Discussion paper

Official Statistics based on the Dutch Health Survey during Covid-19 Pandemic

Marc Smeets and
Jan van den Brakel

November 2021

Content

Summary
The Dutch Health Survey, conducted by Statistics Netherlands, is designed to produce reliable direct estimates at an annual frequency. Data collection is based on web interviewing (CAWI) and face-to-face interviewing (CAPI). During the Covid-19 lockdown CAPI partially stopped, which resulted in a sudden change in measurement and selection effects in the survey outcomes. Furthermore, the production of annual data about the effect of Covid-19 on health-related themes with a delay of about one year compromises the relevance of this survey. The sample size of the survey does not allow the production of figures for shorter reference periods. Both issues are solved by developing a bivariate structural time series model (STM) to estimate quarterly figures for a selection of eight key variables. The input series are quarterly direct estimates based on the complete response of CAPI and CAWI and a series of direct estimates based on the CAWI response only. During the lockdown, the direct estimates for the complete response are missing and the time series model provides an optimal nowcast for this figure. The model is also used as a form of small area estimation that borrows sample information observed in previous reference periods. In this way timely and relevant statistics that describe the effects of the corona crisis on the development of health, care use and lifestyle in the Netherlands are published. In this paper the method based on the bivariate STM is compared with two alternative methods. The first one is a univariate STM where no correction for the lack of CAPI is applied to the estimates by means of modelling the difference of the web series and the complete series. The second one is a univariate STM that also contains an intervention variable that models the effect of the loss of CAPI response during the lockdown.

Keywords
Small area estimation, structural time series model, corona crisis

Reviewers
Jan-Willem Bruggink and Rob Willems

1. Introduction

The Dutch Health Survey (DHS) is a continuing survey conducted by Statistics Netherlands that measures health, care use and lifestyle in the Netherlands. Data collection is based on a sequential mixed-mode design where a combination of web participation (CAWI – Computer-assisted web interviewing) and face-to-face interviewing (CAPI – Computer-assisted personal interviewing) is applied. Due to Dutch lockdown measures during the Covid-19 pandemic face-to-face interviewing was not allowed in parts of 2020 and 20201. In the rest of these years, except for the period January to mid-March 2020, there were restrictions on the normal way of data collection. This results in an abrupt change in the composition of selection effects and measurement bias and therefore results in a systematic effect on the outcomes of the DHS. A second issue is that the DHS is designed to produce reliable estimates on an annual basis, using standard direct estimators like the general regression (GREG) estimator (Särndal et al., 1992). The Covid-19 pandemic that started in the beginning of March 2020 made clear that the release of annual data about the effect of Covid-19 on health-related themes with a delay of about one year strongly compromises the relevance of this survey. The DHS normally publishes on an annual basis over year t in the month of March of year t+1. Another disadvantage of annual figures is that the period of the corona crisis is not well-delineated in the reference period of the DHS. In the second quarter of 2020, there was indeed a strong external demand for more timely figures of the DHS, since quarterly figures are more timely and better delineate the corona period. The effect of this crisis on health is more readily apparent in the published figures. The sample size of the DHS, however, does not allow the production of sufficiently precise direct estimates for shorter reference periods.

To solve these issues, a structural time series model is developed for eight important variables of the DHS, defined on a quarterly frequency. This model is used to correct for the changes of measurement and selection errors due to the loss of CAPI response and is used as a form of small area estimation (Rao and Molina, 2015) since the model uses sample information observed in previous reference periods to produce sufficiently reliable model-based estimates for quarterly DHS figures. In this way the period of the corona crisis can be better defined and figures are more timely available. By using sample information from the past the effective sample size can be increased in order to improve the accuracy of the direct estimates. In small area estimation this is commonly called borrowing strength over time.

Buelens and Van den Brakel (2015) proposed a weighting method for sequential mixed mode designs to stabilize the bias in period-to-period changes that arise from fluctuations in the distributions of respondents over the data collection modes in subsequent editions of a repeated survey. This method assumes a fixed distribution of the population over the different data collection modes, which is added as an additional component to the weighting model of the GREG estimators. This method cannot be considered as an alternative to compensate for the loss of CAPI respondents during the lockdown. The method indeed increases the weights of the CAPI respondents but will in this case increase selection bias as well because the CAPI respondents are all observed outside the lockdown period.

