Statistics Netherlands

Discussion Paper

The impact of contact effort and interviewer performance on mode-specific nonresponse and measurement bias

2014 | 05

Barry Schouten
Fannie Cobben
Jan van der Laan
Judit Arends

07-02-2014

The impact of contact effort and interviewer performance on mode-specific nonresponse and measurement bias

Barry Schouten, Fannie Cobben, Jan van der Laan en Judit Arends

Summary:

In 2011, a large-scale mixed-mode experiment was linked to the Crime Victimisation Survey (CVS). This experiment consisted of a randomized allocation of sample persons to the four survey modes Web, mail, telephone and face-to-face, and a follow-up using only interviewer modes face-to-face and telephone. The aim of the experiment was to disentangle mode-specific selection- and measurement effects.

In previous papers, mode-specific coverage, nonresponse and measurement biases have been reported for key variables from the CVS and the Labour Force Survey. In this paper, we go a step further and investigate the size of the nonresponse and measurement biases as a function of contact effort and historic interviewer performance. With contact effort we refer to the number of telephone calls, face-to-face visits and the number of reminders in web and mail. The interviewer performance is implemented as a simple binary indicator, below or above average performance, based on ratings by interviewer coordinators. In all analyses, face-to-face response based on a maximum of six visits, the default face-to-face strategy at Statistics Netherlands, is used as the benchmark.

The analyses show that contact effort has little impact on the size of measurement bias and a modest impact on the size of selection bias. Also, interviewer performance plays just a small role in the size of both biases. From these results, we conclude that contact effort and interviewer performance do not have a simultaneous impact on nonresponse and measurement error.

Keywords: Mixed-mode surveys; Total survey error; Level of effort; Interviewer effects.

1. Introduction

Data collection modes affect multiple sources of non-sampling errors in the data collection phase of a survey process. Under the total survey error framework (Biemer and Lyberg, 2003; Groves, 1989) all these errors are related and decisions to reduce one source of survey error may affect other sources of error. In single mode surveys, all these non-sampling errors together result in a discrepancy between the true value and the estimate for that value based on the survey answers (the survey outcome). In addition, when the same survey is administered using different data collection modes there will be differences in survey outcomes between survey modes. Both the discrepancy between the true and estimated value and the difference between the modes is the result of selection effects and measurement effects. Selection effects arise since different modes have different coverage and response rates. Measurement effects arise because different modes evoke different types of measurement errors during the process of reporting an answer.

In regular data collection work, it is impossible to disentangle mode-specific selection and measurement effects. When a difference between modes is observed, it is often not possible to determine what part of the difference is caused by a different selection of respondents, and what part of the difference can be attributed to different measurement processes between modes. Separation of selection effects from measurement effects in empirical studies requires carefully designed experiments in combination with weighting or regression based inference methods to control for selection effects, see e.g. Jäckle et al. (2010). As an alternative, Vannieuwenhuyze et al. (2010) proposed a method to disentangle measurement and selection effects based on a comparative single mode survey.

From March to June 2011, a large scale experiment was performed by Statistics Netherlands to disentangle mode effects for telephone, web, mail and face-to-face survey modes. The experiment was linked to the Crime Victimisation Survey (CVS), but also target variables from the Labour Force Survey (LFS) were included. The project was initiated as a result of large mode effects in the CVS, see Buelens and Van den Brakel (2011). The experiment consisted of two waves. For the first wave, respondents were randomly assigned to the four different survey modes. The questionnaire for the first wave consisted of questions from the CVS as well as key LFS questions about employment status. For the second wave of the survey, both wave 1 respondents and nonrespondents were again approached, except for nonresponse due to language, physical or mental problems (not able or 'niet in staat' categories) and due to changes of address of residence outside the interviewer region. Wave 2 was mostly administered face-to-face (77%), but a random part of the sample was also administered by telephone (23%) for cost reasons. The questionnaire for the second wave consisted of a repetition of a subset of the CVS variables, various attitudinal questions about the survey topic, about surveys in general and about Statistics Netherlands, the survey sponsor, and a range of evaluation questions about the first

wave including an open question for the main motive to respond or to refuse. The attitudinal questions were partly taken from De Leeuw et al (2010) and have a strong resemblance to attitudinal scales presented by Loosveldt and Storms (2008).

The main objective of this paper is to investigate mode-specific common causes for nonresponse bias and measurement bias. This paper can be seen as a follow-up to Buelens, Van der Laan and Schouten (2012) and Klausch, Hox and Schouten (2013a and b). In Buelens et al (2012) mode effect decompositions are presented for CVS and LFS key variables. In Klausch et al (2013a and b) the focus is on mode differences in nonresponse bias. Here, we detail the mode effect decompositions for different levels of contact effort and for different performance levels of interviewers. Common causes are underlying reasons and motives that affect both types of error, while separate causes affect only one of the two errors. It is important to understand common causes, as efforts to reduce one of the two biases may be counterbalanced by an increase of the other bias. A simple example of a common cause is topic interest. Persons that have no interest in the topic are likely to have a higher probability to refuse participation, but, if they do respond, also have a higher probability of being less motivated and concentrated during the interview. The terminology, separate causes and common causes, was put forward by Groves (2006) and Olson (2007). Olson (2012) gives a recent overview of empirical evidence in the survey literature about common causes. Figure 1.1 presents a schematic overview.

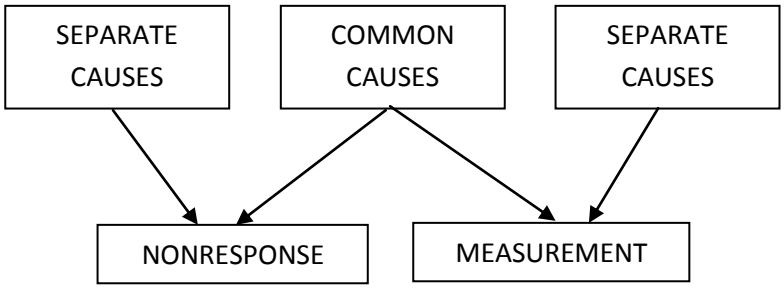

Figure 1.1: Separate causes and common causes for nonresponse and measurement error.

The attitudinal questions in wave 2 can be seen as direct measures of common causes for nonresponse and measurement bias. They are, however, measured in an interviewer-administered survey and may be prone to social desirability. Furthermore, the attitudes may have been formed partly by the experiences in wave 1. Inspection of the questions revealed relatively high item-nonresponse for some of the attitudes, no or little dispersion for some of the scales with 75% or more of respondents providing the same answer, and some dependence on the mode of wave 1 for attitudes about survey burden. For these reasons, we did not use the questions in our investigation. They will, however, be part of a separate analysis.

A surrogate that is often used for respondent motivation is the level of effort required to get a response. It is conjectured that hard to get respondents are less motivated and would produce more measurement error. The level of effort is traditionally applied to

interviewer-administered surveys where effort is a mix of contact attempts and refusal conversion. For web and mail, the level of effort may be reflected by the type, the number and the content of reminders. The level of effort in these modes is then a mix of contact effort and refusal conversion effort that cannot easily be disentangled, as a reminder may lead to first contact but also to refusal conversion. In the experiment we did not vary the type and content of the reminders, but we are able to investigate the impact of the number of reminders on nonresponse and measurement bias. We refer to this as contact effort, despite the fact that a reminder may also be perceived as a refusal conversion attempt. For telephone and face-to-face, we are only able to investigate the impact of contact effort on nonresponse and measurement bias as no refusal conversion was conducted in wave 1. Refusal conversion, other than making appointments for another, more convenient time, is not standard practice at Statistics Netherlands. Once a persons or household refuses explicitly, then no further call or visit is made. We conjecture that contact effort is not a common cause for interviewer-administered surveys as the sampled persons are mostly unaware of the contact efforts. For the non-interviewer administered surveys, we conjecture that contact effort is a common cause but with a relatively small impact as the reminders may only have been noted by some of the respondents.

Another potential common cause which is discussed in the literature is the interviewer who may affect both response and measurement. The impact of interviewers is, however, hard to derive when interviewer assignment is not random, which is usually the case in face-to-face surveys. In face-to-face surveys one can only compare interviewer performance under two assumptions: 1) The lack of interviewer randomization can be neutralized by conditioning on available auxiliary variables and 2) the interviewer samples overlap on these auxiliary variables. The data of the experiment allow us to evaluate these assumptions and to investigate interviewer performance as a common cause to nonresponse and measurement bias. We have no conjecture about the size and direction of the interviewer as a common cause, but we do conjecture it exists.

The wave 2 evaluation questions about wave 1 open up the possibility to investigate mode-specific nonresponse bias in more detail. For nonrespondents to web, mail and CATI, wave 2 can be viewed as a follow-up survey. In the experiment, the nonrespondents did not receive a new advance letter at the start of wave 2 and interviewers were instructed to introduce wave 2 as if it still was the same wave, i.e. like in a sequential mixed-mode design. The evaluation questions in wave 2 can be used to reproduce the decisions made by the respondents and nonrespondents to the wave 1 requests to participate. With this reproduced process we can distinguish different levels of accessibility and propensity to participate and evaluate how nonresponse bias evolves depending on these levels. We conjecture that accessibility produces nonresponse bias for variables that relate to at-home patterns, but differently for each of the modes as the modes use different channels to contact respondents. For the propensity to participate we expect that all variables can be

affected, and we conjecture that persons with negative opinions about safety or with negative experiences have a higher propensity to participate.

In figure 1.2, we sketch the three common causes that we evaluate in the paper and the proxy measures that we have available and employ. We investigate accessibility, motivation and interviewer behaviour as the common causes to nonresponse and measurement error. We have no direct measures of these causes but instead use the numbers of calls, the number of reminders, the reconstructed decision process via wave 2 evaluation questions and the coordinator interviewer ratings as proxies. The wave 2 survey attitudes are not employed for reasons given earlier. Note that, when taken together, accessibility, motivation and interviewer tell us whether a survey mode is a common or separate cause to both errors.

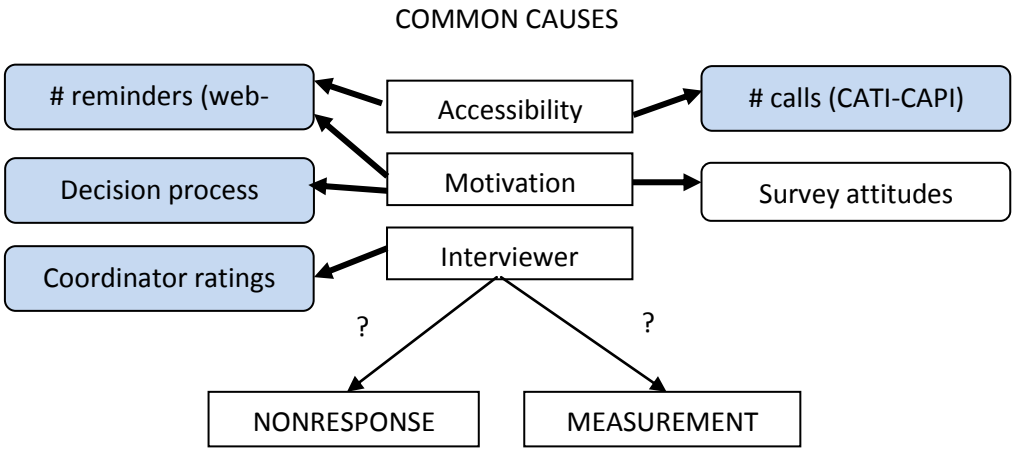

Figure 1.2: Hypothesized common causes for nonresponse and measurement error and proxy measures.

The research questions that we aim to answer in this paper are:

1. Do higher levels of contact effort impact mode-specific nonresponse- and measurement bias?
2. Does interviewer performance impact nonresponse- and measurement biases within the interviewer-assisted modes?
3. Is mode-specific nonresponse bias dependent on acccessability and propensity to participate?

The answers to these questions form input to improvements of mixed-mode survey methodology. If indeed contact effort affects nonresponse bias and measurement bias differently for different modes, then this may be incentive to re-evaluate the contact effort per mode. The results may be used in adaptive survey designs to find optimal numbers of calls, visits and reminders. If interviewer performance affects bias, then this may be a starting point to re-evaluate interviewer training or even interviewer assignment in adaptive survey designs.

This paper is outlined as follows. In section 2, we present the experiment and the data that are used for the analyses in more detail, and perform an exploration of the mode-

specific reasons for response and nonresponse. In sections 3 to 5, we answer the three main research questions. In section 3, we focus on contact effort. In section 4, we investigate interviewer performance. In section 5, we detail the analysis for mode-specific nonresponse bias based on the wave 2 evaluation questions. In section 6, we summarize findings and make recommendations for mixed-mode survey methodology.

2. The Crime Victimisation Survey experiment

In this section we describe the experimental design, analyse mode-specific response behaviour, and discuss the validity of assumptions that allow for a decomposition of mode effects.

2.1 The experimental design

From March to June 2011, the experiment was conducted. The experiment consisted of two waves. The CVS is based on a person sample from the Dutch population of 15 years and older. For the first wave, sample elements were randomly assigned to one of four modes: CAPI, CATI, mail or Web. The questionnaire for the first wave consisted of the regular Crime Victimisation Survey (CVS) questionnaire, in which the last two blocks of questions were replaced by one block of questions on employment situation from the Labour Force Survey (LFS) and one block of questions from the European Social Survey (ESS). Figure 2.1 presents the design of the experiment.

