Discussion Paper

Experiences with mixed-mode surveys in times of COVID-19 at Statistics Netherlands

Kees van Berkel (CBS, Zuyd)

Jan van den Brakel (CBS, UM)

Daniëlle Groffen (CBS)

Joep Burger (CBS)

CBS: Statistics Netherlands

UM: Maastricht University

Zuyd: Zuyd University of Applied Sciences

August 26, 2024

Statistics Netherlands' social surveys are based on a sequential mixed-mode data collection approach using web, telephone, and face-to-face interviewing. This article illustrates how Statistics Netherlands addressed the sudden, unforeseen loss of face-to-face interviews in social surveys amidst the COVID-19 pandemic. At the beginning of the pandemic, survey processes were immediately adjusted in several ways to mitigate the negative effects of respondent attrition. Where possible, sampled people initially assigned to face-to-face interviewing were motivated to respond through web or telephone to minimize the loss of response. At the same time regression analysis and simulation were conducted to obtain quantitative insight into the effects of losing face-to-face responses in the sequential mixed-mode designs. Furthermore, alternative model-based estimation procedures based on structural time series models were implemented to compensate for the bias that is a result of the loss of face-to-face responses. These initiatives are illustrated with applications to the Dutch Labor Force Survey, the Housing Survey, and the Health Survey.

Keywords: Coronavirus, pandemic, measurement error, selection error, mode effect

This paper is published in The Statistical Journal of the International Association for Official Statistics, Berkel et al. 2024.

1 Introduction

The COVID-19 pandemic has significantly impacted data collection for social surveys conducted by national statistical offices. To ensure the safety of both survey participants and interviewers, adjustments were made to data collection strategies to comply with social distancing guidelines and other pandemic-related measures. Rapid adaptations were implemented to facilitate the transition from computer-assisted personal (face-to-face) interviewing (CAPI) to alternative data collection methods. A commonly employed approach among national statistical offices involved the use of remote data collection methods, including computer-assisted web interviewing (CAWI), computer-assisted telephone interviewing (CATI), computer-assisted video interviewing (CAVI), postal surveys, crowdsourcing, and internet panels. These methods allowed for the continuation of data collection while minimizing in-person interactions.

The UK Office for National Statistics (ONS) implemented a ”knock-to-nudge” approach in several social surveys, where interviewers visited sampled addresses to obtain telephone numbers from people at their doorsteps Kastberg and Siegler 2022. This necessitated the adaptation of questionnaires for telephone interviews, communication of changes in data collection to sampled people and users, and equipping interviewers to work from home. Telematching was extended, and sampled people were encouraged to enter their telephone numbers on an online portal in advance. Despite these efforts, survey response rates decreased, and changes in respondent characteristics were observed. To counteract these challenges, sample sizes were increased, incentive values were adjusted, and the knock-to-nudge approach was intensified.

During the pandemic, efforts were made to incorporate other data sources, either in combination with survey data or independently. For instance, the Australian Bureau of Statistics explored the use of administrative and transactional data from both the public and private sectors to enhance official social and economic statistics ABS 2021. The Destatis website showcases an interactive map highlighting successful measures taken by European National Statistical Institutes in response to the COVID-19 pandemic German Presidency of Council Working Party on Statistics 2020. Statistics Netherlands developed an experimental Corona sentiment indicator based on open data from Twitter and Facebook CBS 2020. A broader view on the effects of the COVID-19 pandemic on National Statistical Offices can be found in Wollburg et al. 2022.

Data collection of most of Statistics Netherlands’ household surveys is based on a sequential mixed-mode design using CAWI, CATI, and CAPI. In strict lockdown periods, no contact attempts were made by face-to-face interviewers. The general strategy was as follows. For surveys with CATI and CAPI in the mixed-mode design, the CAPI-sampled people for whom a telephone number was available were contacted via CATI. For surveys without CATI in the mixed-mode design, the observation period for CAWI was extended. In times of COVID-19 without strict lockdown, face-to-face interviewers were allowed to visit sampled people. In doing so, they had to adhere to the 1.5-meter distance rule. If distancing was not possible or if sampled people refused a face-to-face interview, interviewers were instructed to ask for a telephone number and make an appointment to interview by telephone. For surveys for which telephone interviewing is not allowed, including the Health Survey, the option of completing the questionnaire via the internet was offered again.

The sudden loss of CAPI response combined with the above-mentioned adapted approach strategy for face-to-face interviewers may affect survey results, as mode-specific measurement and selection errors may change Dillman and Christian 2005; Dillman, Phelps, et al. 2009; Dillman, Smyth, et al. 2014; De Leeuw 2005; Couper 2011; Schouten, Brakel, Buelens, Giesen, et al. 2022. To quantify these effects on social survey results, Statistics Netherlands used three techniques: logistic regression, simulation, and time series modeling. This paper demonstrates these techniques for three major surveys at Statistics Netherlands: the Labor Force Survey, the Health Survey, and the Housing Survey.

For the Labor Force Survey, logistic regression was employed to assess the extent to which CATI, replacing CAPI, affected the main survey results. Additionally, the existing time series of direct estimates were recalculated by excluding CAPI responses. This yielded two sets of direct estimates for each target variable: one with and one without CAPI respondents. Both sets were incorporated into multivariate state-space models to generate estimates unaffected by the sudden loss of CAPI respondents.

For the Housing Survey, a lockdown was simulated on a previous edition of the survey to evaluate the potential impact of reduced CAPI responses on the accuracy of estimates for key variables. By gradually removing CAPI responses, short to long lockdown scenarios were simulated. The remaining responses were weighted, and variable estimates were compared with the original ones.

For the Dutch health survey, time series modeling was introduced to produce statistics on a quarterly rather than annual basis. This allowed a faster assessment of the impact of COVID-19 on health indicators. At the same time, the model was used to correct for bias due to a reduction of face-to-face respondents, and an increase in web respondents.

The purpose of this paper is to illustrate how Statistics Netherlands handled the sudden, unforeseen loss of face-to-face respondents in social surveys. In particular, it is described which attempts were undertaken to reduce the loss of response as much as possible, to quantify the impact of switching from CAPI to CATI, and to implement model-based inference methods that correct for bias resulting from the sudden loss of CAPI respondents.

The paper is organized as follows: Section 2 outlines the general mixed-mode strategy for social surveys at Statistics Netherlands and describes the survey designs of the Dutch Labor Force Survey, Housing Survey, and Health Survey. Section 3 presents the adjustments to data collection at Statistics Netherlands during lockdown periods due to COVID-19. Sections 4, 5 and 6 delve into the analysis of observation mode effects for the Labor Force Survey, the Housing Survey, and the Health Survey during the pandemic, respectively. Finally, Section 7 summarizes the key insights gained during the pandemic.