The net effect of the lack of CAPI is computed based on the response of previous years. This is done by removing CAPI from the response and by reweighting the remaining response. This leads to two direct estimates for one target variable: one based on the complete response (CAWI and CAPI) and one based on only web response (CAWI). In this way quarterly time series can be constructed for DHS that start in the first quarter of 2014: the complete series based on full response and the web series based on web response. Both series are the input for a bivariate structural time series model (STM). The web series is available in all quarters, also during the lockdown. In the quarters during the lockdown no estimates are available for the complete series, because CAPI is completely or partly missing. With the bivariate STM nowcasts are obtained for the missing figures based on the web series.

In this paper the time series method based on the bivariate model is compared with two alternative and more straightforward methods. The first one is a univariate STM where no correction for the lack of CAPI is applied to the estimates by means of modelling the difference of the web series and the complete series. This method applies a univariate STM to the series of direct estimates based on all available response in every quarter. In quarters where CAPI is available the direct estimates are based on both CAWI and CAPI, so they are equal to the estimates of the complete series. In quarters where no CAPI is available the direct estimates are based on only CAWI and are thus equal to the estimates of the web series. The second one is a univariate STM that also contains an intervention variable that models the effect of the loss of CAPI response during the lockdown.

For the eight selected DHS indicators official quarterly figures have been published based on the bivariate STM. The published series starts in 2017 and will probably run until the end of 2021. Annual DHS figures for 2020 are benchmarked with the quarterly figures by adding the model-based quarterly figures of the eight variables to the weighting model of the GREG estimator of the DHS. Annual figures for 2021 will be benchmarked in the same way. Statistics Netherlands have decided not to make any revisions on published figures. This means that annual figures for 2017, 2018 and 2019 have not been made consistent with the model-based quarterly figures in these years.

In Section 2 a description of the Dutch Health Survey is given. In section 3 the time series method is developed and both the univariate and bivariate structural time series models are introduced. Section 4 explores the results and section 5 discusses the officially published DHS figures by Statistics Netherlands in 2020 and 2021. The paper ends with a discussion in section 6.

2. Dutch Health Survey

The Dutch Health Survey (DHS) is a continuing survey that measures health, care use and lifestyle in the Netherlands on a yearly basis. The target population is the Dutch population living in private households. Each month a single-stage stratified sample of approximately 1250 persons is drawn from the Dutch Personal Records Database. The strata are defined by the municipalities.

Sampled persons are asked to participate via web interviewing (CAWI). Non-respondents are re-approached to participate in a face-to-face interview (CAPI). For the re-approach a target group strategy has been used since 2018, implying that not all CAWI non-respondents are re-approached. The non-respondents are first divided into subgroups, so-called target groups, and second in every subgroup only a sample is re-approached. To achieve higher CAWI response rates smaller samples are selected from the target groups for a re-approach. Target groups are based on age, income and migration background.

Until 2020 there was a yearly response of approximately 10,000 persons, of whom 6500 responded by CAWI and 3500 by CAPI. The response is more or less evenly divided over the months. So the overall response rate was about 65%. Due to the Covid-19 pandemic that started in 2020 there was a lockdown in the Netherlands that started mid-March 2020. The first relaxations were implemented in May 2020. Due to this lockdown no face-to-face interviews were al-lowed from mid-March 2020 to the end of July 2021. A second lockdown started in mid-December 2020, which was gradually relaxed from March 2021. This lockdown resulted in a stop of face-to-face interviewing from mid-December 2020 until the end of March 2021. From April 2021 face-to-face interviews were possible again. In order to increase response, persons selected for CAPI during the pandemic were given the opportunity to respond via the internet. This was done by sending an invitation letter and by handing over an invitation letter in the last two weeks of December and from the second quarter of 2021. In 2020 only few people used this option and they were considered as CAWI respondents. In 2021 there more people that were selected for CAPI and responded via the internet. This third response mode will be called CAPI/CAWI response. The resulting response sizes per month and response mode are shown in tables 2.1 and 2.2.

Table 2.1 shows that in 2020 CAPI response is lower in March and December and is completely missing from April to July. The relatively large CAWI response size in May is the result of compensation measures carried out by Statistics Netherlands for the resulting response shortages. In 2021 CAPI response is completely missing in the first quarter and is lower in April and May. From June CAPI response seems to recover.

Annual figures are obtained by weighting the response by means of the general regression estimator (Särndal et al. 1992). In this way it is corrected, at least partially, for selective non-response. Covariates used in the weighting model are gender, age, migration background, marital status, urban environment, province, household size, income, wealth and survey season, and from 2018 also target group (Boonstra, 2019).