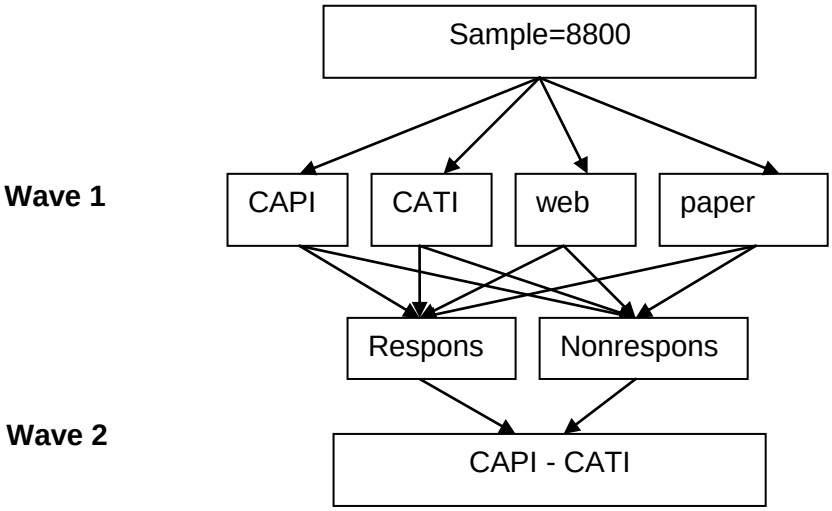

Figure 2.1: Design of the experiment

The second wave was conducted between 4 and 8 weeks later, depending on the survey mode of wave 1 and the time between contacts in both waves. For the second wave of the survey, both wave 1 respondents and nonrespondents were again approached, except for nonresponse due to language, physical or mental problems and due to changes of address of residence outside the interviewer region. A total of 9.9% of the nonresponse (5.8% of the eligible sample) was not approached in wave 2. A majority of 77% of the sampled persons were administered to CAPI in the second wave. The remaining 23% of the sample persons was interviewed by CATI. When a sample person had a registered telephone number (77% of the cases), then the allocation to CATI or CAPI was random with probabilities 70% to CAPI and 30% to CATI. Sampled persons without a registered number were always allocated to CAPI. The 70%-30% CAPI-CATI distribution was chosen such that anticipated mode effects between CATI and CAPI were much smaller than the sampling errors.

The questionnaire for the second wave consisted of a repetition of the key statistics from the CVS, and additional questions on:

- General attitudes towards safety and politics;
- General attitudes towards surveys;
- General attitudes towards Statistics Netherlands;
- Evaluation of survey participation in wave 1 and survey design features like the advance letter and the interviewer;
- Evaluation of the CVS questionnaire (wave 1 respondents only);
- Access to web and mode preferences.

Throughout the paper we consider a number of important LFS and CVS target variables. For the LFS we focus on employment status and highest obtained educational level and for the CVS on number of crimes per 100 inhabitants, being a victim of a crime (yes/no), feeling unsafe at times (yes/no) and an aggregated neighbourhood nuisance scale (five-point). An overview of the LFS and CVS variables in the experiment is given in Appendix A.

The additional questionnaire blocks after the repeated CVS blocks in wave 2 are important; they allow for an analysis of general attitudes towards the CVS topic, towards surveys and towards Statistics Netherlands, and they allow for an analysis of the various design features that are intrinsic to survey modes like the interviewer or the advance letter. The additional questions in wave 2 about access to web are necessary to identify wave 1 nonrespondents to Web who were not able to respond because of undercoverage.

The wave 1 evaluation questions that were asked in wave 2 consisted of:

- Recall questions (Web and mail only): Does the (non)respondent remember the wave 1 advance letters and, if so, did he/she decide to participate directly or postpone a decision. In the evaluation it was not asked how many of the letters (advance and reminders) the (non)respondent recalled, in order to avoid confusion in other evaluation questions.
- Recall questions (CATI and CAPI only): Does the (non)respondent remember contact with the agent or interviewer. Does the (non)respondent recall receiving an advance letter and did it affect his/her decision.
- An open question about the main motive to respond or to refuse (not asked to non-contacts in CAPI and CATI).

Additional evaluation questions were asked about the role of the agent/interviewer and lay-out and length of the questionnaire, but we do not consider these questions here.

2.2 Mode-specific response behaviour

In this section, we investigate the differences and similarities in the response process for the four survey modes. In Table 2.1 the sample sizes and the response to both waves of the experiment are displayed. For CAPI and Web the response rate to wave 1 was close to the anticipated response rate. A remarkable high response was obtained for mail and for CATI. In the following, we investigate underlying motives that may cause these differences, especially between mail and Web.

All respondents to wave 2, except those that were not contacted in CATI or CAPI during wave 1, were asked whether they remember the first wave of the experiment. For the interviewer modes, recall was linked to contact with the interviewer. For the non-interviewer modes, recall was linked to the advance letter (plus paper questionnaire for mail). When the wave 2 respondents did recall the Web or mail advance letter, then they were asked whether they made their decision on participation directly or postponed the decision, and when they postponed whether they looked at the letter later. All respondents to wave 2 that recalled wave 1 received an open question after the main motive to respond to wave 1 (respondents) or to refuse to wave 1 (nonrespondents). The open question was coded by two cognitive questionnaire lab researchers.

Table 2.1: Sample sizes and response rates to wave 1, wave 2 and both waves. The projected wave 1 response rate is also given.

	CAPI	*CATI*	*Mail*	*Web*	*Total*
Sample size	2182	2200	2200	2199	8781
Wave 1 response	61%	45%	49%	29%	46%
Anticipated wave 1 response	58%	38%	40%	29%	
Wave 2 response	49%	47%	50%	49%	49%
Response to both waves	43%	32%	33%	20%	32%

We start with an analysis of the open questions given by wave 2 respondents after the main motive (not) to respond in wave 1. The resulting categories after coding are given in Table 2.2 for nonrespondents and Table 2.3 for respondents. Some of the categories are relatively similar and led to smaller coder reliability. This holds especially true for the coding of response. We ordered the reasons for (non)response in Tables 2.2 and 2.3 based on similarity. For wave 1 nonrespondents, we excluded the interviewer modes, because numbers of wave 1 CATI and CAPI nonrespondents in wave 2 response were relatively small. The number of wave 1 mail and web nonrespondents responding in wave 2 was 375 and 640, respectively, and was sufficiently large to evaluate. The number of wave 1 respondents responding again to wave 2 was 724, 444, 700 and 933, respectively, for mail, Web, CATI and CAPI.

Table 2.2: Main motive to refuse for nonrespondents to wave 1 in mail and Web.

Main motive	*Mail*	*Web*
No time/too busy	25%	16%
Forgotten	15%	23%
Personal circumstances	10%	7%
Not meaningful/low priority	10%	12%
Dislikes surveys	4%	4%
Privacy/topics of the CVS	4%	3%
Questionnaire is lengthy/difficult	16%	11%
Advance letter/reminders	5%	20%
Don't know	1%	2%
Claims participation in wave 1	10%	3%
Number of nonrespondents responding to wave 2	375	640

For wave 1 Web and mail nonrespondents, there are a number of striking differences. Web nonrespondents more often claim they simply have forgotten to respond. The second most frequent motive in Web is the combination of advance letter and (two) reminders, which web nonrespondents did not find appealing or even harassing. For mail this is one of the smaller motives to refuse. Mail nonrespondents most often put forward that they were too busy. They also often mention that the questionnaire was too long or difficult. This result was anticipated as mail allows for an easy direct scan of the questionnaire, while for Web one would have to go the website and login first. Mail nonrespondents also often claim that they did participate, while there was no record of response after four weeks of data collection. A part of the claimed response comes from late responders that sent in their questionnaire after the processing deadline. It is unknown how large this part is as questionnaires for late responders have not been processed. However, given the size of this group, it is likely that part of the nonrespondents forgot to mail the filled-in questionnaire or recalled erroneously that they did respond.

Table 2.3. Main motive to respond for respondents to wave 1.

Main motive	*Mail*	*Web*	*CATI*	*CAPI*
Meaningful	7%	10%	7%	8%
Wanted to give opinion	8%	8%	6%	5%
Important to society	15%	11%	10%	10%
Civic duty	10%	10%	6%	9%
Subject/topic	31%	35%	35%	27%
Statistics Netherlands	10%	7%	9%	9%
Interviewer/advance letter/reminders	2%	3%	7%	10%
Wanted to experience CVS survey	13%	14%	19%	21%
Don't know	1%	1%	1%	1%
Other person/did not participate	3%	2%	0%	0%
Number of respondents participating to wave 2	724	444	700	933

For respondents, differences between modes are less striking and are remarkably similar. Many of the categories have comparable reported frequencies, despite the fact that response rates over modes are very different. The strongest difference comes from the interviewer; respondents to interviewer modes often mention that it was the interviewer that was the main motive to participate or that they wanted to experience the survey. The latter motive is not directly related to the interviewer, but we believe that the assistance of an interviewer is part of the experience. Especially, face-to-face surveys are rare in the Netherlands and may be considered an experience. Of course, given the data, it would be speculation to make conclusions as to why respondents mention the interviewer as a main motive; they may want to help the interviewer doing his/her job but the interviewer may also be able to convince respondents of the importance of the other motives at the doorstep. A proportion of between 2% and 3% of the Web and mail respondents mentions that they did not participate or a household member did. This is the first time an estimate is reported, as far as we know, about the proportion of responses from other household members. This number is relatively small and smaller than was conjectured or speculated.

We turn to the evaluation questions for non-interviewer modes about the timing of response. The main difference between mail and Web is the attachment of the paper questionnaire. Since Web nonrespondents often claim to have forgotten to respond, it is useful to further disentangle the response process for the two modes. Table 2.4 contains the percentages of mail and web respondents to wave 2 that recalled wave 1, that decided immediately to participate or not (given recall), that participated given they made this decision directly, that looked at the letter later when they did not decide directly, and that participated given that they postponed their decision.

Table 2.4 shows that mail and Web differ in every step of the response process. Persons that received a mailed questionnaire more often recall the advance letter, they more often decide directly, and if they do participate more often, they more often look at the letter later, and again they participate more often if they do. Together, these steps support the large difference in response rate between the two modes. Remarkably, 49% of the mail group that did not recall the letter, did actually respond, while for Web this proportion is only 24%. So a lack of memory about reading the letter does not imply that the letter was not opened or that no conscious decision was made about participation. If we ignore this difference and treat recall as having opened the letter, then the estimated response rates based on the evaluation questions for mail and web are 55% and 33%, respectively. These estimates are surprisingly close to the realised response rates of 49% and 29% and only slightly overestimate them. We can also estimate the probability to forget to respond based on Table 2.4. For mail this probability is 7% (= 78% x (100% - 47%) x (100% - 84%)) and for Web it is 15% (= 72% x (100% - 40%) x (100% - 66%)). Both probabilities are much lower than the reported frequencies in Table 2.2. Hence, the two sets of evaluation questions are not fully consistent. The inconsistency may come from social desirable answer behaviour and from recall effects.

Table 2.4: The response processes for mail and Web.

	Mail	*Web*
Recalls advance letter	78%	72%
Decides directly to participate or not	47%	40%
Participates given direct decision	80%	55%
Looks at letter later given postponement	84%	66%
Participates given postponed decision	75%	65%

Summarizing, we can say that mail and Web differ more strongly in their response processes than expected. The results indicate that for Web still improvements can be made by making clearer how much effort is needed and what the questionnaire is about. This could be done by showing sections of the questionnaire or even the full questionnaire. Such changes may reduce the probability that the letter is put aside without a decision on participation.

What do these exploratory results mean in terms of our research questions? Since we investigated reasons for (non)response, they tell us mostly about mode-specific nonresponse bias. Nonrespondents to Web often state that the advance letter and reminders had an adverse effect. In mail this reason is not frequently given. This finding indicates that higher levels of contact effort in Web may lead to larger levels of mode-specific nonresponse bias (research question 1). In the interviewer modes, especially in CAPI, the interviewer is frequently given as the main reason for participation. This finding implies that there is some leverage for interviewers and there might be differences between interviewers in arousing nonresponse bias (research question 2). In the next sections, we will investigate whether contact effort and interviewer performance indeed lead to differences in nonresponse and measurement bias. Before we do, we first discuss the assumptions underlying the mode effect decompositions.

2.3 Contact effort, interviewer performance and mode effect assumptions

Buelens et al (2012) describe the estimation strategy for mode-specific coverage bias, mode-specific nonresponse bias and mode-specific measurement bias. The total bias between a mode and CAPI, i.e. the mode effect, is derived using the parallel samples in the experiment. The mode-specific coverage bias between a mode and CAPI is estimated by the difference between the CAPI response with access to the mode and all CAPI response. The mode-specific nonresponse bias is estimated by weighting the mode wave 1 response that participated also in wave 2, to all wave 2 responses with access to the mode. The mode-specific measurement bias is the remaining term.

The experiment allows for the estimation of mode-specific selection- and measurement effects under four assumptions: 1) the mode of wave 1 did not impact the response to wave 2, 2) the mode of wave 1 did not impact the answers given in wave 2, 3) the repeated CVS questions allow for a full adjustment of differences in nonresponse between the modes, and 4) mode effects between the two modes in wave 2 (telephone and face-to-face) are negligible. Buelens et al (2012) and Schouten et al (2013) give a detailed account of the estimation strategy and analyse the validity

of the underlying assumptions. They conclude that there is no evidence that the mode of wave 1 impacted the response or the answering behaviour to wave 2. Table 3.1 shows that indeed response rates in wave 2 are very similar for all wave 1 groups. They also concluded that the repeated wave 2 questions are generally good predictors of the wave 1 answers so that they are powerful variables in the adjustment for mode-specific nonresponse. Finally, they concluded that the impact of mode effects between CAPI and CATI in wave 2 is much smaller than the sampling variation.

When mode effect decompositions are made as a function of contact effort and interviewer performance, as we will do in this paper, then the important question is whether these assumptions still hold. In sections 3 and 4 we will discuss the validity of the assumptions.

