Survey Research Methods (2013)
Vol. 8, No. 2, pp. 81-98
ISSN 1864-3361
http://www.surveymethods.org

The effect of different survey designs on nonresponse in surveys among non-Western minorities in The Netherlands.

Joost. W.S. Kappelhof
The Netherlands institute for Social Research/SCP
The Hague, The Netherlands

The present study investigates the impact of survey design choices on the representativity and the potential for nonresponse bias on survey estimates of eight sub-surveys conducted among non-Western minorities in The Netherlands. The surveys were part of a repeated, cross-sectional measurement conducted separately among each of the four largest non-Western minority groups. This study utilizes fieldwork disposition codes in conjunction with the R-indicator and maximal absolute standardized bias to show the impact of survey design choices - such as the period and length of fieldwork, the use of bilingual interviewers, the number of face-to-face call attempts and a re-issue of nonresponding sampled persons - on the potential for nonresponse bias on survey estimates. Partial R-indicators are used to detect which socio-demographic subgroups contribute the most to a nonrepresentative response, conditional on ethnic group and survey design. The results indicate that long fieldwork periods increase the potential for nonresponse bias on survey estimates among non-Western minorities due to moving and that the timing of fieldwork has an impact on the number of sampled persons who are unavailable during the fieldwork period. Furthermore, the use of bilingual interviewers is necessary to conduct a survey among Turkish and Moroccans due to language problems; otherwise the potential for nonresponse bias on survey estimates can be quite severe. Also, the use of a re-issue phase reduces the potential for nonresponse bias on survey estimates in surveys among non-Western minorities in The Netherlands. Finally, partial R-indicator analyses provide further insight on how future surveys can be improved in order to reduce the potential for nonresponse bias on survey estimates among each of the four non-Western minority groups.
Keywords: survey design choices, nonresponse bias, non-Western minorities, representativity, quality indicators

1. Introduction

In general population surveys, non-Western minorities – or ethnic minorities as they are sometimes referred to – tend to be underrepresented (Feskens, 2009; Groves & Couper, 1998; Schmeets, 2005; Stoop, 2005). At the same time, there is a great need for specific information about this group, especially on issues such as socio-economic and cultural integration in The Netherlands and elsewhere (Bijl & Verweij, 2012). That is why separate surveys among non-Western minorities continue to be necessary. However, large scale surveys are costly, and surveys among minorities are even more expensive per completed interview than general surveys, due to the lower response rates among non-Western minorities. It is therefore of great importance to determine which strategies are effective for surveying non-Western minorities, while maintaining a certain level of quality and minimizing the costs.

Joost W.S. Kappelhof, The Netherlands Institute for Social Research/SCP, The Hague, P.O. Box 16164, The Netherlands (j.kappelhof@scp.nl)

This paper sets out to investigate how different survey design choices affect the composition of the *response* sample (i.e., the composition of the group of respondents) and how this might relate to the occurrence of nonresponse bias on survey estimates in surveys conducted among non-Western minorities in The Netherlands. We shall compare eight sub-surveys – four separate sub-surveys in two different survey rounds – that vary in these choices and we shall try to ascertain which set of design choices leads to the sample with the lowest potential for nonresponse bias on survey estimates.

A standard measure for judging the quality of a response sample is still the response rate, despite the fact that it is not a direct measure of nonresponse bias (Biemer & Lyberg, 2003; Groves, 2006). In the last few years several other quality indicators have been developed that - under assumptions – provide a more direct insight in the existence of nonresponse bias and allow us to estimate its size (see for instance Andridge & Little, 2011; Särndal, 2011; Särndal & Lundström, 2005; Schouten, Cobben, & Bethlehem, 2009; Wagner, 2010). In this study, next to the response rate, we shall make use of two methods to evaluate the quality of the response samples of both surveys among non-Western minorities and its potential for nonresponse bias on survey estimates.

The first method is based on studying different reasons for

nonresponse by analysing the final disposition code of the sample units (see for instance, De Heer, 1999). The second method utilizes the representativity indicator (R-indicator) and the related maximal absolute standardized bias estimator ($\widehat{B'}_m$) to study nonresponse (J. Bethlehem, Cobben, & Schouten, 2011; Schouten et al., 2009). We also analyse the impact of separate survey design choices, such as the number of face-to-face contact attempts, the re-issue of nonresponding sampled persons and the use of bilingual interviewers with a common ethnic background. To this end we use the R-indicator and $\widehat{B}_m$ to show the impact of these design choices on the quality of the response sample. We conduct a detailed analysis of the under- and overrepresented sociodemographic subgroups within each survey design, separately for each minority group, using partial R-indicators (Schouten et al., 2013). This will allow us to further develop tailor-made approach strategies for future surveys among non-Western minorities in The Netherlands.

The article starts with a brief overview of the main data collection difficulties resulting in nonresponse when surveying non-Western minorities. The data and methods section describes the surveys, the survey design choices and the methods used to answer our research aim. This is followed by the results of the analysis and the subsequent conclusion and discussion.

2. Why are non-Western minorities underrepresented in population surveys in The Netherlands?

In 2011, non-Western minorities made up about 11% of the population in The Netherlands (CBS-statline, 2013). Statistics Netherlands uses the following official definition: "Every person residing in The Netherlands of whom one or both parents were born in Africa, Latin- America and Asia (excluding Indonesia and Japan) or Turkey" (Reep, 2003).[1]

The main reason for the underrepresentation of non-Western minorities in population surveys in The Netherlands is nonresponse. One can make a distinction between direct causes and correlates for nonresponse on the one hand, and between characteristics of the person and the survey design features on the other hand. A direct cause would be language problems or the higher rate of illiteracy especially among older non-Western immigrants (Feskens, Kappelhof, Dagevos, & Stoop, 2010). A correlate would be that non-Western minorities tend to live more often in the larger cities in The Netherlands. Big city dwellers in general are more difficult to contact and refuse more often (Groves & Couper, 1998; Stoop, 2005).

Adapting the survey design in such a way that the direct causes of nonresponse are addressed may reduce the specific nonresponse among non-Western minorities. Language problems stop being a problem when one of the design features is a translated questionnaire. Functional illiteracy ceases to be a problem when the interviews are conducted by interviewers who read out the questionnaire. Also, the use of the telephone for interviews increases the number of refusals among non-Western minorities to an incomparable degree in comparison to native Dutch or to a face-to-face mode and should therefore be avoided (Schothorst, 2002; Korte & Dagevos, 2011).

Other cultural differences influencing nonresponse may also be reduced by specific survey design choices. For example, the use of interviewers with a common ethnic background: they do not only speak the language, but are also aware of the proper etiquette to approach sampled persons. An often overlooked cause is the timing and length of the fieldwork. Especially among some of the ethnic minority groups, it is not uncommon to go on an extended holiday to their country of origin during summer time. Sometimes, there is also a mismatch between religious holidays of ethnic groups and the way the fieldwork agency plan their fieldwork (Kemper, 1998; Schothorst, 2002; Veenman, 2002).

Sampling frame errors and especially undercoverage provide other reasons why non-Western minorities are underrepresented in population surveys in The Netherlands. Undercoverage is what happens when not all elements of the target population can be found in the sampling frame (Groves, 1989). In The Netherlands, (semi)-governmental and scientific institutes mainly use the postal data service (delivery sequence file) or municipal personal data records database (population register) as a sampling frame. Both frames suffer from frame errors, such as moving of the sample unit, no known address of the sample unit, slow registration of the sample unit or death of the sample unit. Some of these causes seem to occur far more often among non-Western minorities, such as moving or no known address of sample units (Feskens, 2009; Kappelhof, 2010).

3. Data & Methods

Data

The Survey on the Integration of Minorities (SIM) sets out to measure the socio-economic position of non-Western minorities as well as their socio-cultural integration. This survey is a nationwide, cross-sectional survey which started in 2006 and was repeated in 2011. In the present study both face-to-face CAPI survey rounds are included (SIM2006 and SIM2011).

In both SIM-rounds Statistics Netherlands drew a random sample of named individuals from each of five mutually exclusive population strata; Dutch of Turkish, Moroccan, Surinamese, and Antillean[2] descent and the remainder of the population (mostly native Dutch) living in The Netherlands, in

[1] A further distinction is made between first generation (born in Africa, Latin- America and Asia (excluding Indonesia and Japan) or Turkey and moved to the Netherlands) and second generation (born in the Netherlands, but parents were born in Africa, Latin- America and Asia (excluding Indonesia and Japan) or Turkey). Indonesian and Japanese immigrants are seen as (more similar to) Western minorities based on their socio-economic and socio-cultural position. It mainly involves persons born in former Dutch-Indie (Indonesia) and employees working for Japanese companies with their families.

[2] Including Aruba

the age of 15 years and above. The present study focusses on how different survey design choices affect the potential for nonresponse bias on survey estimates in surveys conducted among non-Western minorities. This is why the samples containing native Dutch are excluded from this study resulting in eight samples for analysis.

The official definition of Dutch of Turkish, Moroccan, Surinamese, and Antillean descent includes persons that were either born in Turkey, Morocco, Surinam or the Dutch Antilles[3] or have at least one parent who was born there. In case the father and mother were born in different countries, the mother's country of birth is dominant unless the mother was born in The Netherlands in which case the father's country of birth is dominant. These four ethnic groups make up about two-thirds of the total non-Western population in The Netherlands (CBS-statline, 2013). For the purpose of brevity, they will be referred to as Turkish, Moroccans, Surinamese and Antilleans in the remainder of this article.

Both SIM-rounds used the population register as a sampling frame and the same stratified two stage probability sampling design in all four population strata. In the first stage municipalities were selected and in the second stage named individuals were selected. The strata variable used was municipality size and consisted of three strata: the four largest municipalities, all with a population of over 250 000 (self-selecting); midsize municipalities with a population of between 50 000 and 250 000 and small municipalities with a population of less than 50 000. For each target group, the sample size was proportionally allocated across different municipality size strata (Table 1).

In this study we used the fieldwork and response data files from SIM2006 and SIM2011. The fieldwork data files contain both process data, such as number, time, date and outcome of contact attempt, and auxiliary information from the sampling frame about each sample unit, such as ethnicity, age, sex, first or second generation immigrants, municipality, etc. Process data and auxiliary information, also known as paradata, are potentially useful for increasing participation, for nonresponse adjustment or for evaluating potential nonresponse bias (Couper, 2005; Kreuter, 2013; Maitland, Casas-Cordero, & Kreuter, 2009). The response data files contain the answers of the respondents to the survey questions, but also interviewer observations about respondents, such as their ability to speak Dutch.