	CAPI	CAWI	total
January	265	584	849
February	261	586	847
March	104	917	1021
April	0	455	455
May	0	1118	1118
June	0	708	708
July	0	483	483
August	193	527	720
September	286	259	545
October	149	763	912
November	181	587	768
December	53	286	339
total	1492	7273	8765

Table 2.1: response DHS 2020 per mode and month

	CAPI	CAPI/CAWI	CAWI	total
January	0	48	738	786
February	0	36	546	582
March	0	22	655	677
April	38	77	460	575
May	51	62	738	851
June	109	62	283	454
total	198	307	3420	3925

Table 2.2: response DHS 2021 (first six months) per mode and month

In 2020 the data collection of the DHS was disrupted because of the corona crisis. In consultation with the Dutch authorities the National Institute for Public Health and Environmental Protection, the Ministry of Health, and the Welfare and Sports and Socio-cultural Planning Department, it was decided in June 2020 to publish quarterly figures for a selection of eight variables of the DHS. For these variables there was the desire to monitor whether quarterly changes occur in the outcomes during the corona crisis. The variables cover the three main topics of the survey (health, care use and lifestyle) and are given by

- Perceived health
- Mentally unhealthy
- GP consult
- Dental visit
- Specialist consult
- Daily smoking
- Overweight
- Excessive alcohol consumption

Perceived health is measured for people of all ages. There are five possible answers: very good, good, fair, poor and very poor. Perceived health is the percentage of people that have given one of the positive answers very good or good. *Mentally unhealthy* is measured by the so-called Mental Health Inventory-5, consisting of 5 questions on mental well-being in the past four weeks. Persons with a MHI-score lower than 60 (on a 0-100 scale) were considered 'unhealthy'. This indicator is related to people aged 12 years or older. The three care use variables (*GP consult*, *Dental visit*, *Specialist consult*) measure the percentage of people that have made any use of health care over the past four weeks and is related to all age groups in the population. *Daily smoking* concerns the percentage of people with a daily smoking habit and is measured for people aged 18 years or older. Overweight is based on the Body Mass Index (BMI). BMI is the ratio of somebody's body mass in kilograms and the square of the height in metres. *Overweight* measures the percentage of people with a BMI of 25 or more and is related to people aged 18 years or older. *Excessive alcohol consumption* is measured for the population aged 18 years or older and measures the percentage of people that report a consumption of 21 or more units per week for men or a consumption of 14 or more units per week for women.

3. Structural time series method

For the computation of the quarterly figures three structural time series methods are considered. The first two methods are based on a univariate STM and the second method uses a bivariate STM. The models are introduced in the subsections 3.1 and 3.2, respectively. The computation of the direct estimates that serve as the input series of the models is discussed in subsection 3.3. Subsection 3.4 is about the computation of the model-based estimators based on a given STM.

3.1 Univariate models

Two univariate STM's are applied. The first univariate model ignores the loss of CAPI during the lockdown periods and is given by

$$\hat{y}_t^{\mathrm{A}} = \vartheta_t + e_t^{\mathrm{A}}, \text{ with} \qquad (3.1)$$
$$\vartheta_t = L_t + S_t \text{ and quarter } t.$$

Here is

- $\hat{y}_t^{\mathrm{A}}$: direct estimate in quarter t based on the available response.
- ϑ_t: population parameter in quarter t.
- L_t: trend, modelled by a smooth trend model with flexible slope.
- S_t: seasonal component with quarter as period, modelled by a trigonometric seasonal model.

– e_t^{A}: measurement error of $\hat{y}_t^{\mathrm{A}}$ modelled as $e_t^{\mathrm{A}} = \sqrt{\widehat{\mathrm{V}}(\hat{y}_t^{\mathrm{A}})}\ \tilde{e}_t^{\mathrm{A}}$, with $\widehat{\mathrm{V}}(\hat{y}_t^{\mathrm{A}})$ the variance estimate of $\hat{y}_t^{\mathrm{A}}$ and white noise $\tilde{e}_t^{\mathrm{A}}$.

The underlying models of the different components of model (3.1) are given in Appendix A.