3. Mode effects as a function of contact effort

In this section, we answer the first research question: Do higher levels of contact effort impact mode-specific nonresponse- and measurement bias? We, first, give some descriptives of the contact effort per mode, evaluate the validity of the assumptions underlying the mode effect estimates, and then decompose the mode effect with respect to CAPI for different levels of contact effort. Figure 3.1 depicts the part of figure 1.2 that is investigated in this section. In section 5, we attempt to disentangle accessibility and motivation for mail and Web by adding wave 2 the evaluation questions.

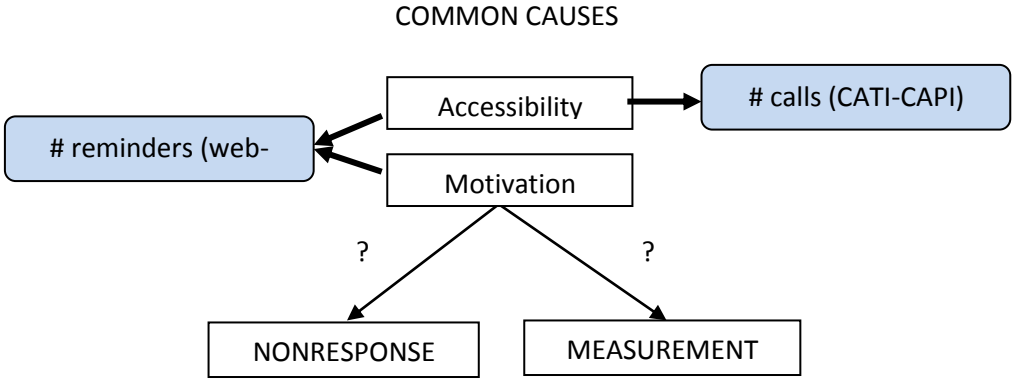

Figure 3.1: Contact effort as a proxy common cause for nonresponse and measurement error.

3.1 Contact effort per mode

We measure contact effort by the number contact attempts up and to first contact. For the interviewer-assisted modes CAPI and CATI, the number of calls in the first wave of the experiment is recorded. For the self-administered modes mail and Web the number of reminders is recorded. These measures differ considerably between the four modes. For CAPI, calls reflect house visits made by the field interviewer. The timing and frequency of calls are largely prescribed, but there is some room for the interviewer to choose the contact strategy. For CATI, calls reflect actual phone calls made by the telephone interviewer. The timing and frequency of the calls is determined by the central system that administers calls to interviewers. The distribution of calls/sample elements to interviewers, thus, is random and the interviewer has no influence on the contact strategy.

Reminders are send to those sample elements that are approached with a mail questionnaire or request to fill in the questionnaire on-line. It is not clear whether these reminders are read. Up to two reminders are send. This is very different from the call attempts in either CAPI or CATI. Hence, the level of effort heavily depends on the mode in which persons are approached.

The contact effort per mode is displayed in table 3.1. For the interviewer-assisted modes there are a number of sample elements that received no calls. These are cases

that were not processed in the field due to illness or other exceptional reasons. For CAPI, sample elements called the help desk to state that they did not want to be approached. For CATI, these are mainly sample elements that could not be called because they do not have a listed telephone number. Hence, these are cases of undercoverage. For mail and Web, all sampled elements received an invitation in the mail to either fill in the mail questionnaire or go to a secured Web site to answer the survey questions on-line. In table 3.1, the number of sample units that received two reminders in mail or Web, respectively, 1497 and 1770 cases, consist of respondents and nonrespondents. For mail, 19.8% of the cases needed two reminders before responding. For Web, this percentage was 10.4%.

Table 3.1 .Distribution of call attempts and percentage of the total response realised in each attempt for the different modes in the experiment.

	CAPI		*CATI*		*Mail*		*Web*	
Number of calls	*n*	*% R*	*N*	*% R*	*n*	*% R*	*n*	*% R*
0	57		483		0		0	
1	537	13,9%	493	9,7%	313	13,5%	211	8,8%
2	641	20,5%	341	11,6%	390	15,6%	218	9,5%
3	417	13,9%	200	6,6%	1497	19,8%	1770	10,4%
4	240	7,8%	110	3,5%				
5	121	3,4%	104	3,1%				
6	97	1,1%	71	2,2%				
>6	72	0,6%	398	8,3%				
Total	2182	61,3%	2200	45,1%	2200	48,9%	2199	28,7%

3.2 Contact effort and mode effect

As described in section 3.1, we evaluate contact effort by restricting response to one visit (CAPI), one call attempt (CATI) and no reminder (Web and mail), and then adding visits, call attempts and reminders. Assumptions 3 and 4 of section 2.3 are not affected by this restriction, but assumptions 1 and 2 of section 2.3 can potentially become invalid. This would happen if response and answers to wave 2 would be different if we had really restricted the contact strategy of wave 1 to fewer visits, call attempts or reminders. For noncontacts in wave 1 such an impact is unlikely, but for respondents or refusals it is imaginable. For instance, if we had stopped after two CATI call attempts, would a person, that had not yet been contacted, react differently to wave 2 than if we had proceeded and had contacted this person later on and forced a decision to participate. The validity of assumptions 1 and 2 cannot be tested because contact effort in wave 1 was not randomized; we cannot simulate persons' reactions to less effort. Since we found no overall impact of mode on wave 2 behaviour, we believe that reduced effort cannot have a large impact. However, since response rates to wave 2 are lower than for CAPI wave 1 (see table 3.1), there seems to be some attrition that can only be caused by wave 1. This attrition means that we have to be careful when drawing strong conclusions about the impact of contact effort.

3.3 Mode-specific nonresponse and measurement bias

In this section, we decompose the mode effect after different numbers of calls, visits and reminders following Buelens et al (2012). The mode effect is decomposed into a mode-specific coverage bias, a mode-specific nonresponse bias and a mode-specific measurement bias. Since coverage of a mode is independent of the number of calls/visits/reminders, this component will be the same after each call and is, therefore, combined with the nonresponse bias. In the following, we will refer simply to nonresponse bias, but it represents the total selection bias due to the sum of undercoverage and nonresponse differences with respect to a CAPI approach where the maximum of visits is six (or seven in case of an appointment at the sixth visit).

In Appendix B, detailed tables of the estimates are provided. Here, we display the results graphically and discuss the significance of differences. Standard errors for the mode effect components are estimated using bootstrap resampling procedures.

Figures 3.2 to 3.4 show the selection bias (the compound of nonresponse and coverage bias), measurement bias and the total mode effect as a function of the number of contact attempts for each of the modes. Filled symbols indicate a significant difference with respect to the maximum number of contact attempts in the same mode. Square symbols indicate a significant difference with respect to the reference mode of CAPI with the full number of contact attempts. A filled square symbol implies that both the difference to CAPI and to the maximum number of contact attempts in the same mode are significant. All significances are evaluated at the 5% level. The white line in the graphs is the zero line and corresponds to no selection bias, no measurement bias or a zero mode effect. In the figures, CAWI stands for the web wave 1 sample. For the LFS variables (figures 3.2 and 3.3) no estimates are given for mail. In wave 1 mail, the LFS questions were not asked due to the complexity of the questionnaire.

The CAPI results in figures 3.2 to 3.4 (column 1) have to be interpreted somewhat differently. In the mode effect decomposition, the mode-specific nonresponse bias is estimated by weighting the sample that responds both to wave 1 in a certain mode and to wave 2, to the full wave 2 response. As a result, some respondents in wave 1 are discarded when they did not respond to wave 2. Although this is a natural choice, for CAPI it implies that the selection bias with respect to itself does not have to be equal to zero. Since the measurement bias is its complement, also this bias does not have to be equal to zero for CAPI. This is an artefact of the estimation method and the distinction between circles and squares is meaningless for CAPI.

It is important to note that the three figures contain many tests and that standard errors of the estimates are sometimes relatively large. For this reason, we investigate only general patterns, i.e. increasing or decreasing biases as a function of contact effort. We will only conclude that contact effort is a proxy common cause to nonresponse and measurement error when the two bias terms are (significantly) increasing or decreasing simultaneously. This means that they should show filled symbols on both selection bias and measurement bias for small numbers of calls or

reminders and the direction of change, i.e. decrease or increase, should be the same in absolute sense.

Figure 3.2 shows the effects for some of the work related questions from the LFS. For all three modes in which these questions were asked, an increase in the number of employed persons is seen and this is mainly related to the selection bias. For CATI and web there is a decrease in the percentage of unemployed persons with increased contact attempts and this seems to be caused by an increasing measurement bias. The differences are significant up to the last contact attempt, so the maximum number of contact attempts is needed.

The results on education level are shown in figure 3.3. For CAPI, mainly an effect is seen for primary education. After the first contact attempt there is an overrepresentation which decreases with increased contact. This is both a measurement and selection effect. CATI shows large effects for pre-vocational and higher professional education. Both are measurement biases. For web, there are no clear effects visible, except that the fraction with missing education increases. This is mainly a measurement effect: persons who need more reminders more often do not fill in their education.

Figure 3.4 shows the results for the target variables of the victimisation survey. The results on the selection bias show that, for CATI and paper, victims of crimes and persons who feel less secure need more effort to collect. The measurement bias does not seem to depend on the amount of effort. The results for Web show no clear patterns.

Combining the results from figures 3.2 – 3.4, we can conclude that an increased contact effort does affect mode-specific nonresponse bias and mode-specific measurement bias for some of the variables. However, for none of the variables for none of the modes did we find a simultaneous significant selection bias and measurement bias for small numbers of calls and reminders. Hence, there is no strong indication that contact effort is a proxy common cause to both types of biases.

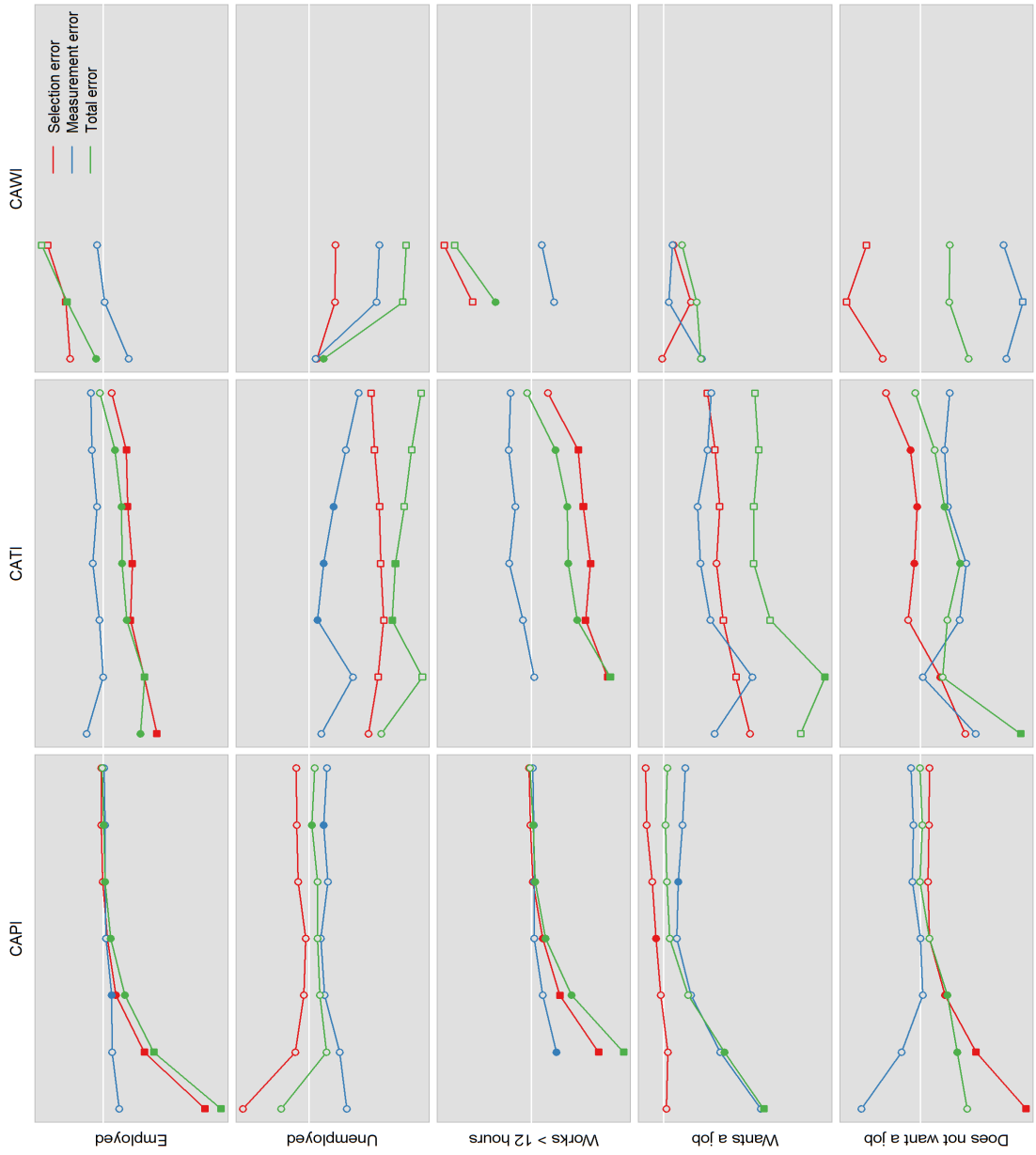

Figure 3.2. Effect of increasing level of effort on selection, measurement and total mode effect for the LFS variables on work. Filled symbols indicate a significant difference from the last contact attempt of the same mode; square symbols indicate a significant difference from the last contact attempt of CAPI.

