, S., & Slattery, M. (2004). Contacting Controls: Are we Working Harder for Similar Response Rates, and Does it Make a Difference? *American Journal of Epidemiology*, *160*(1), 85–90.

Rosenbaum, P., & Rubin, D. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*, *70*(1), 41-55.

Särndal, C.-E. (2010). *The Probability Sampling Tradition in a Period of Crisis.* (Keynote Speech at the Q2010 European Conference on Quality in Official Statistics, Helsinki, May 4-6)

Särndal, C.-E., & Lundström, S. (2006). *Estimation in Surveys with Nonresponse.* Chichester: Wiley.

Särndal, C.-E., & Lundström, S. (2008). Assessing Auxiliary Vectors for Control of Nonresponse Bias in the Calibration Estimator. *Journal of Official Statistics*, *24*(2), 167–191.

Saßenroth, D. (2010). *Understanding Survey Participation in the Context of Personality: A Theoretical Framework.* (Paper Presented at 21st International Workshop on Household Survey Nonresponse, Nürnberg, Germany, August 30 - September 1)

Schouten, B., Bethlehem, J., Beullens, K., Kleven, Ø., Loosveldt, G., Luiten, A., et al. (2012). Evaluating, Comparing, Monitoring, and Improving Representativeness of Survey Response Through R-indicators and Partial R-indicators. *International Statistical Review*,

80(3), 382-399.

Schouten, B., Calinescu, M., & Luiten, A. (2011). *Optimizing Quality of Response Through Adaptive Survey Designs* (Tech. Rep.). Statistics Netherlands.

Schouten, B., Cobben, F., & Bethlehem, J. (2009). Indicators for the Representativeness of Survey Response. *Survey Methodology*, *35*, 101–113.

Shlomo, N., Skinner, C., Schouten, B., Bethlehem, J., & Zhang, L. (2008). *Statistical Properties of R-indicators, RISQ Deliverable 2.1, available at www.r-indicator.eu* (Tech. Rep.).

Shlomo, N., Skinner, C., Schouten, B., Carolina, B., & Morren, M. (2009). *Partial Indicators for Representative Response, RISQ Deliverable 4.2, available at www.r-indicator.eu* (Tech. Rep.).

Singer, E. (2006). Introduction: Nonresponse Bias in Household Surveys. *Public Opinion Quarterly*, *70*(5), 637–645.

Sinibaldi, J., Kreuter, F., & Durrant, G. (2011). *Evaluating the Measurement Error of Interviewer Observed Paradata.* (Paper Presented at 22nd International Workshop on Household Survey Nonresponse, Bilbao, Spain, September 5-7)

Skinner, C. (1991). On the Efficiency of Raking Ratio Estimates for Multiple Frame Surveys. *Journal of the American Statistical Association*, *86*, 779–784.

Skinner, C. (1996). Introduction to Part A. In C. Skinner, D. Holt, & T. Smith (Eds.), *Analysis of Complex Surveys* (pp. 23–58). New York: Wiley.

Smith, T. (1984). Estimating Nonresponse Bias with Temporary Refusals. *Sociological Perspectives*, *27*(4), 473–489.

Squire, P. (1988). Why the 1936 Literary Digest Poll Failed. *Public Opinion Quarterly*, *52*(1), 125-133.

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

Stoop, I., Billiet, J., Koch, A., & Fitzgerald, R. (2010). *Improving Survey Response. Lessons Learned from the European Social Survey.* Chichester: Wiley.

Triplett, T. (2002). *What is Gained From Additional Call Attempts and Refusal. Conversion and What are the Cost Implications* (Tech. Rep.). Washington D.C.: Urban Institute.

Triplett, T., Scheib, J., & Blair, T. (2001). How Long Should you Wait Before Attempting to Convert a Telephone Refusal? In *Proceedings of the Survey Research Methods Section* (p. 5-9). American Statistical Association.

Vandecasteele, L., & Debels, A. (2007). Attrition in Panel Data: The Effectiveness of Weighting. *European Sociological Review*, *23*(1), 81–97.

Vehovar, V. (2007). Non-response Bias in the European Social Survey. In G. Loosveldt, M. Swyngedouw, & B. Cambré (Eds.), *Measuring Meaningful Data in Social Research* (pp. 335–356). Leuven: Acco.

Vercruyssen, A., Roose, H., & Van de Putte, B. (2011). Underestimating Busyness : Indications of Nonresponse Bias Due to Work-Family Conflict and Time Pressure. *Social Science Research*, *40*(6), 1691–1701.

Vercruyssen, A., Van de Putte, B., & Stoop, I. (2011). Are they Really too Busy for Survey Participation? The Evolution of Busyness and Busyness Claims in Flanders. *Journal of Official Statistics*, *27*(4), 619–632.

Wagner, J. (2010). The Fraction of Missing Information as a Tool for Monitoring the Quality of Survey Data. *Public Opinion Quarterly*, *74*(2), 223–243.

Wang, K., & Biemer, P. (2010). The Accuracy of Interview Paradata: Results from a Field Investigation. In *Paper Presented at the Annual Meeting of the American Association for Public Opinion Research, Chicago, IL (May).*

West, B. (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 (Statistics in Society), *176*(1), 211–225.

Abstract - Samenvatting - Résumé

Abstract

This dissertation looks at survey nonresponse both from the perspective of the fieldwork process (second chapter) as well as from the output perspective (first chapter). In the output perspective, nonresponse is already a fact and survey researchers need to find ways to measure the effects of nonresponse and should adequately deal with this issue. In the process perspective, specific attention is given to the fieldwork objective and how it affects the quality of the obtained respondent sample.

In order to measure the effects of nonresponse, the researcher can use both the fixed as well as the random response model. In the fixed response model, there are basically two main strata in the population (respondents and nonrespondents) between which differences can be observed with respect to statistics of interest (that are usually unknown) or auxiliary variables (known, but not always relevant). These differences, also termed contrasts, can be transformed in an estimate of bias, expressing the distance between the respondent-only statistic and its full-sample counterpart. If there are many auxiliary variables available for both respondents and nonrespondents, an aggregated distribution of biases can be assumed, by which statistics of interest may also be affected. The random response model considers nonresponse as an individual latent probability or propensity of positively responding to a survey request. If these propensities can

be determined reasonably well, it is possible to estimate how nonresponse affects target statistics.

As nonresponse cannot be observed by definition, it is important to use as many methods and perspectives as possible in order to get some grip on the effects of nonresponse. Therefore assumptions and their implied uncertainties should be taken into account when estimating survey statistics in the presence of nonresponse. In particular, the fixed response model forces the researcher into a defensive attitude, since bias due to nonresponse can be easily found. Bias, in this perspective, is a relatively frequently occurring event, instead of an exception. This means that making inferences from a survey without taking nonresponse into consideration may lead to a substantive increase of the risk of making a type I error.

The second chapter suggests that part of the unfavorable effects of nonresponse may be reduced by altering the process through which the sample of respondents is obtained. Still, the widely held view on response is to maximize the response rate. And although some survey researchers support the idea of so called 'balanced response rates', making equal response rates with regard to a predefined set of auxiliary variables, the low hanging fruit still seems to be prioritized. A simulation study suggests that the maximization of response rates has a bias-creating effect instead of a bias-combating effect. Instead of following the line of least resistance, survey fieldwork should be more focused on the high hanging fruit, even if this implies more efforts per completed interview. Alternatively, instead of reverting the line of least resistance, smaller sample sizes may be chosen, lowering the cost of a survey without drastically affecting the (already bad) quality of the obtained respondent set. Anyway, the option of smaller sample sizes seems to be an inevitable choice if one is interested in improving survey quality or reducing survey costs.

Samenvatting

Dit proefschrift gaat in op nonrespons in survey onderzoek zowel vanuit het perspectief van het veldwerk proces (tweede hoofdstuk) alsook vanuit

het output perspectief (eerste hoofdstuk). Bij het output perspectief is nonrespons reeds een feit waardoor onderzoekers manieren moeten vinden om de effecten van nonrespons te meten en er gepast mee om te gaan. In het proces perspectief wordt vooral aandacht besteed aan de veldwerk doelstellingen en de gevolgen ervan op de kwaliteit van de verkregen steekproef van respondenten.

Om de effecten van nonrespons te meten kan de onderzoeker zowel het *fixed response model* als het *random response model* volgen. In het fixed response model, gaat men uit van twee strata in de bevolking (respondenten en niet-respondenten) waartussen verschillen kunnen waargenomen worden met betrekking tot de statistieken waarvoor het onderzoek is opgezet (deze zijn meestal onbekend) of zogenaamde *auxiliary* variabelen (bekend, maar niet altijd relevant). Deze verschillen, ook wel contrasten genoemd, kunnen worden omgezet naar een schatting van *bias* of vertekening, of de afstand tussen de respondenten en de volledige steekproef. Als er veel hulpvariabelen beschikbaar zijn voor zowel de respondenten en nonrespondenten, kan een geaggregeerde verdeling van de vertekeningen worden opgesteld, waarvan men verondersteld dat ze de doelvariabelen op een gelijkaardige manieren aantast. Het random response model beschouwt nonrespons als een individuele kans of *propensity* om positief te reageren op een survey verzoek. Als deze propensities redelijk goed worden bepaald, is het mogelijk te achterhalen hoe nonrespons de surveyresultaten beïnvloedt.

Aangezien nonrespons per definitie niet kan worden waargenomen, is het belangrijk om zoveel methoden en perspectieven als mogelijk te gebruiken. Daarom zijn assumpties en hun geïmpliceerde onzekerheden noodzakelijk en dient met hiermee rekening te houden bij het schatten van parameters van surveys. Met name het fixed response model dwingt de onderzoeker tot een defensieve houding. Vertekening, in dit perspectief, is een relatief veel voorkomende gebeurtenis, in plaats van een uitzondering. Dit betekent dat besluiten trekken op basis van surveydata zonder rekening te houden met nonrespons tot een substantiële verhoging van het risico op een type I fout zal leiden.

In het tweede hoofdstuk wordt nagegaan of een deel van de ongunstige effecten van nonrespons kan worden verminderd door het veranderen van het veldwerkproces. Het is een gangbare opvatting om de responsgraad te maximaliseren, eventueel aangepast volgens het principe van 'gebalanceerde respons', waarbij gelijke responsgraden worden nagestreefd tussen verschillende strata. Echter, het veldwerkproces blijkt nog steeds de weg van de minste weerstand te bewandelen en geeft dus systematisch voorrang aan de *low hanging fruit*.

Simulaties suggereren dat responsgraad-maximalisatie eerder vertekening creëert dan het te bestrijden. In plaats van het volgen van de weg van de minste weerstand, moet het veldwerk meer gericht zijn op het *high hanging fruit*, zelfs als dit meer inspanningen vraagt voor het binnenhalen van afgewerkte interviews. Als alternatief kan men er ook voor kiezen om de steekproefomvang te verkleinen. Dit verlaagt de kosten en heeft geen drastische vermindering van de (reeds lage) steekproefkwaliteit tot gevolg. Hoe dan ook, de optie van een kleinere steekproefomvang lijkt een onvermijdelijke keuze te zijn, hetzij door het verlagen van de kosten, hetzij door het verhogen van de kwaliteit.

Résumé

Cette dissertation porte sur la non-réponse d'enquête. Elle se concentre sur le processus de récolte des données sur le terrain (deuxième chapitre) ainsi que sur la qualité des données obtenues (premier chapitre). Dans ce dernier cas, la non-réponse est déjà un fait accompli. Les chercheurs doivent trouver le moyen de mesurer ses effets et les traiter de manière appropriée. En ce qui concerne la récolte des données sur le terrain, cette dissertation accorde une attention particulière à la façon dont l'objectif que l'on s'est fixé affecte ce travail ainsi que la qualité de l'échantillon obtenu. Afin de mesurer les effets de la non-réponse, les chercheurs peuvent utiliser à la fois le fixed ainsi que le random response model.

Dans le modèle fixe, deux groupes de la population sont essentiels: répondants et non-répondants. Entre ces deux groupes des différences

peuvent exister en ce qui concerne les variables qui ont de l'intérêt pour la recherche (ces differences ne sont pas nécessairement connues au préalable) ou en ce qui concerne les variables auxiliaires (qui ont donc moins d'intérêt pour la recherché mais dont les differences sont souvent connues au préalable). Ces différences, aussi appelées 'contrastes', peuvent être exprimées en une estimation la distance entre les statistiques des répondants et leurs homologues dans l'échantillon complet. S' il y a beaucoup de variables auxiliaires disponibles pour les répondants et les non-répondants, une distribution agrégée de biais peut être construite. On assume ensuite que les statistiques d'intérêt peuvent également être affectées de la même façon. Le modèle de réponse random part du principe que la non-réponse latente individuelle est question de probabilité. Elle depend donc de la propension de l'individu à vouloir participer à l'enquête. Si ces propensions peuvent être déterminées raisonnablement bien, il est possible d'estimer comment la non-réponse affecte les statistiques cibles.

Comme par définition la non-réponse ne peut pas être observée, il est important d'utiliser autant de méthodes et de perspectives que possible pour l'estimer. Par conséquent, les hypothèses et les incertitudes implicites doivent être prises en compte lors de l'estimation des statistiques de l'enquête. Le modèle de réponse fixe en particulier oblige le chercheur à adopter une attitude conservatrice, car un biais dû à la non-réponse se présente régulièrement avec ce genre de modèle. Le biais, dans cette perspective, est un événement relativement frequent plutôt qu'une exception. Cela signifie que les inférences tirées d'une enquête sans prendre en considération la non-réponse peuvent augmenter de manière significative le risque de faire une erreur de type I.