2 Mixed-mode surveys at Statistics Netherlands

Statistics Netherlands employs a sequential mixed-mode strategy for most of its social surveys, starting with CAWI for all sampled people and follow-ups for non-respondents by CATI if a telephone number is available and by CAPI otherwise. This general strategy is denoted by CAWI → CATI/CAPI. Telephone interviewing is not always applied as a follow-up mode, e.g. if a part of the questionnaire has to be completed by the respondent himself. This is the case in the Health Survey.

The CAWI observation strategy involves sending a postal letter in advance to a selected sample of people, inviting them to complete an online questionnaire. The invitation letter includes information about the particular survey and instructions on how to access the personalized questionnaire. Non-respondents receive a maximum of two postal appeals at intervals of one or two weeks.

The CATI observation strategy comprises for each CAWI non-respondent with a known telephone number a maximum of 18 contact attempts, evenly distributed over the observation period. The distribution of contact attempts is automated in the CATI management system. A computer program is used to administer the questionnaires, with the interviewer reading the questions to the respondent and entering their answers into the computer.

In the CAPI observation strategy, each CAWI non-respondent without a known telephone number is visited a maximum of 6 times, evenly distributed over the observation period. The interviewers schedule the visits themselves, taking travel costs into account. The first three visits are unannounced, although cards are put in the letterbox notifying the visits. After that, appointments can be made by telephone because the third and subsequent cards show the interviewer's telephone number. The interviewers use laptops to fill out the questionnaires.

2.1 Survey design for the Labor Force Survey

The Labor Force Survey (LFS) aims to provide statistics about the participation of the Dutch population in the labor market. The target population consists of all people aged 14–89 years living in the Netherlands who do not belong to the institutional population. Core indicators are unemployed, employed, and total labor force. The survey applies a rotating panel design with five waves at three-month intervals.

The sampling design for the first wave is a stratified sample of people aged 14–89 years with unequal probabilities. The strata are the municipalities. In each municipality, a simple random sample is drawn from people aged 14–89 years, with a sample size proportional to its number of inhabitants. This initially yields a self-weighting sample. Thereafter, to improve the precision of unemployment figures, people registered at the Netherlands Employees Insurance Agency as job seeker are given a higher inclusion probability. To compensate for relatively low response rates, non-western migrants and 15–24-year-olds are also given a higher inclusion probability. People aged 65 or over and 14-year-olds get a lower inclusion probability because they are less relevant to the survey. The sampling design aims to provide about 935 respondents per week in the fifth wave.

The observation strategy for the first wave is CAWI → CATI/CAPI, with different CATI- and CAPI-sampling fractions for different groups of CAWI-non-respondents. This adaptive design is described in detail by Van Berkel 2022. The observation strategy for the subsequent waves is CAWI → CATI, meaning that after everyone is asked to complete a questionnaire via the internet, all non-respondents will be contacted by telephone, provided a phone number is available. As a response-increasing measure, gift cards and iPads are raffled among the sampled people in the first wave.

2.2 Survey design for the Housing Survey

The Housing Survey provides information on the housing situation and housing preferences of Dutch households. The target population consists of all people aged 18 years or over living in the Netherlands who do not belong to the institutional population. Core indicators are housing situation and housing costs of tenants and homeowners, moving dynamics, housing quality, living experience, sustainability, and energy consumption. The results of the survey are an important basis for the Dutch government's housing policy. The survey is conducted once every three years with a six-month observation period and follows a cross-sectional design. In addition to the national survey, provinces, municipalities, and housing associations are offered the possibility to obtain additional responses, to get reliable estimates at the local level. In the remainder of this section, only the national part is considered.

The sample is a stratified sample in which people aged 18 years or over are selected with unequal probabilities. The selection probability of people registered as 'partner' in the Personal Records Database is half of the selection probability of the other people. For stratification, a partition of the Netherlands is made into 19 areas. Using response probabilities from previous editions, the sample is initially constructed so that the expected numbers of respondents per stratum are proportional to the population numbers per stratum, and that the total number of respondents is 40,000. Thereafter, the sample is increased in areas where fewer than 1,200 respondents are expected. This is at the expense of the sample in the other areas which is reduced proportionally to the size per area, returning the expected total response to 40,000.

The observation strategy is CAWI → CATI/CAPI, with a CATI-sampling fraction and a CAPI-sampling fraction to obtain the pre-agreed sample sizes to be contacted by telephone and face-to-face respectively. As a response-increasing measure, all sampled people receive a five-euro gift card along with the advance letter.

2.3 Survey design for the Health Survey

The Dutch Health Survey aims to provide as complete an overview as possible of developments in health, medical contacts, lifestyle, and preventive behavior of the population in the Netherlands. The target population consists of all people living in the Netherlands who do not belong to the institutional population. The survey applies an ongoing cross-sectional design.

The sample is a stratified sample in which people are selected with equal probabilities. The strata are the municipalities. In each municipality, a simple random sample is drawn from inhabitants, with a sample size proportional to its number of inhabitants. This yields a self-weighting sample. The target number of respondents is about 10,000 per year.

The observation strategy is CAWI → CAPI meaning that every sampled person is asked to participate in the survey via the internet, and a specific sample of non-respondents will be visited at home by an interviewer to complete a questionnaire. For a detailed description of the adaptive survey design for the Health Survey Van Berkel et al. 2020. As a response increasing measure, gift cards and iPads are raffled among the sampled people.

3 Data collection at Statistics Netherlands during lockdown

At Statistics Netherlands, CAPI was completely stopped due to lockdown restrictions from mid-March to the end of August 2020, from mid-December 2020 to the end of January 2021, and from mid-December 2021 to mid-January 2022. In general, for surveys with CATI in the observation strategy, the CAPI-sampled people for whom a telephone number was available were contacted via CATI. For surveys without telephone interviewing in the observation strategy, online observation was extended.

From September 2020, face-to-face interviews were allowed again, adhering to the 1.5-meter distance rule. If distancing was not possible or if sampled people refused a face-to-face interview, interviewers were instructed to ask for a telephone number and make an appointment to interview by telephone. For surveys for which telephone observation is not allowed, the option of completing the questionnaire via the internet was offered again. This strategy turned out to be successful and the response rates increased, improving the precision of the survey outcomes. However, this did not completely solve the problem of mode-specific measurement and selection bias.