Estimates based on model (3.1) borrow strength from the past in order to improve the accuracy of the direct estimates. Here it is assumed that the absolute values of the period-to-period changes observed in the past are not affected by the Covid-19 pandemic. For some variables the Covid-19 pandemic causes strong deviations in the quarterly series of direct estimates in the second quarter of 2020 and the first quarter of 2021, especially for the care use variables. For these variables the assumption that the volatility of the period-to-period changes is not affected by the Covid-19 pandemic is no longer valid. To account for strong deviations in the direct estimates the trend in model (3.1) is therefore modelled by a smooth trend model with a flexible trend (Appendix A, equation (A.1)). The trend consists of level L_t and slope R_t with disturbance term η_t^{R}. The trend is made more flexible by using a time-dependent variance $f_t\sigma_{\mathrm{R}}^2$ of the disturbance term of the slope for a time-dependent factor $f_t \geq 1$.

Model (3.1) borrows strength from the past through both the trend L_t and the seasonal pattern S_t. The Covid-19 pandemic may influence both the trend and the seasonal pattern. Since it is not possible to estimate a structural change in the seasonal pattern due to the Covid-19 pandemic, with less than one year of observations during the Covid-19 pandemic it is assumed that there is only an effect on the development of the trend. The seasonal component S_t is therefore modelled by a trigonometric seasonal model with a time-independent variance (Appendix A, equation (A.2)). In this way the seasonal pattern is modelled dynamically and therefore has the flexibility to accommodate effects of the Covid-19 pandemic on the seasonal pattern.

The measurement error e_t^{A} is a combination of sampling noise and noise in the population parameter. Since DHS is based on a cross-sectional survey it is not possible to distinguish these two terms by means of a structural time series model. The model does account for changes in the variances of the direct estimates caused by changes in response size and the sample design. The measurement error model is given by equation (A.4) in Appendix A.

To account in the univariate model for sudden changes in measurement and selection errors due to the loss of CAPI response during the lockdown, model (3.1) is extended with an intervention variable:

$$\hat{y}_t^{\mathrm{A}} = \vartheta_t + \beta\frac{x_t}{3} + e_t^{\mathrm{A}}, \text{ with} \qquad (3.2)$$
$$\vartheta_t = L_t + S_t \text{ and quarter } t.$$

Here x_t is the number of months in quarter t without CAPI response and β a regression coefficient that can be interpreted as the net effect of the change in measurement and selection bias due to the loss of CAPI. In a quarter with full CAPI response, $x_t/3 = 0$ and β is switched off. In a quarter without any CAPI respondents, $x_t/3 = 1$ and β absorbs the effect of the loss of CAPI and avoids that the model estimates for the population parameter ϑ_t are affected, at least partially. If a quarter only contains one or two months without CAPI, then $x_t/3 = 1/3$ or $x_t/3 = 2/3$ respectively and the correction of β contributes proportionally to the number of months without CAPI in that quarter. Compared to model (3.1) it is expected that this model better accommodates for the loss of CAPI during the lockdown. Model (3.2), however, assumes no structural change in the evolution of the population parameter ϑ_t. If the lockdown results in e.g. strong turning points in the population parameter, it can be expected that this is partially and incorrectly absorbed in the regression coefficient of the intervention variable. To accommodate for this risk, the bivariate model, proposed in the next section is developed.

3.2 Bivariate model

The third approach is the bivariate STM given by

$$\begin{pmatrix} \hat{y}_t^{\mathrm{C}} \\ \hat{y}_t^{\mathrm{W}} \end{pmatrix} = \begin{pmatrix} \vartheta_t \\ \vartheta_t \end{pmatrix} + \begin{pmatrix} 0 \\ \lambda_t \end{pmatrix} + \begin{pmatrix} e_t^{\mathrm{C}} \\ e_t^{\mathrm{W}} \end{pmatrix}, \text{ with} \qquad (3.3)$$
$$\vartheta_t = L_t + S_t \text{ and quarter } t.$$