Le deuxième chapitre essaie de voir si les effets défavorables de la non-réponse peuvent être réduits en modifiant le processus par lequel l'échantillon de répondants est obtenu. Ce chapitre met en question l'opinion largement répandue sur la réponse qui consiste à maximiser le taux de réponse. Certains chercheurs appuient l'idée de la 'réponse équilibrée'. Il faut pour cela des taux de réponse égaux pour un ensemble prédéfini de variables auxiliaries. Le plus facile est alors de faire appèl à des individus

dont la probabilité de participer à l'enquête est élevée. Pourtant, une étude de simulation montre que cette maximisation des taux de réponse renforce le biais plutôt que de le diminuer.

Au lieu de suivre la ligne de la moindre résistance la collection des données devrait donc être axée sur les individus les plus difficiles à faire participer, même si cela demande plus d'efforts. Alternativement, si abandonner la ligne de la moindre résistance n'est pas possible, on peut choisir de travailler avec des échantillons plus petits. Ceci abaisse le coût de l'enquête sans pourtant radicalement affecter négativement la qualité de l'échantillon obtenu, qui de toute façon ne sera pas de qualité optimale. Quoi qu'il en soit, adherer ou abondonner la ligne de la moindre résistance, implique de travailler avec des échantillons de petite taille. Cela semble inévitable si l'on s'intéresse à l'amélioration de la qualité des enquêtes ou si l'on veut réduire les coûts de l'enquête.

Appendix

In this dissertation, the European Social Survey has primarily been used as an empirical reference. To a lesser extent, the Flemish Housing Survey and the Dutch dataset General Population Survey have also been deployed. These datasets are documented in this appendix.

The European Social Survey - ESS

The European Social Survey is a biennial multi-country survey. Its main goal is to screen and explain Europe's changing institutions, its political and economic structures, and the populations' beliefs, attitudes, and behavior. It is funded via the European Commission's 6^{th} Framework Programme, the European Science Foundation, and national funding bodies in each participating country. It involves strict random probability sampling, a minimal target response rate of 70%, and rigorous translation protocols. The noncontact rate should not exceed 3%. Survey topics include media, social trust, political interest and participation, socio-political orientations, social exclusion, national, ethnic and religious allegiances, timing of key life events and the life course, personal and social well-being and satisfaction with work and life, demographics, and socio economics. Currently, the sixth round of the ESS is being fielded and its datasets are expected to be released by the end of 2013. The first round (ESS1) was fielded in 2002-03. Subsequent rounds ESS2, ESS3, ESS4, ESS5, and ESS6 have been fielded every two years.

All participants have to be aged 15 or over and resident within private households, regardless of their nationality, citizenship, language, or

legal status, in the country of residence. Auxiliary information is not so widely available in the ESS, although it offers a large fund of contact data in order to assess the causes and effects of nonresponse. The contact sheets, datasets, and fieldwork documentation can be found at `http://ess.nsd.uib.no/`.

The next three pages are an example of the contact sheets that needed to be filled out by an interviewer during the fieldwork. This data is used to determine the response rates and follow whether the specific fieldwork requirements have been met, for example whether four noncontact outcomes have been realized before a sample unit can be considered as a final nonrespondent. For this dissertation, contact sheet data is of crucial importance as it monitors the actions that have been taken to contact and identify the (non)respondents, the outcome of each visit, the reasons for occasional refusal, the times and dates of the visits, the neighborhood information as observed by the interviewer, and the interviewer ID.

European Social Survey

Individual_Named (Round 5)

Type of sample:
Individual named

SAMPLE UNIT LABEL: PERSONAL

Respondent ID:

Respondent's name:

Respondent's telephone number:

…… refused no phone

Calls	Interviewer Number				
1 → ….					
… → …..					
… → …..					
… → …..					

VISIT RECORD *(Visit = every attempt made to reach the respondent/ household)*

Visit No.	1. Date dd/mm	2. Day of the week	3. Time 24 hr clock	4. Mode of visit 1 = personal visit 2 = telephone 3 =personal visit, but only intercom 4 = info through survey organisation 5 = other	5. RESULTS of the visit 1= Completed interview 2= Partial Interview 3 = Contact with someone, don't know if target respondent 4 = Contact with Target Respondent but NO interview 5 = Contact with somebody other than Target Respondent 6 = No contact at all 7 = Address is not valid (unoccupied, demolished, institutional,…) 8 = Other information about sample unit
1	/		:		
2	/		:		
3	/		:		
4	/		:		
5	/		:		
6	/		:		
7	/		:		
8	/		:		
9	/		:		
10	/		:		

Notes on time ….

If result of visit is code:
1,2,6→ Go to N1
3,4,5,8→ Go to 6 = OUTCOME CONTACT
7 → Go to 12 = OUTCOME ADDRESS INVALID

ESS DOCUMENT DATE 25/06/2010

6. OUTCOME CONTACT — **ONLY IF CONTACT but NO INTERVIEW**

		Visit 1	*Visit 2*	*Visit 3*	*Visit 4*	*Visit 5*	*Visit 6*	*Visit 7*	*Visit 8*	*Visit 9*	*Visit 10*
1. Appointment	→ N1	□1	□1	□1	□1	□1	□1	□1	□1	□1	□1
2. Refusal of respondent	→ 7	□2	□2	□2	□2	□2	□2	□2	□2	□2	□2
3. Refusal by proxy	→ 7	□3	□3	□3	□3	□3	□3	□3	□3	□3	□3
4. Refusal. Don't know if target respondent	→ 7	□4	□4	□4	□4	□4	□4	□4	□4	□4	□4
5. Respondent is unavailable/not at home until …./…….	→ N1	□5 …../……	□5 …../……	□5 …../……	□5 …../……	□5 …../……	□5 …../……	□5 …../……	□5 …../……	□5 …../……	□5 …../……
6. Respondent is mentally or physically unable to participate	→ N1	□6	□6	□6	□6	□6	□6	□6	□6	□6	□6
7. Respondent is deceased	→ END	□7	□7	□7	□7	□7	□7	□7	□7	□7	□7
8. Respondent has moved out of country	→ END	□8	□8	□8	□8	□8	□8	□8	□8	□8	□8
9. Respondent moved to unknown destination*	→ END	□9	□9	□9	□9	□9	□9	□9	□9	□9	□9
10. Respondent has moved, still in country	→ 13	□10	□10	□10	□10	□10	□10	□10	□10	□10	□10
11. Language Barrier	→ 6b	□11	□11	□11	□11	□11	□11	□11	□11	□11	□11
12. Other ……	→ N1	□12	□12	□12	□12	□12	□12	□12	□12	□12	□12

6b in case of language barrier: What is the language of the respondent? **……→ N1, p.5**

*Only use this category when interviewers really do not know whether the selected sampling unit has moved within or outside the country. Otherwise use codes 8 or 10.

IF REFUSAL (code 2, 3 or 4 at Q. 6)			
7. The refusal occurred at visit number *(write in)*	***VISIT …..***	***VISIT …..***	***VISIT …..***
8. REASON for REFUSAL? *(code all that apply)*			
1 Bad timing (e.g. sick, children,…) , otherwise engaged (e.g. visit)	□1	□1	□1
2 Not interested	□2	□2	□2
3 Don't know enough/anything about subject, too difficult for me	□3	□3	□3
4 Waste of time	□4	□4	□4
5 Waste of money	□5	□5	□5
6 Interferes with my privacy / I give no personal information	□6	□6	□6
7 Never do surveys	□7	□7	□7
8 Co-operated too often	□8	□8	□8
9 Do not trust surveys	□9	□9	□9
10 Previous bad experience	□10	□10	□10
11 Don't like subject	□11	□11	□11
12 R refuses because partner/family/HH gives no approval to co-operate	□12	□12	□12
13 Do not admit strangers to my house/afraid to let them in	□13	□13	□13
14. Other: ……	□14	□14	□14
9. Give your own estimation of the likely co-operation in the future of the selected respondent:			
1 will DEFINITELY NOT co-operate in the future	□1	□1	□1
2 will PROBABLY NOT co-operate in the future	□2	□2	□2
3 may PERHAPS co-operate in the future	□3	□3	□3
4 WILL co-operate in the future	□4	□4	□4
8 Don't know	□8	□8	□8

10. How old do you think the respondent (or the person who refused on their behalf) is?	***VISIT …..***	***VISIT …..***	***VISIT …..***
1 Under 20	□1	□1	□1
2 20 up to 39	□2	□2	□2
3 40 up to 59	□3	□3	□3
4 60 or more	□4	□4	□4
8 Don't know	□8	□8	□8
11. The respondent/contacted person is			
1 Male	□1	□1	□1
2 Female	□2	□2	□2
8 Don't know	□8	□8	□8

→ **Go to N1, p.5**

12. OUTCOME ADDRESS INVALID	**ONLY IF ADDRESS WAS NOT TRACEABLE, RESIDENTIAL OR OCCUPIED**
□ 1 Derelict or demolished house/ address	□ 5 Address is not residential: Institution (retirement home, hospital, military unit, monastery, …)
□ 2 Not yet built/ not yet ready for occupation	
□ 3 Address is not occupied (empty, second home, seasonal…)	□ 6 Address is not traceable, address was not sufficient
□ 4 Address is not residential: only business/ industrial purpose.	□ 7 Other (*please give details)* ……

→ **END**

ONLY IF R HAS MOVED and still in country

13. New Address

□ 1: The new address is:

Street: ……Number: …… Box: ……
City: …… City code: ……
State/ county: …… Country: …… → **Go to 14**

□ 2: Moved to an institution → **END**

14. Is this new address still in your interviewer-area?

□ 1 Yes → Skip N1, try to reach the respondent at this new address, fill in as next 'visit'

□ 2 No → **END**

NEIGHBOURHOOD CHARACTERISTICS FORM

- **ONE FORM TO BE COMPLETED FOR EACH ADDRESS**
- **COMPLETE DURING DAYLIGHT WHEREVER POSSIBLE**
- **MUST BE COMPLETED FOR ALL SAMPLE UNITS INCLUDING ALL NON CONTACTS, ALL REFUSALS, ALL OTHER TYPES OF NONRESPONSE UNITS AS WELL AS ALL INTERVIEWS**

N1. What type of house does the (target) respondent live in?

1 Farm

Single-unit:

2 Detached house
3 Semi-detached house
4 Terraced house
5 The only housing unit in a building with another purpose (commercial property)

Multi-unit:

6 Multi-unit house, flat
7 Student apartments, rooms
8 Retirement house

Other:

9 House-trailer or boat
10 Other (SPECIFY)……
88 Don't know

N2. Before reaching the (target) respondent's individual door, is there an entry phone system or locked gate / door?

INTERVIEWER: Record whether there is a gate / door that is locked at the time that the neighbourhood characteristics form is completed.

1. Yes – entry phone system
2. Yes – locked gate / door
3. Yes – entry phone system AND locked gate / door
4. No – neither of these

N3. What is your assessment of the overall physical condition of this building/house?

NOTE TO INTERVIEWER:

Consider the following issues when assessing the overall physical condition of this building/house.

1. Roof problems (e.g. sagging roof, missing roofing material)
2. Problems with windows (e.g. boarded up or broken windows)
3. Other problems (e.g. sloping outside walls, broken plaster or peeling paint, guttering problems)

1. Very good
2. Good
3. Satisfactory
4. Bad
5. Very bad

NOTE TO INTERVIEWER:

For the remaining two questions (N4 & N5) please give your overall opinion about the 'immediate vicinity' of the building/house of the target respondent. Look to the left and the right of the building/house taking into account a distance of about 2 normal sized houses on either side (approximately 15 metres on either side). Only include this area and the property of the target respondent when answering these questions.

There may not be other properties on either side of the building so just estimate the space that about 2 'normal' size houses on either side would take up.

Note that in the case of blocks of flats refer to the space on either side of the whole building and NOT just the individual flat where the target respondent lives.

N4. In the immediate vicinity, how much litter and rubbish is there?

1 Very large amount
2 Large amount
3 Small amount
4 None or almost none

N5. In the immediate vicinity, how much vandalism and graffiti is there?

1 Very large amount
2 Large amount
3 Small amount
4 None or almost none

The third round of the ESS in Belgium was fielded from October 23, 2006 until February 19, 2007, using TNS Dimarso as a subcontractor. Fieldworkers comprised 118 experienced interviewers working on a freelance basis, who were personally briefed about the ESS for half a day or less. They were paid per completed interview and all received some refusal conversion training. In Belgium, the basis is the commercial database of 'Orgassim'. Using the National Register, Orgassim has developed a database with 'Statistics of inhabitants per building'. With this database it is possible to construct an individual database including age, gender, and address for each person. Names are not available in this database. Then, the individual database is linked with another commercial database and 'enriched' with names (65% matches). A person is identified by his or her name or the combination of gender and age. Next, the gross sample is drawn from the frame. The Belgian sample is a result of a stratified two-stage probability sampling design. The ten provinces and Brussels are used for regional stratification. At stage 1, the primary sampling units (PSU's) are 'virtual' clusters located in municipalities, which means that the clusters within the municipalities are not further defined regionally. The number of clusters for each province is proportional to the size of the population in each province. For that, a list of municipalities with a population distribution (+15 years) for each province is used. The number of clusters in a municipality is proportional to the size of its population. The total number of clusters equals 338. At stage 2 in each of the 338 clusters, nine people are selected for the gross sample by simple random sampling, implying that the number of contacted people in each municipality equals the number of clusters in the municipality times nine.

From the register, some information about the sample cases is available such as *age* and *gender*. The age variable is divided into four classes: 14-20, 21-40, 41-60, and 60+. This classification is used, as these categories coincide with the age classes on the contact form (in some cases, interviewers needed to estimate the age of the (non)respondent). From the identification numbers, the Belgian province the sample case lives in can be derived. The eleven provinces are then reduced to the three constitutional

Categorization of three of the auxiliary variables

Population density in municipality (inh./km^2)	Percentage of non-Belgians in municipality in %	Average annual per capita income in municipality in €
1 <200	1 <2	1 <12000
2 201-400	2 2-5	2 12000-14000
3 401-700	3 5-15	3 14000-16000
4 701-2500	4 >15	4 >16000
5 >2501		

regions (Flanders, Wallonia, and Brussels). The register data (though not publicly available) also contains the postal codes of the sample units. It should be mentioned that this information provides access to a relatively large fund of external data. Postal codes can be linked to administrative or census data (these can be found at `http://statbel.fgov.be/`) such as the *population density* or the *percentage of non-Belgians living in the municipality*. The *average income of the municipalities* is also linked to the postal codes.