Survey design choices and response enhancing measures. There are differences in the survey design between both SIM-rounds with respect to the fieldwork and the questionnaire. In SIM2006, the main part of the fieldwork lay outside the winter period, whereas for the SIM2011 survey the main part of the fieldwork was conducted during the winter period. The length of the SIM2006 fieldwork was also about twice that of the SIM2011 measurement: nine months versus five months. Also, the fieldwork agencies differed across rounds. Bureau Veldkamp conducted the fieldwork in 2006 and Gfk Netherlands[4] in 2011.

The main difference about the questionnaire between the surveys resided in the length. The research topics were identical, but the questionnaire length was reduced. The reason for this reduction was based on interviewer reports after the completion of the SIM2006 survey, but also on opinions of fieldwork experts and experts on minority research (Feskens et al., 2010). They all believed the questionnaire was too long which could potentially harm the response rate. This resulted in a reduced questionnaire length between the SIM2006 and the SIM2011 measurement from an estimated, based on CAPI timers, average of 55 minutes to 44 minutes.

Response enhancing measures such as the use of incentives and advance letters have a proven positive effect on the response rates (Dillman, 2007; Groves & Couper, 1998; Singer, Van Hoewyk, Gebler, Raghunathan, & McGonagle, 1999; Singer, Van Hoewyk, & Maher, 2000; Singer, 2001). These measures may therefore also affect the response composition and the quality of the response sample.

The type of measures that were used varied between the two SIM-rounds and ethnic groups. There were also differences in the extent to which the same measures were used in 2006 and 2011, and in the ethnic groups. An unconditional non-monetary incentive (stamps) was used in SIM2006 among all groups, whereas no unconditional incentives were used in the SIM2011 measurement.

Conditional incentives were used among all groups in both surveys. All respondents received a gift certificate (€ 10) after completion. In SIM2011, respondents were also offered the option to donate € 10 to charity.

A recent survey conducted by Statistics Netherlands among the four largest non-Western minorities discovered that approximately 14% of the sample were nonrespondents due to language problems (Feskens, 2009). Results from other surveys among the same minorities groups in The Netherlands showed that nonrespondents who are not able to read or speak Dutch are mostly found among the Turkish and Moroccan population (Kappelhof, 2010). For both SIM2006 and SIM2011, auxiliary information about ethnicity, age, sex, municipality and status as first or second generation immigrants was available in the sample frame data for all sampled persons. This allowed for a tailored approach of the sampled persons. Two types of tailoring were used to increase response. They mainly have to do with anticipated language problems, but also with anticipated cultural differences. Research has shown that the greater cultural familiarity due to the common ethnic background between interviewer and respondent may also be a factor in increasing the willingness to respond (see, for instance Moorman, Newman, Millikan, Tse, & Sandler, 1999).

The first type of tailoring was the use of translated questionnaires and advance letters. These were used in SIM2006 and SIM2011, but only among Moroccan and Turkish. Also a phonetically translated Berber version was available as an aid for the interviewer. This is a spoken (i.e., not written) language that many Moroccans living in The Netherlands have

[3] or Aruba

[4] Gfk also made use of a subcontractor (Labyrinth) to ensure enough interviewers with a shared ethnic background were available to conduct the fieldwork among all ethnic groups.

Table 1 Gross sample sizes per ethnic group and survey year across municipality strata.

	Turkish		Moroccans		Surinamese		Antilleans	
	2006	2011	2006	2011	2006	2011	2006	2011
Large Municipalities	802	554	1218	812	1563	1020	867	695
Midsize Municipalities	928	727	771	674	714	662	947	945
Small Municipalities	432	284	401	254	401	248	398	334
Total	2162	1565	2390	1740	2678	1930	2212	1974

as their mother tongue. The answers were recorded in the CAPI program in either Dutch or Moroccan Arabic. There was no need to translate questionnaires or advance letters for Surinamese or Antilleans. Dutch is the mother tongue for many, if not all persons of Surinamese or Antillean origin.

The second type of tailoring is the assignment of sample units to an interviewer with a common ethnic background. Both surveys used interviewers with a shared ethnic background with the sampled person, but the intensity in which they were used varied between SIM2006 and SIM2011 *and* between target groups.

In both SIM-rounds *bilingual* interviewers with a common ethnic background approached sampled persons of Moroccans or Turkish origin. In SIM2006 there was a limited and systematic use of bilingual interviewers with a common ethnic background among a part of this group. They mainly contacted older, first generation immigrants who lived in the larger cities, because that is where the language problems were mostly anticipated. For respondents that were interviewed by *non-bilingual* interviewers without a common ethnic background, the translated questionnaire was also made available. The questionnaire could be shown on request of the respondent or in case a question posed in Dutch was unclear to the respondents. Interviewers with a common ethnic background were hardly used at all among sampled persons of Surinamese or Antillean origin in the SIM2006 study.

In SIM2011, all sampled persons of Moroccan or Turkish origin were contacted by a bilingual interviewer with a common ethnic background. In SIM2011 about half of the sampled persons of Surinamese or Antillean origin were approached by interviewers with a common ethnic background, the other half were approached by either Dutch interviewers or interviewers with another ethnic background. The allocation of Surinamese and Antillean sample units to interviewers with a common ethnic background was based on the availability of an interviewer with a common ethnic background in the area.

In 2006 and 2011, potential respondents could call a toll free number in case of questions or to reschedule an appointment for an interview. Finally, interviewer bonuses to increase interviewer productivity were used in SIM2006, but not in SIM2011. Unfortunately, there was no information available on the identity of the interviewers who received these bonuses in 2006, so as to analyse the effectiveness of this measure.

The reality of fieldwork: deviations from the planned survey design. Both SIMs used a responsive design approach where non-responding sampled persons in the first phase of fieldwork are taken 'out of the field' and re-issued again by the fieldwork agency (Groves & Heeringa, 2006). This approach provides the opportunity to introduce other design choices after the first phase, such as an increased incentive or another interviewer.

In the first phase of SIM2006, a minimum of four contact attempts (CA) had to be made to a sampled person before the sampled person could be registered as a noncontact and returned to the fieldwork office for potential re-issuing. The CAs had to be made on different days and at different times in the day.

In SIM2011, there had to be at least three CAs on different days of the week and at different times during the day before the sampled person could be registered as a noncontact and returned to the fieldwork office. However, interviewers were encouraged to conduct more CAs. Only after three unsuccessful CAs in the first phase, the interviewer was allowed to try and reach the sampled person by telephone (if available) and set up an appointment or leave a "sorry I missed you" card.

The way unsuccessful sampling units were selected to be re-issued in the second phase varied between both SIMs. In SIM2006, the planned second phase of fieldwork involved only the re-issue of soft refusals and noncontacts among underrepresented non-Western minority subgroups, such as young males living in urban areas. These re-issued sampled persons were offered the same conditional incentive worth € 10 and a minimum of four CAs had to be made by another interviewer.

Unfortunately, during the second phase of the SIM2006 fieldwork not all sample units selected for re-issuing were re-contacted with a minimum of four contact attempts for noncontacts. The difference in selection and re-issue of unsuccessful sample units back into the field was based on the availability of another interviewer in the area and costs. This meant that, if a sample unit was selected to be re-issued but no other interviewer was available in the area, none would be sent in case there were less than three re-issued sample units.

In total 1,143 sample units were selected for re-issuing in 2006. Unfortunately fieldwork ended before all sample units selected to be re-issued were actually re-issued or re-contacted at least four times. This resulted in 522 second phase sampled persons that were either not re-issued or where no final disposition code was achieved. Only for 621

sample units a final disposition code was declared (see Table 2).

In SIM2011 the plan was to select all first phase nonrespondents and to re-issue them for the second phase. A minimum of three face-to-face contact attempts had to be made by another interviewer. Furthermore the amount of the promised or conditional non-monetary incentive (gift certificates) was increased from € 10 to € 15.

Unfortunately, again, due to time constraints, only very few sample units were actually re-contacted by another interviewer (Table 2). In this case, the difference in selection and re-issue of unsuccessful sample units was based on the availability of another interviewer in the area within the remainder of the fieldwork period. In case no other interviewer was available in the area, the original interviewer had to conduct at least six contact attempts.

Methods

The analysis of data from nationwide, cross sectional surveys among hard to reach populations for which specific measures were undertaken imposes limits on the use of analysis methods, such as logistic regression. Both SIMs are not set up as an experiment to assess the effectiveness of separate response enhancing measures on the probability of participation among various socio-demographic subgroups. They were designed to be as efficient as possible in increasing the probability of response among various, very difficult to survey populations by using auxiliary information available on the sampled persons. This meant, for instance, a non-random allocation of sampled persons with specific characteristics to ethnic interviewers in SIM2006. Also, in both SIM-rounds only certain nonresponding sampled persons were selected and actually re-issued. As a result, the effect of socio-demographic variables such as age, immigration generation, municipality size, ethnic group on the odds to participate is confounded with the non-random allocation of a (bilingual) interviewer with a common ethnic background and with whether or not a sampled person has been re-issued, in which case they were usually contacted by a more successful interviewer.

Another potential confounding factor is the possible change in perception of surveys and in general willingness to participate in surveys that may take place in the interval between both survey rounds among the hard to reach minorities. For instance, in the five year interval, a continuing shift towards the right was noticeable in Dutch society, combined with the rise of a more populist discourse on migrants in The Netherlands. This might negatively affect the willingness to participate of non-Western minorities.

The representativity-indicator (R-indicator) and the maximal absolute standardized bias are quality indicators that allow for a comparison between surveys using different, targeted designs and/or a comparison across time (Schouten et al., 2009). Recently, both indicators have been developed as a result of a large European project to assess the effects of nonresponse on the quality of statistics (RISQ-project.eu).These indicators are not dependent on a random allocation of sample units, but allow for an assessment of the quality of the response sample in which targeted response enhancing measures were used. They also allow for an estimation of the impact of separate response enhancing measures on the quality of the response sample.