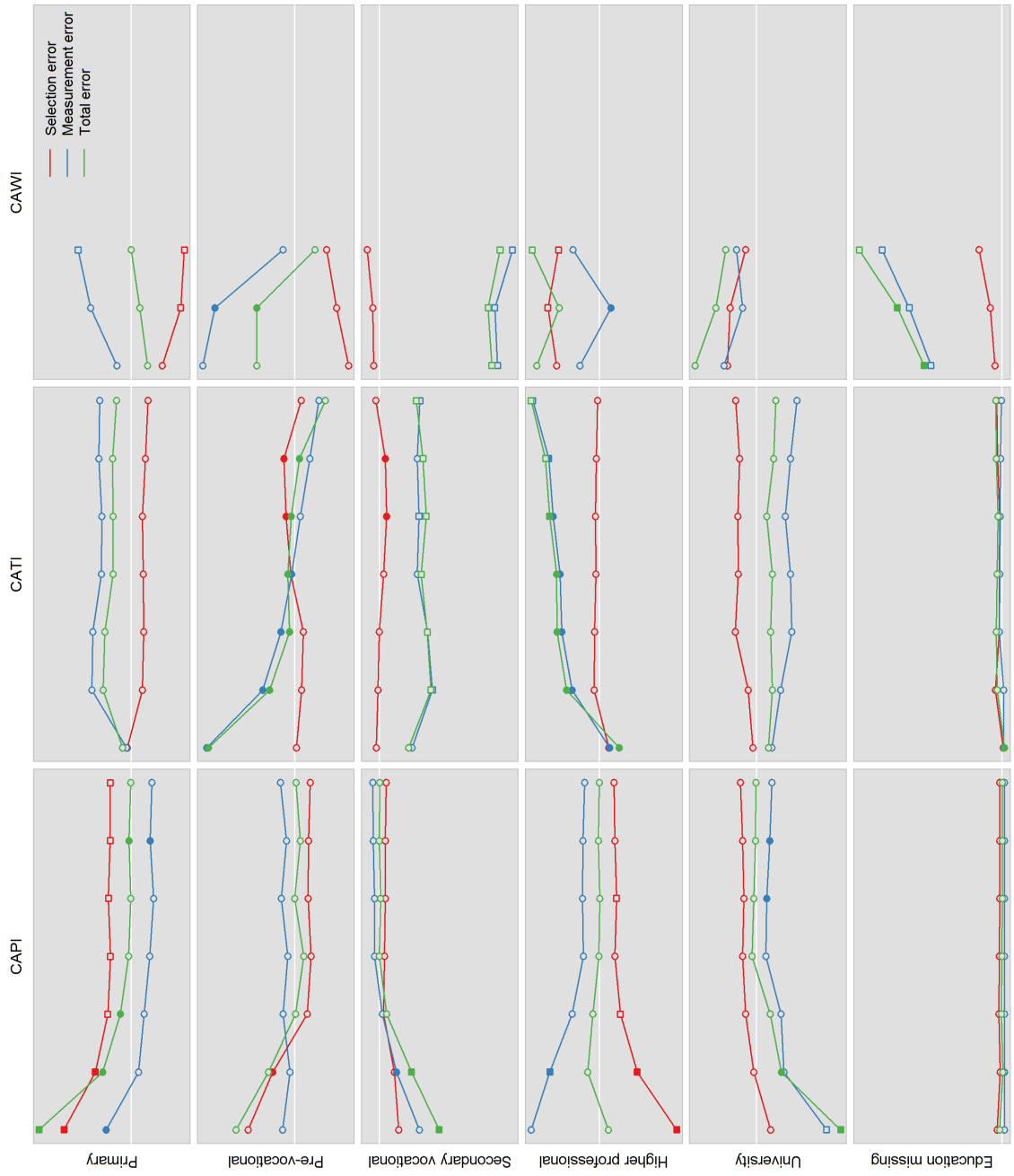

Figure 3.3 Effect of increasing level of effort on selection, measurement and total mode effect for the LFS variables on education.

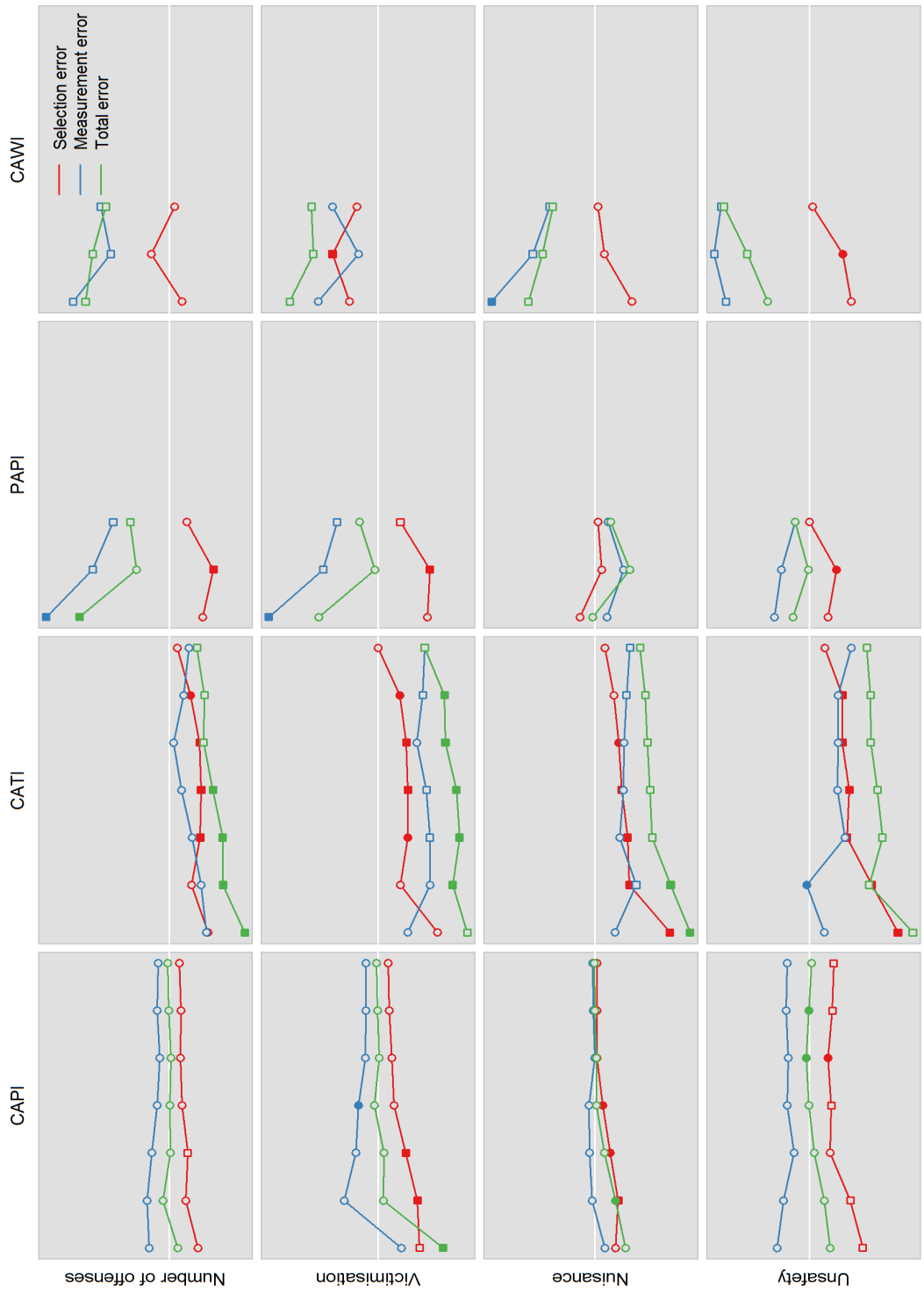

Figure 3.4. Effect of increasing level of effort on selection, measurement and total mode effect for the CVS target variables.

4. Mode effects as a function of interviewer performance

In this section, we answer the second research question: Does interviewer performance impact nonresponse- and measurement biases within the interviewer-assisted modes? First, we explore the rating of interviewer performance by interviewer coordinators. Next, we evaluate the allocation of interviewers in wave 1 and 2 of the experiment. Finally, we decompose the mode effect as a function of interviewer performance in CAPI and in CATI. Figure 4.1 depicts the part of figure 1.2 that is investigated in this section.

4.1 Interviewer performance and allocation

For CAPI, the assignment of sample elements to interviewers is not random. This means demographic differences over regions may be confounded with interviewer performance. For CATI, the assignment is random but there may be more interviewers involved in recruiting the same sample unit, i.e. one interviewer may produce an appointment while another may do the actual interview. Here, we chose to use the interviewer that did the interview, because appointments are often missed and even after an appointment some convincing is often needed.

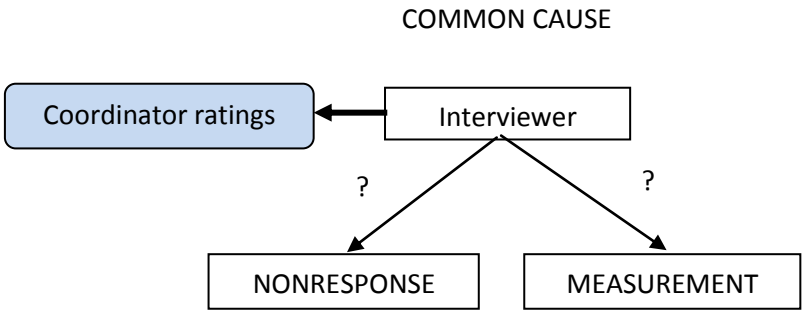

Figure 4.1: Interviewer performance ratings as a proxy common cause for nonresponse and measurement error.

In total, 217 CAPI interviewers conducted 2.145 interviews, and 58 CATI interviewers conducted 1.494 interviews. The CAPI interviewers had on average ten addresses that they approached for the experiment. The CATI interviewers had on average 24 sample elements for which they obtained the final result. The variation over CATI interviewers is large, ranging from 1 to 121. For CAPI, this is 1 to 27.

As a measure for the quality of the interviewers, the fieldwork managers and the call centre managers were asked to fill in a questionnaire on all the interviewers that were involved in the experiment. The questionnaire contained questions on the manager's general opinion and also the opinion on a number of aspects related to interviewing, such as making contact, persuading sample elements and asking the actual survey questions.

In addition, historic response rates were compared to the assessment of the managers. The historic response rates confirmed the managers' opinion to a large extent.

The quality measure that we use in this analysis is the manager's general opinion on the interviewers' performance on a 7-point scale. A score of 1 indicates very low quality and a score of 7 a very high quality. For the analysis, we grouped the interviewers in two categories: below average (1-4) and above average (5-7).

In Table 4.1, the distribution of the interviewer performance levels is shown for both CATI and CAPI. For both CAPI and CATI, the distributions are slightly skewed to higher performance levels. This makes sense; interviewers with lower performance probably self-select out of the job or are trained so that they perform better. In CATI, according to the managers there are less low performance level interviewers than in CAPI. There is less variation in the performance level in CATI. This is probably because CATI interviewers do not have to make contact and can hence only distinguish themselves in persuasion techniques and conversational skills.

Table 4.1. Distribution of the interviewer performance levels for CAPI and CATI.

	CAPI		*CATI*	
Level	*n*	*%*	*n*	*%*
1	2	1%	0	0%
2	13	6%	2	3%
3	21	10%	4	7%
4	66	30%	15	26%
5	57	26%	13	22%
6	42	19%	19	33%
7	16	7%	5	9%
Total	217		58	

In CAPI, the contact strategy is an important ingredient of interviewing. Differences in performance with respect to making contact will be reflected in the nonresponse effect. For CATI, the nonresponse effect merely reflects the quality of the interviewer in persuading respondents. But in CATI, it is possible that the interviewer who finally conducts the interview followed up on an appointment made by another interviewer, sort of heading the ball into the goal. The success of the interviewer making a successful appointment is then wrongfully ascribed to the interviewer following up on the appointment. Here, we will ignore this.

4.2 Interviewer performance and mode effect assumptions

We evaluate interviewer performance by splitting the CATI and CAPI interviewers into two groups, above average and below average, based on ratings given by (regional) interviewer coordinators. The two resulting groups are treated as if they were two different modes in the mode effect decomposition. Again assumption 4 of section 2.3, a negligible CATI-CAPI mode effect in wave 2, is unlikely to be impacted by this choice, but the other assumptions may be violated for CAPI.

Assumption 3 of section 2.3, the MAR assumption when weighting to wave 2, now extends to the allocation of above average performing interviewers to sample persons. If the allocation probabilities were unequal for different persons, then the resulting allocation effect ends up in the decomposed mode effect as a spurious nonresponse effect between above average and below average interviewers. For CATI, the allocation is random. Interviewers operate from one central telephone central. Respondents that need to be contacted are assigned to the next available interviewer. For CAPI the allocation is not random and may lead to bias. The resulting bias can be partially adjusted for by estimating allocation propensities based on register characteristics and using the estimated propensities as weights. The weighting is needed to get unbiased estimates for the nonresponse effect. However, the estimated measurement effect between above average and below average interviewers will not be affected by the selective allocation. It is important to stress that for CAPI, the interviewers were ranked by their regional coordinators. If some coordinators are systematically more positive or negative, i.e. would rank the same interviewer more positively or negatively, then a coordinator effect is present in the interviewer performance that cannot be separated from true differences between interviewers.

In the derivation of interviewer ratings, we also considered ratings based on historic performance, the individual interviewer response rates in the previous year. Although such ratings would not suffer from a coordinator effect, we would have to isolate the component in the response rate that is independent of sample differences because of regional allocation of addresses to interviewers. Hence, historic performance ratings introduce a new assumption that interacts in a complicated way with the assumption of random interviewer allocation.