4 Estimating mode effects

Mode effects refer to systematic differences in survey outcomes that arise when using different modes of data collection Dillman and Christian 2005; Dillman, Phelps, et al. 2009; Dillman, Smyth, et al. 2014; De Leeuw 2005; Couper 2011; Schouten, Brakel, Buelens, Giesen, et al. 2022. Mode effects are the combined result of selection and measurement effects. Mode-dependent selection effects relate to differences in coverage and response propensity. If not all sampling units have access to the mode, then the mode suffers from undercoverage. The response rate of the sampling units depends on the mode, an effect known as mode-dependent non-response behavior De Leeuw 2005. The coverage and non-response effects are collectively referred to as the mode-dependent selection effect Schouten, Brakel, Buelens, Giesen, et al. 2022, Chapter 4. Most agencies do not have access to an integral telephone register. Therefore, people in the target population whose phone number is not known to the agency cannot be contacted by telephone, leading to undercoverage for CATI. If the agency has access to the Personal Records Database consisting of all people registered at a municipality, the undercoverage for CAPI is small. Moreover, some people prefer to participate via telephone rather than in person, or vice versa. This need not cause problems, except if the response propensities are related to one or more target variables of the survey.

Mode-dependent measurement effects refer to differences in answers that the same respondent provides when questions are posed in different modes Dillman, Smyth, et al. 2014 and Schouten, Brakel, Buelens, Giesen, et al. 2022, Chapter 3. Measurement differences between different observation methods may be caused by psychological implications such as the presence or absence of an interviewer or the difference between aural and visual questioning. For instance, respondents may provide more socially desirable answers to interviewers than via the internet or in writing. Measurement errors cause a difference between the true value of a target variable and the value processed by the survey.

During the COVID-19 pandemic, face-to-face interviewers interviewed some of the people in their sample by telephone. This may have led to selection and measurement effects. It was investigated whether these effects existed and whether they could be explained by variables in the weighting process.

Estimating mode effects and classifying them into mode-specific selection effects and mode-specific measurement effects requires sophisticated experiments Schouten, Brakel, Buelens, Van der Laan, et al. 2013; Brakel 2008. These experiments are expensive and time-consuming. During the pandemic, there was no time to conduct such an experiment. Nevertheless, immediate action was needed to interpret the effect of changes in data collection on the survey results. Switching from face-to-face to telephone interviewing required rapid analysis of survey results. As an alternative, logistic regression analyses were applied to get an idea of the errors introduced by the switch of modes. In line with the regression analysis according to Jäckle et al. 2010, two logistic regression models were considered.

In the first model, target variables are only explained by the mode of response applying four categories: (1) CAWI, (2) CATI, (3) CAPI-tel, and (4) CAPI-ftf. Here CAPI-tel means telephone response to CAPI-sampled people, and CAPI-ftf means face-to-face response to CAPI-sampled people. Category 4 was taken as the benchmark category. It is assumed that the difference between categories 3 and 4 can be used as an approximation of the overall mode effect for telephone versus face-to-face interviewing, that is the mode-specific selection and measurement effects together.

In the second model, target variables are explained by both mode of response with the above categories, and auxiliary variables used in the regular weighting model for non-response correction. Weighting models for the production of official publications are carefully selected models that, given the available auxiliary information, correct for selective non-response as best as possible. It is therefore assumed that the selection effect is explained by these auxiliary weighting variables and that the remaining difference between categories 3 and 4 represents the mode-specific measurement effect for telephone versus in-person interviewing.

For the analysis, data from the Dutch LFS collected from July 2021 to December 2021 was used. The data from first-wave respondents include information on the mode of observation, important target variables such as employment status, working hours per week, highest attained educational level, labor market position, and variables that were used during the weighting procedure: sex, age, migration, type of household, applied for unemployment benefit or looking for a job, gross monthly personal income, most important source of income, and region.

From July to December 2021, 1203 CAPI-sampled people switched from face-to-face to telephone interviewing. This concerns 40.2% of the respondents coming from CAPI interviewers. Every month, this proportion ranges from 32% in July to 65% in December. Of all 32040 interviews over the whole period, 79.8% was conducted via CAWI, 10.9% via CATI, 3.8% switched from face-to-face to telephone, and 5.6% was face-to-face.

Four target variables produced an odds ratio (OR) significantly different from 1 in the first logistic regression model (Table 4.1). An OR greater/less than 1 indicates that there is a higher/smaller probability of the particular outcome appearing among respondents who switched from face-to-face to telephone, compared to the reference category of face-to-face interviews. Such a difference is due to the overall mode effect of selection and measurement effects together. An OR equal to 1 indicates that there is no difference between the two categories. If 1 belongs to the 95% confidence interval (CI) of an OR, the effect found is due to randomness with a probability of at least 95%.

Table 4.1 Logistic regression analyses of mode [1] on target variables, LFS 2021.

	Model 1 [2]	Model 2 [3]
	OR (95% CI)	OR (95% CI)
Employed labor force	1.66 (1.41-1.96)	1.45 (1.17-1.81)
Permanent position	1.28 (1.10-1.50)	1.14 (0.93-1.41)
Working more than 28 hours per week	1.42 (1.23-1.65)	1.27 (1.04-1.55)
Highest attained educational level	1.31 (1.11-1.54)	1.11 (0.92-1.33)

[1] Odds Ratio's (OR) and Confidence Intervals (CI) of the CAPI sampled people who responded by telephone are shown. The CAPI face-to-face mode was used as a benchmark category. OR for the original CATI and CAWI modes are not shown.

[2] Unadjusted for weighting variables.

[3] Adjusted for sex, age, migration background, type of household, applied for unemployment benefit or looking for job, gross monthly personal income (salary and benefit), most important source of income, region. By means of likelihoodratio tests, non-significant variables were excluded.

In logistic regression model 2 explanatory weighting variables were added, making an OR unequal to 1 attributable to mode-specific measurement effects only. Two variables still show significant differences. People interviewed by telephone were significantly more likely to belong to the employed labor force (OR = 1.45, CI = 1.17–1.81) and they were more likely to work more than 28 hours per week (OR = 1.27, CI = 1.04–1.55). This indicates that, for these variables, there is evidence for mode-specific measurement effects that cannot be adjusted for during the weighting procedure. Researchers should therefore be aware that the adapted observation strategy increases the precision of survey results but may introduce biases differing from those in previous periods.

5 Simulation

The Housing Survey consists of two parts. The first part is a national sample aiming for 40 thousand respondents in 19 areas. The second part consists of samples that third parties like municipalities or provinces can buy extra to get more precise regional estimates. These third parties are referred to as oversampling parties.

On 15 December 2020, the Netherlands went into a second lockdown to prevent the spread of the coronavirus. This was about a quarter into the observation period. Oversampling parties were allowed to back out of the agreement. At that point, it was unclear how long the lockdown would last and when or if CAPI could be resumed before the end of the observation period. To better inform the oversampling parties about the consequences of the duration of the lockdown on the quality of the estimates, a simulation was conducted on an earlier (2018) edition of the Housing Survey. Oversampling parties that participated in 2018 but not in 2021 were removed.