Interviewer-observed data is also available in the ESS. Whether the sample unit lives in an *apartment* is dichotomized in an auxiliary variable. Further, a composite index is constructed reflecting the *quality of the neighborhood*. This latter variable indicates to what extent the interviewer felt the neighborhood shows traces of vandalism, graffiti, litter, rubbish, deliberate damage, or the state of the building and dwellings in the area. Unfortunately, this information about the apartments and the quality of the neighborhood is not available for the entire gross sample (about 5% is missing). Therefore, it was decided to impute these values, conditional on the other available auxiliary variables. The variable reflecting the quality of the neighborhood comprises three categories ('poor neighborhood conditions', 22%; 'good neighborhood conditions', 31%; 'excellent neighborhood conditions', 47%).

Flemish Housing Survey 2005-06 FHS

The Flemish Housing Survey was conducted by the Research Network on Sustainable Housing Policy commissioned by the Housing Policy Department of the Ministry of the Flemish Community. The target population consisted of all private dwellings in Flanders.

Preceding the actual survey, an evaluation of the quality of the dwellings by experts took place. Ten experts were trained and their inspections were predominantly based on strongly objectified and pre-specified criteria. For this part of the research project, no cooperation (or even contact) was required from the occupants. This technical inspection generated a large inventory of highly relevant auxiliary information about the dwellings, particularly because a subsequent face-to-face survey was carried out with the occupants of the houses. The actual survey screened the profiles, expectations, and needs of the population as housing consumers. The fieldwork period spanned the period from April 2005 to February 2006 and was conducted by 187 experienced (at least one year) interviewers. Of the 8400 screened dwellings, some 7770 (93%) were selected for the face-to-face survey. The selection of cases to be attempted is believed to have been randomly determined and mainly driven by budget considerations. Within the attempted sample, some elements could not be contacted, despite the mandatory four contact attempts (of which the first needed to be personal, and at least one had to be in the evening while another had to take place at the weekend). These requirements were met relatively well, as in only 14% of the final noncontact cases were one of these conditions violated. In instances where the reference person (usually the head of the family) was deceased, or if the address was not valid, the sample case was considered as ineligible. Availability was decided based on if the reference respondent was abroad or simply not at home. After these response barriers, some 77% of the sample was found to be available for survey participation. The cooperation rate among these sample cases was about 80%, leading to a final response rate of about 70% among the 7770 attempted cases.

General Population Survey

The General Population Survey (GPS) is a real survey data set and accompanies the 'Handbook of Nonresponse in Household Survey' by Bethlehem et al. (2011). The fieldwork period covered about two months, of which the first was used to contact the selected individuals in person (CAPI). The second month of the fieldwork was used to re-approach the initial non-contacts and refusals. These initial nonrespondents were only re-issued if a listed phone number was available. A stratified two-stage sample was conducted to obtain the gross sample from the population register of the municipalities.

The whole sample comprises 32,019 people, out of whom 18,792 responded positively to the survey request. For both respondents and non-respondents, the dataset covers various auxiliary variables, though call record data is are not accessible for the GPS.

A system called Social Statistics Database (SSD) contains a wide range of characteristics on each individual in the Netherlands, including information on demography, geography, income, labor, education, health, and social protection. SSD records are then linked to the survey data records using a unique personal identification key so that auxiliary variables are made available for both respondents and nonrespondents. By linking municipality-level information to the dataset, additional sources of auxiliary variables have also become available. The next table provides an overview of the available auxiliary variables.

Auxiliary variables for the General Population Survey (GPS)

Auxiliary variable	**number of categories**
Gender	2
Marital status	4
Is married?	2
Age	13
Is non-native?	2
Type of non-native	5
Size of the household	5
Type of household	5
Children in household	2
Has listed phone number?	2
Has a job?	2
Employment situation	3
Has social allowance?	2
Has disability allowance?	2
Has unemployment allowance?	2
Has an allowance?	2
Region of the country	5
Degree of urbanization	5
Average house value in neighborhood	12
Percentage non-natives in neighborhood	8
Percentage non-western non-natives in neighborhood	7

Source: `http://www.survey-nonresponse.com`

DOCTORATEN IN DE SOCIALE WETENSCHAPPEN
EN
DOCTORATEN IN DE SOCIALE EN CULTURELE ANTROPOLOGIE

I. REEKS VAN DOCTORATEN IN DE SOCIALE WETENSCHAPPEN(1)

1. CLAEYS, U., *De sociale mobiliteit van de universitair afgestudeerden te Leuven. Het universitair onderwijs als mobiliteitskanaal*, 1971, 2 delen 398 blz.

2. VANHESTE, G., *Literatuur en revolutie*, 1971, 2 delen, 500 blz.

3. DELANGHE, L., *Differentiële sterfte in België. Een sociaal-demografische analyse*, 1971, 3 delen, 773 blz.

4. BEGHIN, P., *Geleide verandering in een Afrikaanse samenleving. De Bushi in de koloniale periode*, 1971, 316 blz.

5. BENOIT, A., *Changing the education system. A Colombian case-study*, 1972, 382 blz.

6. DEFEVER, M., *De huisartssituatie in België*, 1972, 374 blz.

7. LAUWERS, J., *Kritische studie van de secularisatietheorieën in de sociologie*, 1972, 364 blz.

8. GHOOS, A., *Sociologisch onderzoek naar de gevolgen van industrialisering in een rekonversiegebied*, 1972, 256 blz. + bijlagen.

9. SLEDSENS, G., *Mariage et vie conjugale du moniteur rwandais. Enquête sociologique par interview dirigée parmi les moniteurs mariés rwandais*, 1972, 2 delen, 549 blz.

10. TSAI, C., *La chambre de commerce internationale. Un groupe de pression international. Son action et son rôle dans l'élaboration, la conclusion et l'application des conventions internationales établies au sein des organisations intergouvernementales à vocation mondiale (1945-1969)*, 1972, 442 blz.

11. DEPRE, R., *De topambtenaren van de ministeries in België. Een bestuurssociologisch onderzoek*, 1973, 2 delen, 423 blz. + bijlagen.

12. VAN DER BIESEN, W., *De verkiezingspropaganda in de democratische maatschappij. Een literatuurkritische studie en een inhoudsanalyse van de verkiezingscampagne van 1958 in de katholieke pers en in de propagandapublikaties van de C.V.P.*, 1973, 434 blz.

13. BANGO, J., *Changements dans les communautés villageoises de l'Europe de l'Est. Exemple : la Hongarie*, 1973, 434 blz.

14. VAN PELT, H., *De omroep in revisie. Structurering en ontwikkelingsmogelijkheden van het radio- en televisiebestel in Nederland en België. Een vergelijkende studie*, Leuven, Acco, 1973, 398 blz.

15. MARTENS, A., *25 jaar wegwerparbeiders. Het Belgisch immigratiebeleid na 1945*, 1973, 319 blz.

16. BILLET, M., *Het verenigingsleven in Vlaanderen. Een sociologische typologieformulering en hypothesetoetsing*, 1973, 695 blz. + bijlagen.

17. BRUYNOOGHE, R., *De sociale structurering van de gezinsverplegingssituatie vanuit kostgezinnen en patiënten*, 1973, 205 blz. + bijlagen.

---

(1) EEN EERSTE SERIE DOCTORATEN VORMT DE REEKS VAN DE SCHOOL VOOR POLITIEKE EN SOCIALE WETENSCHAPPEN (NRS. 1 TOT EN MET 185). DE INTEGRALE LIJST KAN WORDEN GEVONDEN IN NADIEN GEPUBLICEERDE DOCTORATEN, ZOALS G. DOOGHE, "DE STRUCTUUR VAN HET GEZIN EN DE SOCIALE RELATIES VAN DE BEJAARDEN". ANTWERPEN, DE NEDERLANDSE BOEKHANDEL, 1970, 290 BLZ.
EEN TWEEDE SERIE DOCTORATEN IS VERMELD IN DE "NIEUWE REEKS VAN DE FACULTEIT DER ECONOMISCHE EN SOCIALE WETENSCHAPPEN". DE INTEGRALE LIJST KAN WORDEN GEVONDEN IN O.M. M. PEETERS, "GODSDIENST EN TOLERANTIE IN HET SOCIALISTISCH DENKEN". EEN HISTORISCH-DOCTRINAIRE STUDIE, 1970, 2 DELEN, 568 BLZ.

18. BUNDERVOET, J., *Het doorstromingsprobleem in de hedendaagse vakbeweging. Kritische literatuurstudie en verkennend onderzoek in de Belgische vakbonden*, 1973, 420 blz. + bijlagen.

19. GEVERS, P., *Ondernemingsraden, randverschijnselen in de Belgische industriële democratiseringsbeweging. Een sociologische studie*, 1973, 314 blz.

20. MBELA, H., *L'intégration de l'éducation permanente dans les objectifs socio-économiques de développement. Analyse de quelques politiques éducationnelles en vue du développement du milieu rural traditionnel en Afrique noire francophone*, 1974, 250 blz.

21. CROLLEN, L., *Small powers in international systems*, 1974, 250 blz.

22. VAN HASSEL, H., *Het ministrieel kabinet. Peilen naar een sociologische duiding*, 1974, 460 blz. + bijlagen.

23. MARCK, P., *Public relations voor de landbouw in de Europese Economische Gemeenschap*, 1974, 384 blz.

24. LAMBRECHTS, E., *Vrouwenarbeid in België. Een analyse van het tewerkstellingsbeleid inzake vrouwelijke arbeidskrachten sinds 1930*, 1975, 260 blz.

25. LEMMEN, M.H.W., *Rationaliteit bij Max Weber. Een godsdienstsociologische studie*, 1975, 2 delen, 354 blz.

26. BOON, G., *Ontstaan, ontwikkeling en werking van de radio-omroep in Zaïre tijdens het Belgisch Koloniale Bewind (1937-1960)*, 1975, 2 delen, 617 blz.

27. WUYTS, H., *De participatie van de burgers in de besluitvorming op het gebied van de gemeentelijke plannen van aanleg. Analyse toegespitst op het Nederlandstalige deel van België*, 1975, 200 blz. + bijlage.

28. VERRIEST, F., *Joris Helleputte en het corporatisme*, 1975, 2 delen, 404 blz.

29. DELMARTINO, F., *Schaalvergroting en bestuurskracht. Een beleidsanalystische benadering van de herstrukturering van de lokale besturen*, 1975, 3 delen, 433 blz. + bijlagen.

30. BILLIET, J., *Secularisering en verzuiling in het Belgisch onderwijs*, 1975, 3 delen, 433 blz. + bijlagen.

31. DEVISCH, R., *L'institution rituelle Khita chez les Yaka au Kwaango du Nord. Une analyse séméiologique*, 1976, 3 volumes.

32. LAMMERTYN, F., *Arbeidsbemiddeling en werkloosheid. Een sociologische verkenning van het optreden van de diensten voor openbare arbeidsbemiddeling van de R.V.A.*, 1976, 406 blz.

33. GOVAERTS, F., *Zwitserland en de E.E.G. Een case-study inzake Europese integratie*, 1976, 337 blz.

34. JACOBS, T., *Het uit de echt scheiden. Een typologiserend onderzoek, aan de hand van de analyse van rechtsplegingsdossiers in echtscheiding.* 1976, 333 blz. + bijlage.

35. KIM DAI WON, *Au delà de l'institutionalisation des rapports professionnels. Analyse du mouvement spontané ouvrier belge.* 1977, 282 blz.

36. COLSON, F., *Sociale indicatoren van enkele aspecten van bevolkingsgroei.* 1977, 341 blz. + bijlagen.

37. BAECK, A., *Het professionaliseringsproces van de Nederlandse huisarts.* 1978, 721 blz. + bibliografie.

38. VLOEBERGHS, D., *Feedback, communicatie en organisatie. Onderzoek naar de betekenis en de toepassing van het begrip "feedback" in de communicatiewetenschap en de organisatietheorieën.* 1978, 326 blz.

39. DIERICKX, G., *De ideologische factor in de Belgische politieke besluitvorming.* 1978, 609 blz. + bijvoegsels.

40. VAN DE KERCKHOVE, J., *Sociologie. Maatschappelijke relevantie en arbeidersemancipatie.* 1978, 551 blz.

41. DE MEYER A., *De populaire muziekindustrie. Een terreinverkennende studie.* 1979, 578 blz.

42. UDDIN, M., *Some Social Factors influencing Age at Death in the situation of Bangladesh.* 1979, 316 blz. + bijlagen.

43. MEULEMANS, E., *De ethische problematiek van het lijden aan het leven en aan het samen-leven in het oeuvre van Albert Camus. De mogelijke levensstijlen van luciditeit, menselijkheid en solidariteit.* 1979, 413 blz.

44. HUYPENS, J., *De plaatselijke nieuwsfabriek. Regionaal nieuws. Analyse van inhoud en structuur in de krant.* 494 blz.

45. CEULEMANS, M.J., *Women and Mass Media: a feminist perpective. A review of the research to date the image and status of women in American mass media.* 1980, 541 blz. + bijlagen.

46. VANDEKERCKHOVE, L., *Gemaakt van asse. Een sociologische studie van de westerse somatische kultuur.* 1980, 383 blz.

47. MIN, J.K., *Political Development in Korea, 1945-1972.* 1980, 2 delen, 466 blz.

48. MASUI, M., *Ongehuwd moeder. Sociologische analyse van een wordingsproces.* 1980, 257 blz.

49. LEDOUX, M., *Op zoek naar de rest ...; Genealogische lezing van het psychiatrisch discours.* 1981, 511 blz.

50. VEYS, D., *De generatie-sterftetafels in België.* 1981, 3 delen, 326 blz. + bijlagen.

51. TACQ, J., *Kausaliteit in sociologisch onderzoek. Een beoordeling van de zgn. 'causal modeling'-technieken in het licht van verschillende wijsgerige opvattingen over kausaliteit.* 1981, 337 blz.