The following two approaches, which we will present in more detail, are used to ascertain the quality of the response sample. The first approach is the final disposition code of the sample unit and the second approach is the representativity indicator (R-indicator) in conjunction with the maximal absolute standardized bias ($\widehat{B}_m$). Furthermore, the impact of the different survey designs on the balance of the response across different subgroups in each ethnic group will be assessed via partial R-indicator analysis (Shlomo, Skinner, Schouten, Carolina, & Morren, 2009). These results will be used to gain insight on how to further improve fieldwork. It is important to note that the study of underrepresented subgroups in a response sample, given a certain survey design, is different from estimating the effect of separate response enhancing measures on the propensity to respond among various subgroups.

Final disposition codes. The complement of the response rate is the nonresponse rate. The nonresponse rate can be used to gauge at the potential for nonresponse bias, specifically the underlying mechanism for nonresponse (Groves, 1989; Lynn, Sturgis, Clarke, & Martin, 2001; Stoop, 2005). Refusing to participate or not being able to participate are two different causes of nonresponse and offer an additional insight on the potential for nonresponse bias. Process or paradata information can be used to evaluate how well a specific set of survey design features is able to accurately survey our population of interest.

One way to gain insight is by analysing the final disposition code of nonresponding sample units. There are several main reasons for nonresponse, such as refusal, noncontact, not available, not able, language problems, moved, etc. Each of these reasons may be caused by a specific difficulty of surveying non-Western or ethnic minorities in The Netherlands, which in turn provides insight in the way the response sample reflects our population of interest. Furthermore, this specific information can be used to assess the probability of nonresponse bias for survey items if there is a known relation between the topic of interest and a specific cause for nonresponse. An example might be the correlation that exists between employment status and language problems or functional illiteracy. If persons are not able to speak and/or write Dutch, their chance on having a job in The Netherlands decreases. Another example would be the correlation between home ownership and high mobility. It is fair to say that the probability of a highly mobile person being a home owner is rather low. Nonresponse due to moved sample units varies between non-Western minorities and native Dutch. Non-Western minorities, especially Antilleans, move around more often than native Dutch (Feskens, 2009). This difference will increase if the fieldwork period is longer. So, if a specific set of survey design choices leads to an underrepresentation or exclusion of certain subgroups, the response

Table 2 Sample units selected for face-to-face CAPI re-issue in SIM2006 and SIM2011

	Number of nonresponding sampled persons selected for re-issue		Number of sampled persons not re-issued or with no declared final disposition code		Number of sampled persons re-issued with a final disposition code	
	2006	2011	2006	2011	2006	2011
Turkish	250	346	108	288	142	58
Moroccans	217	242	102	234	115	8
Surinamese	413	453	214	227	199	226
Antilleans	263	485	98	303	165	182
Total	1143	1526	522	1052	621	474

sample will not give an accurate reflection of the population of interest. Survey design choices such as the decision not to use bilingual interviewers or translated questionnaires will cause a high nonresponse rate due to language problems or functional illiteracy. Even if the composition of the response sample is similar to the population of interest with respect to correlated background characteristics, such as age and immigration generation, the underrepresentation of subgroups with language problems may cause biased estimates.

Analysing final disposition codes is straightforward and the appeal of this method is the ease with which it can point out potential nonresponse biases as well as provide insight for the development of new tailor made approach strategies. Furthermore it uses more information than just the response rate in order to judge the quality of the response sample.

Representativity-indicator and the maximal absolute standardized bias. The representativity-indicator (R-indicator) is a measure that describes how well the response sample reflects (i.e., how representative it is for) the population of interest, based on a certain number of background variables (J. Bethlehem et al., 2011; Schouten & Cobben, 2007, 2008; Schouten et al., 2009). Obviously, this representativity only applies to the variables included in the model for estimating this measure. One very important prerequisite is that the R-indicator needs complete (frame) data on all sample members: respondents and nonrespondents. This might not always be available. The R-indicator evaluates the differences in the estimated average response propensities between all strata, based on the variables included in the model from the available frame data. Obviously, the individual response propensities are unknown and the fewer distinct strata used to estimate the average response propensities, the less informative the R-indicator tends to be. Response is considered representative if the response propensities are constant across the sample which corresponds to a missing completely at random mechanism (Andridge & Little, 2011, 154).

In essence one can view it as a measure that uses the variability between nonresponse adjustment weights. The larger the variability is in nonresponse adjustment weights, the lower the R-indicator will be.

The R-indicator is useful in a variety of ways. First of all, it allows for the comparison of surveys, provided the same variables are available to estimate the model for each survey. Secondly, it is easy to interpret. It is one single estimate between zero and one (or 0% and 100%). Zero means a complete lack of representativity and one means a perfect fit. Thirdly, it can be used to monitor the progress of fieldwork and make more informed decisions on when and how to intervene. Fourthly, it can assist in designing a survey and provide an estimate of the quality while constraining other important parameters such as time and budget. Finally, Schouten et al. (2009, 107) show that "the R-indicator can also be used to set upper bounds to the non-response bias and to the root mean square error (RMSE) of adjusted response means."

For the estimation of the maximal absolute standardized bias ($\widehat{B}_m$) Schouten et al. (2009) make use of the proof provided by J. G. Bethlehem (1988) and Särndal and Lundström (2005) that the bias of the Horvitz-Thompson estimator is approximately equal to the population covariance between survey items and the response probabilities divided by the mean response probability. The following equation (1) from J. Bethlehem et al. (2011) shows the relation between the (estimated) average response probabilities ($\widehat{\overline{\rho}}$), the R-indicator $\widehat{R}\left(\widehat{\overline{\rho}}\right)$, the estimated standard deviation of the survey item, $\widehat{S}(y)$ and the maximal absolute bias $\widehat{B}_m\left(\widehat{\overline{\rho}}, y\right)$.

$$\widehat{B}_m\left(\widehat{\overline{\rho}}, y\right) = \frac{\left(1 - \widehat{R}\left(\widehat{\overline{\rho}}\right)\right)\widehat{S}(y)}{2\widehat{\overline{\rho}}} \tag{1}$$

For an unambiguous comparison, J. Bethlehem et al. (2011) propose to use a hypothetic survey item with a known and equal variance, for example $\widehat{S}(y) = 1$. This results in the estimated maximal absolute *standardized* bias:

$$\widehat{B'}_m\left(\widehat{\overline{\rho}}, y\right) = \frac{\left(1 - \widehat{R}\left(\widehat{\overline{\rho}}\right)\right)}{2\widehat{\overline{\rho}}} \tag{2}$$

The $\widehat{B}_m\left(\widehat{\overline{\rho}}, y\right)$ presented in equations (1) and (2) is an estimate of the upper non-response bias for a hypothetical survey item under the scenario that nonresponse correlates maximally with the selected auxiliary variables (J. Bethlehem et al., 2011, 186).

Unconditional and Conditional partial R-indicators. Sometimes certain socio-demographic subpopulations can be

expected to have a different position or opinion on important research topics such as having a job or the attitude towards socio-cultural integration. When they are underrepresented in the response sample, the results with respect to these research questions may be biased. It is therefore important to see how such subpopulations are represented in the response sample, given a certain survey design. We shall use partial R-indicators to check for the over or underrepresentation of subpopulations in the response sample (Schouten, Luiten, Loosveldt, Beullens, & Kleven, 2010; Schouten, Shlomo, & Skinner, 2011; Schouten et al., 2013; Shlomo et al., 2009). These subpopulations can be determined based on variables included in the model used to estimate the R-indicator. A partial R-indicator on a variable level shows the contribution of a specific background variable to the overall lack of representativity of the response sample.

There are unconditional and conditional partial R-indicators for discrete variables. The unconditional partial R-indicator on a variable level can be used to compare between surveys (Shlomo et al., 2009, 7). It measures the variability of the response propensities between the different categories of a variable. The larger the variability, the greater the contribution to the lack of representativity. This indicator is nonnegative and bounded above by 0.5 (Schouten et al., 2011, 236).

The conditional partial R-indicator on a variable level measures the contribution of a variable to the lack of representative response, adjusted for the impact of the other variables included in the model (Schouten et al., 2011, 237). It tries to isolate the part of the nonrepresentative response that is attributable to a specific variable. The conditional partial R-indicator on a variable level can take on any value in the interval [0, 0.5.]

Both partial R-indicators can also be calculated on a category level to ascertain the contribution to the lack of representative response separately for each category. The values of the unconditional partial R-indicators on a category level can be positive and negative. A negative value indicates an underrepresented category and a positive value indicates an overrepresented category. The unconditional partial R-indicators on the category level may take values between -0.5 and 0.5, where 0 means no contribution (Schouten et al., 2011, 236).

The values of the conditional partial R-indicator on the category level are always positive and show the conditional contribution of a category to the lack of representative response. The higher the value the larger the contribution of the category to the lack of representativity; the values range from 0 to 0.5.

4. Results of the different quality indicators

Final disposition codes: response rate and nonresponse composition

In this part, the paradata used are the final disposition code of the sample units. Table 3 presents the breakdown for ethnicity in final disposition code of the sample units for each survey. Here we use the AAPOR definition 1 (RR1), the minimum response rate (AAPOR, 2011).[5] Among Moroccans, there is a significantly higher response rate in SIM2011 compared to SIM2006. The other three ethnic groups show no significant difference in response rates over time. This indicates that the survey design used in the SIM2011 measurement might have successfully counteracted the general trend of decreasing response rates (De Heer & De Leeuw, 2001).

When we use the information from the final disposition code to judge which of the samples reflects the population of interest, we can draw four general conclusions with respect to the (planned) different survey design choices. First of all, the survey with the longest fieldwork period (SIM2006) suffers more from an outdated sample frame due to *moving* (Table 3). This can cause quite significant nonresponse among non-Western minorities. The second conclusion is that the targeted use of bilingual interviewers with a common ethnic background in SIM2006 still resulted in a higher exclusion of sampled persons among Turkish and Moroccans due to *language problems* compared to the complete use of bilingual interviewers with a common ethnic background in the SIM2011 survey. Thirdly, the timing of the fieldwork in SIM2006 caused a greater number of Turkish and Moroccans sampled persons to be *unavailable during fieldwork*, despite the longer fieldwork time and the larger number of re-issued unsuccessful sampled units.