Here is

- $\hat{y}_t^{\mathrm{C}}$: direct estimate in quarter t based on complete response (complete series).
- $\hat{y}_t^{\mathrm{W}}$: direct estimate in quarter t based on web response (web series).
- ϑ_t: population parameter in quarter t.
- L_t: trend, modelled by a smooth trend model with flexible slope.
- S_t: seasonal component with quarter as period, modelled by a trigonometric seasonal model.
- λ_t: systematic difference between the web series and the regular series, modelled as a random walk.
- e_t^j: measurement error of $\hat{y}_t^j$ for $j \in \{\mathrm{C}, \mathrm{W}\}$ modelled as $e_t^j = \sqrt{\hat{\mathbb{V}}(\hat{y}_t^j)}\, \tilde{e}_t^j$, with $\hat{\mathbb{V}}(\hat{y}_t^j)$ the variance estimate of $\hat{y}_t^j$ and white noise $\tilde{e}_t^j$. Furthermore it is proved in Appendix A that
$$\mathrm{Cov}(e_t^{\mathrm{C}}, e_t^{\mathrm{W}}) = \frac{\sqrt{n_t^{\mathrm{W}}}}{\sqrt{n_t^{\mathrm{C}}}} \sqrt{\hat{\mathbb{V}}(\hat{y}_t^{\mathrm{C}})} \sqrt{\hat{\mathbb{V}}(\hat{y}_t^{\mathrm{W}})}, \qquad (3.4)$$
with n_t^{W} the size of the web response in quarter t and n_t^{C} the size of the complete response in quarter t.

The underlying models of the different components of model (3.3) are given in Appendix A.

Concerning the measurement errors e_t^j, the model accounts for changes in the variances of the direct estimates caused by changes in response size and the sample design, similarly to the univariate model. The bivariate model also accounts for the correlation between the measurement errors of the input series that arise from the overlap in the response on which the direct estimates $\hat{y}_t^{\mathrm{C}}$ and $\hat{y}_t^{\mathrm{W}}$ are based.

3.3 Direct estimates

For the DHS direct quarterly estimates can be computed starting in the first quarter of 2014. From the first quarter of 2014 up to the last quarter of 2019 these direct estimates are based on the weighted annual DHS response files obtained by applying the GREG estimator. Quarterly estimates $\hat{y}_t^{\mathrm{C}}$ for the complete series are obtained by computing the domain estimator based on the GREG estimator with quarter t as domain. Quarterly estimates $\hat{y}_t^{\mathrm{W}}$are obtained by recalculating the GREG estimator using the CAWI response only and subsequently computing the domain estimator based on the GREG estimator with quarter t as the domain. Since there is no loss of CAPI response before 2020, the direct estimates $\hat{y}_t^{\mathrm{A}}$ are equal to $\hat{y}_t^{\mathrm{C}}$. Standard errors are computed in R (R Core Team, 2015) with the package 'Survey' (Lumley, 2014). For the estimation of the standard errors the sample design of the DHS is taken into account, where the stratification is based on the cross-classification of months and provinces. Here provinces are used, because the subdivision into municipalities leads to strata with too little response.

Since it was decided to construct the quarterly figures in July 2020, direct estimates for the first two quarters of 2020 are based on the weighted response based on the GREG estimator available from January to June 2020. Estimates for quarter 3 are based on the weighted response available from January to September and quarter 4 is based on the weighted annual response file over 2020. The GREG estimators by which the six and nine months response files are weighted use the same weighting model and population totals of the covariates as the GREG estimator that is used for weighting the annual response. Direct estimates for the first quarter in 2021 are computed in a similar way based on the response from January to March 2021. Estimates for quarter 2 in 2021 are based on the response from January to June 2021. For the web series estimates $\hat{y}_t^{\mathrm{W}}$ are then obtained for all quarters in 2020 and for the first two quarters in 2021.

For the complete series $\hat{y}_t^{\mathrm{C}}$ in 2020 the second quarter is missing and the other quarters are based on response where CAPI is partially missing (Table 2.1). In quarter 1 CAPI is only missing in the last two weeks of March and for this quarter it is assumed that sufficient CAPI response is available to obtain plausible estimates. So in quarter 1 the estimates $\hat{y}_t^{\mathrm{C}} = \hat{y}_t^{\mathrm{A}}$ are based on the available CAWI and CAPI response and in quarter two $\hat{y}_t^{\mathrm{C}}$ is missing and $\hat{y}_t^{\mathrm{A}} = \hat{y}_t^{\mathrm{W}}$. For quarter 3 CAPI response is only available in August and September. Here a correction is applied to $\hat{y}_t^{\mathrm{C}}$ based on the bivariate model (3.3). The direct estimate $\hat{y}_t^{\mathrm{C}}$ for quarter 3 is obtained by computing the domain estimator of the GREG applied to the available response in August and September minus 1/3 of the difference $\hat{\lambda}_t$ estimated by model (3.3) in quarter 2. No correction is applied to the corresponding standard errors. The direct estimate $\hat{y}_t^{\mathrm{A}}$ in quarter 3 is equal to the uncorrected weighted mean of the available response in August and September. In quarter 4 CAPI is also missing for only two weeks and it is also assumed that there is enough CAPI response available to obtain plausible estimates, so $\hat{y}_t^{\mathrm{C}} = \hat{y}_t^{\mathrm{A}}$.