52. NKUNDABAGENZI, F., *Le système politique et son environnement. Contribution à l'étude de leur interaction à partir du cas des pays est-africains : le Kenya et la Tanzanie.* 1981, 348 blz.

53. GOOSSENS, L., *Het sociaal huisvestingsbeleid in België. Een historisch-sociologische analyse van de maatschappelijke probleembehandeling op het gebied van het wonen.* 1982, 3 delen.

54. SCHEPERS, R., *De opkomst van het Belgisch medisch beroep. De evolutie van de wetgeving en de beroepsorganisatie in de 19de eeuw.* 1983, 553 blz.

55. VANSTEENKISTE, J., *Bejaardzijn als maatschappelijk gebeuren.* 1983, 166 blz.

56. MATTHIJS, K., *Zelfmoord en zelfmoordpoging.* 1983, 3 delen, 464 blz.

57. CHUNG-WON, Choue, *Peaceful Unification of Korea. Towards Korean Integration.* 1984, 338 blz.

58. PEETERS, R., *Ziekte en gezondheid bij Marokkaanse immigranten.* 1983, 349 blz.

59. HESLING, W., *Retorica en film. Een onderzoek naar de structuur en functie van klassieke overtuigingsstrategieën in fictionele, audiovisuele teksten.* 1985, 515 blz.

60. WELLEN, J., *Van probleem tot hulpverlening. Een exploratie van de betrekkingen tussen huisartsen en ambulante geestelijke gezondheidszorg in Vlaanderen.* 1984, 476 blz.

61. LOOSVELDT, G., *De effecten van een interviewtraining op de kwaliteit van gegevens bekomen via het survey-interview.* 1985, 311 blz. + bijlagen.

62. FOETS, M., *Ziekte en gezondheidsgedrag : de ontwikkeling van de sociologische theorievorming en van het sociologisch onderzoek.* 1985, 339 blz.

63. BRANCKAERTS, J., *Zelfhulporganisaties. Literatuuranalyse en explorerend onderzoek in Vlaanderen.* 1985.

64. DE GROOFF, D., *De elektronische krant. Een onderzoek naar de mogelijkheden van nieuwsverspreiding via elektronische tekstmedia en naar de mogelijke gevolgen daarvan voor de krant als bedrijf en als massamedium.* 1986, 568 blz.

65. VERMEULEN, D., *De maatschappelijke beheersingsprocessen inzake de sociaal-culturele sector in Vlaanderen. Een sociologische studie van de "verzuiling", de professionalisering en het overheidsbeleid.* 1983, 447 blz.

66. OTSHOMANPITA, Aloki, *Administration locale et développement au Zaïre. Critiques et perspectives de l'organisation politico-administrative à partir du cas de la zone de Lodja.* 1988, 507 blz.

67. SERVAES, J., *Communicatie en ontwikkeling. Een verkennende literatuurstudie naar de mogelijkheden van een communicatiebeleid voor ontwikkelingslanden.* 1987, 364 blz.

68. HELLEMANS, G., *Verzuiling. Een historische en vergelijkende analyse.* 1989, 302 blz.

II. NIEUWE REEKS VAN DOCTORATEN IN DE SOCIALE WETENSCHAPPEN EN IN DE SOCIALE EN CULTURELE ANTROPOLOGIE

1. LIU BOLONG, *Western Europe - China. A comparative analysis of the foreign policies of the European Community, Great Britain and Belgium towards China (1970-1986).* Leuven, Departement Politieke Wetenschappen, 1988, 335 blz.

2. EERDEKENS, J., *Chronische ziekte en rolverandering. Een sociologisch onderzoek bij M.S.-patiënten.* Leuven, Acco, 1989, 164 blz. + bijlagen.

3. HOUBEN, P., *Formele beslissingsmodellen en speltheorie met toepassingen en onderzoek naar activiteiten en uitgaven van locale welzijnsinstellingen en coalities.* Leuven, Departement Sociologie, 1988, 631 blz. (5 delen).

4. HOOGHE, L., *Separatisme. Conflict tussen twee projecten voor natievorming. Een onderzoek op basis van drie succesvolle separatismen.* Leuven, Departement Politieke Wetenschappen, 1989, 451 blz. + bijlagen.

5. SWYNGEDOUW, M., *De keuze van de kiezer. Naar een verbetering van de schattingen van verschuivingen en partijvoorkeur bij opeenvolgende verkiezingen en peilingen.* Leuven, Sociologisch Onderzoeksinstituut, 1989, 333 blz.

6. BOUCKAERT, G., *Productiviteit in de overheid.* Leuven, Vervolmakingscentrum voor Overheidsbeleid en Bestuur, 1990, 394 blz.

7. RUEBENS, M., *Sociologie van het alledaagse leven.* Leuven, Acco, 1990, 266 blz.

8. HONDEGHEM, A., *De loopbaan van de ambtenaar. Tussen droom en werkelijkheid.* Leuven, Vervolmakingscentrum voor Overheidsbeleid en Bestuur, 1990, 498 blz. + bijlage.

9. WINNUBST, M., *Wetenschapspopularisering in Vlaanderen. Profiel, zelfbeeld en werkwijze van de Vlaamse wetenschapsjournalist.* Leuven, Departement Communicatiewetenschap, 1990.

10. LAERMANS, R., *In de greep van de "moderne tijd". Modernisering en verzuiling, individualisering en het naoorlogse publieke discours van de ACW-vormingsorganisaties : een proeve tot cultuursociologische duiding*. Leuven, Garant, 1992.

11. LUYTEN, D., *OCMW en Armenzorg. Een sociologische studie van de sociale grenzen van het recht op bijstand.* Leuven, S.O.I. Departement Sociologie, 1993, 487 blz.

12. VAN DONINCK, B., *De landbouwcoöperatie in Zimbabwe. Bouwsteen van een nieuwe samenleving ?* Grimbergen, vzw Belgium-Zimbabwe Friendship Association, 1993. 331 blz.

13. OPDEBEECK, S., *Afhankelijkheid en het beëindigen van partnergeweld*. Leuven, Garant, 1993. 299 blz. + bijlagen.

14. DELHAYE, C., *Mode geleefd en gedragen*. Leuven, Acco, 1993, 228 blz.

15. MADDENS, B., *Kiesgedrag en partijstrategie*. Leuven, Departement Politieke Wetenschappen, Afdeling Politologie, K.U.Leuven, 1994, 453 blz.

16. DE WIT, H., *Cijfers en hun achterliggende realiteit*. *De MTMM-kwaliteitsparameters op hun kwaliteit onderzocht*. Leuven, Departement Sociologie, K.U.Leuven, 1994, 241 blz.

17. DEVELTERE, P., *Co-operation and development with special reference to the experience of the Commonwealth Carribean*. Leuven, Acco, 1994, 241 blz.

18. WALGRAVE, S., *Tussen loyauteit en selectiviteit. Een sociologisch onderzoek naar de ambivalente verhouding tussen nieuwe sociale bewegingen en groene partij in Vlaanderen*. Leuven, Garant, 1994, 361 blz.

19. CASIER, T., *Over oude en nieuwe mythen. Ideologische achtergronden en repercussies van de politieke omwentelingen in Centraal- en Oost-Europa sinds 1985*. Leuven, Departement Politieke Wetenschappen, K.U.Leuven, 1994, 365 blz.

20. DE RYNCK, F., *Streekontwikkeling in Vlaanderen. Besturingsverhoudingen en beleidsnetwerken in bovenlokale ruimtes*. Leuven, Departement Politieke Wetenschappen, Afdeling Bestuurswetenschap, K.U.Leuven, 1995, 432 blz.

21. DEVOS, G., *De flexibilisering van het secundair onderwijs in Vlaanderen. Een organisatie-sociologische studie van macht en institutionalisering*. Leuven, Acco, 1995, 447 blz.

22. VAN TRIER, W., *Everyone A King? An investigation into the meaning and significance of the debate on basic incomes with special references to three episodes from the British inter-War experience*. Leuven, Departement Sociologie, K.U.Leuven, 1995, vi+501 blz.

23. SELS, L., *De overheid viert de teugels. De effecten op organisatie en personeelsbeleid in de autonome overheidsbedrijven*. Leuven, Acco, 1995, 454 blz.

24. HONG, K.J., *The C.S.C.E. Security Regime Formation: From Helsinky to Budapest*. Leuven, Acco, 1996, 350 blz.

25. RAMEZANZADEH, A., *Internal and international dynamics of ethnic conflict. The Case of Iran*. Leuven, Acco, 1996, 273 blz.

26. HUYSMANS, J., *Making/Unmaking European Disorder. Meta-Theoretical, Theoretical and Empirical Questions of Military Stability after the Cold War*. Leuven, Acco, 1996, 250 blz.

27. VAN DEN BULCK J., *Kijkbuiskennis. De rol van televisie in de sociale en cognitieve constructie van de realiteit*. Leuven, Acco, 1996, 242 blz.

28. JEMADU Aleksius, *Sustainable Forest Management in the Context of Multi-level and Multi-actor Policy Processes*. Leuven, Departement Politieke Wetenschappen, Afdeling Bestuur en Overheidsmanagement, K.U.Leuven, 1996, 310 blz.

29. HENDRAWAN Sanerya, *Reform and Modernization of State Enterprises. The Case of Indonesia*. Leuven, Departement Politieke Wetenschappen, Afdeling Bestuur en Overheidsmanagement, K.U.Leuven, 1996, 372 blz.

30. MUIJS Roland Daniël, *Self, School and Media: A Longitudinal Study of Media Use, Self-Concept, School Achievement and Peer Relations among Primary School Children*. Leuven, Departement Communicatiewetenschap, K.U.Leuven, 1997, 316 blz.

31. WAEGE Hans, *Vertogen over de relatie tussen individu en gemeenschap.* Leuven, Acco, 1997, 382 blz.

32. FIERS Stefaan, *Partijvoorzitters in België of 'Le parti, c'est moi'?* Leuven, Acco, 1998, 419 blz.

33. SAMOY Erik, *Ongeschikt of ongewenst? Een halve eeuw arbeidsmarktbeleid voor gehandicapten*. Leuven, Departement Sociologie, K.U.Leuven, 1998, 640 blz.

34. KEUKELEIRE Stephan, *Het Gemeenschappelijk Buitenlands en Veiligheidsbeleid (GBVB): het buitenlands beleid van de Europese Unie op een dwaalspoor*. Leuven, Departement Politieke Wetenschappen, Afdeling Internationale Betrekkingen, K.U.Leuven, 1998, 452 blz.

35. VERLINDEN Ann, *Het ongewone alledaagse: over zwarte katten, horoscopen, miraculeuze genezingen en andere geloofselementen en praktijken. Een sociologie van het zogenaamde bijgeloof*. Leuven, Departement Sociologie, K.U.Leuven, 1999, 387 blz. + bijlagen.

36. CARTON Ann, *Een interviewernetwerk: uitwerking van een evaluatieprocedure voor interviewers*. Leuven, Departement Sociologie, 1999, 379 blz. + bijlagen.

37. WANG Wan-Li, *Undestanding Taiwan-EU Relations: An Analysis of the Years from 1958 to 1998*. Leuven, Departement Politieke Wetenschappen, Afdeling Internationale Betrekkingen, K.U.Leuven, 1999, 326 blz. + bijlagen.

38. WALRAVE Michel, *Direct Marketing en Privacy. De verhouding tussen direct marketingscommunicatie en de bescherming van de informationele en de relationele privacy van consumenten*. Leuven, Departement Communicatiewetenschap, K.U.Leuven, 1999, 480 blz. + bijlagen.

39. KOCHUYT Thierry, *Over een ondercultuur. Een cultuursociologische studie naar de relatieve deprivatie van arme gezinnen*. Leuven, Departement Sociologie, K.U.Leuven, 1999, 386 blz. + bijlagen.

40. WETS Johan, *Waarom onderweg? Een analyse van de oorzaken van grootschalige migratie- en vluchtelingenstromen.* Leuven, Departement Politieke Wetenschappen, Afdeling Internationale Betrekkingen, K.U.Leuven, 1999, 321 blz. + bijlagen.

41. VAN HOOTEGEM Geert, *De draaglijke traagheid van het management. Productie- en Personeelsbeleid in de industrie.* Leuven, Departement Sociologie, K.U.Leuven, 1999, 471 blz. + bijlagen.

42. VANDEBOSCH Heidi, *Een geboeid publiek? Het gebruik van massamedia door gedetineerden*. Leuven, Departement Communicatiewetenschap, K.U.Leuven, 1999, 375 blz. + bijlagen.

43. VAN HOVE Hildegard, *De weg naar binnen. Spiritualiteit en zelfontplooiing*. Leuven, Departement Sociologie, K.U.Leuven, 2000, 369 blz. + bijlagen.

44. HUYS Rik, *Uit de band? De structuur van arbeidsverdeling in de Belgische autoassemblagebedrijven.* Leuven, Departement Sociologie, K.U.Leuven, 2000, 464 blz. + bijlagen.

45. VAN RUYSSEVELDT Joris, *Het belang van overleg. Voorwaarden voor macroresponsieve CAO-onderhandelingen in de marktsector.* Leuven, Departement Sociologie, K.U.Leuven, 2000, 349 blz. + bijlagen.

46. DEPAUW Sam, *Cohesie in de parlementsfracties van de regeringsmeerderheid. Een vergelijkend onderzoek in België, Frankrijk en het Verenigd Koninkrijk (1987-97).* Leuven, Departement Politieke Wetenschappen, K.U.Leuven, 2000, 510 blz. + bijlagen.