These specific design choices made for the survey SIM2006 caused nonresponse among approximately 10.7% (5.7 plus 2.6 plus 2.4) of the eligible sample among Turkish compared to 3.6% in SIM2011 (Table 3). The same goes for the Moroccan sample which misses out on 10.5% in SIM2006 because of nonresponse due to survey design choices versus 5% of the total eligible sample in SIM2011. The difference is smaller, but similar for the Surinamese and there is hardly any difference between both samples for the Antilleans.

Fourthly, there are also large and unexpected differences found in both *noncontact* rates and the final disposition code *'other, no final disposition code'* for all groups between both surveys. These outcomes are related. The main reason for the correlation is that in the SIM2006 re-issue phase a high number of non-contacts were re-issued, but never got exhaustively re-contacted before fieldwork ended. For those cases, the final disposition code 'no final disposition code' was declared.

Also, in the first fieldwork phase in SIM2006 a few sampled persons never received a final disposition code, because they were not contacted the minimum number of times. The majority of these 'still not exhaustively contacted' outcomes were noncontacts up to that point. The main reasons for not following up these cases completely was due to lack of capacity (too high a workload for the interviewer) and inter-

[5] This definition was slightly adapted for the Dutch situation since the AAPOR guidelines do not provide for In Person Surveys of Specifically Named Persons.

Table 3 Final disposition code (in %) per ethnic group per survey year

	Turkish		Moroccans		Surinamese		Antilleans	
	2006	2011	2006	2011	2006	2011	2006	2011
Interview (RR1)	52.9	52.1	43.8*	48.0*	40.1	41.0	46.2	44.2
Moved	5.7	2.9	5.8	4.2	6.6	4.7	8.4	7.3
Language problem	2.6	0.1	1.8	0.3	0.2	0.2	0.3	0.1
Unavail. during fieldw. period	2.4	0.6	2.7	0.5	2.2	1.6	1.5	2.0
Non-contact	10.1	20.7	16.3	22.4	20.6	28.4	19.2	24.7
Refusal	19.3	21.6	21.1	22.1	21.8	20.7	17.7	18.6
Sick, not able	1.4	0.9	1.4	1.0	2.1	2.0	1.2	1.1
Other, no final disposition code	5.6	1.1	7.2	1.5	6.4	1.6	5.6	2.0
Total eligible sample size	2142	1564	2359	1737	2656	1929	2181	1973
Ineligibles	20	1	31	3	22	1	31	1
Total sample	2162	1565	2390	1740	2678	1930	2212	1974

*$p<0.05$. Rounding differences can cause some columns not to add up to 100%.

Table 4 Variables and interaction terms included in the R-indicator model

Variables
Age group (15-24; 25-34; 35-44; 45-54; 55-64; 64+)
Sex (male; female)
Municipality size (large; midsize; small)
Immigration generation (first; second)
Interaction terms
Age × Municipality Size
Immigration generation × Sex
Immigration generation × Municipality Size

viewer unavailability (illness, holidays).

Finally, there are varying numbers of ineligibles between both surveys. The main reason for this is the pre-fieldwork check conducted by the fieldwork agency on the SIM2011 gross sample. Before the gross sample was issued to the interviewers, it was enriched with phone numbers of the sample units, if any could be found. This check also revealed ineligible sample units such as sample units that moved abroad, frame errors etc.

Representativity and the upper bounds of nonresponse bias among the response samples

In this section, the paradata used are the auxiliary sample frame variables. The R-indicator tells us how representative the response composition of the net sample is compared to the gross sample with respect to several specific background variables (Schouten et al., 2009). This representativity is then expressed as a single number. The variables and interaction terms used in our R-indicator model are presented in Table 4.

The choice of variables included in the model was based on the availability of socio-demographic variables in the sample frame. No other complete frame data was available to be included in the analysis. The inclusion of interactions was based on our interest in whether or not specific difficult to survey subgroups, such as young persons living in large cities, first generation male immigrants and first generation immigrants living in large cities, were better represented using the set of design choices present in the survey design of SIM2011.

The results of the 'representativity' analysis of the response composition of the response samples show that achieving a higher response rate (RR_1) does not necessarily result in a more representative sample ($\widehat{R}$); see Table 5.

The $\widehat{B}_m$ takes into account both the response rate and the response composition with respect to the variables in the model (equation 2). The combination of both indicators shows that the SIM2006 design leads to a more representative sample with a lower maximal absolute bias among the Turkish. The SIM2011 design leads to a more representative sample with lower maximal absolute standardized bias among Moroccans, Surinamese and Antilleans.

The evolution of sample representativity in the first and second phase of fieldwork.

In this section, the paradata used are the intermediary fieldwork disposition code of the sample units and the auxiliary sample frame variables. The evolution of the sample representativity after each face-to-face contact attempt (CA) in the first phase and the effect of the re-issue phase (RI) can be monitored to assess the impact or usefulness of each additional CA on the sample representativity. Of course this representativity is conditional on the previous steps, but since this is done for both surveys and separately for each of the four ethnic groups, consistent outcomes can be interpreted with more certainty.

Figure 1 shows the evolution of the response rate and the R-indicator for both surveys after each face-to-face CA including the RI separately for each of the four non-Western minority groups. The first contact attempt is indicated by

Table 5 AAPOR Response rate 1 (RR_1), R-indicator ($\widehat{R}$) and 95% CI, and estimates for the maximal absolute standardized bias ($\widehat{B}_m$) for each ethnic group in SIM2006 and SIM2011 (in %) based on the model presented in Table 4

	Turkish		Moroccans		Surinamese		Antilleans	
	2006	2011	2006	2011	2006	2011	2006	2011
RR_1	52.9	52.1	43.8	48.0	40.1	41.0	46.2	44.2
$\widehat{R}$	86.0	80.5	81.7	85.7	83.6	86.6	80.3	85.6
$CI_{95}^{\widehat{R}}$	85.4–86.6	79.5–81.4	81.1–82.2	84.5–87.0	83.0–84.1	85.5–87.8	79.6–80.9	84.9–86.2
$\widehat{B}_m$	13.2	18.8	21.0	14.8	20.6	16.4	21.4	16.4
N[a]	2142	1564	2359	1737	2656	1929	2181	1973

[a]Based on all eligible cases

1st, the second by 2nd, etc. Five or more contact attempts are indicated by 5+ and the re-issue is indicated by RI. The corresponding cumulative response rate and R-indicator are presented as dots for SIM2006 and SIM2011.

For the Turkish sample in SIM2006 an interesting pattern is revealed. Each additional contact attempt (CA) in the SIM2006 increases the representativity of the sample. In this case a higher response rate does seem to indicate a better quality sample. Also, the targeted re-issuing was successful, improving the representativity of the sample as well. The effect of additional CAs among Turkish in SIM2011 is somewhat different. After each additional CA during the 1st phase, the representativity decreases slightly to end a little under 80%, despite the increase in response rate after each CA. Also for this survey the re-issue (RI) has a positive effect on the representativity of the response sample. The effect during the first phase, starting at a high level, followed by a slow descent and then stabilizing is not uncommon for the evolution of the R-indicator (see for example Schouten & Cobben, 2007, 2008). As there are only few respondents, none of the subgroups based on the model used to estimate the R-indicator can be very over- or underrepresented in comparison with the other strata.

For the SIM2006 study among Moroccans, the fourth CA and the RI clearly have a positive effect on the R-indicator. This pattern is different from the SIM2011 pattern with its quick convergence. Among Moroccans in SIM2011, the additional CAs during the first phase after the second CA do not increase the R-indicator by much and the optimum seems to be just below 86%. Since there was hardly a RI among Moroccans in the SIM2011, it is clear that the impact is marginal (see also Table 1).

Among Surinamese, both SIM2006 and SIM2011 show the same pattern. After each of the first three CAs in the first phase, there is a significant increase in response rate, but also a drop in representativity. From the fourth CA the representativity stabilises and reaches its optimum, given the design features in this phase. The RI only increases the representativity slightly.

Both SIM2006 and SIM2011 show a similar pattern among the Antilleans. It is also very similar to the pattern among Surinamese. After each CA during the first phase the response rate increases, but the representativity decreases. It looks as if the interviewers are focusing their attention on the 'easy' respondents during the first fieldwork phase. The second phase clearly has a stabilising effect here.

Overall this analysis shows that a re-issue has a positive or at least stabilizing effect on the representativity of the sample in comparison to the level of representativity at the end of phase one. This already happens with quite modest re-issuing. It seems that an extended first phase makes interviewers eventually target cases with the highest probability of success, which increases response rate, but does not (necessarily) increase the representativity of the sample. A re-issue increases the representativity, probably because equal attention is again given by the new interviewer to all available sample units in the interviewers assignment.

The re-issue strategies differed between SIM2006 and SIM2011. In 2006 only underrepresented subgroups got re-issued to another interviewer and they received the same conditional incentive. In 2011 there was no targeted selection of underrepresented subgroups in the RI and the incentive was increased.

Despite the limited RI in both SIMs, there are some interesting differences caused by the different RI strategies (Table 6). It is quite clear that, in terms of response rate, the RI was much more successful in SIM2011. Also, re-issuing seems to have been more successful among Turkish than the other ethnic groups. Almost half of the re-issued cases were converted among the Turkish in 2011. However, since the increased incentives and non-targeted RI in 2011 are confounded, it cannot be determined which of the two contributed more to the increased response.

The more successful RI in 2011, in terms of response rate, does not seem to result in an equal increase in representativity. In relative terms, it appears that the less successful RI in 2006 actually had a slightly larger, positive impact on the representativity.