Six scenarios were compared with a null scenario (Table 5.1). In the null scenario, there was no lockdown, resulting in the published estimates that served as the benchmark. These estimates are based on 50 thousand respondents, 14% of which were obtained through CAPI. The first scenario simulates the most positive scenario of a short lockdown from the second half of December through February, after which CAPI could be resumed. This would only affect the national sample because CAPI for the oversampling parties started in March. In the other scenarios, the lockdown was prolonged stepwise by a month until the lockdown would take longer than the observation period and CAPI would be limited to the observation period before the lockdown (22% of the planned CAPI response, or 4% of 44 thousand respondents). Thus, the lockdown not only reduced the response rate but also changed the ratio between modes. In the simulation, the CAPI deficit was not compensated with other modes, because that would change the mode ratio even more and rather increase the bias.

Removing the CAPI respondents by month would only yield a single realization per scenario. Instead, we randomly sampled the number of CAPI respondents shown in Table 5.1 across all months, stratified by sub-province (NUTS-3) and oversampling domain. These CAPI respondents were removed from the response and some key population parameters were re-estimated (see below). The sampling of CAPI respondents across months allowed us to repeat the process and get a more robust estimate of the effect by averaging the results across samples. Scenarios with a longer lockdown are sub-samples from scenarios with a shorter lockdown.

Table 5.1 Number of CAPI respondents per scenario. Gray cells are lost due to a fictitious lockdown.

	2017		begin	end	2018						
Scenario	Oct	Nov	Dec	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul
0	602	602	361	241	602	602	813	813	813	813	813
1	602	602	361	241	602	602	813	813	813	813	813
2	602	602	361	241	602	602	813	813	813	813	813
3	602	602	361	241	602	602	813	813	813	813	813
4	602	602	361	241	602	602	813	813	813	813	813
5	602	602	361	241	602	602	813	813	813	813	813
6	602	602	361	241	602	602	813	813	813	813	813

For each scenario, design weights were recalculated by dividing per sampling stratum the total number of people by the number of respondents. Correction weights were calculated by linear weighting using a somewhat simplified version of the original weighting model. The original weighting model had to be simplified because the lockdown reduced the number of (CAPI) respondents. This is the simplified weighting model where for each categorical variable, the number of classes is given in parentheses: stratum (62) $\times$ partner (2) + municipality (376) + gender (2) $\times$ age (15) + gender (2) $\times$ origin (3) + age (7) $\times$ origin (3) + region (8) $\times$ origin (3) + province (12) $\times$ income (5) + real estate value (23) + household position (5) + household size + region (68) $\times$ gender (2) + region (68) $\times$ age (7) + region (68) $\times$ property ownership (3) + region (68) $\times$ dwelling type (2).

Generalized regression estimators (GREG) were applied to estimate a number of key population parameters: the number of households wishing to move out of a rental or owner-occupied property, the fraction of households in a rental or owner-occupied property that is (very) satisfied with the neighborhood, and the fraction of their income spent on housing costs. The accuracy of the estimates was assessed by the relative root mean squared error ($RRMSE(\hat{\theta}_s) = \frac{\sqrt{[B(\hat{\theta}_s)]^2+[SE(\hat{\theta}_s)]^2}}{\hat{\theta}_0}$), relative bias ($RB(\hat{\theta}_s) = \frac{B(\hat{\theta}_s)}{\hat{\theta}_0}$) and relative standard error, also known as coefficient of variation ($CV(\hat{\theta}_s) = \frac{SE(\hat{\theta}_s)}{\hat{\theta}_0}$), where $\hat{\theta}_s$ is the point estimate of a population parameter in scenario $s = 0, \dots, 6$, $B(\hat{\theta}_s) = \hat{\theta}_s - \hat{\theta}_0$ the bias of $\hat{\theta}_s$ relative to the estimate in scenario 0, and $SE(\hat{\theta}_s)$ the standard error of $\hat{\theta}_s$.

Figure 5.1 shows the effect of the duration of the lockdown on the relative accuracy of the three population parameter estimates by home ownership. Note that the estimates under the null scenario are unbiased by definition so that the accuracy is completely determined by the standard error. The relative accuracy (RRMSE) increases more rapidly as the lockdown lasts longer. In the first hundred days, the effect remains relatively small. Reduced CAPI has mainly an effect on the relative bias (RB) and only a minor effect on the relative standard error (CV). The increase in bias implies that the weighting model cannot correct for any remaining selectivity in the CAPI response or that respondents answer differently in CAPI than in the other modes. The decrease in precision is caused by a reduced sample size and a larger variation in weights. Moreover, the effects depend on the target variable. The lockdown has almost no effect on the accuracy of the fraction of homeowners that are (very) satisfied with the neighborhood and the relative housing costs. Possible explanations are that some questions are more sensitive to mode effects than others and that CAPI has a larger share in some domains than in others. The effects also varied across oversampling domains (not shown). In particular, the bias was positive (overestimation) for some oversampling domains and negative (underestimation) for others. Based on the results of this simulation, all oversampling parties decided to pursue, i.e. none contracted out of the agreement with Statistics Netherlands.

The real lockdown turned out to run through May (scenario 4). In an attempt to make up for the missed CAPI, extra letters were sent out from January through March, requesting non-respondents once more to respond via CAWI. In April, CAPI sampling units were visited and asked for a telephone number to hold the interview by CATI (CAPI-tel). This CAPI-tel mode was prolonged after the lockdown. In addition, the fieldwork was prolonged through September. In the 2018 simulation, these compensation measures were not included, because they were unknown at the time. They have helped to meet the agreed number of responses and to increase the precision of the estimates. Still, they have distorted the mode ratio even further, which may have increased the bias and reduced the overall accuracy of the estimates.

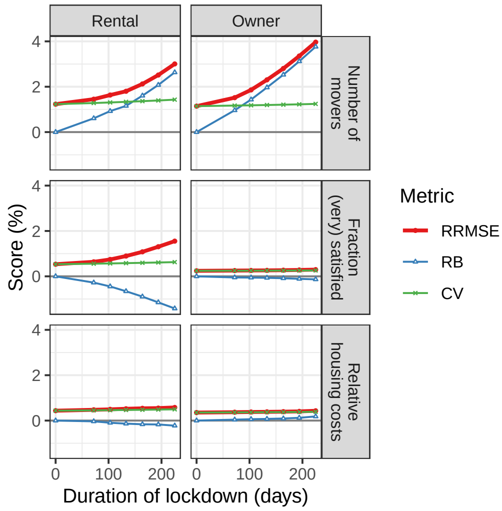


Figure 5.1 Effect of simulated lockdown duration on accuracy of three population parameter estimates (row panels) by home ownership (column panels).