47. BEYERS Jan, *Het maatschappelijk draagvlak van het Europees beleid en het einde van de permissieve consensus. Een empirisch onderzoek over politiek handelen in een meerlagig politiek stelsel.* Leuven, Departement Politieke Wetenschappen, K.U.Leuven, 2000, 269 blz. + bijlagen.

48. VAN DEN BULCK Hilde, *De rol van de publieke omroep in het project van de moderniteit. Een analyse van de bijdrage van de Vlaamse publieke televisie tot de creatie van een nationale cultuur en identiteit (1953-1973).* Leuven, Departement Communicatiewetenschap, K.U.Leuven, 2000, 329 blz. + bijlagen.

49. STEEN Trui, *Krachtlijnen voor een nieuw personeelsbeleid in de Vlaamse gemeenten. Een studie naar de sturing en implementatie van veranderingsprocessen bij de overheid.* Leuven, Departement Politieke Wetenschappen, K.U.Leuven, 2000, 340 blz. + bijlagen.

50. PICKERY Jan, *Applications of Multilevel Analysis in Survey Data Quality Research. Random Coefficient Models for Respondent and Interviewer Effects.* Leuven, Departement Sociologie, K.U.Leuven, 2000, 200 blz. + bijlagen.

51. DECLERCQ Aniana (Anja), *De complexe zoektocht tussen orde en chaos. Een sociologische studie naar de differentiatie in de institutionele zorgregimes voor dementerende ouderen.* Leuven, Departement Sociologie, K.U.Leuven, 2000, 260 blz. + bijlagen.

52. VERSCHRAEGEN Gert, *De maatschappij zonder eigenschappen. Systeemtheorie, sociale differentiatie en moraal.* Leuven, Departement Sociologie, K.U.Leuven, 2000, 256 blz. + bijlagen.

53. DWIKARDANA Sapta, *The Political Economy of Development and Industrial Relations in Indonesia under the New Order Government.* Leuven, Departement Sociologie, K.U.Leuven, 2001, 315 blz. + bijlagen.

54. SAUER Tom, *Nuclear Inertia. US Nuclear Weapons Policy after the Cold War (1990-2000).* Leuven, Departement Politieke Wetenschappen, K.U.Leuven, 2001, 358 blz. + bijlagen.

55. HAJNAL Istvan, *Classificatie in de sociale wetenschappen. Een evaluatie van de nauwkeurigheid van een aantal clusteranalysemethoden door middel van simulaties.* Leuven, Departement Sociologie, K.U.Leuven, 2001, 340 blz. + bijlagen.

56. VAN MEERBEECK Anne, *Het doopsel: een familieritueel. Een sociologische analyse van de betekenissen van dopen in Vlaanderen.* Leuven, Departement Sociologie, K.U.Leuven, 2001, 338 blz. + bijlagen.

57. DE PRINS Peggy, *Zorgen om zorg(arbeid). Een vergelijkend onderzoek naar oorzaken van stress en maatzorg in Vlaamse rusthuizen.* Leuven, Departement Sociologie, K.U.Leuven, 2001, 363 blz. + bijlagen.

58. VAN BAVEL Jan, *Demografische reproductie en sociale evolutie: geboortebeperking in Leuven 1840-1910.* Leuven, Departement Sociologie, K.U.Leuven, 2001, 362 blz. + bijlagen.

59. PRINSLOO Riana, *Subnationalism in a Cleavaged Society with Reference to the Flemish Movement since 1945.* Leuven, Departement Politieke Wetenschappen, K.U.Leuven, 2001, 265 blz. + bijlagen.

60. DE LA HAYE Jos, *Missed Opportunities in Conflict Management. The Case of Bosnia-Herzegovina (1987-1996).* Leuven, Departement Politieke Wetenschappen, K.U.Leuven, 2001, 283 blz. + bijlagen.

61. ROMMEL Ward, *Heeft de sociologie nood aan Darwin? Op zoek naar de verhouding tussen evolutiepsychologie en sociologie.* Leuven, Departement Sociologie, K.U.Leuven, 2002, 287 blz. + bijlagen.

62. VERVLIET Chris, *Vergelijking tussen Duits en Belgisch federalisme, ter toetsing van een neofunctionalistisch verklaringsmodel voor bevoegdheidsverschuivingen tussen nationale en subnationale overheden: een analyse in het economisch beleidsdomein.* Leuven, Departement Politieke Wetenschappen, K.U.Leuven, 2002, 265 blz. + bijlagen.

63. DHOEST Alexander, *De verbeelde gemeenschap: Vlaamse tv-fictie en de constructie van een nationale identiteit.* Leuven, Departement Communicatiewetenschap, K.U.Leuven, 2002, 384 blz. + bijlagen.

64. VAN REETH Wouter, *The Bearable Lightness of Budgeting. The Uneven Implementation of Performance Oriented Budget Reform Across Agencies.* Leuven, Departement Politieke Wetenschappen, K.U.Leuven, 2002, 380 blz. + bijlagen.

65. CAMBRé Bart, *De relatie tussen religiositeit en etnocentrisme. Een contextuele benadering met cross-culturele data.* Leuven, Departement Sociologie, K.U.Leuven, 2002, 257 blz. + bijlagen.

66. SCHEERS Joris, *Koffie en het aroma van de stad. Tropische (re-)productiestructuren in ruimtelijk perspectief. Casus centrale kustvlakte van Ecuador.* Leuven, Departement Sociologie, K.U.Leuven, 2002, 294 blz. + bijlagen.

67. VAN ROMPAEY Veerle, *Media on / Family off? An integrated quantitative and qualitative investigation into the implications of Information and Communication Technologies (ICT) for family life.* Leuven, Departement Communicatiewetenschap, K.U.Leuven, 2002, 232 blz. + bijlagen.

68. VERMEERSCH Peter, *Roma and the Politics of Ethnicity in Central Europe. A Comparative Study of Ethnic Minority Mobilisation in the Czech Republic, Hungary and Slovakia in the 1990s.* Leuven, Departement Politieke Wetenschappen, K.U.Leuven, 2002, 317 blz. + bijlagen.

69. GIELEN Pascal, *Pleidooi voor een symmetrische kunstsociologie. Een sociologische analyse van artistieke selectieprocessen in de sectoren van de hedendaagse dans en de beeldende kunst in Vlaanderen.* Leuven, Departement Sociologie, K.U.Leuven, 2002, 355 blz. + bijlagen.

70. VERHOEST Koen, *Resultaatgericht verzelfstandigen. Een analyse vanuit een verruimd principaal-agent perspectief.* Leuven, Departement Politieke Wetenschappen, K.U.Leuven, 2002, 352 blz. + bijlagen.

71. LEFèVRE Pascal, *Willy Vandersteens Suske en Wiske in de krant (1945-1971). Een theoretisch kader voor een vormelijke analyse van strips.* Leuven, Departement Communicatiewetenschap, K.U.Leuven, 2003, 186 blz. (A3) + bijlagen.

72. WELKENHUYSEN-GYBELS Jerry, *The Detection of Differential Item Functioning in Likert Score Items.* Leuven, Departement Sociologie, K.U.Leuven, 2003, 222 blz. + bijlagen.

73. VAN DE PUTTE Bart, *Het belang van de toegeschreven positie in een moderniserende wereld. Partnerkeuze in 19de-eeuwse Vlaamse steden (Leuven, Aalst en Gent).* Leuven, Departement Sociologie, K.U.Leuven, 2003, 425 blz. + bijlagen.

74. HUSTINX Lesley, *Reflexive modernity and styles of volunteering: The case of the Flemish Red Cross volunteers.* Leuven, Departement Sociologie, K.U.Leuven, 2003, 363 blz. + bijlagen.

75. BEKE Wouter, *De Christelijke Volkspartij tussen 1945 en 1968. Breuklijnen en pacificatiemechanismen in een catch-allpartij.* Leuven, Departement Politieke Wetenschappen, K.U.Leuven, 2004, 423 blz. + bijlagen.

76. WAYENBERG Ellen, *Vernieuwingen in de Vlaamse centrale - lokale verhoudingen: op weg naar partnerschap? Een kwalitatieve studie van de totstandkoming en uitvoering van het sociale impulsbeleid.* Leuven, Departement Politieke Wetenschappen, K.U.Leuven, 2004, 449 blz. + bijlagen.

77. MAESSCHALCK Jeroen, *Towards a Public Administration Theory on Public Servants' Ethics. A Comparative Study.* Leuven, Departement Politieke Wetenschappen, K.U.Leuven, 2004, 374 blz. + bijlagen.

78. VAN HOYWEGHEN Ine, *Making Risks. Travels in Life Insurance and Genetics*. Leuven, Departement Sociologie, K.U.Leuven, 2004, 248 blz. + bijlagen.

79. VAN DE WALLE Steven, *Perceptions of Administrative Performance: The Key to Trust in Government?* Leuven, Departement Politieke Wetenschappen, K.U.Leuven, 2004, 261 blz. + bijlagen.

80. WAUTERS Bram, *Verkiezingen in organisaties.* Leuven, Departement Politieke Wetenschappen, K.U.Leuven, 2004, 707 blz. + bijlagen.

81. VANDERLEYDEN Lieve, *Het Belgische/Vlaamse ouderenbeleid in de periode 1970-1999 gewikt en gewogen.* Leuven, Departement Sociologie, K.U.Leuven, 2004, 386 blz. + bijlagen.

82. HERMANS Koen, *De actieve welvaartsstaat in werking. Een sociologische studie naar de implementatie van het activeringsbeleid op de werkvloer van de Vlaamse OCMW's*. Leuven, Departement Sociologie, K.U.Leuven, 2005, 300 blz. + bijlagen.

83. BEVIGLIA ZAMPETTI Americo, *The Notion of 'Fairness' in International Trade Relations: the US Perspective.* Leuven, Departement Politieke Wetenschappen, K.U.Leuven, 2005, 253 blz. + bijlagen.

84. ENGELEN Leen, *De verbeelding van de Eerste Wereldoorlog in de Belgische speelfilm (1913-1939).* Leuven, Departement Communicatiewetenschap, K.U.Leuven, 2005, 290 blz. + bijlagen.

85. VANDER WEYDEN Patrick, *Effecten van kiessystemen op partijsystemen in nieuwe democratieën.* Leuven, Departement Sociologie, K.U.Leuven/K.U.Brussel, 2005, 320 blz. + bijlagen.

86. VAN HECKE Steven, *Christen-democraten en conservatieven in de Europese Volkspartij. Ideologische verschillen, nationale tegenstellingen en transnationale conflicten.* Leuven, Departement Politieke Wetenschappen, K.U.Leuven, 2005, 306 blz. + bijlagen.

87. VAN DEN VONDER Kurt, *"The Front Page" in Hollywood. Een geïntegreerde historisch-poëticale analyse.* Leuven, Departement Communicatiewetenschap, K.U.Leuven, 2005, 517 blz. + bijlagen.

88. VAN DEN TROOST Ann, *Marriage in Motion. A Study on the Social Context and Processes of Marital Satisfaction.* Leuven, Departement Sociologie, K.U.Leuven/R.U.Nijmegen, Nederland, 2005, 319 blz. + bijlagen.

89. ERTUGAL Ebru, *Prospects for regional governance in Turkey on the road to EU membership: Comparison of three regions.* Leuven, Departement Politieke Wetenschappen, K.U.Leuven, 2005, 384 blz. + bijlagen.

90. BENIJTS Tim, *De keuze van beleidsinstrumenten. Een vergelijkend onderzoek naar duurzaam sparen en beleggen in België en Nederland.* Leuven, Onderzoekseenheid: Instituut voor de Overheid [IO], K.U.Leuven, 2005, 501 blz. + bijlagen

91. MOLLICA Marcello, *The Management of Death and the Dynamics of an Ethnic Conflict: The Case of the 1980-81 Irish National Liberation Army (INLA) Hunger Strikes in Northern Ireland.* Leuven, Onderzoekseenheid: Instituut voor Internationaal en Europees Beleid [IIEB], K.U.Leuven, 2005, 168 blz. + bijlagen

92. HEERWEGH Dirk, *Web surveys. Explaining and reducing unit nonresponse, item nonresponse and partial nonresponse.* Leuven, Onderzoekseenheid: Centrum voor Sociologie [CeSO], K.U.Leuven, 2005, 350 blz. + bijlagen

93. GELDERS David (Dave), *Communicatie over nog niet aanvaard beleid: een uitdaging voor de overheid?* Leuven, Onderzoekseenheid: School voor Massacommunicatieresearch [SMC], K.U.Leuven, 2005, (Boekdeel 1 en 2) 502 blz. + bijlagenboek

94. PUT Vital, *Normen in performance audits van rekenkamers. Een casestudie bij de Algemene Rekenkamer en het National Audit Office.* Leuven, Onderzoekseenheid: Instituut voor de Overheid [IO], K.U.Leuven, 2005, 209 blz. + bijlagen

95. MINNEBO Jurgen, *Trauma recovery in victims of crime: the role of television use.* Leuven, Onderzoekseenheid: School voor Massacommunicatieresearch [SMC], K.U.Leuven, 2006, 187 blz. + bijlagen

96. VAN DOOREN Wouter, *Performance Measurement in the Flemish Public Sector: A Supply and Demand Approach.* Leuven, Onderzoekseenheid: Instituut voor de Overheid [IO], K.U.Leuven, 2006, 245 blz. + bijlagen

97. GIJSELINCKX Caroline, *Kritisch Realisme en Sociologisch Onderzoek. Een analyse aan de hand van studies naar socialisatie in multi-etnische samenlevingen.* Leuven, Onderzoekseenheid: Centrum voor Sociologie [CeSO], K.U.Leuven, 2006, 305 blz. + bijlagen

98. ACKAERT Johan, *De burgemeestersfunctie in België. Analyse van haar legitimering en van de bestaande rolpatronen en conflicten.* Leuven, Onderzoekseenheid: Instituut voor de Overheid [IO], K.U.Leuven, 2006, 289 blz. + bijlagen