The evolution of the maximal absolute standardized bias in the first and second phase of fieldwork

The R-indicator shows one part of the picture, but the response rate needs to be taken into account as well in order to get an appreciation of the potential nonresponse related bias for a particular survey item. The R-indicator and the

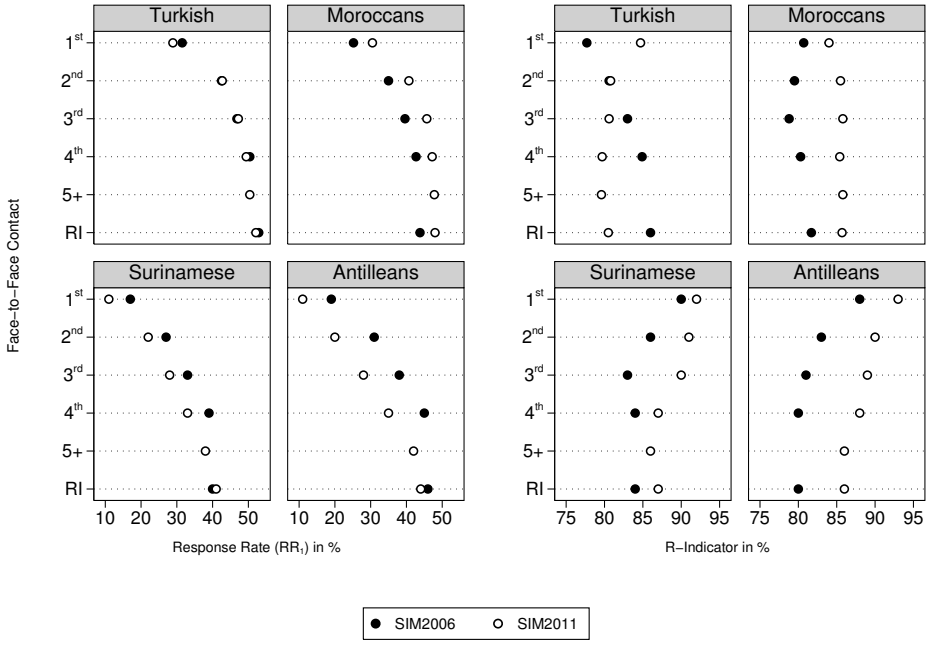

Figure 1. The evolution of the response rate and R-indicator after each face-to-face contact attempt in both surveys separately for Turkish, Moroccans, Surinamese and Antilleans.

Table 6 The actual number of re-issued sample persons and the number of successful interviews in SIM2006 and SIM2011 per ethnic group.

	Actual number of sample units re-issued with a final disposition code		Number of achieved interviews	
	2006	2011	2006	2011
Turkish	142	58	53	25
Moroccans	115	8	24	4
Surinamese	199	226	27	53
Antilleans	165	182	36	53
Total	621	474	140	135

response rate are used to calculate the $\widehat{B}_m$ (see formulae 1 and 2), which serves as an estimate for the upper bound nonresponse bias on a particular survey item given the sample. Here the $\widehat{B}_m$ estimate is calculated after each contact attempt during the first phase and after the RI to show how these design features influence the upper bound nonresponse bias on a particular survey estimate. Since all these measures are part of a system of design features, the impact can only be assessed depending on the sequence preceding the measure. However, similar changes in surveys with different designs offer additional weight in evaluating the effect of each CA and a RI on the potential for nonresponse bias on survey estimates among ethnic groups.

Figure 2 presents how the $\widehat{B'}_m$ estimate in both SIM designs changes after each face-to-face CA and the RI separate for each ethnic group. The first contact attempt is indicated by 1, the second by 2 etc. Five or more CAs are indicated by 5+ and the re-issue is indicated by RI.

The evolution of the $\widehat{B'}_m$ estimate during the first phase of fieldwork in SIM2006 shows a slightly different picture for all four the non-Western minority groups. In this survey design, each additional CA during the first phase results in a reduced $\widehat{B'}_m$ estimate and there seems to be no converging to a local minimum in the first phase. The evolution of the $\widehat{B'}_m$ estimate also shows a positive effect of the RI among all groups.

Figure 2 shows that the call strategy of SIM2011 stabilises to a local minimum in the first phase after the third or fourth CA among all groups. The subsequent contact attempts – up to 15 in the SIM2011 during the first phase – do not result in a much reduced potential for nonresponse bias on survey estimates despite the additional response. If figure 2 is compared with Figure 1 one can see this effect quite clearly among the Surinamese. Stopping after the third CA in the first phase and then starting the RI seems to be a more fruitful endeavour if one wants to reduce the upper bound nonresponse bias, given a fixed number of contact attempts. Also in this design the evolution of the $\widehat{B'}_m$ estimate shows a positive effect of the

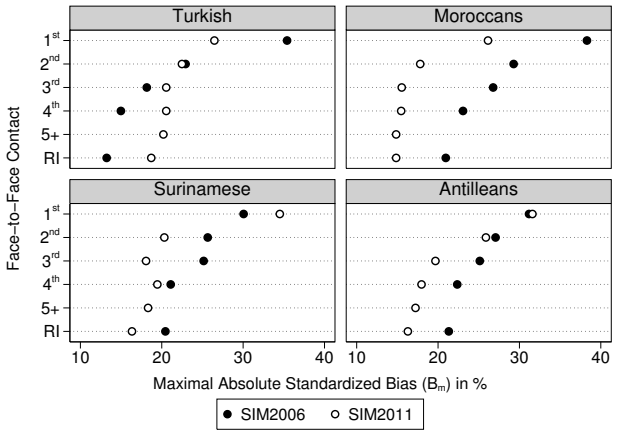

Figure 2. The evolution of the $\widehat{B'}_m$ estimate after each face-to-face contact attempt in both surveys separately for Turkish, Moroccans, Surinamese and Antilleans

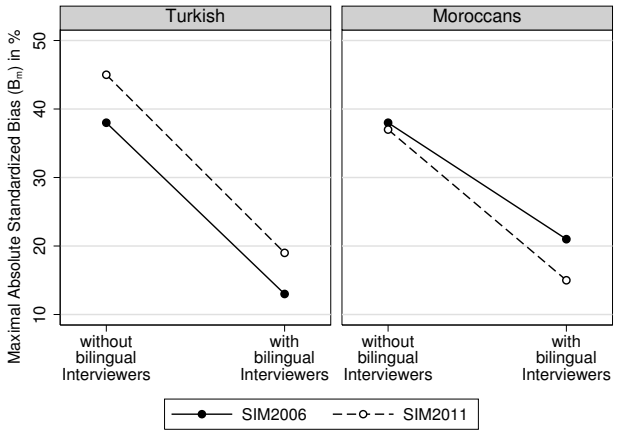

Figure 3. The estimated maximal absolute standardized bias $\widehat{B'}_m$ among the Turkish and Moroccans response sample with or without the use of bilingual interviewers.

re-issue, although among the Moroccans the re-issue phase was hardly implemented (Table 6).

The effect of bilingual interviewers with a common ethnic background on the potential for nonresponse bias on survey estimates among Turkish and Moroccans

The use of bilingual face-to-face CAPI interviewers with a common ethnic background was meant to reduce nonresponse due to language problems and functional illiteracy. Both reasons can still cause response rates to drop quite significantly especially among the first generation Turkish and Moroccans in The Netherlands. This can lead to biased estimates, since it excludes a very specific group. For Surinamese and Antilleans language problems are not seen as an important cause for nonresponse since, for many, Dutch is their mother tongue.

In this section, the paradata used are the interviewer observations about the respondent's ability to read or speak Dutch and the auxiliary sample frame variables. To find out to what extent bilingual interviewers are still necessary among Turkish and Moroccans, the interviewers were asked to fill out a short questionnaire. After each successful interview, they had to answer several questions about the language in which the survey was conducted, how they assessed the respondent's proficiency in Dutch, etc. These assessments on the respondent's ability to understand Dutch were used to estimate the number of respondents that would have been missed due to language problems if no bilingual interviewers were used. In our situation, if the interview was conducted (almost) completely in their native language and the interviewer also assessed that the level of Dutch of the respondent was (very) poor, we assumed that a respondent would have been a nonrespondent due to language problems in the absence of a bilingual interviewer. This corrected response rate excluding the potential language problems, in combination with the re-estimated R-indicator enables us to re-calculate the $\widehat{B'}_m$. The difference between the original and re-estimated $\widehat{B'}_m$ serves as an indicator for the effect that bilingual interviewers have on the potential for language problems related nonresponse bias on survey estimates (see Figure 3).

There is a marked increase in the potential for nonresponse bias on survey estimates if bilingual interviewers are not used. This holds across both ethnic groups and surveys. Without bilingual interviewers, the $\widehat{B}_m$ increased about 25 percentage points on average among Turkish and about 20 percentage points among Moroccans. If the representativity of the response sample (as indicated by the R-indicator) remained equal, the increase in $\widehat{B}_m$ should have been less than the decrease in response rate (see equation 2). However, the drop in response rates was on average about 13 percentage points among Turkish and 6 percentage points among Moroccans. This suggests that the increased $\widehat{B}_m$ is largely the result of a much more unbalanced sample. This, in turn, results in an increased potential for nonresponse bias on survey estimates.

The contribution of different subgroups to the lack of overall representativity

The lack of representativity as expressed by the R-indicator can also be partitioned into the contribution to lack of representativity of each variable included in the model to estimate the R-indicator. This is done by unconditional and conditional partial R-indicators. In this case, the larger the variation in the response propensities of a variable, the greater the contribution to the overall lack of representativity.

The unconditional and conditional variable level partial R-indicators were calculated for the variables age group, sex, municipality size and immigration generation. The unconditional partial R-indicators allow for a comparison between

surveys and the conditional partial R-indicators show the unique contribution of a variable to the variability in response propensities within a survey and ethnic group, after controlling for the other variables in the model. For both indicators, the contribution of each variable to the lack of representative response is shown separately for each survey and ethnic group (Table 7).

Among Turkish and Moroccans, the unconditional partial R-indicator shows the largest variation in response propensities for age group. This is true for both SIM2006 and SIM2011. In the Turkish samples in both surveys, the second largest contribution comes from sex. Among the Moroccans, it comes from immigration generation, that is the imbalance of response propensities between first and second immigration generation.

In the Surinamese and Antillean samples in both surveys, the unconditional partial R-indicator shows municipality size as the largest contributor to the variation in the response. In both surveys the second largest contribution comes from age group among Surinamese, whereas among Antilleans, it varies per survey: in SIM2006 it is sex and in SIM2011 it is age group.

The unconditional and conditional partial R-indicators at the variable level differ in size among Turkish and Moroccans for both surveys. This means that the variables included in the model are correlated among the Turkish and Moroccan samples. In SIM2006 the contribution of immigration generation to the variation in response propensities decreases among the Turkish and Moroccans after conditioning on the other variables. Also the contribution of age group is less after conditioning, especially among the Moroccans. The conditional partial R-indicators show that, after conditioning, the two largest contributions come from age group and sex in both groups.