6 Time series modeling

Due to the first lockdown, resulting from the COVID-19 pandemic, CAPI data collection stopped in the second quarter of 2020. Compared to a single CAPI data collection design, the sequential mixed-mode design had the advantage that CAWI and CATI data collection were still operational. Normally 30%–35% of the sampling units respond via CAPI. The loss of the CAPI response, nevertheless reduced the overall response size and created sudden shocks in the official figures due to a change in the measurement and selection bias since a specific subpopulation was interviewed using CAPI.

Furthermore, the COVID-19 pandemic made clear that besides precision and bias, timeliness is at least as important quality indicator for official statistics. The Dutch Health Survey (DHS) is designed to publish annual figures, which implies that the first official figures on the impact of the COVID-19 pandemic on health-related themes would become available in the first quarter of 2021, which compromises the relevance of the survey dramatically. This resulted in a strong demand for more timely quarterly figures for the key variables of the DHS from several external data users. The sample size of the DHS is, however, not large enough to apply a standard direct inference method like the general regression (GREG) estimator Särndal et al. 1992 to produce sufficiently precise quarterly figures.

A solution for both aforementioned problems was found by developing a bivariate structural time series model. With a structural time series model, an observed time series is decomposed into a trend component, a seasonal component, other cyclic components, a regression component, and an irregular component. For each component, a stochastic model is assumed. This allows the trend, seasonal, and cyclic components but also the regression coefficients to be time-dependent. See Durbin and Koopman 2012 for details about structural time series modeling.

The input for the bivariate model is two series containing quarterly GREG estimates for the variable of interest on a quarterly frequency from the first quarter of 2014 until the second quarter of 2021.

The first series, say y_t^C, contains the quarterly GREG estimates that are based on the full response of CAWI and CAPI (superscript C stands for Complete response). This series is recalculated using the CAWI response only, denoted y_t^I (superscript I stands for Internet response), which is used as the second series in the model. In this way a parallel run is created for six years where data are collected using the complete data collection approach versus CAWI only. It is understood that the systematic difference between both series is the relative bias that arises in the DHS estimates due to the loss of CAPI during the data collection. This difference is the net result of mode-dependent selection bias and measurement bias. The bivariate model for both series is defined as:

$$\begin{pmatrix} y_t^C \\ y_t^I \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} (L_t + S_t) + \begin{pmatrix} 0 \\ \lambda_t \end{pmatrix} + \begin{pmatrix} e_t^C \\ e_t^I \end{pmatrix}. \tag{1}$$

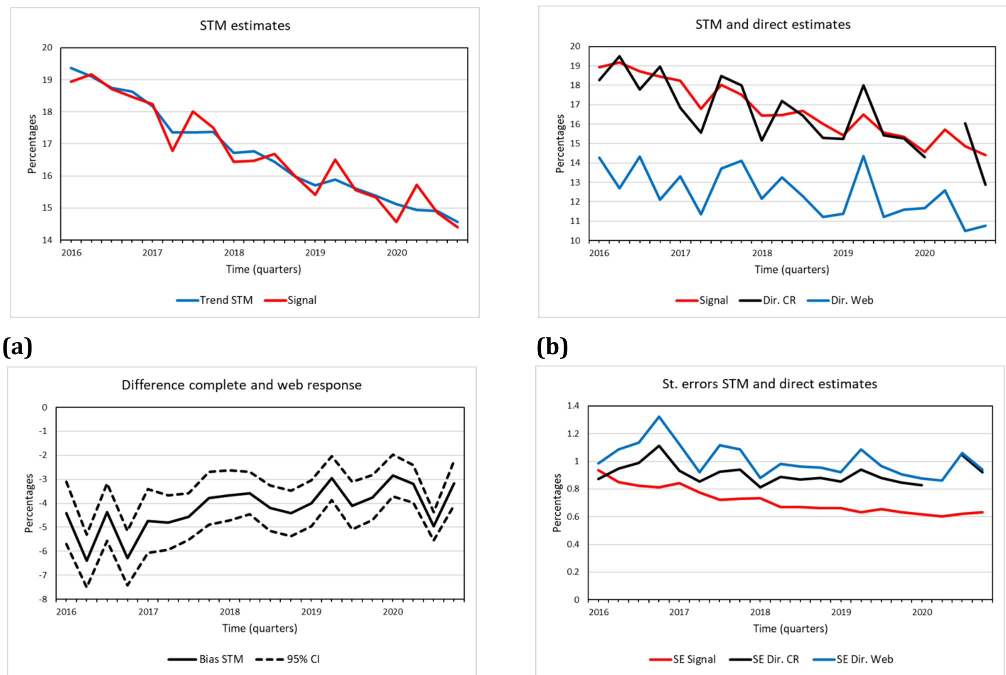
Both series share a common trend, say L_t, and a common seasonal component, say S_t. For L_t the smooth trend model and for S_t the trigonometric seasonal model is assumed, Durbin and Koopman 2012 for details. Furthermore, λ_t denotes the systematic difference or relative bias between y_t^C and y_t^I, which is modeled with a random walk. As a result, λ_t is time-dependent, which gives model (1) the flexibility to let the relative bias between y_t^C and y_t^I gradually change over time. Finally, e_t^C and e_t^I denote the measurement errors of the input series. The model accounts for heteroscedasticity in the measurement errors by making the variance of e_t^C and e_t^I proportional to the variance of the GREG estimates of y_t^C and y_t^I, respectively. The model also accounts for the correlation between the input series, which arises since the GREG estimates of both input series are partially based on the same CAWI respondents. The variances and covariances of y_t^C and y_t^I are estimated from the survey data. Model (1) is expressed in state space form and fitted with the Kalman filter. Details of this approach are described in Brakel and Smeets 2023.

The parameter λ_t measures the difference between the two input series and can be interpreted as the bias in the estimates based on CAWI only compared to the series based on the complete response. It can be interpreted as the additional bias in the outcomes of the DHS due to the sudden loss of the CAPI respondents in the data collection. It is called a relative bias since it is the systematic difference between estimates based on CAWI respondents only and estimates based on the complete response in CAWI and CAPI. A negative value for λ_t implies an underestimation of the variable of interest if the estimate is based on CAWI respondents only. A positive value for λ_t implies an overestimation of the variable of interest if the estimate is based on CAWI respondents only.