99. VLEMINCKX Koen, *Towards a New Certainty: A Study into the Recalibration of the Northern-Tier Conservative Welfare States from an Active Citizens Perspective.* Leuven, Onderzoekseenheid: Centrum voor Sociologie [CeSO], K.U.Leuven, 2006, 381 blz. + bijlagen

100. VIZI Balázs, *Hungarian Minority Policy and European Union Membership. An Interpretation of Minority Protection Conditionality in EU Enlargement.* Leuven, Onderzoekseenheid: Insituut voor Internationaal en Europees Beleid [IIEB], K.U.Leuven, 2006, 227 blz. + bijlagen

101. GEERARDYN Aagje, *Het goede doel als thema in de externe communicatie. Bedrijfscommunicatie met een sociaal gezicht?* Leuven, Onderzoekseenheid: School voor Massacommunicatieresearch [SMC], K.U.Leuven, 2006, 272 blz. + bijlagen

102. VANCOPPENOLLE Diederik, *De ambtelijke beleidsvormingsrol verkend en getoetst in meervoudig vergelijkend perspectief. Een two-level analyse van de rol van Vlaamse ambtenaren in de Vlaamse beleidsvorming.* Leuven, Onderzoekseenheid: Instituut voor de Overheid [IO], K.U.Leuven, 2006, 331 blz. + bijlagenboek

103. DOM Leen, *Ouders en scholen: partnerschap of (ongelijke) strijd? Een kwalitatief onderzoek naar de relatie tussen ouders en scholen in het lager onderwijs.* Leuven, Onderzoekseenheid: Centrum voor Sociologisch Onderzoek [CeSO], K.U.Leuven, 2006, 372 blz. + bijlagen

104. NOPPE Jo, *Van kiesprogramma tot regeerakkoord. De beleidsonderhandelingen tussen de politieke partijen bij de vorming van de Belgische federale regering in 1991-1992 en in 2003.* Leuven, Onderzoekseenheid: Centrum voor Politicologie [CePO], K.U.Leuven, 2006, 364 blz. + bijlagen

105. YASUTOMI Atsushi, *Alliance Enlargement: An Analysis of the NATO Experience.* Leuven, Onderzoekseenheid: Instituut voor Internationaal en Europees Beleid [IIEB], K.U.Leuven, 2006, 294 blz. + bijlagen

106. VENTURINI Gian Lorenzo, *Poor Children in Europe. An Analytical Approach to the Study of Poverty in the European Union 1994-2000.* Dipartimento di Scienze Sociali, Università degli studi di Torino, Torino (Italië) / Onderzoekseenheid: Centrum voor Sociologisch Onderzoek [CeSO], K.U.Leuven, 2006, 192 blz. + bijlagen

107. EGGERMONT Steven, *The impact of television viewing on adolescents' sexual socialization.* Onderzoekseenheid: School voor Massacommunicatieresearch [SMC], K.U.Leuven, 2006, 244 blz. + bijlagen

108. STRUYVEN Ludovicus, *Hervormingen tussen drang en dwang. Een sociologisch onderzoek naar de komst en de gevolgen van marktwerking op het terrein van arbeidsbemiddeling.* Onderzoekseenheid: Centrum voor Sociologisch Onderzoek [CeSO], K.U.Leuven, 2006, 323 blz. + bijlagen

109. BROOS Agnetha, *De digitale kloof in de computergeneratie: ICT-exclusie bij adolescenten.* School voor Massa-communicatieresearch [SMC], K.U.Leuven, 2006, 215 blz. + bijlagen

110. PASPALANOVA Mila, *Undocumented and Legal Eastern European Immigrants in Brussels*. Onderzoekseenheid: Centrum voor Sociologisch Onderzoek [CeSO], K.U.Leuven/K.U.Brussel, 383 blz. + bijlagen

111. CHUN Kwang Ho, *Democratic Peace Building in East Asia in Post-Cold War Era. A Comparative Study*. Onderzoekseenheid: Instituut voor Internationaal en Europees Beleid [IIEB], K.U.Leuven, 2006, 297 blz. + bijlagen

112. VERSCHUERE Bram, *Autonomy & Control in Arm's Length Public Agencies: Exploring the Determinants of Policy Autonomy.* Onderzoekseenheid: Instituut voor de Overheid [IO], K.U.Leuven, 2006, 363 blz. + bijlagenboek

113. VAN MIERLO Jan, *De rol van televisie in de cultivatie van percepties en attitudes in verband met geneeskunde en gezondheid.* Onderzoekseenheid: School voor Massa-communicatieresearch [SMC], K.U.Leuven, 2007, 363 blz. + bijlagen

114. VENCATO Maria Francesca, *The Development Policy of the CEECs: the EU Political Rationale between the Fight Against Poverty and the Near Abroad.* Onderzoekseenheid: Instituut voor Internationaal en Europees Beleid [IIEB], K.U.Leuven, 2007, 276 blz. + bijlagen

115. GUTSCHOVEN Klaas, *Gezondheidsempowerment en de paradigmaverschuiving in de gezondheidszorg: de rol van het Internet.* Onderzoekseenheid: School voor Massa-communicatieresearch [SMC], K.U.Leuven, 2007, 330 blz. + bijlagen

116. OKEMWA James, *Political Leadership and Democratization in the Horn of Africa (1990-2000)* Onderzoekseenheid: Instituut voor Internationaal en Europees Beleid [IIEB], K.U.Leuven, 2007, 268 blz. + bijlagen

117. DE COCK Rozane, *Trieste Vedetten? Assisenverslaggeving in Vlaamse kranten.* Onderzoekseenheid: School voor Massa-communicatieresearch [SMC], K.U.Leuven, 2007, 257 blz. + bijlagen

118. MALLIET Steven, *The Challenge of Videogames to Media Effect Theory.* Onderzoekseenheid: Centrum voor Mediacultuur en communicatietechnologie [CMC], K.U.Leuven, 2007, 187 blz. + bijlagen

119. VANDECASTEELE Leen, *Dynamic Inequalities. The Impact of Social Stratification Determinants on Poverty Dynamics in Europe.* Onderzoekseenheid: Centrum voor Sociologisch Onderzoek [CeSO], K.U.Leuven, 246 blz. + bijlagen

120. DONOSO Veronica, *Adolescents and the Internet: Implications for Home, School and Social Life.* Onderzoekseenheid: School voor Massa-communicatieresearch [SMC], K.U.Leuven, 2007, 264 blz. + bijlagen

121. DOBRE Ana Maria, *Europeanisation From A Neo-Institutionalist Perspective: Experiencing Territorial Politics in Spain and Romania.* Onderzoekseenheid: Instituut voor Internationaal en Europees Beleid [IIEB], K.U.Leuven, 2007, 455 blz. + bijlagen

122. DE WIT Kurt, *Universiteiten in Europa in de 21e eeuw. Netwerken in een veranderende samenleving.* Onderzoekseenheid: Centrum voor Sociologisch Onderzoek [CeSO], K.U.Leuven, 2007,362 blz. + bijlagen

123. CORTVRIENDT Dieter, *The Becoming of a Global World: Technology / Networks / Power / Life.* Onderzoekseenheid: Centrum voor Sociologisch Onderzoek [CeSO], K.U.Leuven, 2008, 346 blz. + bijlagen

124. VANDER STICHELE Alexander, *De culturele alleseter? Een kwantitatief en kwalitatief onderzoek naar 'culturele omnivoriteit' in Vlaanderen.* Onderzoekseenheid: Centrum voor Sociologisch Onderzoek [CeSO], K.U.Leuven, 2008, 414 blz. + bijlagen(boek)

125. LIU HUANG Li-chuan, *A Biographical Study of Chinese Restaurant People in Belgium: Strategies for Localisation.* Onderzoekseenheid: Centrum voor Sociologisch Onderzoek [CeSO], K.U.Leuven, 2008, 365 blz. + bijlagen

126. DEVILLé Aleidis, *Schuilen in de schaduw. Een sociologisch onderzoek naar de sociale constructie van verblijfsillegaliteit.* Onderzoekseenheid: Centrum voor Sociologisch Onderzoek [CeSO], K.U.Leuven, 2008, 469 blz. + bijlagen

127. FABRE Elodie, *Party Organisation in a multi-level setting: Spain and the United Kingdom*. Onderzoekseenheid: Centrum voor Politicologie [CePO], K.U.Leuven, 2008, 282 blz. + bijlagen

128. PELGRIMS Christophe, *Politieke actoren en bestuurlijke hervormingen. Een stakeholder benadering van Beter Bestuurlijk Beleid en Copernicus.* Onderzoekseenheid: Instituut voor de Overheid [IO], K.U.Leuven, 2008, 374 blz. + bijlagen

129. DEBELS Annelies, *Flexibility and Insecurity. The Impact of European Variants of Labour Market Flexibility on Employment, Income and Poverty Dynamics.* Onderzoekseenheid: Centrum voor Sociologisch Onderzoek [CeSO], K.U.Leuven, 2008, 366 blz. + bijlagen

130. VANDENABEELE Wouter, *Towards a public administration theory of public service motivation.* Onderzoekseenheid: Instituut voor de Overheid [IO], K.U.Leuven, 2008, 306 blz. + bijlagen

131. DELREUX Tom, *The European union negotiates multilateral environmental agreements: an analysis of the internal decision-making process.* Onderzoekseenheid: Instituut voor Internationaal en Europees Beleid [IIEB], K.U.Leuven, 2008, 306 blz. + bijlagen

132. HERTOG Katrien, *Religious Peacebuilding: Resources and Obstacles in the Russian Orthodox Church for Sustainable Peacebuilding in Chechnya.* Onderzoekseenheid: Instituut voor Internationaal en Europees Beleid [IIEB], K.U.Leuven, 2008, 515 blz. + bijlagen

133. PYPE Katrien, *The Making of the Pentecostal Melodrama. Mimesis, Agency and Power in Kinshasa's Media World (DR Congo).* Onderzoekseenheid: Instituut voor Antropologie in Afrika [IARA], K.U.Leuven, 2008, 401 blz. + bijlagen + dvd

134. VERPOEST Lien, *State Isomorphism in the Slavic Core of the Commonwealth of Independent States (CIS). A Comparative Study of Postcommunist Geopolitical Pluralism in Russia, Ukraine and Belarus.* Onderzoekseenheid: Instituut voor Internationaal en Europees Beleid [IIEB], K.U.Leuven, 2008, 412 blz. + bijlagen

135. VOETS Joris, *Intergovernmental relations in multi-level arrangements: Collaborative public management in Flanders.* Onderzoekseenheid: Instituut voor de Overheid [IO], K.U.Leuven, 2008, 260 blz. + bijlagen

136. LAENEN Ria, *Russia's 'Near Abroad' Policy and Its Compatriots (1991-2001). A Former Empire In Search for a New Identity.* Onderzoekseenheid: Instituut voor Internationaal en Europees Beleid [IIEB], K.U.Leuven, 2008, 293 blz. + bijlagen

137. PEDZIWIATR Konrad Tomasz, *The New Muslim Elites in European Cities: Religion and Active Social Citizenship Amongst Young Organized Muslims in Brussels and London.* Onderzoekseenheid: Centrum voor Sociologisch Onderzoek [CeSO], K.U.Leuven, 2008, 483 blz. + bijlagen

138. DE WEERDT Yve, *Jobkenmerken en collectieve deprivatie als verklaring voor de band tussen de sociale klasse en de economische attitudes van werknemers in Vlaanderen.* Onderzoekseenheden: Centrum voor Sociologisch Onderzoek [CeSO] en Onderzoeksgroep Arbeids-, Organisatie- en Personeelspsychologie, K.U.Leuven, 2008, 155 blz. + bijlagen

139. FADIL Nadia, *Submitting to God, submitting to the Self. Secular and religious trajectories of second generation Maghrebi in Belgium.* Onderzoekseenheid: Centrum voor Sociologisch Onderzoek [CeSO], K.U.Leuven, 2008, 370 blz. + bijlagen

140. BEUSELINCK Eva, *Shifting public sector coordination and the underlying drivers of change: a neo-institutional perspective.* Onderzoekseenheid: Instituut voor de Overheid [IO], K.U.Leuven, 2008, 283 blz. + bijlagen

141. MARIS Ulrike, *Newspaper Representations of Food Safety in Flanders, The Netherlands and The United Kingdom. Conceptualizations of and Within a 'Risk Society'.* Onderzoekseenheid: School voor Massa-communicatieresearch [SMC], K.U.Leuven, 2008, 159 blz. + bijlagen

142. WEEKERS Karolien, *Het systeem van partij- en campagnefinanciering in België: een analyse vanuit vergelijkend perspectief.* Onderzoekseenheid: Centrum voor Politicologie [CePO], K.U.Leuven, 2008, 248 blz. + bijlagen

143. DRIESKENS Edith, *National or European Agents? An Exploration into the Representation Behaviour of the EU Member States at the UN Security Council.* Onderzoekseenheid: Instituut voor Internationaal en Europees Beleid [IIEB], K.U.Leuven, 2008, 221 blz. + bijlagen

144. DELARUE Anne, *Teamwerk: de stress getemd? Een multilevelonderzoek naar het effect van organisatieontwerp en teamwerk op het welbevinden bij werknemers in de metaalindustrie.* Onderzoekseenheid: Centrum voor Sociologisch Onderzoek [CeSO], K.U.Leuven, 2009, 454 blz. + bijlagen

145. MROZOWICKI Adam, *Coping with Social Change. Life strategies of workers in Poland after the end of state socialism.* Onderzoekseenheid: Centrum voor Sociologisch Onderzoek [CeSO], K.U.Leuven, 2009, 383 blz. + bijlagen

146. LIBBRECHT Liselotte, *The profile of state-wide parties in regional elections. A study of party manifestos: the case of Spain.* Onderzoekseenheid: Centrum voor Politicologie [CePO], K.U.Leuven, 2009, 293 blz. + bijlagen