In SIM2011 the variables also show collinear response behaviour among the Turkish and the Moroccans. However, in this instance, the contribution of age group and immigration generation to the variation in response propensities increases after conditioning on the other variables. After conditioning, the two largest contributors among Turkish and Moroccans are age group and immigration generation.

Among Surinamese and Antilleans there is not much difference in contribution between the unconditional and the conditional partial R-indicators at the variable level in both surveys. This means there is no strong collinear response behaviour and the variables have a unique and separate impact on the representativity of the response samples among Surinamese and Antilleans.

Partial R-indicators were also estimated at the category level. These estimates can provide additional insight on how to improve on an existing survey design for a specific ethnic group by identifying under and overrepresented subgroups. The category level can also consist of categories based on an interaction of variables (Schouten et al., 2011, 236). We analysed a mix of two separate, single variable category level indicators and one category level indicator based on a combination of two variables (Table 8). Based on the conditional variable level results, the two variables that contributed the most to the variation in the response propensities were included in the interaction (Table 7). As a result, the interaction-category level indicators can vary between surveys and or between ethnic groups. The remaining two variable category level indicators were calculated separately for each variable.

Table 8 Overview of the models used to estimate the partial R-indicators by survey and ethnic group

SIM	Estimation model
Turkish	
2006	Immigration generation + Municipality size + Age group × Sex
2011	Sex + Municipality size + Age group × Immigration generation
Moroccans	
2006	Immigration generation + Municipality size + Age group × Sex
2011	Sex + Municipality size + Age group × Immigration generation
Surinamese	
2006	Sex + Immigration generation + Age group × Municipality size
2011	Sex + Immigration generation + Age group × Municipality size
Antilleans	
2006	Age group + Immigration generation + Sex × Municipality size
2011	Sex + Immigration generation + Age group × Municipality size

For ease of interpretation, the category level results are shown separately for Turkish and Moroccans, on the one hand, and Surinamese and Antilleans, on the other hand. This is done because models to estimate the partial R-indicators at the category level are similar between Turkish and Moroccan in both SIM2006 and SIM2011 (Table 8). There is also great similarity in the models used to estimate the partial R-indicators at the category level between the Surinamese and Antillean samples in both SIM2006 and SIM2011.

The unconditional and conditional category level results show that the single largest contribution to the variation in response propensities among the Turkish 2006 sample comes from the overrepresentation of women in the age category of 35 to 44 (Table 9)[6].

Among the Moroccan 2006 sample there are more subgroups with a relatively large contribution (with a conditional contribution of over 20) to the variation in response propensities. These are the 15 to 34 year old men and 55 to 64 year old women, who are underrepresented, and the 35 to 54 year old women and men aged above 64, who are overrepresented.

It is interesting to note that, while the SIM2006 design sometimes leads to similar subgroups in the Turkish and Mo-

[6] Confidence intervals were also approximated using 1000 bootstrap replicates of the estimates and excluding the 25 highest en lowest estimates and can be delivered upon request by the author.

Table 7 The unconditional and conditional variable level partial R-indicators, separate for each ethnic group and time of the survey (multiplied by 1000)Sample units selected for face-to-face CAPI re-issue in SIM2006 and SIM2011

	Turkish		Moroccans		Surinamese		Antilleans	
	2006	2011	2006	2011	2006	2011	2006	2011
Unconditional								
Age group	33.2∗	51.4	63.4∗	20.0	42.5∗	29.8	37.5	34.8
Sex	26.1∗	37.6	37.9∗	6.9	36.2∗	4.6	48.5∗	11.0
Municipality size	7.3∗	26.3	22.4∗	15.8	53.8∗	44.9	62.9∗	45.6
Immigration generation	24.7∗	32.0	38.2∗	17.5	12.1∗	1.1	14.4∗	3.4
Conditional								
Age group	28.0	60.5	51.7	23.8	41.3	31.1	36.3	37.6
Sex	27.3	36.5	40.0	5.3	35.7	5.2	50.6	11.9
Municipality size	6.4	30.5	22.3	15.3	53.4	45.5	62.4	46.8
Immigration generation	15.2	45.0	4.7	21.6	2.0	1.4	13.2	3.1
N[a]	2142	1564	2359	1737	2656	1929	2181	1973

[*] $p<0.05$ between SIM2006 and SIM2011 within ethnic groups based on confidence intervals (not included here) that were approximated using 1000 bootstrap replicates of the estimates and excluding the 25 highest en lowest estimates.
[a] Based on all eligible cases.

roccan sample, such as, for instance, 15 to 24 year old males, being under (or over)represented, it also shows differences in representation of certain subgroups, such as 55 to 64 year old females, between the two samples.

As expected, the Turkish 2011 sample shows more subgroups with a large contribution to the variation in response propensities. These groups are the overrepresented women, persons living in midsize cities and first generation Turkish in the age of 15 to 24 and the underrepresented men, persons living in small municipalities and second generation Turkish in the age of 25 to 34.

Among the Moroccan 2011 sample, the underrepresented first generation immigrants aged 25 to 34 contribute the most. The complete lack of similar under and overrepresented subgroups between the Turkish and the Moroccan 2011 sample is also quite notable.

Table 9 also shows that the Turkish and Moroccan sample did not contain any second generation immigrant above the age of 44. This was to be expected since the Turkish and Moroccan immigration only started in the mid-sixties of the last century. The immigrants were mostly men who came to The Netherlands for work. Partner reunification only started in the mid-seventies.

The Surinamese 2006 sample shows that the largest contributions to the nonrepresentative response come from the overrepresentation of women and youngsters living in midsize and small municipalities and the underrepresentation of men and 25 to 44 year old big city dwellers (Table 10)[7]. In the Surinamese 2011 sample, the largest contributions come from the underrepresentation of 25 to 44 year old big city dwellers and the overrepresentation of youngsters living in small cities.

The category level indicators also reveal that not only the sex balance has improved in the SIM2011 sample, but also that 25 to 34 year old big city inhabitants are less underrepresented and youngsters living in midsize cities are less overrepresented compared to the SIM2006 sample. The different survey design choices made for the SIM2011 survey seem to be effective in reducing heavily over and underrepresented subgroups.

The results for the Antillean 2006 sample show that the largest contributions to the variations in response propensities come from the underrepresentation of men living in the big cities and 25 to 34 year old persons and the overrepresentation of women living in midsize and small municipalities (Table 10).

The 2011 sample shows the largest contribution coming from the underrepresented big city dwellers aged 25 to 34 and the overrepresented youngsters and persons between the ages 35 to 54 living in midsize cities. It is also interesting to see that the SIM2011 design leads to quite a few differences in the over and underrepresented subgroups among Antilleans compared to the Surinamese.

The results of the variable and category level partial R-indicators analysis have shown which groups are over- and underrepresented among the different ethnic groups in the SIM2006 and SIM2011 survey. The analyses have shown that different subgroups are under and overrepresented across the various ethnic groups and surveys. This means that the survey design and the characteristics of the population under study cannot be viewed as separate entities that affect the likelihood of response, but should be viewed as an interactive system. For instance, if one takes the SIM2011 design as a basis to conduct another survey among the same four ethnic groups, varying targeted data collection strategies should be developed depending on the ethnic group, but these strategies for the same ethnic groups would be different if the SIM2006

[7] Confidence intervals were approximated using 1000 bootstrap replicates of the estimates and excluding the 25 highest en lowest estimates.

Table 9 Unconditional partial R-indicators on category level, separate for Turkish and Moroccans for SIM2006 and SIM2011 (multiplied by 1000)

	Unconditional				Conditional			
	2006		2011		2006		2011	
	Turkish	Moroccans	Turkish	Moroccans	Turkish	Moroccans	Turkish	Moroccans
Age × Sex								
15-24 Male	−22.2	−36.3	–	–	14.0	22.4	–	–
15-24 Female	−2.5	−5.3	–	–	5.6	10.1	–	–
25-34 Male	−15.6	−28.2	–	–	13.9	29.2	–	–
25-34 Female	5.9	−0.9	–	–	7.2	12.2	–	–
35-44 Male	8.1	6.7	–	–	5.1	8.1	–	–
35-44 Female	25.2	30.8	–	–	22.4	27.3	–	–
45-54 Male	3.9	11.8	–	–	0.9	5.7	–	–
45-54 Female	14.2	32.0	–	–	10.6	24.4	–	–
55-64 Male	−7.9	15.1	–	–	11.6	7.9	–	–
55-64 Female	2.9	−27.5	–	–	1.9	21.2	–	–
64+ Male	−10.4	18.4	–	–	14.3	24.3	–	–
64+ Female	−0.8	−3.4	–	–	4.0	9.0	–	–
Age× Immig.Gen								
15-24 1G	–	–	32.4	−8.1	–	–	32.2	8.5
15-24 2G	–	–	7.9	13.6	–	–	8.1	13.8
25-34 1G	–	–	−3.9	−23.5	–	–	4.9	22.4
25-34 2G	–	–	−49.3	4.1	–	–	49.4	3.2
35-44 1G	–	–	−8.6	3.4	–	–	8.7	3.4
35-44 2G	–	–	−18.8	6.9	–	–	18.4	7.2
45-54 1G	–	–	13.8	5.4	–	–	15.2	5.1
45-54 2G	–	–	*n.a.*	*n.a.*	–	–	*n.a.*	*n.a*
55-64 1G	–	–	14.9	−0.6	–	–	13.8	0.6
55-64 2G	–	–	*n.a*	*n.a*	–	–	*n.a*	*n.a*
64+ 1G	–	–	17.6	−0.3	–	–	17.2	1.1
64+ 2G	–	–	*n.a*	*n.a*	–	–	*n.a*	*n.a*
Municipality Size[a]								
Large	−5.7	−0.7	−4.7	0.1	5.0	1.0	8.4	1.1
Medium	2.9	−12.9	16.5	8.2	3.5	13.4	20.3	7.9
Small	3.5	18.3	−20.0	−13.4	1.8	17.8	21.2	13.0
Immigration Generation[b]								
1G	16.5	25.4	–	–	10.6	3.3	–	–
2G	−18.5	−28.5	–	–	10.9	3.4	–	–
Sex								
Male	–	–	−26.3	4.8	–	–	25.5	3.7
Female	–	–	26.9	−5.0	–	–	26.1	3.8
N[c]	2142	2359	1564	1737	2142	2359	1564	1737

[a]Large: municipality size >250,000; Medium: 250,000 -50,000; Small: <50000.
[b]1G: first generation immigrant; 2G: second generation immigrant.
[c]Based on eligible cases.

were to be used as a basis. In addition, when developing group dependent data collection strategies, one should not only look at the characteristics of the underrepresented subgroup, but also at their cause for nonresponse.