In 2021 there is besides CAPI and CAWI also CAPI/CAWI response (section 2). To find out how to use the CAPI/CAWI response in the best possible way, two scenarios were elaborated. In the first scenario quarterly figures are computed where CAPI/CAWI response is considered as CAPI and in the second scenario CAPI/CAWI response is considered as CAWI. Since there were no major differences in the results of both scenarios the CAPI/CAWI response is considered as CAWI. Results of this comparison are not shown in this paper. In the first quarter of 2021 $\hat{y}_t^{\mathrm{C}}$ is missing and $\hat{y}_t^{\mathrm{A}} = \hat{y}_t^{\mathrm{W}}$ and in the second quarter of 2021 CAPI is available and so $\hat{y}_t^{\mathrm{C}} = \hat{y}_t^{\mathrm{A}}$.

In this way input series for models (3.1), (3.2) and (3.3) are obtained. The series run from the first quarter in 2014 up to the second quarter in 2021. The series $\hat{y}_t^{\mathrm{A}}$ and $\hat{y}_t^{\mathrm{W}}$ are available for all quarters and for the series $\hat{y}_t^{\mathrm{C}}$ estimates are missing in quarter 2 of 2020 and quarter 1 of 2021.

3.4 Model-based estimates

Given the series of direct estimates $\hat{y}_t^{\mathrm{A}}$, $\hat{y}_t^{\mathrm{C}}$ and $\hat{y}_t^{\mathrm{W}}$ model-based estimates based on one of the models (3.1), (3.2) or (3.3) can be produced by means of the Kalman filter (Durbin and Koopman, 2012). To this end the filtered estimates $\hat{L}_t + \hat{S}_t$ of the population parameter ϑ_t are computed. In order to apply the Kalman filter the structural time series model is written in terms of a state space model, given by

$$\boldsymbol{y}_t = \mathbf{Z}_t \boldsymbol{\alpha}_t \tag{3.5}$$
$$\boldsymbol{\alpha}_t = \mathbf{T}\boldsymbol{\alpha}_{t-1} + \boldsymbol{\eta}_t \text{ for quarter } t\text{, where} \tag{3.6}$$
$$\boldsymbol{\eta}_t \sim \mathcal{N}(0, \mathbf{H}).$$

The state space model contains a measurement equation (3.5), which expresses the vector $\boldsymbol{y}_t$ of direct estimates in terms of a vector of non-observed state variables $\boldsymbol{\alpha}_t$, like the trend and the seasonal components. For the univariate model we have $\boldsymbol{y}_t = y_t = \hat{y}_t^{\mathrm{A}}$ and for the bivariate model $\boldsymbol{y}_t = (\hat{y}_t^{\mathrm{C}}, \hat{y}_t^{\mathrm{W}})'$. The state space model further contains a transition equation (3.6), which describes the development of the state variables over time. The covariance matrix $\mathbf{H}$ of the vector of disturbance terms $\boldsymbol{\eta}_t$ contains the hyperparameters of the model, which are estimated by Maximum Likelihood (ML). The computations are done with Ssfpack 3.0 (Koopman et al, 2008) in combination with Ox (Doornik, 2009). Note that the measurement errors of the measurement equation are included in the vector of disturbance terms of the transition equation, which is necessary to make the variance of the measurement errors proportional to the variance of the direct estimates.

In quarters where CAPI is missing the model-based estimates rely on strong assumptions. For the univariate model (3.1) it is assumed that there are no mode effects between CAPI and CAWI. For the univariate model (3.2) it is assumed that the trend and the seasonal component correctly describe the evolution of the population parameter and that sudden strong changes such as turning points are not partially absorbed in the level intervention component. These assumptions are evaluated in section 4. For the bivariate model it is assumed that the composition of the web response does not change during the corona crisis. This assumption has been evaluated through a response analysis and it indeed turned out to be the case. A second assumption is that the difference between CAWI and CAPI response does not change due to the corona crisis. The latter assumption cannot be evaluated.