147. SOENEN Ruth, *De connecties van korte contacten. Een etnografie en antropologische reflectie betreffende transacties, horizontale bewegingen, stedelijke relaties en kritische indicatoren.* Onderzoekseenheid: Interculturalism, Migration and Minorities Research Centre [IMMRC], K.U.Leuven, 2009, 231 blz. + bijlagen

148. GEERTS David, *Sociability Heuristics for Interactive TV. Supporting the Social Uses of Television.* Onderzoekseenheid: Centrum voor Mediacultuur en Communicatietechnologie [CMC], K.U.Leuven, 2009, 201 blz. + bijlagen

149. NEEFS Hans, *Between sin and disease. A historical-sociological study of the prevention of syphilis and AIDS in Belgium (1880-2000).* Onderzoekseenheid: Centrum voor Sociologisch Onderzoek [CeSO], K.U.Leuven, 2009, 398 blz. + bijlagen

150. BROUCKER Bruno, *Externe opleidingen in overheidsmanangement en de transfer van verworven kennis. Casestudie van de federale overheid.* Onderzoekseenheid: Instituut voor de Overheid [IO], K.U.Leuven, 2009, 278 blz. + bijlagen

151. KASZA Artur, *Policy Networks and the Regional Development Strategies in Poland. Comparative case studies from three regions.* Onderzoekseenheid: Instituut voor Internationaal en Europees Beleid [IIEB], K.U.Leuven, 2009, 485 blz. + bijlagen

152. BEULLENS Kathleen, *Stuurloos? Een onderzoek naar het verband tussen mediagebruik en risicogedrag in het verkeer bij jongeren.* Onderzoekseenheid: School voor Massacommunicatieresearch [SMC], K.U.Leuven, 2009, 271 blz. + bijlagen

153. OPGENHAFFEN Michaël, *Multimedia, Interactivity, and Hypertext in Online News: Effect on News Processing and Objective and Subjective Knowledge.* Onderzoekseenheid: Centrum voor Mediacultuur en Communicatietechnologie [CMC], K.U.Leuven, 2009, 233 blz. + bijlagen

154. MEULEMAN Bart, *The influence of macro-sociological factors on attitudes toward immigration in Europe. A cross-cultural and contextual approach.* Onderzoekseenheid: Centrum voor Sociologisch Onderzoek [CeSO], K.U.Leuven, 2009, 276 blz. + bijlagen

155. TRAPPERS Ann, *Relations, Reputations, Regulations: An Anthropological Study of the Integration of Romanian Immigrants in Brussels, Lisbon and Stockholm.* Onderzoekseenheid: Interculturalism, Migration and Minorities Research Centre [IMMRC], K.U.Leuven, 2009, 228 blz. + bijlagen

156. QUINTELIER Ellen, *Political participation in late adolescence. Political socialization patterns in the Belgian Political Panel Survey.* Onderzoekseenheid: Centrum voor Politicologie [CePO], K.U.Leuven, 2009, 288 blz. + bijlagen

157. REESKENS Tim, *Ethnic and Cultural Diversity, Integration Policies and Social Cohesion in Europe. A Comparative Analysis of the Relation between Cultural Diversity and Generalized Trust in Europe.* Onderzoekseenheid: Centrum voor Politicologie [CePO], K.U.Leuven, 2009, 298 blz. + bijlagen

158. DOSSCHE Dorien, *How the research method affects cultivation outcomes.* Onderzoekseenheid: School voor Massacommunicatieresearch [SMC], K.U.Leuven, 2010, 254 blz. + bijlagen

159. DEJAEGHERE Yves, *The Political Socialization of Adolescents. An Exploration of Citizenship among Sixteen to Eighteen Year Old Belgians.* Onderzoekseenheid: Centrum voor Politicologie [CePO], K.U.Leuven, 2010, 240 blz. + bijlagen

160. GRYP Stijn, *Flexibiliteit in bedrijf - Balanceren tussen contractuele en functionele flexibiliteit.* Onderzoekseenheid: Centrum voor Sociologisch Onderzoek [CeSO], K.U.Leuven, 2010, 377 blz. + bijlagen

161. SONCK Nathalie, *Opinion formation: the measurement of opinions and the impact of the media.* Onderzoekseenheid: Centrum voor Sociologisch Onderzoek [CeSO], K.U.Leuven, 2010, 420 blz. + bijlagen

162. VISSERS Sara, *Internet and Political Mobilization. The Effects of Internet on Political Participation and Political Equality.* Onderzoekseenheid: Centrum voor Politicologie [CePO], K.U.Leuven, 2010, 374 blz. + bijlagen

163. PLANCKE Carine, « J'irai avec toi » : désirs et dynamiques du maternel dans les chants et les danses punu (Congo-Brazzaville). Onderzoekseenheden: Instituut voor Antropologie in Afrika [IARA], K.U.Leuven / Laboratoire d'Anthropologie Sociale [LAS, Parijs], EHESS, 2010, 398 blz. + bijlagenboek + DVD + CD

164. CLAES Ellen, *Schools and Citizenship Education. A Comparative Investigation of Socialization Effects of Citizenship Education on Adolescents.* Onderzoekseenheid: Centrum voor Politicologie [CePO], K.U.Leuven, 2010, 331 blz. + bijlagen

165. LEMAL Marijke, *"It could happen to you." Television and health risk perception.* Onderzoekseenheid: School voor Massacommunicatieresearch [SMC], K.U.Leuven, 2010, 316 blz. + bijlagen

166. LAMLE Nankap Elias, *Laughter and conflicts. An anthropological exploration into the role of joking relationships in conflict mediation in Nigeria: A case study of Funyallang in Tarokland.* Onderzoekseenheid: Instituut voor Antropologie in Afrika [IARA], K.U.Leuven, 2010, 250 blz. + bijlagen

167. DOGRUEL Fulya, *Social Transition Across Multiple Boundaries: The Case of Antakya on The Turkish-Syrian Border.* Onderzoekseenheid: Interculturalism, Migration and Minorities Research Centre [IMMRC], K.U.Leuven, 2010, 270 blz. + bijlagen

168. JANSOVA Eva, *Minimum Income Schemes in Central and Eastern Europe.* Onderzoekseenheid: Centrum voor Sociologisch Onderzoek [CeSO], K.U.Leuven, 2010, 195 blz. + bijlagen

169. IYAKA Buntine (François-Xavier), *Les Politiques des Réformes Administratives en République Démocratique du Congo (1990-2010).* Onderzoekseenheid: Instituut voor de Overheid [IO], K.U.Leuven, 2010, 269 blz. + bijlagen

170. MAENEN Seth, *Organizations in the Offshore Movement. A Comparative Study on Cross-Border Software Development and Maintenance Projects.* Onderzoekseenheid: Centrum voor Sociologisch Onderzoek [CeSO], K.U.Leuven, 2010, 296 blz. + bijlagen

171. FERRARO Gianluca *Domestic Implementation of International Regimes in Developing Countries. The Case of Marine Fisheries in P.R. China.* Onderzoekseenheid: Instituut voor de Overheid [IO], K.U.Leuven, 2010, 252 blz. + bijlagen

172. van SCHAIK Louise, *Is the Sum More than Its Parts? A Comparative Case Study on the Relationship between EU Unity and its Effectiveness in International Negotiations.* Onderzoekseenheid: Instituut voor Internationaal en Europees Beleid [IIEB], K.U.Leuven, 2010, 219 blz. + bijlagen

173. SCHUNZ Simon, *European Union foreign policy and its effects - a longitudinal study of the EU's influence on the United Nations climate change regime (1991-2009).* Onderzoekseenheid: Instituut voor Internationaal en Europees Beleid [IIEB], K.U.Leuven, 2010, 415 blz. + bijlagen

174. KHEGAI Janna, *Shaping the institutions of presidency in the post-Soviet states of Central Asia: a comparative study of three countries..* Onderzoekseenheid: Instituut voor Internationaal en Europees Beleid [IIEB], K.U.Leuven, 2010, 193 blz. + bijlagen

175. HARTUNG Anne, *Structural Integration of Immigrants and the Second Generation in Europe: A Study of Unemployment Durations and Job Destinations in Luxembourg, Belgium and Germany.* Onderzoekseenheid: Centrum voor Sociologisch Onderzoek [CeSO], K.U.Leuven, 2010, 285 blz. + bijlagen

176. STERLING Sara, *Becoming Chinese: Ethnic Chinese-Venezuelan Education Migrants and the Construction of Chineseness.* Onderzoekseenheid: Interculturalism, Migration and Minorities Research Centre [IMMRC], K.U.Leuven, 2010, 225 blz. + bijlagen

177. CUVELIER Jeroen, *Men, mines and masculinities in Katanga: the lives and practices of artisanal miners in Lwambo (Katanga province, DR Congo).* Onderzoekseenheid: Instituut voor Antropologie in Afrika [IARA], K.U.Leuven, 2011, 302 blz. + bijlagen

178. DEWACHTER Sara, *Civil Society Participation in the Honduran Poverty Reduction Strategy: Who takes a seat at the pro-poor table?* Onderzoekseenheid: Instituut voor Internationaal en Europees Beleid [IIEB], K.U.Leuven, 2011, 360 blz. + bijlagen

179. ZAMAN Bieke, *Laddering method with preschoolers. Understanding preschoolers' user experience with digital media.* Onderzoekseenheid: Centrum voor Mediacultuur en Communicatietechnologie [CMC], K.U.Leuven, 2011, 222 blz. + bijlagen

180. SULLE Andrew, *Agencification of Public Service Management in Tanzania: The Causes and Control of Executive Agencies.* Onderzoekseenheid: Instituut voor de Overheid [IO], K.U.Leuven, 2011, 473 blz. + bijlagen

181. KOEMAN Joyce, *Tussen commercie en cultuur: Reclamepercepties van autochtone en allochtone jongeren in Vlaanderen.* Onderzoekseenheid: Centrum voor Mediacultuur en Communicatietechnologie [CMC], K.U.Leuven, 2011, 231 blz. + bijlagen

182. GONZALEZ GARIBAY Montserrat, *Turtles and teamsters at the GATT/WTO. An analysis of the developing countries' trade-labor and trade-environment policies during the 1990s.* Onderzoekseenheid: Instituut voor Internationaal en Europees Beleid [IIEB], K.U.Leuven, 2011, 403 blz. + bijlagen

183. VANDEN ABEELE Veronika, *Motives for Motion-based Play. Less flow, more fun.* Onderzoekseenheid: Centrum voor Mediacultuur en Communicatietechnologie [CMC], K.U.Leuven, 2011, 227 blz. + bijlagen

184. MARIEN Sofie, *Political Trust. An Empirical Investigation of the Causes and Consequences of Trust in Political Institutions in Europe.* Onderzoekseenheid: Centrum voor Politicologie [CePO], K.U.Leuven, 2011, 211 blz. + bijlagen

185. JANSSENS Kim, *Living in a material world: The effect of advertising on materialism.* Onderzoekseenheid: School voor Massacommunicatieresearch [SMC], K.U.Leuven, 2011, 197 blz. + bijlagen

186. DE SCHUTTER Bob, *De betekenis van digitale spellen voor een ouder publiek.* Onderzoekseenheid: Centrum voor Mediacultuur en Communicatietechnologie [CMC], K.U.Leuven, 2011, 339 blz. + bijlagen

187. MARX Axel, *Global Governance and Certification. Assessing the Impact of Non-State Market Governance.* Onderzoekseenheid: Centrum voor Sociologisch Onderzoek [CeSO], K.U.Leuven, 2011, 140 blz. + bijlagen

188. HESTERS Delphine, *Identity, culture talk & culture. Bridging cultural sociology and integration research - a study on second generation Moroccan and native Belgian residents of Brussels and Antwerp.* Onderzoekseenheid: Centrum voor Sociologisch Onderzoek [CeSO], K.U.Leuven, 2011, 440 blz. + bijlagen

189. AL-FATTAL Rouba, *Transatlantic Trends of Democracy Promotion in the Mediterranean: A Comparative Study of EU, US and Canada Electoral Assistance in the Palestinian Territories (1995-2010).* Onderzoekseenheid: Instituut voor Internationaal en Europees Beleid [IIEB], K.U.Leuven, 2011, 369 blz. + bijlagen

190. MASUY Amandine, *How does elderly family care evolve over time? An analysis of the care provided to the elderly by their spouse and children in the Panel Study of Belgian Households 1992-2002.* Onderzoekseenheden: Centrum voor Sociologisch Onderzoek [CeSO], K.U.Leuven / Institute of Analysis of Change in Contemporary and Historical Societies [IACCHOS], Université Catholique de Louvain, 2011, 421 blz. + bijlagen

191. BOUTELIGIER Sofie, *Global Cities and Networks for Global Environmental Governance.* Onderzoekseenheid: Instituut voor Internationaal en Europees Beleid [IIEB], K.U.Leuven, 2011, 263 blz. + bijlagen

192. GÖKSEL Asuman, *Domestic Change in Turkey: An Analysis of the Extent and Direction of Turkish Social Policy Adaptation to the Pressures of European Integration in the 2000s.* Onderzoekseenheid: Instituut voor Internationaal en Europees Beleid [IIEB], K.U.Leuven, 2011, 429 blz. + bijlagen

193. HAPPAERTS Sander, *Sustainable development between international and domestic forces. A comparative analysis of subnational policies.* Onderzoekseenheid: Instituut voor Internationaal en Europees Beleid [IIEB], K.U.Leuven, 2011, 334 blz. + bijlagen

194. VANHOUTTE Bram, *Social Capital and Well-Being in Belgium (Flanders). Identifying the Role of Networks and Context.* Onderzoekseenheid: Centrum voor Politicologie [CePO], K.U.Leuven, 2011, 165 blz. + bijlagen

195. VANHEE Dieter, *Bevoegdheidsoverdrachten in België: een analyse van de vijfde staatshervorming van 2001.* Onderzoekseenheid: Instituut voor de Overheid [IO], K.U.Leuven, 2011, 269 blz. + bijlagen