For example, to increase the representativity among a sample of Moroccans using the SIM2011 design, it is likely that one needs to increase the response among first generation immigrants in the age of 25 to 34. The characteristics of the subgroup tell us that these are people who have come to a new country and could be unfamiliar with the Dutch culture or language. They could have come to The Netherlands to get married or to find work. In order to improve the probability of response among this subgroup one can choose different methods, such as using a different data collection mode (i.e., CAWI in case the potential respondent is away during interviewer working hours or in case the potential respondent is

Table 10 Unconditional partial R-indicators on category level separate for Surinamese and Antilleans for SIM2006 and SIM2011(multiplied by 1000)

	Unconditional				Conditional			
	2006		2011		2006		2011	
	Surinames	Antilleans	Surinames	Antilleans	Surinames	Antilleans	Surinames	Antilleans
Immigration Generation[b]								
1G	8.6	−8.8	−0.7	−1.8	1.4	8.8	1.0	1.9
2G	−8.5	11.4	0.9	2.9	1.3	9.9	1.0	2.4
Age × Municipality size[a]								
15-24 in L	−3.8	–	−3.5	−12.4	1.6	–	4.0	12.5
15-24 in M	20.6	–	11.1	24.2	22.8	–	10.7	23.6
15-24 in S	21.2	–	21.3	0.0	23.1	–	21.1	0.2
25-34 in L	−40.4	–	−21.6	−33.7	38.5	–	21.8	33.6
25-34 in M	−5.9	–	−5.9	−5.9	5.5	–	6.0	6.5
25-34 in S	1.0	–	9.3	−13.3	1.9	–	9.3	13.8
35-44 in L	−21.4	–	−21.1	−8.2	21.8	–	21.1	8.3
35-44 in M	8.2	–	−4.2	20.6	9.1	–	4.1	20.8
35-44 in S	13.1	–	13.4	13.6	13.7	–	13.4	11.4
45-54 in L	−3.1	–	−8.1	−3.8	5.9	–	7.8	3.5
45-54 in M	19.8	–	6.8	22.5	16.8	–	6.9	23.1
45-54 in S	21.5	–	18.5	3.8	19.8	–	18.8	4.1
55-64 in L	−6.4	–	−6.6	−12.4	8.5	–	6.5	12.1
55-64 in M	12.1	–	4.3	8.9	10.3	–	4.5	9.2
55-64 in S	12.1	–	13.3	−3.3	11.4	–	13.4	3.1
64+ in L	−1.8	–	7.4	−5.9	4.0	–	7.6	5.4
64+ in M	11.9	–	12.9	8.6	10.3	–	13.1	8.4
64+ in S	11.1	–	15.2	−0.4	11.4	–	15.3	0.4
Sex								
Male	−26.3	–	3.3	−7.6	25.9	–	3.7	8.2
Female	24.9	–	−3.2	8.0	24.6	–	3.6	8.6
Sex × Municipality size								
Male in L	–	−55.9	–	–	–	54.5	–	–
Male in M	–	−4.2	–	–	–	4.9	–	–
Male in S	–	5.6	–	–	–	4.1	–	–
Female in L	–	−14.1	–	–	–	13.2	–	–
Female in M	–	44.7	–	–	–	44.7	–	–
Female in S	–	34.7	–	–	–	33.0	–	–
Age Group								
15-24	–	12.1	–	–	–	10.5	–	–
25-34	–	−29.9	–	–	–	30.6	–	–
35-44	–	5.3	–	–	–	10.3	–	–
45-54	–	9.3	–	–	–	15.6	–	–
55-64	–	14.1	–	–	–	3.2	–	–
64+	–	−7.5	–	–	–	2.4	–	–
N[c]	2656	1929	1973	2181	2656	1929	1973	2181

[a]Large: municipality size >250,000; Medium: 250,000-50,000; Small: <50,000.
[b]1G: first generation immigrant; 2G: second generation immigrant.
[c]Based on eligible cases.

unwilling to communicate with an interviewer), increasing the number of contact attempts, using higher incentives or sex matching the interviewer to the potential respondent. The relatively high noncontact rate among this subgroup would suggest that a sex match or increased incentives might not be the preferred tailored approach, but that another data collection mode or increasing the number of contact attempts might be more applicable.

Targeting a different subgroup using a different method would have been appropriate among the Turkish in the 2011 sample. In that case, 25 to 34 year old second generation immigrants were underrepresented and the refusal rate was relatively high. In the 2006 sample various other subgroups were underrepresented among the Moroccans and also the cause for nonresponse differed between the various subgroups.

5. Conclusion and Discussion

Surveying among non-Western minorities continues to be difficult, but focussing on other indicators instead of only the response rate as measures of quality might prove insightful in the pursuit of a more representative sample among non-Western minorities – or other populations for that matter. In this paper we focused on how different survey design choices affect the composition of the response sample and how this might relate to the occurrence of nonresponse bias on survey estimates in surveys among non-Western minorities in The Netherlands.

It is important to know about the survey related characteristics of the population of interest when designing a survey. Each design choice can potentially lead to the exclusion of target population members, therefore the more aware one is of these survey related characteristics, the more informed the tradeoff decision. Fieldwork disposition codes show that basic survey design decisions, such as the intended length and timing of fieldwork, can result in increased nonresponse among ethnic groups in The Netherlands, because of higher mobility among non-Western minorities and unavailability due to long holidays in the country of origin. Especially the use of a long fieldwork period increases the likelihood of nonresponse due to the fact a sampled person may have moved.

The results from the R-indicator analysis show that different survey designs lead to different levels of representativity of the response samples among non-Western minority groups in The Netherlands. Furthermore, the level of representativity seems to be uncorrelated with the response rate when the difference between response rates is significant. A higher response rate under these conditions does not necessarily result in a more representative sample.

The estimated maximal absolute standardized bias, where the R-indicator is used in conjunction with the response rate, shows that the potential for nonresponse bias on substantive outcomes can be quite substantial. This result raises concerns on the validity of results concerning non-Western minorities obtained from non-Western respondents in general population surveys, because less extensive measures are usually undertaken to reach non-Western minority groups.

All in all, the results have shown that it is possible, given the right survey design, to combat declining response rates and increase the quality of response samples in surveys among hard-to-reach populations, such as non-Western minorities This is even possible despite the potentially harmful effect of a more populist discourse on migrants on the willingness to participate in The Netherlands.

The impact of several survey design choices on the potential for nonresponse bias on survey estimates was also analysed in more detail. The results showed that the optimal number of face-to-face contact attempts in a multi-phase approach of non-Western minorities in The Netherlands is about three to four in the first phase. More contact attempts made in the first phase by the same interviewer do increase the response, but do not decrease the potential for nonresponse bias on survey estimates. Limiting the number of contact attempts to a maximum of four during the first phase before moving to a re-issue phase in which other design features can be introduced can potentially result in significant time and/or financial gain.

The re-issue phase, in which non-responding sampled persons were contacted by another interviewer, reduced the potential for nonresponse bias on survey estimates and increased the representativity of the response sample composition among all non-Western minorities samples. This is despite the fact of some serious shortcomings in the execution of the re-issue phase among all samples in the current study. All samples used in this study had a far more limited re-issue phase than initially planned. Let this serve as a reminder to always plan enough time to conduct a re-issue phase and to ensure the availability of enough interviewers. Even so, the analysis showed that even a quite modest re-issue had a positive effect on the sample composition. It is self-evident that if the re-issue had been fielded as intended, the response would have been higher and based on this analysis, the nonresponse bias of the survey estimates should also have been reduced.

Our study suggests that an increased conditional nonmonetary incentive during the re-issue phase does not necessarily result in a larger decrease of potential nonresponse bias among non-Western minority groups compared to keeping the conditional incentives at the same level. However, the effect of an increased incentive is most likely confounded with the way the re-issue in the SIM2011 design was designed. In the 2011design, re-issued persons did not necessarily belong to underrepresented subgroups. This is different from the targeted re-issue that was applied in the SIM2006 design. From a cost perspective and bias reduction point of view, it may be better not to use an increased conditional nonmonetary incentive and re-issue all non-responding sampled persons in the second phase of surveys among non-Western minorities, but rather to target underrepresented subgroups. Of course, one needs to be careful and realize that the maximal absolute standardized bias is only an indicator for nonresponse bias on survey estimates. Also, when targeting underrepresented subgroups a different payment scheme for interviewers might be in order to keep them motivated.

Interviewers with a common ethnic background remain of great importance in order to reach a balanced or representative sample among non-Western minorities. Obviously, the use of bilingual interviewers with a common ethnic background reduces the nonresponse due to language problems and also the potential for nonresponse bias on survey estimates. Especially among ethnic groups with known language problems, the possibility of quite severe nonresponse bias on survey estimates exists if one does not use bilingual interviewers. Reducing the potential for nonresponse bias on survey estimates by minimizing language problems is only one of the benefits of using interviewers with a common ethnic background. The results of the partial R-indicators also suggest that other difficult subgroups without any known language problems, such as young second generation Moroccan immigrants or Antillean men living in large cities, are also better represented and sometimes even overrepresented in the response samples among non-Western minorities when inter-

viewers with a common ethnic background are more extensively used. Of course, the effectiveness of interviewers with a common ethnic background is evaluated here in terms of potential for nonresponse bias on survey estimates, but this is only part of the survey cycle. Interviewers with a common ethnic background may also have a greater influence on the way respondent answers to survey questions compared to interviewers without a common ethnic background, which can lead to increased measurement bias. One should be aware of this trade-off.