For all three models the model assumptions are evaluated by analysing the standardized innovations of the direct estimates. The innovations are the errors of the predictions $\boldsymbol{y}_{t|t-1}$ of $\boldsymbol{y}_t$ that are produced by the Kalman filter based on the data available until the previous quarter. Under the model the standardized innovations, defined by the innovations divided by their standard errors, follow a standard normal distribution. Through the standardized innovations the model is checked for normality, occurrance of heteroscedasticity and for autocorrelation. To this end the following statistical tests are carried out: Bowman-Shenton normality test, F-test for heteroscedasticity, QQ-plot, plot of standardized innovations and sample correlogram, Durbin Watson test. For more details on these tests it is referred to Durbin and Koopman (2012, Chapter 2).

4. Results

The models (3.1), (3.2), and (3.3) are fitted to the direct quarterly series as described in Subsection 3.3. Due to the Covid-19 pandemic some DHS variables show a strong increase in the quarter-to-quarter changes, especially at the beginning of the two lockdown periods. In these periods, the smooth trend model is not flexible enough to follow the increased period-to-period movements of the input series. This can be expected since the flexibility of the trend, which is determined by the variance of the slope disturbance terms of the trend model, is based on the quarter-to-quarter movements observed in the period before the Covid-19 crisis. A sudden increase in the dynamics of the population parameter results in temporary miss-specification of the STM, which is visible in large values for the standardized innovations in these periods. To accommodate in the STM for the suddenly increased dynamics of the population parameters, the flexibility of the smooth trend in the models (3.1), (3.2), and (3.3) is temporarily increased by using a time-dependent variance for the slope disturbance terms (Appendix A, equation (A.1)). This is realized by multiplying the maximum likelihood estimate for the variance of the slope disturbance terms (σ_{R}^2) by a time-dependent factor $f_t \geq 1$. As a result, the variance of the slope disturbance terms is equal to $f_t\sigma_{\mathrm{R}}^2$. Values for f_t are determined outside the model, as explained below, and used in the Kalman filter as values that are known in advance, similar to the maximum likelihood estimates of the other hyperparameters. This approach is initially proposed by Van den Brakel et al. (2021) and compared with alternative approaches to account for sudden shocks in the input series of an STM due to the Covid-19 crisis.

The values of the factor f_t are determined by analysing the standardized innovations. The values for f_t are chosen in such a way that the standardized innovations in the period during the start of the corona crisis have values within or just outside the admissible range of 1.96 in absolute terms. In this way, the value of the factor $f_t > 1$ is kept as small as possible, so that the model can still borrow strength from the past. Note that adjusting the variance σ_{R}^2 in quarter t influences the slope disturbance term from quarter t+1 and the trend only from quarter t+2. So there is a lag of two quarters in the effect of the outcomes after adjustment of σ_{R}^2. To avoid a large sudden change in the variance of the slope disturbance terms, the values of f_t are just slightly increased. In the subsequent quarters the values of f_t are reduced to 1 as soon as possible.

From the analysis of the standardized innovations it follows that for most variables it is necessary to make the slope more flexible during the corona crisis. Table 4.1 shows the values of the factors $f_t > 1$ for models (3.1), (3.2) and (3.3). In quarters where $f_t = 1$ no values are shown. Variables for which it was not necessary to make the slope more flexible are not shown in the tables either. For perceived health and the three care use variables a flexible slope is applied in the quarters before the first lockdown in the second quarter of 2020. For mentally unhealthy the slope has been made more flexible only in the quarters before the second lockdown in the first quarter of 2021. For the univariate model without intervention (3.1), $f_t > 1$ also before the second lockdown for the variable daily smoking.

For the variable perceived health Figure 4.1 displays the standardized innovations estimated by the univariate and bivariate model with and without a flexible slope. For models with a flexible slope the values of parameter f_t are given in tables 4.1 and 4.2. For the models without a flexible slope the parameter f_t has value 1 in all quarters. For all three series the innovations estimated by the model without flexible slope (black dashed line) exceed the admissible interval of (-1.96,1.96) implying that the model is miss-specified. By making the slope more flexible the standardized innovations (red solid line) get plausible values. For the other variables the standardized innovations are not shown in this paper. After setting the flexibility parameters, the underlying model assumptions are evaluated by testing whether the standardized innovations are standard normally and independently distributed. For all three models the performed tests (section 3.4) show some small violations of these assumptions for some of the variables.