Model (1) is used to produce quarterly estimates for the trend L_t and the signal $L_t + S_t$ of the parameter of interest. The contribution of the model is twofold. First, during the lockdown, there is only an observation for y_t^I but not for y_t^C. Based on the relation between y_t^C and y_t^I, observed in the period before the lockdown, an estimate for the trend and the signal that is corrected for the relative bias between y_t^C and y_t^I is obtained since this effect is captured by λ_t. The Kalman filter computes a prediction for the unobserved components (L_t, S_t, and λ_t) in the presence of missing observations for the respective observable variables. Under the assumption that the relative bias between the input series is not affected by the lockdown, a bias-corrected estimate is obtained based on the internet response only. Second, the time series model is used as a form of small area estimation by using sample information from preceding periods to produce stable quarterly estimates that are more precise than the quarterly direct GREG estimates.

This approach is applied to eight key variables of the DHS, which are listed in Table 6.1. These 8 variables are published as timely official health indicators during the COVID-19 pandemic. The quarterly model-based estimates of these eight variables are also used in the weighting scheme of the GREG estimator for the annual figures to obtain annual estimates that are numerically consistent with the quarterly figures. In this way the annual GREG estimates for these target variables including more detailed breakdowns are corrected for the bias due to the loss of the CAPI response. GREG estimates for variables that are related to these eight target variables are possibly also corrected, at least partially, for the loss of the CAPI response.

Results for one variable, the fraction of people with a daily smoking habit, are presented in Figure 6.1. Figure 6.1.b and 6.1.c show that there is a significant relative bias between the input series. For smoking the estimates based on the internet response are significantly smaller compared to the complete response, resulting in a negative estimate for λ_t that is significantly different from zero at a 5% significance level. Ignoring this bias would result in an unrealistic small estimate for smoking during the lockdown in the second quarter of 2020. The Kalman smoother estimates for trend and signal, based on the state space model, are corrected for this bias since the model (1) accommodates this bias in a separate parameter λ_t. Figures 6.1.a and 6.1.b show that the Kalman smoother estimates for trend and signal are clearly less volatile than the quarterly GREG estimates based on the complete response. Indeed, the standard errors of the smoothed signal are about 25% smaller compared to the standard errors of y_t^C. This exercise shows that model (1) successfully corrects for the bias that is introduced due to the loss of CAPI response during the lockdown and at the same time can be used to produce sufficiently precise timely quarterly estimates.



(a) (b) (c) (d)

Figure 6.1 Results daily smoking. Panel (a): Kalman smoother estimates for trend (L_t) and signal $(L_t + S_t)$. Panel (b): Kalman smoother estimates signal, input series complete response (y_t^C) and web response (y_t^I). Panel (c): Kalman smoother estimates relative bias between input series (λ_t) with 95% confidence int. Panel (d): standard errors Kalman smoother estimates signal, input series complete response, and web response.

Table 6.1 provides an overview of the estimates for the bias parameter λ_t for the eight variables of the DHS to which Model (1) is applied. These values correspond to the period before the start of the first lockdown, i.e. the first quarter of 2020. In addition, the level of the trend (L_t) and the bias relative to the trend ($100 * \lambda_t / L_t$) are presented. It follows that the impact of the loss of CAPI respondents differs between the eight target variables. The bias is significantly different from zero at a 5% significance level for all variables except overweight. Failing to correct for the bias results in a substantial overestimation of dental visits, specialist visits, and feeling mentally unhealthy and a substantial underestimation of daily smoking and excessive alcohol consumption. The effects of perceived health and overweight are minor. As explained before, the observed effects are the net result of a change in the composition of mode-dependent measurement bias and selection effects. In this application to the DHS, it can be anticipated that with the CAWI approach the more easily persuaded respondents are selected, who might be more interested in and concerned with their health. This could explain why the estimates based on the CAWI respondents only are lower for smoking, overweight, and alcohol consumption and higher for good perceived health, GP visits, dental visits, and specialist visits. Explaining the selection mechanism for the mentally unhealthy is less obvious. It should, however, be noted that other effects like providing more socially desirable answers under the CAPI mode and other measurement error sources also play a role. Details of this time series modeling approach and the results for other variables are described in Brakel and Smeets 2023.

A similar state space approach is developed for the Dutch LFS to produce official monthly labor force figures that are corrected for the loss of CAPI response during the lockdown periods Brakel, Souren, et al. 2022. Table 6.2 contains the results for the relative bias for the monthly direct estimates with and without CAPI for the unemployed, employed, and total labor force. The presented results are based on the monthly series observed from 2013 until March 2020. One series is based on the complete response and one series is based on CAWI and CATI response only. This result is in line with the logistic regression results for the employed labor force in Table 4.1, where it is concluded that there is evidence for mode-specific measurement error resulting in a higher amount of employed people among the CAPI respondents. The relative bias is significantly different from zero at a 5% significance level for all three variables. For the unemployed labor force, the loss of CAPI would result in an underestimation of about 7%. For the employed and total labor force, the loss of CAPI respondents results in a marginal overestimation. The official figures of the monthly labor force figures are also itemized by age and gender in six domains. The effect of the loss of CAPI for some of these domains is more substantial. The inference method used for the publication of monthly labor force figures, however, corrects for these effects in a similar way as explained for the DHS.

Table 6.1 Differences complete response (CAWI+CAPI) and CAWI response for eight target variables DHS.

Variable	Bias (λ)	(St. Er.)	Level (L)	Relative bias $(100 * \lambda/L)$
Percentage (very) good perceived health	0.7	(0.31)	82.0	0.8
Percentage feeling mentally unhealthy	0.8	(0.15)	11.8	6.8
Percentage GP consult over past 4 weeks	1.1	(0.20)	24.5	4.5
Percentage daily smoking	−3.5	(0.46)	14.5	−24.1
Percentage overweight	−0.5	(0.5)	50.0	−1.0
Percentage excessive alcohol cons.	−1.3	(0.3)	7.3	−17.8
Percentage dental visit over past 4 weeks	2.2	(0.35)	19.3	11.4
Percentage specialist visit over past 4 weeks	1.5	(0.34)	16.0	9.4

Table 6.2 Differences complete response (CAWI+CAPI+CATI) and CAWI+CATI response for monthly LFS figures.

Variable	Bias (λ)	(St. Er.)	Level (L)	Relative bias $(100 * \lambda/L)$
Unemployed labor force	−21,000	(2,600)	300,000	−7.0
Employed labor force	35,000	(4,200)	9,000,000	0.4
Total labor force	14,000	(4,100)	9,300,000	0.2

7 Discussion

For decades, CAPI uni-mode data collection was the standard approach in many national statistical institutes. The COVID-19 pandemic suddenly created a new risk factor for this fieldwork strategy and increased the awareness that fieldwork strategies that are more robust for sudden lockdowns are required. Non-interviewer-driven data collection modes are of course most robust against the sudden loss of data due to a lockdown, but they generally result in lower response rates. Besides reducing data collection costs, this additional risk is an important motivation to move towards mixed-mode designs that apply a strategic combination of interviewer-driven and non-interviewer-driven modes. The sequential mixed-mode data collection approach implemented at Statistics Netherlands proved to be more robust for the risks of a sudden loss of data due to a lockdown, at least compared with a uni-mode CAPI data collection approach. The sudden loss of CAPI nevertheless reduced the data quality of Statistics Netherlands' surveys. To this end, several ad hoc projects were launched during the first lockdown.