Assumptions 1 and 2 of section 2.3, no impact on response and answering behaviour, now read as: the assignment of an above or below average performing interviewer does not affect response and answering behaviour in wave 2. For wave 1 CATI, these assumptions seem fair due to the time lag between waves 1 and 2 and the change of mode and interviewer. For wave 1 CAPI, these assumptions are not guaranteed to hold, because CAPI interviewers are assigned to sample persons based on their vicinity to the addresses, and, as a result, for part of the sampled persons there was no switch of interviewer from wave 1 to wave 2. This lack of randomization is mitigated by the size of wave 2. Since wave 2 CAPI was much larger in size than wave 1 CAPI, more CAPI interviewers were active. However, for CAPI, we need to evaluate assumptions 1 and 2 before drawing conclusions.

4.3 A comparison of below and above average interviewers

In the following subsections, we evaluate the effect of the interviewer performance level on mode specific selection- and measurement effects. We make three comparisons: above average CAPI interviewers to below average CAPI interviewers, above average CATI interviewers to all CAPI interviewers, and below average CATI interviewers to all CAPI interviewers For CATI, we do not compare the below average interviewers to the above average interviewers directly. Since CAPI is the benchmark

with respect to which selection- and measurement effects are determined, we need to compare relative mode effects to CAPI for the two groups of interviewers.

4.3.1 An evaluation of mode effect assumptions for CAPI

In section 4.2, we concluded that for CAPI the mode effect assumptions need to be checked. There may be an allocation effect in wave 1, a coordinator effect in the rating of interviewers, and a lack of randomization in wave 2. The first two effects cannot be disentangled, however, without additional data.

As was done in Schouten et al., (2013), we may investigate if the assumptions underlying the analyses are violated. First, when the assignment of the higher and lower quality interviewers in wave 1 is random there is no problem. When the assignment is random, it should not be possible to predict the interviewer quality using background properties of the persons the interviewer has been assigned to. In order to test this we attempted to predict interviewer quality using properties of the persons the interviewers have been assigned to. This can only be done using properties which are available for respondent and non-respondents. Table 4.2 shows for each of the available register variables whether or not in a logistic regression model they explain interviewer quality. Significance is assessed based on likelihood ratio (LR) tests. For comparison, we include the same evaluation for CATI, although no effect is expected there.

Table 4.2. Significance of models predicting interviewer quality in wave 1 using register variables.

Covariate	*Df*	*CAPI*		*CATI*	
		LR	*Sign.*	*LR*	*Sign.*
Degree of urbanisation	4	119.72	***	3.78	
Province	11	299.41	***	9.71	
Registered unemployed	1	0.12		0.00	
Gender	1	0.33		0.45	
Etnicity (3 levels)	2	13.6	**	2.00	
Age (7 categories)	6	9.31		7.81	
Number of persons in household	9	17.38	*	7.24	
Position in household	8	3.75		8.54	
Type of household	6	7.16		8.37	
Income	5	4.75		8.31	
Type of income	2	0.70		3.06	
Owns registered telephone	1	3.96	*	-[a]	-[a]

[a] *Only for CATI respondents that own a registered telephone the interviewer quality is known*
*Signif. codes: '***': p < 0.001 ; '**' : p < 0.01 ; '*': p < 0.05 ; '.': p < 0.1.*

For CATI, as expected, none of the models was significant, but for CAPI the results indicate selective allocation. For a number of variables there are significant effects. Therefore, assignment of CAPI interviewers to respondents is not random and it is

possible that the interviewer quality is correlated with the target variables just because of the assignment. It could for example be possible that persons who feel less safe have a larger probability of being interviewed by a lower quality interviewer, and the other way around lower quality interviewers are more likely to interview persons who feel less safe. Therefore, even if lower quality interviewers perform just as well as higher quality interviewers (response and measurement error are equal on average), we will see a difference between the two, which is solely caused by the non-random assignment.

The second validation we perform is a test of random allocation of above and below average interviewers. We test the allocation by comparing the distributions of repeated CVS variables in wave 2 and register variables for respondents in wave 2 that had been assigned to an above average interviewer in wave 1 and for respondents in wave 2 that had been assigned to a below average interviewer in wave 1. These distributions should be the same under the assumption that interviewers were randomly allocated. The test amounts to a χ^2-test for independence. The results are presented in table 4.3.

Table 4.3. Significance of models predicting register variables and wave 2 variables using wave 1 interviewer quality for CAPI.

Variable	χ^2	*Df.*	*P-value*	*Sign.*
Degree of urbanisation	6.57	1	0.010	*
Province	58.82	4	0.000	***
Registered as unemployed	139.79	11	0.000	***
Gender	0.00	1	1.000	
Etnicity (3 levels)	0.02	1	0.889	
Age (7 categories)	4.71	2	0.095	.
Number of persons in household	10.51	6	0.105	
Position in household	8.59	7	0.284	
Type of household	7.82	8	0.451	
Income	8.42	6	0.209	
Type of income	8.90	5	0.113	
Degree of urbanisation	0.87	2	0.647	
Interest in politics (wave 2)	3.24	1	0.072	.
Number of offences (wave 2)	8.60	9	0.475	
Victimisation (wave 2)	0.09	1	0.762	
Nuisance (wave 2)	9.88	7	0.196	
Unsafety (wave 2)	2.24	1	0.135	
Has had contact with police (wave 2)	0.33	1	0.563	
Victim of violent crime (wave 2)	0.92	1	0.338	

*Signif. codes: '***': p < 0.001 ; '**' : p < 0.01 ; '*': p < 0.05 ; '.': p < 0.1.*

From table 4.3 we can conclude that again the same regional variables turn out to be significantly different. Furthermore, the variable registered as unemployed (according to "UWV Werkbedrijf") also shows a significant difference. The dependence on the

regional variables is not surprising given that these variables also determined allocation in wave 1 and given that interviewers can only be assigned to sample units in their interviewer region. The dependence for registered unemployment is, however, unexpected, but is the result of a collinearity with region. We also performed a logistic regression to model response to wave 2. It turned out (results not shown) that after adding province and degree of urbanisation the performance of the wave 1 interviewer becomes insignificant, i.e., when conditioning on the two regional variables, interviewer allocation to wave 2 appears to be random. However, this result should be treated with care as sample sizes are limited.

4.3.2 A comparison of below and above average CAPI interviewers

The measurement and non-response effects between below and above average interviewers are estimated by weighting the wave 1 response to the CAPI response of wave 2. Based on the findings of section 4.3.1, we decided to always include province and degree of urbanisation into the weighting models. Doing so, we adjust for the net impact of a selective allocation of interviewers and a coordinator effect in scoring interviewer performance. The adjustment is likely not to be fully effective. If some non-random allocation or non-random scoring of performance remains unadjusted, then this ends up in the selection effect between interviewers. So especially the selection effect between above and below average CAPI interviewers should be interpreted with care.

Table 4.4. Decomposition of mode effects for above average CAPI interviewers compared to below average CAPI interviewers.

Variable	*NR*	*ME*		*Total*
Employed	1.1%	2.5%		3.7%
Unemployed	-0.2%	0.8%		0.6%
Primary	1.2%	-4.3%	*	-3.1%
Pre-vocational	0.6%	-2.0%		-1.4%
Secondary vocational	-0.1%	3.0%		2.8%
Higher professional	-1.8%	4.8%	.	3.0%
University	-0.1%	-1.5%		-1.6%
Education missing	0.3%	0.0%		0.3%
Works > 12 hours	1.1%	2.5%		3.7%
Wants a job	0.1%	3.7%		3.8%
Does not want a job	0.4%	-4.9%		-4.4%
Is available	-0.2%	4.1%		3.9%
Is not available	-0.1%	-1.2%		-1.3%
Searches for a job	-1.5%	-5.8%		-7.3%
Number of offenses	0.8	2.4		3.1
Victimisation	-1.3%	2.1%		0.8%
Nuisance	-0.04	0.06		0.02
Unsafety	-1.9%	0.90%		-1.0%

*Signif. codes: '***': p < 0.001 ; '**' : p < 0.01 ; '*': p < 0.05 ; '.': p < 0.1.*

Table 4.4 presents the decomposition of selection and measurement effects for the two groups of CAPI interviewers. Since in CAPI there is no undercoverage, the selection effect can fully be attributed to a difference in nonresponse. From the decompositions we can conclude that the measurement effect is generally the larger of the two components. It is, however, rarely significant. Above average interviewers more often conclude that the educational level was higher for the same respondents. There are also some differences for employment variables but these are not tested as significant at the 5% level.

Based on the results we conclude there is no strong indication that the interviewer is a common cause to nonresponse and measurement error; some interviewers produce more nonresponse error and some more measurement error, but there is no indication that interviewers cause more of both simultaneously.

4.3.3 A comparison of above and below average CATI interviewers to CAPI

Table 4.5 shows the decomposition of mode effects for below and above average CATI interviewers. It is important to note that these mode effects are relative to CAPI using all interviewers as wave 1 is first weighted to wave 2. In table 4.2, significant differences to CAPI are marked, but we also tested the differences between above and below average interviewers. In order to avoid confusion these were not marked in the table.

We first compare CATI interviewers to CAPI interviewers and then above average CATI interviewers to below average CATI interviewers.

For the nonresponse effect between CATI interviewers and CAPI interviewers, there is no clear pattern effect whether this depends on the performance of the CATI interviewers; sometimes the above average CATI interviewers show a larger difference, sometimes the below average CATI interviewers. There are two significant differences, both for above average CATI interviewers, but this can merely be seen as a pattern given the large number of tests.

The measurement effects between CATI and CAPI interviewers are generally larger than the nonresponse effect. This result confirms the findings in Buelens et al (2012). If there are significant effects, then they are for below average CATI interviewers, suggesting that below average CATI interviewers show more measurement errors relative to CAPI interviewers than above average CATI interviewers do. However, for many variables, although not significant, the measurement effect for above average CATI interviewers is the largest. Given these results, we find it too risky to conclude that below average CATI interviewers produce more measurement errors with respect to CAPI interviewers.

If we look at the difference between above average and below average CATI interviewers, then we find the same pattern as for the CAPI interviewers. Above average interviewers give on average a higher educational level for the same

respondent than below average interviewers. Hence, there seems to be a pattern that educational level is coded differently by above and below average interviewers.

Again we have to conclude that we find no strong evidence that interviewers are a common cause to nonresponse and measurement error.

Table 4.5. Decomposition of mode effects for below and above average CATI interviewers compared to all CAPI interviewers.

Variable	*Performance*	*NR*		*ME*		*Total*	
Employed	Above average	-0.90%		1.90%		-0.30%	
	Below average	0.50%		1.40%		0.70%	
Unemployed	Above average	-0.60%		-1.60%		-3.10%	*
	Below average	-0.90%		-1.20%		-2.90%	**
Primary	Above average	-1.30%		0.30%		-1.30%	
	Below average	-1.10%		4.10%	*	2.70%	
Pre-vocational	Above average	2.70%		-3.20%		-0.50%	
	Below average	-1.50%		-0.90%		-2.40%	
Secondary vocational	Above average	-4.00%	*	0.50%		-3.50%	
	Below average	2.40%	.	-6.70%	*	-4.30%	.
Higher professional	Above average	2.20%		4.00%		6.30%	*
	Below average	-1.00%		5.40%	*	4.40%	*
University	Above average	0.20%		-2.00%		-1.50%	
	Below average	0.80%		-1.80%		-0.80%	
Education missing	Above average	0.20%		0.30%		0.40%	
	Below average	0.50%	.	-0.10%		0.30%	
Works > 12 hours	Above average	-0.80%		1.60%		-0.40%	
	Below average	0.40%		1.40%		0.60%	
Wants a job	Above average	-0.70%		-2.10%		-4.20%	
	Below average	-1.20%		-2.60%		-5.20%	*
Does not want a job	Above average	5.70%	*	-2.90%		2.60%	
	Below average	1.30%		-1.90%		-0.80%	
Is available	Above average	-2.00%		-2.10%		-6.30%	
	Below average	-2.50%		-2.00%		-6.80%	*
Is not available	Above average	1.10%		0.50%		3.40%	
	Below average	2.00%		2.80%		6.60%	*
Searches for a job	Above average	-2.40%		-6.90%		-8.20%	
	Below average	2.00%		1.10%		4.30%	
Number of offenses	Above average	-0.50		-7.40		-7.60	
	Below average	-2.00		-3.80		-5.60	
Victimisation	Above average	0.30%		-3.90%		-3.60%	
	Below average	0.20%		-4.10%	.	-3.90%	*
Nuisance	Above average	-0.09		-0.05		-0.18	
	Below average	0.01		-0.19	*	-0.22	*
Unsafety	Above average	-0.30%		-4.00%		-4.40%	.
	Below average	-1.30%		-2.40%		-3.90%	*

*Signif. codes: '***': p < 0.001 ; '**' : p < 0.01 ; '*': p < 0.05 ; '.': p < 0.1.*

5. Evaluation of mode-specific nonresponse bias

In this section, we answer the third research question: How does mode-specific nonresponse bias evolve for less accessible persons and for persons with a lower propensity to participate? We restrict ourselves to an investigation of the mode-specific nonresponse bias for Web and mail. We do not detail the analysis of mode-specific nonresponse bias for CATI, since the number of responding CATI refusers in wave 2 turned out to be relatively small. Only 58 CATI refusers responded to wave 2. An analysis would, therefore, amount again to investigating the impact of contact effort. In figure 5.1, we depict the part of figure 1.2 that is investigated in this section.