196. DE VUYSERE Wilfried, *Neither War nor Peace. Civil-Military Cooperation in Complex Peace Operations.* Onderzoekseenheid: Instituut voor Internationaal en Europees Beleid [IIEB], KU Leuven, 2012, 594 blz. + bijlagen

197. TOUQUET Heleen, *Escaping ethnopolis: postethnic mobilization in Bosnia-Herzegovina.* Onderzoekseenheid: Instituut voor Internationaal en Europees Beleid [IIEB], KU Leuven, 2012, 301 blz. + bijlagen

198. ABTS Koenraad, *Maatschappelijk onbehagen en etnopopulisme. Burgers, ressentiment, vreemdelingen, politiek en extreem rechts.* Onderzoekseenheid: Centrum voor Sociologisch Onderzoek [CeSO], KU Leuven, 2012, 1066 blz. + bijlagen

199. VAN DEN BRANDE Karoline, *Multi-Level Interactions for Sustainable Development. The Involvement of Flanders in Global and European Decision-Making.* Onderzoekseenheid: Instituut voor Internationaal en Europees Beleid [IIEB], KU Leuven, 2012, 427 blz. + bijlagen

200. VANDELANOITTE Pascal, *Het spectrum van het verleden. Een visie op de geschiedenis in vier Europese arthousefilms (1965-1975).* Onderzoekseenheid: Centrum voor Mediacultuur en Communicatietechnologie [CMC], KU Leuven, 2012, 341 blz. + bijlagen

201. JUSTAERT Arnout, *The European Union in the Congolese Police Reform: Governance, Coordination and Alignment?.* Onderzoekseenheid: Instituut voor Internationaal en Europees Beleid [IIEB], KU Leuven, 2012, 247 blz. + bijlagen

202. LECHKAR Iman, *Striving and Stumbling in the Name of Allah. Neo-Sunnis and Neo-Shi'ites in a Belgian Context.* Onderzoekseenheid: Interculturalism, Migration and Minorities Research Centre [IMMRC], KU Leuven, 2012, 233 blz. + bijlagen

203. CHOI Priscilla, *How do Muslims convert to Evangelical Christianity? Case studies of Moroccans and Iranians in multicultural Brussels.* Onderzoekseenheid: Interculturalism, Migration and Minorities Research Centre [IMMRC], KU Leuven, 2012, 224 blz. + bijlagen

204. BIRCAN Tuba, *Community Structure and Ethnocentrism. A Multilevel Approach: A case Study of Flanders (Belgium).* Onderzoekseenheid: Centrum voor Politicologie [CePO], KU Leuven, 2012, 221 blz. + bijlagen

205. DESSERS Ezra, *Spatial Data Infrastructures at work. A comparative case study on the spatial enablement of public sector processes.* Onderzoekseenheid: Centrum voor Sociologisch Onderzoek [CeSO], KU Leuven, 2012, 314 blz. + bijlagen

206. PLASQUY Eddy, *La Romería del Rocío: van een lokale celebratie naar een celebratie van lokaliteit. Transformaties en betekenisverschuivingen van een lokale collectieve bedevaart in Andalusië.* Onderzoekseenheid: Institute for Anthropological Research in Africa [IARA], KU Leuven, 2012, 305 blz. + bijlagen

207. BLECKMANN Laura E., *Colonial Trajectories and Moving Memories: Performing Past and Identity in Southern Kaoko (Namibia).* Onderzoekseenheid: Institute for Anthropological Research in Africa [IARA], KU Leuven, 2012, 394 blz. + bijlagen

208. VAN CRAEN Maarten, *The impact of social-cultural integration on ethnic minority group members' attitudes towards society.* Onderzoekseenheid: Centrum voor Politicologie [CePO], KU Leuven, 2012, 248 blz. + bijlagen

209. CHANG Pei-Fei, *The European Union in the Congolese Police Reform: Governance, Coordination and Alignment?.* Onderzoekseenheid: Instituut voor Internationaal en Europees Beleid [IIEB], KU Leuven, 2012, 403 blz. + bijlagen

210. VAN DAMME Jan, *Interactief beleid. Een analyse van organisatie en resultaten van interactieve planning in twee Vlaamse 'hot spots'.* Onderzoekseenheid: Instituut voor de Overheid [IO], KU Leuven, 2012, 256 blz. + bijlagen

211. KEUNEN Gert, *Alternatieve mainstream: een cultuursociologisch onderzoek naar selectielogica's in het Vlaamse popmuziekcircuit.* Onderzoekseenheid: Centrum voor Sociologisch Onderzoek [CeSO], KU Leuven, 2012, 292 blz. + bijlagen

212. FUNK DECKARD Julianne, *'Invisible' Believers for Peace: Religion and Peacebuilding in Postwar Bosnia-Herzegovina.* Onderzoekseenheid: Instituut voor Internationaal en Europees Beleid [IIEB], KU Leuven, 2012, 210 blz. + bijlagen

213. YILDIRIM Esma, *The Triple Challenge: Becoming a Citizen and a Female Pious Muslim. Turkish Muslims and Faith Based Organizations at Work in Belgium..* Onderzoekseenheid: Interculturalism, Migration and Minorities Research Centre [IMMRC], KU Leuven, 2012, 322 blz. + bijlagen

214. ROMMEL Jan, *Organisation and Management of Regulation. Autonomy and Coordination in a Multi-Actor Setting.* Onderzoekseenheid: Instituut voor de Overheid [IO], KU Leuven, 2012, 235 blz. + bijlagen

215. TROUPIN Steve, *Professionalizing Public Administration(s)? The Cases of Performance Audit in Canada and the Netherlands.* Onderzoekseenheid: Instituut voor de Overheid [IO], KU Leuven, 2012, 528 blz. + bijlagen

216. GEENEN Kristien, *The pursuit of pleasure in a war-weary city, Butembo, North Kivu, DRC.* Onderzoekseenheid: Institute for Anthropological Research in Africa [IARA], KU Leuven, 2012, 262 blz. + bijlagen

217. DEMUZERE Sara, *Verklarende factoren van de implementatie van kwaliteitsmanagementtechnieken. Een studie binnen de Vlaamse overheid.* Onderzoekseenheid: Instituut voor de Overheid [IO], KU Leuven, 2012, 222 blz. + bijlagen

218. EL SGHIAR Hatim, *Identificatie, mediagebruik en televisienieuws. Exploratief onderzoek bij gezinnen met Marokkaanse en Turkse voorouders in Vlaanderen.* Onderzoekseenheid: Instituut voor Mediastudies [IMS], KU Leuven, 2012, 418 blz. + bijlagen

219. WEETS Katrien, *Van decreet tot praktijk? Een onderzoek naar de invoering van elementen van prestatiebegroting in Vlaamse gemeenten.* Onderzoekseenheid: Instituut voor de Overheid [IO], KU Leuven, 2012, 343 blz. + bijlagenbundel

220. MAES Guido, *Verborgen krachten in de organisatie: een politiek model van organisatieverandering.* Onderzoekseenheid: Centrum voor Sociologisch Onderzoek [CeSO], KU Leuven, 2012, 304 blz. + bijlagen

221. VANDEN ABEELE Mariek (Maria), *Me, Myself and my Mobile: Status, Identity and Belongingness in the Mobile Youth Culture.* Onderzoekseenheid: School voor Massacommunicatieresearch [SMC], KU Leuven, 2012, 242 blz. + bijlagen

222. RAMIOUL Monique, *The map is not the territory: the role of knowledge in spatial restructuring processes.* Onderzoekseenheid: Centrum voor Sociologisch Onderzoek [CeSO], KU Leuven, 2012, 210 blz. + bijlagen

223. CUSTERS Kathleen, *Television and the cultivation of fear of crime: Unravelling the black box.* Onderzoekseenheid: School voor Massacommunicatieresearch [SMC], KU Leuven, 2012, 216 blz. + bijlagen

224. PEELS Rafael, *Facing the paradigm of non-state actor involvement: the EU-Andean region negotiation process.* Onderzoekseenheid: Instituut voor Internationaal en Europees Beleid [IIEB], KU Leuven, 2012, 239 blz. + bijlagen

225. DIRIKX Astrid, *Good Cop - Bad Cop, Fair Cop - Dirty Cop. Het verband tussen mediagebruik en de houding van jongeren ten aanzien van de politie.* Onderzoekseenheid: School voor Massacommunicatieresearch [SMC], KU Leuven, 2012, 408 blz. + bijlagen

226. VANLANGENAKKER Ine, *Uitstroom in het regionale parlement en het leven na het mandaat. Een verkennend onderzoek in Catalonië, Saksen, Schotland, Vlaanderen en Wallonië.* Onderzoekseenheid: Centrum voor Politicologie [CePO], KU Leuven, 2012, 255 blz. + bijlagen

227. ZHAO Li, *New Co-operative Development in China: An Institutional Approach.* Onderzoekseenheid: Instituut voor Internationaal en Europees Beleid [IIEB], KU Leuven, 2012, 256 blz. + bijlagen

228. LAMOTE Frederik, *Small City, Global Scopes: An Ethnography of Urban Change in Techiman, Ghana.* Onderzoekseenheid: Institute for Anthropological Research in Africa [IARA], KU Leuven, 2012, 261 blz. + bijlagen

229. SEYREK Demir Murat, *Role of the NGOs in the Integration of Turkey to the European Union.* Onderzoekseenheid: Centrum voor Politicologie [CePO], KU Leuven, 2012, 313 blz. + bijlagen

230. VANDEZANDE Mattijs, *Born to die. Death clustering and the intergenerational transmission of infant mortality, the Antwerp district, 1846-1905.* Onderzoekseenheid: Centrum voor Sociologisch Onderzoek [CeSO], KU Leuven, 2012, 179 blz. + bijlagen

231. KUHK Annette, *Means for Change in Urban Policies - Application of the Advocacy Coalition Framework (ACF) to analyse Policy Change and Learning in the field of Urban Policies in Brussels and particularly in the subset of the European Quarter.* Onderzoekseenheid: Instituut voor de Overheid [IO], KU Leuven, 2013, 282 blz. + bijlagen

232. VERLEDEN Frederik, *De 'vertegenwoordigers van de Natie' in partijdienst. De verhouding tussen de Belgische politieke partijen en hun parlementsleden (1918-1970).* Onderzoekseenheid: Centrum voor Politicologie [CePO], KU Leuven, 2013, 377 blz. + bijlagen

233. DELBEKE Karlien, *Analyzing 'Organizational justice'. An explorative study on the specification and differentiation of concepts in the social sciences.* Onderzoekseenheid: Instituut voor de Overheid [IO], KU Leuven, 2013, 274 blz. + bijlagen

234. PLATTEAU Eva, *Generations in organizations. Ageing workforce and personnel policy as context for intergenerational conflict in local government.* Onderzoekseenheid: Instituut voor de Overheid [IO], KU Leuven, 2013, 322 blz. + bijlagen

235. DE JONG Sijbren, *The EU's External Natural Gas Policy – Caught Between National Priorities and Supranationalism.* Onderzoekseenheid: Instituut voor Internationaal en Europees Beleid [IIEB], KU Leuven, 2013, 234 blz. + bijlagen

236. YANASMAYAN Zeynep, *Turkey entangled with Europe? A qualitative exploration of mobility and citizenship accounts of highly educated migrants from Turkey.* Onderzoekseenheid: Instituut voor Internationaal en Europees Beleid [IIEB], KU Leuven, 2013, 346 blz. + bijlagen

237. GOURDIN Gregory, *De evolutie van de verhouding tussen ziekenhuisartsen en ziekenhuismanagement in België sinds de Besluitwet van 28 december 1944.* Onderzoekseenheid: Centrum voor Sociologisch Onderzoek [CeSO], KU Leuven, 2013, 271 blz. + bijlagen

238. VANNIEUWENHUYZE Jorre, *Mixed-mode Data Collection: Basic Concepts and Analysis of Mode Effects.* Onderzoekseenheid: Centrum voor Sociologisch Onderzoek [CeSO], KU Leuven, 2013, 214 blz. + bijlagen

239. RENDERS Frank, *Ruimte maken voor het andere: Auto-etnografische verhalen en zelfreflecties over het leven in een Vlaamse instelling voor personen met een verstandelijke handicap.* Onderzoekseenheid: Interculturalism, Migration and Minorities Research Centre [IMMRC], KU Leuven, 2013, 248 blz. + bijlagen

240. VANCAUWENBERGHE Glenn, *Coördinatie binnen de Geografische Data Infrastructuur: Een analyse van de uitwisseling en het gebruik van geografische informatie in Vlaanderen..* Onderzoekseenheid: Instituut voor de Overheid [IO], KU Leuven, 2013, 236 blz. + bijlagen

241. HENDRIKS Thomas, *Work in the Rainforest: Labour, Race and Desire in a Congolese Logging Camp.* Onderzoekseenheid: Institute for Anthropological Research in Africa [IARA], KU Leuven, 2013, 351 blz. + bijlagen

242. BERGHMAN Michaël, *Context with a capital C. On the symbolic contextualization of artistic artefacts.* Onderzoekseenheid: Centrum voor Sociologisch Onderzoek [CeSO], KU Leuven, 2013, 313 blz. + bijlagen

243. IKIZER Ihsan, *Social Inclusion and Local Authorities. Analysing the Implementation of EU Social Inclusion Principles by Local Authorities in Europe.* Onderzoekseenheid: Centrum voor Sociologisch Onderzoek [CeSO], KU Leuven, 2013, 301 blz. + bijlagen

244. GILLEIR Christien, *Combineren in je eentje. Arbeid en gezin bij werkende alleenstaande ouders in Vlaanderen.* Onderzoekseenheid: Centrum voor Sociologisch Onderzoek [CeSO], KU Leuven, 2013, 250 blz. + bijlagen

ooOoo

Koen BEULLENS

The use of paradata to assess survey representativity
Cracks in the nonresponse paradigm

2013