In the first project, regression analyses were conducted to obtain more insight into which extent mode-specific selection bias and mode-specific measurement bias affect the outcomes of the surveys. Two logistic regression models were used to analyze LFS data in the latter half of 2021. The first model explains the target variables by mode of response, revealing the mode effect for telephone versus in-person interviewing. The second model explains the target variables by both mode of response and auxiliary variables used in the weighting procedure, assuming that the selection effect is explained by the auxiliary weighting variables. The results show that for some variables, including employed labor force and working more than 28 hours per week, there is evidence for mode-specific measurement effects that cannot be adjusted for during the weighting procedure. This suggests that the observation strategy used may increase precision but also introduce bias in certain outcomes. Researchers should take this into account when interpreting the results. Overall, the study highlights the importance of considering mode effects in survey research and the need for researchers to be vigilant in their data analysis. A major limitation of this approach is that it is based on the strong assumption that the covariates from the weighting model correct for selection bias. Experiments based on re-interview designs are more appropriate to disentangle mode-specific measurement bias from mode-specific selection bias, but impracticable in the case of an unexpected crisis.

In the second project, a simulation was conducted. The simulation provided relatively fast some insight into the potential effects of the lockdown on the accuracy of key population parameters, without the need for additional data collection. Removing CAPI respondents, reweighting the remaining responses, and comparing the new estimates to the original estimates provided enough information for the oversampling parties to decide if they wanted to pursue. This approach has limitations, of course. First, the simulation is based on historical data and the results may not generalize to the new edition. Not all oversampling parties participated in both editions. Second, the bias is relative to the null scenario, which may be biased by itself. Some groups may not respond in any mode. Finally, the effects depend on the target variable and oversampling domain, which makes it difficult to draw general conclusions. The simulation nevertheless provides quantitative insight into the comparability of the results over time.

The COVID-19 pandemic increased the awareness that precision is not the only quality concept for official statistics. Quality dimensions such as timeliness and comparability over time are at least as important. Therefore, in a third project, model-based inference methods that are based on structural time series models were developed to produce timely and sufficiently precise estimates that are corrected for the bias that is introduced due to the loss of the CAPI respondents. This approach is implemented for the key variables of the DHS and the LFS. An additional advantage of the time series analysis is that it gives quantitative insight into the relative bias that arises due to the loss of CAPI respondents. The method has some limitations. First, it is less appropriate for multipurpose surveys since a time series model has to be developed for each variable separately. Second, the method is based on the strong assumption that the differences between estimates with and without CAPI respondents are not affected by the lockdown.

A general conclusion that can be drawn from the three aforementioned projects is that the sequential mixed-mode strategy, currently in place at Statistics Netherlands is not sufficiently robust against lockdowns. One approach is to develop fieldwork strategies where interviewers have more flexibility to choose between different modes. In the past three years, some of the interviews that should have been face-to-face were conducted by another observation method, such as telephone or web. In 2022, for the LFS, 16% of the interviews that should have been conducted in person were conducted by telephone. In the first three months of 2023, this proportion was 19%. Although the face-to-face interviewers were instructed to offer the option of interviewing by phone sparsely, it is applied fairly frequently. Interviewers like the ease of a telephone interview, and it is time- and cost-saving as no travel is required. Researchers are aware of possible mode-related measurement effects. For the LFS a time series modeling approach was implemented to correct for this additional bias. To allow more flexibility in the use of modes during data collection, both questionnaires that are less sensitive to mode-dependent measurement bias, and inference methods that at least partially correct for mode-dependent measurement bias are needed.

Another potential approach is to develop and incorporate Computer Assisted Video Interviewing (CAVI) into mixed-mode survey designs. Some arguments for this are: (1) Enhanced data quality. CAVI can improve data quality by reducing errors in data collection. Video interviewing can provide a more engaging and interactive experience for respondents, resulting in more accurate and reliable data. (2) Increased response rates. CAVI can help increase response rates by making surveys more accessible and convenient for respondents. Video interviews can be conducted remotely, which can be especially beneficial for hard-to-reach populations. (3) Cost-effective: CAVI can be a cost-effective method of data collection as it eliminates the need for face-to-face interviews, which can be expensive due to travel costs and travel time. (4) Increased scheduling flexibility. With CAVI, respondents can complete interviews at their convenience, and interviewers can schedule the interviews more efficiently without regard to travel time. (5) Adaptable to different survey designs. CAVI can be used in a variety of survey designs, including longitudinal studies, cross-sectional studies, and panel studies. This flexibility makes it a valuable tool for National Statistical Institutes.

In January-March 2022, Statistics Netherlands conducted a small-scale experiment with video interviewing. Twenty-seven interviewers were trained and equipped with laptops running Zoom software. In case people refused a face-to-face interview, interviewers had the option of interviewing via Zoom. Preliminary findings from this experiment suggest that conducting interviews via Zoom increases response rates, especially in situations considered unsafe. In addition, participants who opted for a Zoom interview were found to be younger and have higher household incomes than those who participated face-to-face.

It is necessary to further develop fieldwork strategies that are less sensitive to mode-dependent measurement and selection bias. One way to reduce the impact of sudden changes in mode effects is to develop questionnaires for mixed-mode surveys, which minimize differences in responses across data collection modes. There is a vast amount of literature on designing questionnaires tailored to mixed-mode data collection Dillman, Smyth, et al. 2014; De Leeuw 2018; Tourangeau 2017; Schouten, Brakel, Buelens, Giesen, et al. 2022.

Obtaining quantitative insight into mode-dependent measurement and selection bias requires controlled experiments with repeated measurements as proposed by Schouten, Brakel, Buelens, Van der Laan, et al. 2013; Klausch et al. 2017 instead of or in addition to the various analyses described in this paper.

Another strand of research is to develop inference methods that attempt to correct and adjust for mode effects. Inference methods that could be considered are the regression modeling approach proposed by Suzer-Gurtekin 2013, the imputation methods proposed by Kolenikov and Kennedey 2014, the adjustment approach of Vannieuwenhuyze 2014 and Klausch et al. 2017, or the unified Bayes-based approach of Pfeffermann 2017. In Buelens and Brakel 2015; Buelens and Brakel 2017 it is proposed to calibrate the responses in repeated sequential mixed-mode designs to a fixed-mode distribution. This calibration technique stabilizes the bias in period-to-period changes that arise from fluctuations in the distribution of the respondents over the data collection modes in subsequent editions of the survey. This method is not a solution to compensate for the loss of CAPI respondents during the lockdown. The weights of the CAPI respondents observed before and after the lockdown period are increased. This might increase the bias if the target variables are affected by the lockdown, which can be expected based on the LFS and DHS figures.