We attempt to answer the research question for Web and mail by reproducing the decision process in wave 1 based on evaluation questions in wave 2. In wave 2, both respondents and nonrespondents to Web and mail were asked whether they recalled the wave 1 advance letter and questionnaire, whether they decided directly to participate or not after opening the letter, and, if they postponed the decision, whether they decided actively later. In the evaluation it was not attempted to reproduce the decision process separately for the advance letter and reminder letters, since it was conjectured that this would lead to confusion and recall effects. The evaluation questions allow us to rank all persons on their motivation. We chose the following categories: direct response, postponed response, postponed refusal and direct refusal. Hence, direct refers to an immediate decision after reading the first letter that was opened, i.e. regardless of whether that was the advance letter or one of the reminders. The evaluation questions obviously do not allow us to derive different levels of accessibility. We, therefore, do this indirectly by comparing the patterns on the decision proccess to patterns on the number of contact attempts. The difference in patterns between the two is determined by the level of accessibility, i.e. the probability to be contacted.

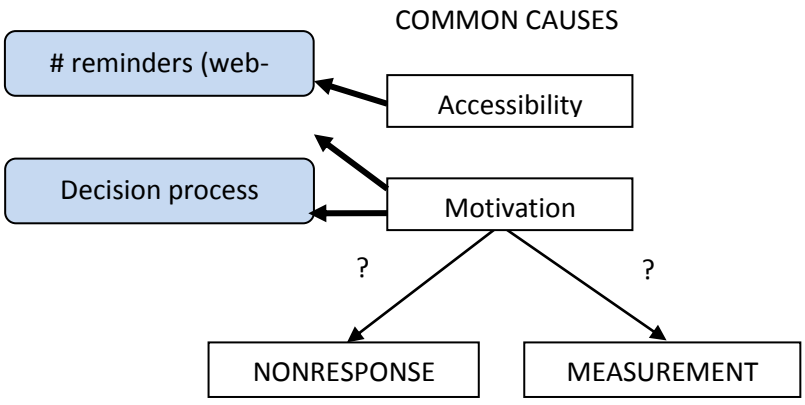

Figure 5.1: Contact effort and reconstructed decision process as proxy common causes for nonresponse and measurement error.

One may clearly debate whether the reconstructed decision process is a proxy measure of motivation. We believe it is, although motivation is an ambiguous concept in which several elements are in play like perceived burden, topic interest and perceived importance of surveys. In our opinion, immediate decisions to participate or to refuse correspond to hard respondents and hard refusers, whereas delayed decisions correspond to soft respondents and soft refusers.

Table 5.1 contains the numbers and proportions for Web and mail. Of the mail sample that responds to wave 2, 78% of the persons recalls reading the letter. For Web this is 72%. Table 5.2 contains the average number of reminders sent for the Web and mail samples, for the samples that responded to wave 2, for the samples that responded to wave 2 and recalled reading the letter, and for the various propensity groups in table 5.1. The average number of reminders is similar for the sample, respondents to wave 2 and respondents to wave 2 with recall. However, the number of reminders clearly increases with a lower propensity to respond. Remarkably, for the direct responders still on average one additional reminder was needed. This may be caused by not opening letters immediately and by mail delivery time lags (especially for the mail survey mode).

Table 5.1: Numbers and proportions of wave 2 respondents in the wave 1 Web and mail samples that recalled the wave 1 approach, responded directly, responded after postponing a decision, refused after postponing a decision, and refused directly.

	No recall	*Direct response*	*Postponed response*	*Postponed refusal*	*Direct refusal*	*All*
Web	28%	16%	18%	25%	13%	100%
	302	174	198	269	141	1084
Mail	22%	29%	26%	16%	8%	100%
	238	323	284	172	82	1099

Table 5.2: Average number of reminders sent for various groups: wave 1 sample, wave 2 respondents, wave 2 respondents with recall, and wave 2 respondents in the propensity groups.

	Wave 1 sample	*Wave 2 response*	*Wave 2 response recall*	*Wave 2 response direct R*	*Wave 2 response post R*	*Wave 2 response post RF*	*Wave 2 response direct RF*
Web	1.7	1.6	1.6	0.9	1.3	2.0	2.0
Mail	1.5	1.4	1.4	1.0	1.3	2.0	2.0

Now, we first search for patterns in the CVS target variables as a function of the propensity to respond. We restrict ourselves to wave 2 respondents that recalled reading the letter. We hypothesize that negative experiences and negative feelings of safety lead to higher propensity to participate. Tables 5.3 and 5.4 shows the means for

the repeated CVS variables in wave 2 supplemented by four wave 2 attitudinal questions. The attitudinal questions are all five-point Lickert scale (1 = fully agree, 2= agree, 3= not agree, not disagree, 4 = disagree, 5= fully disagree). The number of offenses is per 100 inhabitants and nuisance is a scale ranging from 0 to 10.

Table 5.3: Means of various wave 2 repeated CVS variables plus average scores on four attitudes for each of the propensity groups in Web.

Variable	*Wave 2 response direct R*	*Wave 2 response post R*	*Wave 2 response direct RF*	*Wave 2 response post RF*
Safety should be high on political agenda	1.8	1.8	1.8	1.8
Safety is an issue of concern	2.7	2.5	2.4	2.5
The government should do more about safety	2.4	2.3	2.3	2.3
Politicians pay interest in opinions about safety in society	3.0	3.1	3.1	3.3
Number of offenses	41	14	37	33
Victimisation?	10%	6%	15%	9%
Nuisance	1.5	1.3	1.5	1.3
Unsafety?	27%	25%	26%	30%

Table 5.4: Means of various wave 2 repeated CVS variables plus four attitudes for the propensity groups in mail.

Variable	*Wave 2 response direct R*	*Wave 2 response post R*	*Wave 2 response direct RF*	*Wave 2 response post RF*
Safety should be high on political agenda	1.8	1.8	1.9	1.8
Safety is an issue of concern	2.7	2.5	2.7	2.5
The government should do more about safety	2.4	2.4	2.4	2.4
Politics view opinions about safety as important	2.9	3.1	3.1	3.3
Number of offenses	14	19	28	24
Victimisation?	9%	8%	12%	11%
Nuisance	1.1	1.2	1.4	1.5
Unsafety?	25%	29%	22%	33%

The means in tables 5.3 and 5.4 do not show clear overall patterns and do not confirm the hypothesis that persons who are more negative or have more negative experiences have a higher propensity level. The three attitudes about the importance of safety in politics are very stable and uncorrelated with the propensity. The attitude about

politics viewing respondents' opinions important does show a slight tendency towards stronger disagree for persons with a lower propensity. The CVS variables show some variation but not in the direction we hypothesized. The differences in tables 5.3 and 5.4 were not tested as we merely searched for patterns. Some of the cells have relatively large standard errors, which may explain, for instance, the outliers in the number of offenses and victimisation for postponed response in Web.

Table 5.5: Means of various wave 2 repeated CVS variables plus four attitudes for different number of reminders in Web.

Variable	*0*	*1*	*2*
Safety should be high on political agenda	1.8	1.9	1.8
Safety is an issue of concern	2.7	2.4	2.5
The government should do more about safety	2.4	2.3	2.3
Politics view opinions about safety as important	3.0	3.1	3.2
Number of offenses	35	35	30
Victimisation?	10%	9%	11%
Nuisance	1.3	1.4	1.4
Unsafety?	25%	23%	28%

One question remains: does the level of accessibility relate to wave 2 variables. In order to search for patterns, we again have to restrict ourselves to wave 2 respondents who recall the letter of wave 1. We compare means of repeated CVS variables and the same four attitudinal questions for no reminder, one reminder and two reminders to tables 5.3 and 5.4. In tables 5.5 and 5.6 we display the same means but as a function of the number of reminders. Again no clear patterns emerge from the means in Web and in mail, implying that there is no strong relation between accessibility and CVS topics.

From the two investigations, we conclude that there is no strong indication that propensity to respond or accessibility relate to CVS variables. These findings strengthen the conclusions in section 3.3 where it was concluded that mode-specific nonresponse bias for CVS variables does not correlate strongly with contact effort.

Table 5.6: Means of various wave 2 repeated CVS variables plus four attitudes for different number of reminders in mail.

Variable	*0*	*1*	*2*
Safety should be high on political agenda	1.8	1.9	1.8
Safety is an issue of concern	2.6	2.6	2.6
The government should do more about safety	2.4	2.4	2.4
Politics view opinions about safety as important	3.0	2.9	3.1
Number of offenses	16	17	21
Victimisation?	10%	9%	10%
Nuisance	1.3	1.1	1.3
Feeling unsafe at times?	24%	31%	26%

6. Discussion

In this paper, we addressed three research questions about, respectively, the relation of contact effort to mode effects, the relation of interviewer performance to mode effects, and the relation of mode-specific accessibility and participation propensity to nonresponse bias.

We hypothesized that contact effort is not a proxy common cause for nonresponse and measurement error in the interviewer modes, and, further, that it may only be a weak proxy common cause in the non-interviewer modes. From the mode effect decompositions, we indeed cannot determine any pattern that indicates it is a proxy common cause, neither in the interviewer or in the non-interviewer modes; the size of measurement bias is stable while there are changes in the size of the nonresponse bias as contact effort is increased. We did find differences between Web and mail nonrespondents in the frequency they report the advance letter and reminders as a main motive to refuse. However, this difference does not seem to lead to differences in nonresponse bias.

When it comes to interviewer performance, the conclusions are subject to more caution. We hypothesized that interviewers may be a common cause to nonresponse and measurement error. Indeed, around 10% of the respondents indicate that the interviewer is the main motive to participate in CAPI. It is widely known that interviewers achieve different response rates, but now we also find evidence that it is the interviewer explicitly that is given as the reason for response. In CATI we found some differences between above and below average performing interviewers for educational level, mostly at lower educational levels. On the majority of variables, there was no significant difference, however. For CAPI, we found that interviewer allocation in wave 1 and interviewer re-allocation in wave 2 both were selective for some of the available socio-demographic auxiliary variables. This finding implies that we cannot blindly perform a mode effect decomposition as interviewer allocation may end up as a spurious nonresponse bias. We attempted to adjust for the selective (re-)allocation by including the strongest predictors among these variables as weighting variables in the mode effect decomposition. After adjustment, differences in nonresponse bias and measurement bias between above and below average performing CAPI interviewers were not significant for the majority of variables. Again we found a significant difference for educational level. The results do not point at the interviewer as a common cause. We did, however, find some evidence that above and below average interviewers code educational level differently for the same respondent; the educational level is coded higher by both above average CATI and above average CAPI interviewers. We do not have an explanation for this difference, however.

Finally, we detailed the analysis of nonresponse bias by using evaluation questions to reconstruct the response decision process. We restricted ourselves to Web and mail. We conjectured that lower accessibility corresponds to larger nonresponse bias on

variables that relate to at-home patterns, and that the propensity to participate is higher for persons with negative experiences or opinions. Again we found no indication that these conjectures are true for Web and mail; neither the main survey variables nor attitudes that relate to survey topics show any clear pattern for different levels of accessibility or participation propensity.

The findings of this study should be treated and interpreted with some care as the sample size of the mixed-mode experiment was not specifically designed to answer the first two research questions of this paper. The sample size was designed such that acceptable statistical power is obtained for mode effect decompositions at the aggregate level, i.e. without a further stratification on contact effort or interviewer performance. For this reason we restricted ourselves to two performance levels for the interviewers and investigated patterns in contact effort rather than differences at individual numbers of calls, visits or reminders. As discussed, the mode effect decomposition also comes with a number of assumptions which are reasonable for contact effort and CATI interviewer performance. For CAPI interviewer performance caution is required.

What are the implications of these findings for mixed-mode methodology? We see three main implications:

- There is potential to increase response to web surveys: There are large differences in the response decision processes of Web and mail and they lead to a surprisingly large difference in response rate. This difference can only be caused by respondents noticing the mail questionnaire and advance letter but not the web advance letter, by respondents wanting to first scan the questionnaire, and by respondents for which the barrier to log in to a website is too big. At the same time nonrespondents to web often indicate they did not like the amount of reminders. These findings indicate that it may be helpful when the quality of the Web advance letter is improved by making it more visible and more insightful by adding information about the survey and screen dumps of sections of the questionnaire. The findings also indicate that future mobile devices that allow sending survey requests directly to the respondent are promising in terms of response rates; they show a larger resemblance to mail.
- Adaptive survey designs that differentiate contact effort within one mode or differentiate allocation of CATI interviewers can focus mostly on the composition of response: There is no indication that measurement bias changes for different interviewer performance levels and changes in measurement bias for different contact effort levels are generally small and rare within a mode. For the choice of mode this does not hold; the mode may affect both forms of bias simultaneously.
- A varying mix of mode-specific measurement bias due to instability of contact effort in a mode or CATI interviewer performance from one month to the other can be stabilized by calibration: As there is no indication that measurement bias is

strongly linked to contact effort or interviewer performance, any variation in time is likely to lead only to a varying mix of measurement biases per mode. This mix can be stabilized to a large extent by calibration to a fixed distribution of modes in the response.

The last two implications need replication of findings in other surveys in order to generalize from the CVS to other surveys. The finding that above average interviewers code respondents to higher educational level should be evaluated, as far as is possible, by linkage to validation data from registers. We have not performed such an evaluation. However, regardless of the validation, the differences in coding respondents to educational levels may be picked up in interviewer training and evaluations.

As far as we are aware, to date, there still is very little empirical evidence in the literature about common causes for nonresponse and measurement error. The results in this paper are specific to the Crime Victimisation Survey (CVS) and the Dutch survey climate. Although this survey has a wide range of survey questions and is expected to arouse different response styles, it still is a single data set with a specific set of topics set in the Dutch population. Nonetheless, we hope that the findings stimulate others to perform similar exercises and replicate findings.

References

Biemer, P. and Lyberg, L. (2003), Introduction to Survey Quality, Wiley

Brunton-Smith, I., Sturgis, P., Williams, J. (2012), Is success in obtaining contact and cooperation correlated with the magnitude of interviewer variance?, Public Opinion Quarterly, 76 (2), 265 – 286.