When it comes to evaluating the effect of separate response enhancing measures in surveys it is important to note that in many circumstances analysis methods, such as logistic regression, give biased results because of non-random allocation of sample units to 'treatments'. Brehm (1993, 128-130) also sees this inherent problem in applying the continuum of resistance to reluctance. He combined a continuum of resistance with respect to accessibility and to cooperation in his approach to modelling the survey process, in which even more administrative measures (more calls, sending a letter to try and persuade reluctant sample persons, trying to convert a refusal) would increase survey participation. The difficulty he found with this model is that persuasion letters are only sent to reluctant respondents, and therefore seem to have a negative effect (as reluctant respondents more often turn into final refusers and no persuasion letters are sent to respondents who cooperate instantaneously). As he remarks in a footnote (p. 130): 'If one's interest lies in how effective these techniques are [], the persuasion letters and refusal conversions would have to be randomly assigned treatments, not treatments assigned on the basis of an initial refusal.'

It is also important to realize that one size does not fit all when designing a survey among different non-Western minority groups. The results of the unconditional and conditional partial R-indicators showed that there are significant differences in under and overrepresented subgroups depending on survey design and ethnic group. This is important to keep in mind when one is trying to assess whether non-Western minorities are well represented in a general population survey. An underrepresented socio-demographic subgroup among one ethnic group might be cancelled out by the overrepresentation of the same subgroup among another ethnic group. This will lead to a biased result if the two subgroups have different views or attitudes based on their culture or socio-economic status as an ethnic group.

Fieldwork strategies can be improved and tailored to reach hard-to-reach subgroups. The partial R-indicators in conjunction with the final fieldwork disposition codes provide a wealth of information for improving the representativity of a survey among different non-Western minority groups. They can tell us not only who to target, but also how we should target them.

Finally, the approaches used in the analysis provide us with additional insight on the quality of the response sample and on the occurrence of nonresponse bias at survey item level. However, one should keep in mind that these approaches use the information available at survey level to assess the potential for nonresponse bias at item level. However, nonresponse bias is item specific and not survey specific (Groves & Peytcheva, 2008). The predictive value of fieldwork disposition codes or the R-indicator in conjunction with the maximal absolute standardized bias based on auxiliary variables can be quite limited when estimating the actual size of the nonresponse bias, but the combination of these approaches will tell us more about the potential for nonresponse bias on survey estimates than the response rate alone.

Acknowledgements

I would like to thank E.D. de Leeuw, I.A.L. Stoop and some anonymous reviewers for their very helpful comments on previous drafts of this paper.

References

AAPOR. (2011). *Standard definitions: Final dispositions of case codes and outcome rates for surveys.* (7th ed.). Retrieved 20/07/2014, from http://www.aapor.org/AM/Template.cfm?Section=Standard_Definitions2&Template=/CM/ContentDisplay.cfm&ContentID=3156

Andridge, R. R., & Little, R. J. (2011). Proxy pattern-mixture analysis for survey nonresponse. *Journal of Official Statistics*, *27*(2), 153-180.

Bethlehem, J., Cobben, F., & Schouten, B. (Eds.). (2011). *Handbook of nonresponse in household surveys*. New York: Wiley.

Bethlehem, J. G. (1988). Reduction of nonresponse bias through regression estimation. *Journal of Official Statistics*, *4*(3), 251-260.

Biemer, P. P., & Lyberg, L. E. (2003). *Introduction to survey quality.* New York: Wiley.

Bijl, R., & Verweij, A. (Eds.). (2012). *Measuring and monitoring immigrant integration in Europe. 2012-8*. Den Haag: SCP.

Brehm, J. (1993). *The phantom respondents.* Ann Arbor: University of Michigan Press.

Couper, M. P. (2005). Technology trends in survey data collection. *Social Science Computer Review*, *23*, 486-501.

De Heer, W. (1999). International response trends: results of an international survey. *Journal of Official Statistics*, *15*(2), 129-142.

De Heer, W., & De Leeuw, E. D. (2001). Trends in household survey nonresponse: A longitudinal and international comparison. In R. Groves, D. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey nonresponse* (p. 41-54). New York: Wiley-Blackwell.

Dillman, D. A. (2007). *Mail and Internet surveys: The tailored design method.* (2nd ed.). New York: Wiley.

Feskens, R. C. W. (2009). *Difficult groups in survey research and the development of tailor-made approach strategies.* Retrieved 28/07/2014, from http://www.cbs.nl/NR/rdonlyres/8F317AA9-1074-4BF7-84EB-015D013DBB80/0/2009x11feskenspub.pdf (Doctoral dissertation, Centraal Bureau voor de Statistiek en Utrecht University.)

Feskens, R. C. W., Kappelhof, J., Dagevos, J., & Stoop, I. A. L. (2010). *Minderheden in de mixed-mode? Een inventarisatie van voor- en nadelen van het inzetten van verschillende dataverzamelingsmethoden onder niet-westerse migranten.* Den Haag: SCP.

Groves, R. M. (1989). *Survey costs and survey errors*. New York: Wiley.

Groves, R. M. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly*, *70*(5), 646-675.

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

Groves, R. M., & Heeringa, S. G. (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*, 439-457.

Groves, R. M., & Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias. *Public Opinion Quarterly*, *72*(2), 167-89.

Kappelhof, J. W. S. (2010). *Op maat gemaakt? Een evaluatie van enkele responsverbeterende maatregelen onder Nederlanders van niet-westerse afkomst.* Den Haag: SCP.

Kemper, F. (1998). Gezocht: Marokkanen. Methodische problemen bij het verwerven en interviewen van allochtone respondenten. *Migrantenstudies*, *1*, 43-57.

Korte, K., & Dagevos, J. (2011). *Survey Integratie Minderheden 2011 Verantwoording van de opzet en uitvoering van een survey onder Turkse, Marokkaanse, Surinaamse en Antilliaanse Nederlanders en een autochtone vergelijkingsgroep*. Den Haag: SCP.

Kreuter, F. (2013). *Improving surveys with paradata. Analytic uses of process information.* New York: Wiley.

Lynn, P., Sturgis, P., Clarke, P., & Martin, J. (2001). The effects of extended interviewer effort on non-response bias. In R. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey nonresponse.* (p. 135-148). New York: Wiley-Blackwell.

Maitland, A., Casas-Cordero, C., & Kreuter, F. (2009). *An evaluation of nonresponse bias using paradata from a health survey.* (Proceedings of the Section on Survey Research Methods of the American Statistical Association)

Moorman, P. G., Newman, B., Millikan, R. C., Tse, C. K. J., & Sandler, D. P. (1999). Participation rates in a case-control study: The impact of age, race, and race of interviewer. *Annals of epidemiology*, *9*, 188-195.

Reep, C. (2003). *Moeilijk Waarneembare Groepen. Een inventarisatie. Rep. No. H1568-03-s00.* Voorburg/Heerlen: CBS.

Särndal, C. (2011). Three factors to signal non response bias with applications to categorical auxiliary variables. *International Statistical Review*, *79*, 233-254.

Särndal, C., & Lundström, S. (2005). *Estimation in surveys with nonresponse.* Chichester, England: Wiley.

Schmeets, H. (2005). De leefsituatie van allochtonen. In H. Schmeets & R. van der Bie (Eds.), *Enqueteonderzoek onder allochtonen. problemen en oplossingen.* Voorburg/Heerlen: CBS.

Schothorst, Y. (2002). Onderzoek onder allochtonen: wat kan, wat moet en wat kan? In H. Houtkoop & J. Veenman (Eds.), *Interviewen in de multiculturele samenleving: problemen en oplossingen.* (p. 101-116). Assen: Koninklijke Van Gorcum.

Schouten, B., Bethlehem, J., Beullens, K., Kleven, ., Loosveldt, G., & Luiten, A. e. a. (2013). Evaluating, comparing, monitoring, and improving representativity of survey response through R-indicators and partial R-indicators. *International Statistical Review*, *80*, 382-399. Retrieved 20/07/2014, from http://www.risq-project.eu/papers/schouten-et-al-2012.pdf

Schouten, B., & Cobben, F. (2007). *R-indexes for the comparison of different fieldwork strategies and data collection modes.* Voorburg/Heerlen: CBS. Retrieved 20/07/2014, from http://www.risq-project.eu/papers/schouten-cobben-2007-a.pdf (Rep. No. Discussion Paper 07002.)

Schouten, B., & Cobben, F. (2008). *An empirical validation of R-indicators.* Voorburg/Heerlen: Statistics Netherlands. Retrieved 20/07/2014, from http://www.risq-project.eu/papers/cobben-schouten-2008-a.pdf (Rep. No. Discussion Paper 08006.)

Schouten, B., Cobben, F., & Bethlehem, J. (2009). Indicators of representativity of survey nonresponse. *Survey Methodology*, *35*, 101-113. Retrieved 20/07/2014, from http://www.risq-project.eu/papers/schouten-cobben-bethlehem-2009.pdf

Schouten, B., Luiten, A., Loosveldt, G., Beullens, K., & Kleven, . (2010). *Monitoring and changing data collection through R-indicators and partial R-indicators.* (Rep. No. RISQ deliverable 10)

Schouten, B., Shlomo, N., & Skinner, C. (2011). Indicators for monitoring and improving representativity of response. *Journal of Official Statistics,*, *27*, 231-253. Retrieved 20/07/2014, from http://www.risq-project.eu/papers/schouten-shlomo-skinner-2011.pdf

Shlomo, N., Skinner, C., Schouten, B., Carolina, T., & Morren, M. (2009). *Partial indicators for representative response.* Retrieved 20/07/2014, from http://www.risq-project.eu/papers/RISQ-Deliverable-4-V2.pdf (Rep. No. RISQ deliverable 4. version 2.)

Singer, E. (2001). The use of incentives to reduce nonresponse in household surveys. In R. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey nonresponse.* (p. 163-177). New York: Wiley-Blackwell.

Singer, E., Van Hoewyk, J., Gebler, N., Raghunathan, T., & McGonagle, K. (1999). The effect of incentives on response rates in interviewer-mediated surveys. *Journal of Official Statistics*, *15*, 217-230.

Singer, E., Van Hoewyk, J., & Maher, M. P. (2000). Experiments with incentives in telephone surveys. *Public Opinion Quarterly*, *64*, 171-188.

Stoop, I. A. L. (2005). *The hunt for the last respondent: Nonresponse in sample surveys.* Den Haag: The Netherlands Institute for Social Research/SCP,.

Veenman, J. (2002). Interviewen in multicultureel nederland. In H. Houtkoop & J. Veenman (Eds.), *Interviewen in de multiculturele samenleving: problemen en oplossingen.* (p. 1-19). Assen: Koninklijke Van Gorcum.

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.