Acknowledgements

The authors would like to thank the anonymous referees and Associate Editor for carefully reading a former draft of the manuscript and providing comments to further improve our manuscript. The views expressed in this paper are those of the authors and do not necessarily reflect the policies of Statistics Netherlands.

References

ABS (2021). *ABS responds to COVID-19*. URL: https://www.abs.gov.au/abs-responds-covid-19.

Berkel, C. et al. (2024). "Experiences with mixed-mode surveys in times of COVID-19 at Statistics Netherlands". In: *Statistical Journal of the IAOS* 40.2, pp. 361–373. DOI: 10.3233/SJI-230092.

Brakel, J. van den (2008). "Design-based analysis of embedded experiments with applications in the Dutch Labour Force Survey". In: *Journal of the Royal Statistical Society: Series A* 171.3, pp. 581–613. DOI: 10.1111/j.1467-985X.2008.00532.x.

Brakel, J. van den and M. Smeets (2023). "Official statistics based on the Dutch Health Survey during the COVID-19 pandemic". In: *Survey Methodology* 49.1, pp. 1–23.

Brakel, J. van den, M. Souren, and S. Krieg (2022). "Estimating monthly labour force figures during the COVID-19 pandemic in the Netherlands". In: *Journal of the Royal Statistical Society: Series A* 185.4, pp. 1560–1583. DOI: 10.1111/rssa.12869.

Buelens, B. and J. van den Brakel (2015). "Measurement error calibration in mixed-mode sample surveys". In: *Sociological Methods & Research* 44.3, pp. 391–426. DOI: 10.1177/0049124114532444.

– (2017). "Comparing two inferential approaches to handling measurement errors in mixed-mode surveys". In: *Journal of Official Statistics* 43.2, pp. 513–531. DOI: 10.1515/jos-2017-0024.

CBS (2020). *Corona sentimentsindicator*. URL: https://www.cbs.nl/nl-nl/over-ons/innovatie/project/corona-sentimentsindicator.

Couper, M. (2011). "The future of modes of data collection". In: *Public Opinion Quarterly* 75.5, pp. 889–908. DOI: 10.1093/poq/nfr046.

De Leeuw, E.D. (2005). "To mix or not to mix data collection modes in surveys". In: *Journal of Official Statistics* 21.2, pp. 233–255.

De Leeuw, E.D. (2018). “Mixed-mode: past, present and future”. In: *Survey Research Methods* 12.2, pp. 75–89.

Dillman, D.A. and L.M. Christian (2005). “Survey mode as a source of instability in responses across surveys”. In: *Field Methods* 17.1, pp. 30–52. DOI: 10.1177/1525822X04269550.

Dillman, D.A., G. Phelps, et al. (2009). “Response rate and measurement differences in mixed-mode surveys using mail, telephone, interactive voice response (IVR) and the Internet”. In: *Social Science Research* 38.1, pp. 1–18. DOI: 10.1016/j.ssresearch.2008.03.007.

Dillman, D.A., J.D. Smyth, and L.M. Christian (2014). *Internet, phone, mail, and mixed-mode surveys: The tailored design method*. John Wiley & Sons.

Durbin, J. and S.J. Koopman (2012). *Time Series Analysis by State Space Methods*. 2nd ed. Oxford University Press.

German Presidency of Council Working Party on Statistics (2020). *COVID-19 measures taken by national statistical institutes*. URL: https://www.destatis.de/EN/eu2020/digital-conference/EU-map/article.html.

Jäckle, A., C. Roberts, and P. Lynn (2010). “Assessing the effect of data collection on measurement”. In: *International Statistical Review* 78.1, pp. 3–20.

Kastberg, S. and V. Siegler (2022). *Impact of COVID-19 on ONS social survey data collection*. URL: https://tinyurl.com/kastberg2022.

Klausch, L.T. et al. (2017). “Adjusting measurement bias in sequential mixed-mode surveys using re-interview data”. In: *Journal of Survey Statistics and Methodology* 5.4, pp. 409–432.

Kolenikov, S. and C. Kennedey (2014). “Evaluating three approaches to statistically adjust for mode effects”. In: *Journal of Survey Statistics and Methodology* 2.2, pp. 126–158.

Pfeffermann, D.A. (2017). “Bayes-based non-Bayesian inference on finite populations from non-representative samples: a unified approach”. In: *Calcutta Statistical Association Bulletin* 1, pp. 35–63.

Särndal, C.-E., B. Swensson, and J. Wretman (1992). *Model Assisted Survey Sampling*. Springer Verlag.

Schouten, B., J. van den Brakel, B. Buelens, D. Giesen, et al. (2022). *Mixed-Mode Official Surveys: Design and Analysis*. Chapman and Hall.

Schouten, B., J. van den Brakel, B. Buelens, J. Van der Laan, et al. (2013). “Disentangling mode-specific selection and measurement bias in social surveys”. In: *Social Science Research* 42.6, pp. 1555–1570. DOI: 10.1016/j.ssresearch.2013.07.005.

Suzer-Gurtekin (2013). *Investigating the bias properties of alternative statistical inference methods in sequential mixed-mode surveys*. PhD thesis University of Michigan.

Tourangeau, R. (2017). “Tradeoffs among coverage, nonresponse and measurement error”. In: *Total survey error in practice*. Ed. by P.P. Biemer et al. New York: Wiley, pp. 115–132.

Van Berkel, K. (2022). “Adaptive Survey Design for the Dutch Labour Force Survey”. In: *American Journal of Theoretical and Applied Statistics* 11.4, pp. 114–121. DOI: 10.11648/j.ajtas.20221104.12.

Van Berkel, K., S. Van der Doef, and B. Schouten (2020). “Implementing Adaptive Survey Design with an application to the Dutch Health Survey”. In: *Journal of Official Statistics* 36.3, pp. 609–629. DOI: 10.2478/JOS-2020-0031.

Vannieuwenhuyze, J.T. (2014). “On the relative advantage of mixed-mode versus single-mode surveys”. In: *Survey Research Methods* 1, pp. 31–42.

Wollburg, P. et al. (2022). “The uneven effects of the COVID-19 pandemic on National Statistical Offices, Evidence from the Global COVID-19 survey of NSOs”. In: *Statistical Journal of the IAOS* 38.3, pp. 785–803. DOI: 10.3233/SJI-220044.

Colophon

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