Buelens, B., Laan, J. van der, Schouten, B., Brakel, J. van den, Burger, J., Klausch, T. (2012), Disentangling mode-specific selection and measurement bias in social surveys, Discussion paper 201211, CBS, Den Haag.

Buelens, B., Laan, J. van der, Schouten, B. (2012), Project MEPS – decomposition of mode effects for CVS and LFS, Technical paper, PPM-2012-02-25-BBUS-DLAN-BSTN, CBS, Den Haag.

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

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

Jäckle, A., Roberts, C. and Lynn, P. (2010), Assessing the effect of data collection mode on measurement, International Statistical Review, Volume 78, Issue 1, 3- 20.

Klausch, T., Hox, J., Schouten, B. (2013a), Assessing the mode-dependency of sample selectivity across the survey response process, Discussion paper 201303, CBS, Den Haag.

Klausch, L.T., Hox, J., Schouten, B. (2013b), Mode Impact on Selection Error in Single- and Mixed-Mode Surveys of the Dutch General Population, forthcoming.

Laan, J. van der, Beukenhorst, D., Buelens, B., Engelen, F., Meertens, V., Schouten, B. (2013), The impact of survey item characteristics on mode-specific measurement bias in the Crime Victimisation Survey, Discussion paper 2013xxx, CBS, Den Haag.

Leeuw, E. de, Hox, J., Lugtig, P., Scherpenzeel, A., Vis, C., Göritz, A., Bartsch, S., Engel, U., Vehre, H. (2010), Measuring and comparing survey attitude among new and repeat respondents cross-culturally, Paper presented at WAPOR 63rd Annual Conference, Chicago, USA.

Loosveldt, G., Storms, V. (2008), Measuring public opinions about surveys, International Journal of Public Opinion Research, 20 (1), 74 - 88.

Olson, K. (2007), An investigation of the nonresponse – measurement error nexus, PhD thesis, University of Michigan, USA.

Olson, K. (2012), Do non-response follow-ups improve or reduce data quality?: a review of the existing literature, Journal of the Royal Statistical Society A, No 176

Schouten, B., Brakel, J. van den, Buelens, B., Laan, J. van der, Klausch, L.T. (2013), Disentangling mode-specific selection and measurement bias in social surveys, Social Science Research, 42, 1555 – 1570.

Vannieuwenhuyze J., Loosveldt, G., Molenberghs, G. (2010), A method for evaluating mode effects in mixed-mode surveys, Public Opinion Quarterly, No. 74 (5), 1027 – 1045.

Appendix A: Overview of CVS and LFS target variables

Tables A.1 and A.2 contain the analysis survey variables from the LFS (Labour Force Survey) and CVS (Crime Victimization Survey). These variables are observed in the first wave of the experiment. Some of the wave 1 survey variables are repeated, but not all. In particular, the LFS variables are not repeated in wave 2.

Table A.1: Overview of key LFS variables in wave 1

Variable name	*Variable contents*
empstat	Employment status (in LFS bbned with categories employed, unemployed, non-labour force)
educlev	Highest level of education achieved (5 point scale)
hasjob	Indicator for having an employment
wantjob	Indicator for wishing a job or an increase in working hours
searchjob	Indicator for conducting activities to get a job
availjob	Indicator for being available to start a job

Table A.2: Overview of key CVS variables in wave 1

Variable name	*Variable contents*
offtot	Total number of offenses[1]
offbike	Number of bicycle thefts
offprop	Number of property crimes
offviol	Number of violent crimes
victim	Percentage of population victim of crime[2]
victprop	Percentage of population victim of property crime
victviol	Percentage of population victim of violent crime
nuisance	Scale score related to nuisance
degrad	Scale score related to degradation of the neighbourhood
unsafe	Percentage of people feeling unsafe at times
funcpol	Scale score related to functioning of the police
opinpol	Percentage of people generally (very) happy with police
contpol	Percentage of people having had contact with the police
satispol	Percentage of people who where (very) satisfied at contact

[1] Per 100 inhabitants, in the last 12 months.
[2] In the last 12 months.

Appendix B: Results for the relationship between level of effort and mode effects

In the following tables, the decomposition of mode effects into nonresponse effect (NR), coverage effect (CO), measurement effect (ME) and total mode effect (total) is displayed for the number of call attempts in different modes. Significant errors are denoted by an asterisk.

Table B.1: Mode effect decomposition for variable being employed. CAPI average is 56.1%

	Mode	*Average*	*NR*	*CO*	*ME*	*Total*
1	CATI	50,5%	-6,4%	-1,2%	2,0%	-5,7%
2	CATI	50,0%	-4,8%*	-1,2%	-0,2%	-6,1%*
3	CATI	52,6%	-2,8%	-1,2%	0,5%	-3,5%
4	CATI	53,3%	-3,0%	-1,2%	1,4%	-2,8%
5	CATI	53,4%	-2,3%	-1,2%	0,8%	-2,8%
6	CATI	54,3%	-2,2%	-1,2%	1,6%	-1,8%
7+	CATI	56,5%	-0,1%	-1,2%	1,7%	0,4%
1	Web	57,0%	2,2%	2,5%*	-3,9%	0,9%
2	Web	61,2%	2,9%	2,5%*	-0,3%	-4,7%
3	Web	65,1%	5,6%*	2,5%*	0,9%	8,9%*

Table B.2: Mode effect decomposition for variable being unemployed. CAPI average is 7.9%

	Mode	*Average*	*NR*	*CO*	*ME*	*Total*
1	CATI	6,1%	-0,8%	-0,8%	-0,2%	-1,8%
2	CATI	4,9%	-1,0%	-0,8%	-1,2%	-3,0%*
3	CATI	5,7%	-1,2%	-0,8%	-0,2%	-2,2%
4	CATI	5,7%	-1,1%	-0,8%	-0,4%	-2,3%*
5	CATI	5,4%	-1,1%	-0,8%	-0,6%	-2,5%*
6	CATI	5,2%	-1,0%	-0,8%	-1,0%	-2,7%*
7+	CATI	4,9%	-0,9%	-0,8%	-1,3%	-3,0%*
1	Web	7,5%	-0,3%	0,0%	-0,2%	-0,4%
2	Web	5,4%	-0,7%	0,0%	-1,8%	-2,5%
3	Web	5,3%	-0,7%	0,0%	-1,9%	-2,6%*

Table B.3: Mode effect decomposition for variable educational level is primary. CAPI average is 12.7%

	Mode	*Average*	*NR*	*CO*	*ME*	*Total*
1	CATI	13,6%	0,9%	-0,3%	0,3%	0,8%
2	CATI	15,1%	-0,5%	-0,3%	2,3%	2,4%
3	CATI	15,0%	-0,6%	-0,3%	3,3%	2,3%
4	CATI	14,3%	-0,6%	-0,3%	2,6%	1,6%
5	CATI	14,3%	-0,6%	-0,3%	2,5%	1,6%
6	CATI	14,3%	-0,8%	-0,3%	2,8%	1,6%
7+	CATI	14,0%	-1,1%	-0,3%	2,7%	1,3%
1	Web	11,4%	0,6%	-3,2%*	1,3%	-1,3%
2	Web	11,9%	-0,9%	-3,2%*	2,2%	-0,8%
3	Web	12,7%	-1,2%	-3,2%*	4,4%*	0,0%

Table B.4: Mode effect decomposition for variable educational level is pre-vocational. CAPI average is 21.1%

	Mode	*Average*	*NR*	*CO*	*ME*	*Total*
Total	CAPI	21,1%	-	-	-	-
1	CATI	25,7%	-0,2%	-0,1%	4,9%	4,6%
2	CATI	22,4%	-0,4%	-0,1%	1,7%	1,3%
3	CATI	21,3%	-0,5%	-0,1%	0,8%	0,3%
4	CATI	21,4%	0,2%	-0,1%	0,1%	0,3%
5	CATI	21,2%	0,5%	-0,1%	-0,4%	0,1%
6	CATI	20,7%	0,6%	-0,1%	-0,9%	-0,3%
7+	CATI	19,3%	-0,3%	-0,1%	-1,3%	-1,7%
1	Web	22,8%	-2,7%	-0,1%	4,5%	1,7%
2	Web	22,9%	-2,3%	-0,1%	4,2%	1,8%
3	Web	19,8%	-1,7%	-0,1%	0,6%	-1,3%

Table B.5: Mode effect decomposition for variable educational level is secondary vocational. CAPI average is 38.3%

	Mode	*Average*	*NR*	*CO*	*ME*	*Total*
1	CATI	35,0%	0,5%	0,0%	-3,8%	-3,3%
2	CATI	32,6%	0,2%	0,0%	-5,9%*	-5,7%*
3	CATI	33,1%	0,0%	0,0%	-5,4%*	-5,3%*
4	CATI	33,8%	-0,4%	0,0%	-4,1%	-4,5%*
5	CATI	33,2%	-0,8%	0,0%	-4,4%	-5,1%*
6	CATI	33,6%	-0,6%	0,0%	-4,2%	-4,8%*
7+	CATI	34,3%	0,4%	0,0%	-4,5%	-4,0%*
1	Web	25,9%	-0,8%	1,4%	-13,0%*	-12,4%*
2	Web	26,4%	-0,7%	1,4%	-12,6%*	-20,0%*
3	Web	25,0%	-0,1%	1,4%	-14,6%*	-13,3%*

Table B.6: Mode effect decomposition for variable educational level is higher professional. CAPI average is 19.7%

	Mode	*Average*	*NR*	*CO*	*ME*	*Total*
1	CATI	18,2%	-1,0%	0,1%	-0,6%	-1,4%
2	CATI	22,2%	0,2%	0,1%	2,2%	2,5%
3	CATI	22,8%	0,2%	0,1%	2,8%	3,1%
4	CATI	22,8%	0,1%	0,1%	2,9%	3,2%
5	CATI	23,4%	0,1%	0,1%	3,5%	3,7%
6	CATI	23,7%	0,1%	0,1%	3,8%	4,0%*
7+	CATI	24,8%	0,0%	0,1%	5,0%*	5,1%*
1	Web	24,4%	0,9%	2,1%*	1,6%	4,7%
2	Web	22,6%	1,6%	2,1%*	-0,8%	3,0%
3	Web	24,6%	0,9%	2,1%*	1,8%	4,9%*

Table B.7: Mode effect decomposition for variable educational level is university. CAPI is 8.1%

	Mode	*Average*	*NR*	*CO*	*ME*	*Total*
1	CATI	7,5%	-0,2%	0,3%	-0,7%	-0,6%
2	CATI	7,2%	0,1%	0,3%	-1,2%	-0,8%
3	CATI	7,3%	0,6%	0,3%	-1,6%	-0,7%
4	CATI	7,2%	0,5%	0,3%	-1,6%	-0,8%
5	CATI	7,5%	0,5%	0,3%	-1,4%	-0,6%
6	CATI	7,2%	0,4%	0,3%	-1,6%	-0,9%
7+	CATI	7,0%	0,6%	0,3%	-1,9%	-1,0%
1	Web	10,9%	1,3%	-0,1%	1,6%	2,8%

2	Web	10,0%	1,3%	-0,1%	0,7%	1,9%
3	Web	9,5%	0,6%	-0,1%	0,9%	1,4%

Table B.8: Mode effect decomposition for variable educational level is missing. CAPI average is 0.1%

	Mode	*Average*	*NR*	*CO*	*ME*	*Total*
1	CATI	0,0%	0,0%	-0,1%	-0,9%	-0,1%
2	CATI	0,4%	0,4%	-0,1%	-0,1%	0,3%
3	CATI	0,5%	0,2%	-0,1%	0,2%	0,3%
4	CATI	0,4%	0,2%	-0,1%	0,1%	0,3%
5	CATI	0,4%	0,2%	-0,1%	0,1%	0,2%
6	CATI	0,5%	0,3%	-0,1%	0,1%	0,3%
7+	CATI	0,5%	0,4%	-0,1%	0,1%	0,4%
1	Web	4,7%	0,6%	-0,1%	4,0%*	4,5%*
2	Web	6,2%	0,9%	-0,1%	5,3%*	6,1%*
3	Web	8,4%	1,5%	-0,1%	6,9%*	8,2%*

Table B.9: Mode effect decomposition for variable works > 12 hours. CAPI average is 56.1%

	Mode	*Average*	*NR*	*CO*	*ME*	*Total*
1	CATI	-	-	-	-	-
2	CATI	49,9%	-4,8%*	-1,2%	-0,2%	-6,2%*
3	CATI	52,4%	-3,1%	-1,2%	0,6%	-3,7%
4	CATI	53,2%	-3,4%	-1,2%	1,7%	-2,9%
5	CATI	53,2%	-2,9%	-1,2%	1,2%	-2,9%
6	CATI	54,2%	-2,5%	-1,2%	1,8%	-1,9%
7+	CATI	56,4%	-0,1%	-1,2%	1,6%	0,3%
1	Web	-	-	-	-	-
2	Web	59,0%	2,0%	2,5%*	-1,7%	2,8%
3	Web	62,3%	4,4%	2,5%*	-0,7%	6,2%

Table B.10: Mode effect decomposition for variable works < 12 hours. CAPI average is 43.9%

	Mode	*Average*	*NR*	*CO*	*ME*	*Total*
1	CATI	-	-	-	-	-
2	CATI	50,1%	4,8%*	1,2%	0,2%	6,2%*
3	CATI	47,6%	3,1%	1,2%	-0,6%	3,7%
4	CATI	46,8%	3,4%*	1,2%	-1,7%	2,9%
5	CATI	46,8%	2,9%*	1,2%	-1,2%	2,9%
6	CATI	45,8%	2,5%	1,2%	-1,8%	1,9%
7+	CATI	43,6%	0,1%	1,2%	-1,6%	-0,3%
1	Web	-	-	-	-	-
2	Web	41,0%	-2,0%	-2,5%*	1,7%	-2,8%
3	Web	37,7%	-4,4%*	-2,5%*	0,7%	-6,2%*

Table B.11: Mode effect decomposition for variable wants a job is not applicable. CAPI average is 49.6%

	Mode	*Average*	*NR*	*CO*	*ME*	*Total*
1	CATI	64,8%	6,7%	1,6%	6,9%	15,2%*
2	CATI	60,1%	3,8%	1,6%	5,1%	10,5%*
3	CATI	57,4%	0,6%	1,6%	5,6%	7,8%*
4	CATI	57,4%	0,7%	1,6%	5,5%	7,9%*
5	CATI	56,2%	1,2%	1,6%	3,9%	6,7%*
6	CATI	55,7%	0,4%	1,6%	4,1%	6,1%

7+	CATI	54,1%	-1,9%	1,6%	4,7%	4,5%
1	Web	55,1%	3,6%	-6,6%*	8,5%	5,5%
2	Web	53,3%	2,3%	-6,6%*	7,9%	3,7%
3	Web	52,7%	2,8%	-6,6%*	6,8%	3,1%

Table B.12: Mode effect decomposition for variable wants a job. CAPI average is 17.5%

	Mode	*Average*	*NR*	*CO*	*ME*	*Total*
1	CATI	10,2%	-3,3%	-1,4%	-2,7%	-7,3%*
2	CATI	8,8%	-2,4%	-1,4%	-4,9%	-8,7%*
3	CATI	11,8%	-1,7%	-1,4%	-2,6%	-5,7%*
4	CATI	12,8%	-1,4%	-1,4%	-2,0%	-4,8%*
5	CATI	12,7%	-1,6%	-1,4%	-1,8%	-4,8%*
6	CATI	12,5%	-1,3%	-1,4%	-2,3%	-5,1%*
7+	CATI	12,7%	-1,0%	-1,4%	-2,5%	-4,9%*
1	Web	15,7%	-0,5%	0,6%	-1,9%	-1,8%
2	Web	16,0%	-2,1%	0,6%	-0,1%	-1,6%
3	Web	16,7%	-1,1%	0,6%	-0,3%	-0,8%

Table B.13: Mode effect decomposition for variable does not want a job. CAPI average is 32.9%

	Mode	*Average*	*NR*	*CO*	*ME*	*Total*
1	CATI	25,0%	-3,4%	-0,2%	-4,2%	-7,9%
2	CATI	31,1%	-1,4%	-0,2%	-0,2%	-1,8%
3	CATI	30,7%	1,1%	-0,2%	-3,0%	-2,1%
4	CATI	29,8%	0,7%	-0,2%	-3,6%	-3,1%
5	CATI	31,0%	0,4%	-0,2%	-2,1%	-1,9%
6	CATI	31,8%	1,0%	-0,2%	-1,8%	-1,1%
7+	CATI	33,3%	2,9%	-0,2%	-2,3%	0,4%
1	Web	29,2%	-3,0%	5,9%*	-6,6%	-3,7%
2	Web	30,8%	-0,2%	5,9%*	-7,8%	-2,1%
3	Web	30,6%	-1,7%	5,9%*	-6,5%	-2,3%

Table B.14: Mode effect decomposition for variable is available for a job. CAPI average is 22.7%

	Mode	*Average*	*NR*	*CO*	*ME*	*Total*
1	CATI	15,7%	-5,9%	-2,2%*	1,1%	-7,0%
2	CATI	13,0%	-4,9%*	-2,2%*	-2,6%	-9,7%*
3	CATI	16,3%	-4,5%*	-2,2%*	0,3%	-6,4%*
4	CATI	16,2%	-4,5%*	-2,2%*	0,2%	-6,5%*
5	CATI	15,7%	-4,2%*	-2,2%*	-0,5%	-6,9%*
6	CATI	15,6%	-3,8%*	-2,2%*	-1,0%	-7,0%*
7+	CATI	16,1%	-2,3%	-2,2%*	-2,1%	-6,6%*
1	Web	16,7%	-0,1%	4,4%*	-10,3%	-6,0%
2	Web	13,0%	-0,5%	4,4%*	-13,5%*	-9,7%*
3	Web	12,9%	-0,9%	4,4%*	-13,3%*	-9,8%*

Table B.15: Mode effect decomposition for variable is not available for a job. CAPI average is 75.5%

	Mode	*Average*	*NR*	*CO*	*ME*	*Total*
1	CATI	84,3%	5,9%	1,8%	1,1%	8,9%*
2	CATI	86,4%	4,3%	1,8%	4,8%	10,9%*
3	CATI	81,8%	2,6%	1,8%	1,9%	6,4%
4	CATI	81,7%	3,0%	1,8%	1,4%	6,2%
5	CATI	81,5%	3,4%	1,8%	0,9%	6,0%

6	CATI	81,7%	2,9%	1,8%	1,5%	6,2%*
7+	CATI	80,9%	1,7%	1,8%	2,0%	5,5%
1	Web	80,3%	-1,1%	-4,2%*	10,1%	4,8%
2	Web	81,3%	0,1%	-4,2%*	9,9%	5,8%
3	Web	82,0%	0,6%	-4,2%*	10,1%*	6,6%

Table B.16: Mode effect decomposition for variable availability is other category. CAPI average is 1.9%

	Mode	*Average*	*NR*	*CO*	*ME*	*Total*
1	CATI	0,0%	0,0%	0,4%*	-2,3%*	-1,9%*
2	CATI	0,6%	0,6%	0,4%*	-2,3%*	-1,3%
3	CATI	1,9%	1,9%	0,4%*	-2,3%*	0,1%
4	CATI	2,1%	1,5%	0,4%*	-1,6%	0,3%
5	CATI	2,8%	0,9%	0,4%*	-0,4%	0,9%
6	CATI	2,7%	0,9%	0,4%*	-0,5%	0,8%
7+	CATI	3,0%	0,7%	0,4%*	0,1%	1,2%
1	Web	3,0%	1,2%	-0,2%	0,2%	1,2%
2	Web	5,7%	0,4%	-0,2%	3,6%	3,8%
3	Web	5,1%	0,3%	-0,2%	3,1%	3,2%

Table B.17: Mode effect decomposition for variable searches for a job. CAPI average is 45.7%

	Mode	*Average*	*NR*	*CO*	*ME*	*Total*
1	CATI	61,5%	0,5%	1,2%	14,1%	15,9%
2	CATI	44,0%	2,9%	1,2%	-5,8%	-1,7%
3	CATI	45,0%	-0,3%	1,2%	-1,6%	-0,7%
4	CATI	46,8%	-2,6%	1,2%	2,6%	1,1%
5	CATI	45,1%	-2,0%	1,2%	0,2%	-0,6%
6	CATI	46,2%	-1,6%	1,2%	0,9%	0,5%
7+	CATI	45,2%	0,4%	1,2%	-2,1%	-0,5%
1	Web	52,6%	3,0%	4,9%	-1,0%	7,0%
2	Web	57,6%	-0,6%	4,9%	7,5%	11,9%
3	Web	59,6%	-0,7%	4,9%	9,6%	13,9%

Table B.18: Mode effect decomposition for variable does not search for a job. CAPI average is 54.3%

	Mode	*Average*	*NR*	*CO*	*ME*	*Total*
1	CATI	38,5%	-0,5%	-1,2%	-14,1%	-15,9%
2	CATI	56,0%	-2,9%	-1,2%	5,8%	1,7%
3	CATI	55,0%	0,3%	-1,2%	1,6%	0,7%
4	CATI	53,2%	2,6%	-1,2%	-2,6%	-1,1%
5	CATI	54,9%	2,0%	-1,2%	-0,2%	0,6%
6	CATI	53,8%	1,6%	-1,2%	-0,9%	-0,5%
7+	CATI	54,8%	-0,4%	-1,2%	2,1%	0,5%
1	Web	47,4%	-3,0%	-4,9%	1,0%	-7,0%
2	Web	42,4%	0,6%	-4,9%	-7,5%	-11,9%
3	Web	40,4%	0,7%	-4,9%	-9,6%	-13,9%

Table B.19: Mode effect decomposition for variable number of offenses per 100 inhabitants. CAPI average is 41.6.

	Mode	*Average*	*NR*	*CO*	*ME*	*Total*
1	CATI	24,7	-6,0	0,3*	-11,3	-16,9
2	CATI	29,4	-4,9	0,3*	-7,6	-12,2*
3	CATI	29,6	-7,2*	0,3*	-5,1	-12,1*

4	CATI	31,7	-7,4*	0,3	-2,9	-10,0*
5	CATI	33,9	-7,1*	0,3	-0,9	-7,8*
6	CATI	33,7	-4,9	0,3	-3,3	-7,9*
7+	CATI	35,3	-2,0	0,3	-4,5	-6,3
1	Mail	62,5	-7,3	0,0	28,1*	20,9*
2	Mail	49,4	-9,9*	0,0	17,6*	7,8
3	Mail	50,8	-3,6	0,0	12,8*	9,2*
1	Web	61,0	-5,8	2,6	22,5*	19,3*
2	Web	59,1	1,3	2,6	13,6	17,5*
3	Web	56,2	-3,7	2,6	15,7*	14,5*

Table B.20: Mode effect decomposition for variable being victim of a crime. CAPI average is 26.5%

	Mode	*Average*	*NR*	*CO*	*ME*	*Total*
1	CATI	19,2%	-3,5%	0,0%*	-3,9%	-7,3%*
2	CATI	20,3%	-1,2%	0,0%*	-5,0%*	-6,2%*
3	CATI	19,7%	-2,3%*	0,0%*	-4,5%*	-6,8%*
4	CATI	19,9%	-2,4%	0,0%	-4,1%	-6,5%*
5	CATI	20,9%	-2,3%	0,0%	-3,3%	-5,6%*
6	CATI	21,0%	-1,8%	0,0%	-3,7%	-5,5%*
7+	CATI	22,7%	0,0%	0,0%	-3,8%*	-3,8%*
1	Mail	31,8%	-4,0%	0,0%	9,3%	5,3%
2	Mail	26,9%	-4,2%	0,0%	4,6%	0,4%
3	Mail	28,2%	-1,8%	0,0%	3,5%	1,7%
1	Web	33,7%	0,9%	1,3%	5,0%	7,2%*
2	Web	31,8%	2,5%	1,3%	1,6%	5,4%*
3	Web	32,0%	0,4%	1,3%	3,8%	5,6%*

Table B.21: Mode effect decomposition for variable neighbourhood nuisance scale. CAPI average is 1.47.

	Mode	*Average*	*NR*	*CO*	*ME*	*Total*
1	CATI	1,03	-0,30*	-0,04*	-0,10	-0,44*
2	CATI	1,12	-0,12	-0,04*	-0,19	-0,35*
3	CATI	1,21	-0,12*	-0,04*	-0,11	-0,27*
4	CATI	1,22	-0,09	-0,04	-0,13	-0,26*
5	CATI	1,23	-0,07	-0,04	-0,13	-0,24*
6	CATI	1,24	-0,05	-0,04	-0,14	-0,23*
7+	CATI	1,27	-0,01	-0,04	-0,16*	-0,21*
1	Mail	1,50	0,07	0,00	-0,05	0,02
2	Mail	1,32	-0,03	0,00	-0,12	-0,15
3	Mail	1,41	-0,01	0,00	-0,06	-0,07
1	Web	1,78	-0,15	-0,02	0,48*	0,31
2	Web	1,72	-0,02	-0,02	0,28*	0,24
3	Web	1,68	0,01	-0,02	0,21	0,20

Table B.22: Mode effect decomposition for variable feeling unsafe at times. CAPI average is 22.4%.

	Mode	*Average*	*NR*	*CO*	*ME*	*Total*
1	CATI	15,0%	-6,3%*	-0,2%*	-0,9%	-7,5%*
2	CATI	18,1%	-4,3%*	-0,2%*	0,3%	-4,3%*
3	CATI	17,1%	-2,6%*	-0,2%*	-3,0%	-4,1%*
4	CATI	17,5%	-2,7%*	0,4%	-2,0%	-4,9%*
5	CATI	18,0%	-2,2%*	0,4%	-2,0%	-4,4%*
6	CATI	18,0%	-2,2%*	0,4%	-2,0%	-4,4%*

7+	CATI	18,3%	-0,9%	-0,2%	-3,0%	-4,1%*
1	Mail	23,6%	-1,2%	0,0%	2,5%	1,2%
2	Mail	22,5%	-1,9%	0,0%	2,0%	0,1%
3	Mail	23,5%	0,1%	0,0%	1,0%	1,1%
1	Web	25,4%	-3,4%	0,4%	6,0%	3,0%
2	Web	26,9%	-2,7%	0,4%	6,7%*	4,4%
3	Web	28,5%	-0,7%	0,4%	6,3%*	6,1%*

Explanation of symbols

.	Data not available
*	Provisional figure
**	Revised provisional figure (but not definite)
x	Publication prohibited (confidential figure)
–	Nil
–	(Between two figures) inclusive
0 (0.0)	Less than half of unit concerned
empty cell	Not applicable
2013–2014	2013 to 2014 inclusive
2013/2014	Average for 2013 to 2014 inclusive
2013/'14	Crop year, financial year, school year, etc., beginning in 2013 and ending in 2014
2011/'12–2013/'14	Crop year, financial year, etc., 2011/'12 to 2013/'14 inclusive

Due to rounding, some totals may not correspond to the sum of the separate figures..

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

Prepress: Statistics Netherlands, Grafimedia
Design: Edenspiekermann

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

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