Figures 4.1–4.4 display the standardized innovations for perceived health estimated by the three models (3.1), (3.2) and (3.3). The values of the flexibility parameter can be found in Table 4.1. For all series the innovations estimated by the model with no flexible slope (black dashed line) exceed the interval of (-1.96,1.96) implying that the model is miss-specified. By making the slope more flexible the standardized innovations (red solid line) get admissible values. The standardized innovations for the other variables are not shown here. After setting the flexibility parameters the underlying model assumptions are evaluated by testing whether the standardized innovations are standard normally and independently distributed. For all three models the performed tests (section 3.4) show some small violations of these assumptions for some of the variables.

	2018 Q4	2019 Q1	2019 Q2	2019 Q3	2019 Q4	2020 Q1	2020 Q2	2020 Q3	2020 Q4
Univariate model without intervention									
Perceived health			10	100	100	100	10		
Mentally unhealthy							10	50	
GP consult				10	100	100	10		
Dental visit				10	100	100	10		
Specialist consult				10	100	100	10		
Daily smoking								10	50
Univariate model with intervention									
Perceived health			100	200	100	100	10		
Mentally unhealthy						10	100	100	10
GP consult			10	100	200	100	10		
Dental visit	10	100	1000	5000	8000	100	10		
Specialist consult		10	100	500	500	100	10		
Bivariate model									
Perceived health			10	100	100	100	10		
Mentally unhealthy							10	50	
GP consult				10	100	100	10		
Dental visit				10	100	100	10		
Specialist consult				10	100	100	10		

Table 4.1: values of flexibility parameter f_t in quarters where $f_t > 1$ for both univariate models and the bivariate model. In quarters and for variables where no value is displayed, f_t is equal to 1. In quarters 2021 Q1 and 2021 Q2 and in the quarters before 2018 Q3, $f_t = 1$ for all variables.

The tables B.1 – B.4 in Appendix B give the in real-time or concurrent ML estimates of the hyperparameters of the three structural time series models. Concurrent means that the estimates are based on the series observed until the particular quarter in the table. In order to show the values of the hyperparameters before the corona crisis, the estimates are also displayed for quarter 2 in 2019. Even though the variance σ_{R}^2 is multiplied by a factor f_t in the model, it can be seen that in many cases the (square root of the) variance estimate $\hat{\sigma}_{\mathrm{R}}$ increases. The largest increases occur for the care use variables before the first lockdown, especially for dental visit. For mentally unhealthy the variance estimates $\hat{\sigma}_{\mathrm{R}}$ of the slope disturbance terms increase in the first quarter of 2021 for the univariate STM without intervention (3.2) and the bivariate STM (3.3), but not for the univariate STM with intervention (3.1). For daily smoking the variance estimate $\hat{\sigma}_{\mathrm{R}}$ increases in the second quarter of 2021 for the univariate model.

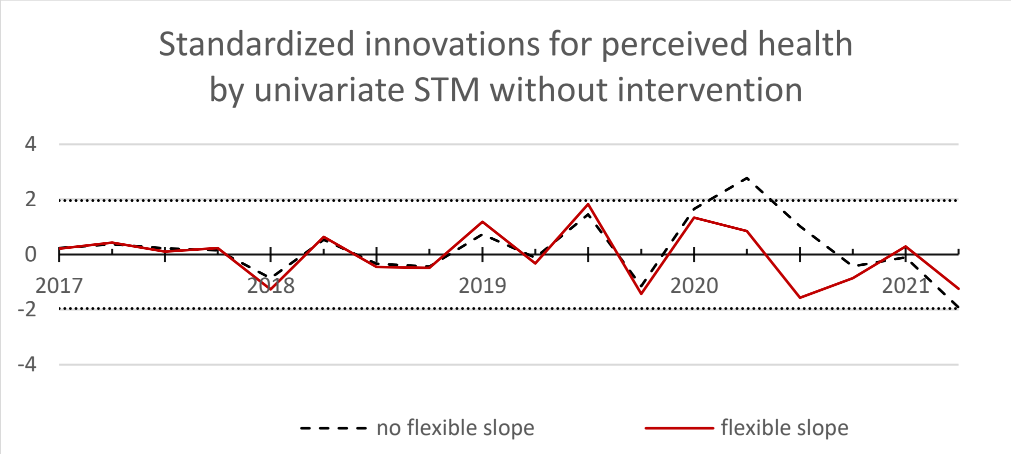


Figure 4.1: standardized innovations for perceived health estimated by univariate STM without intervention, given by (3.1)

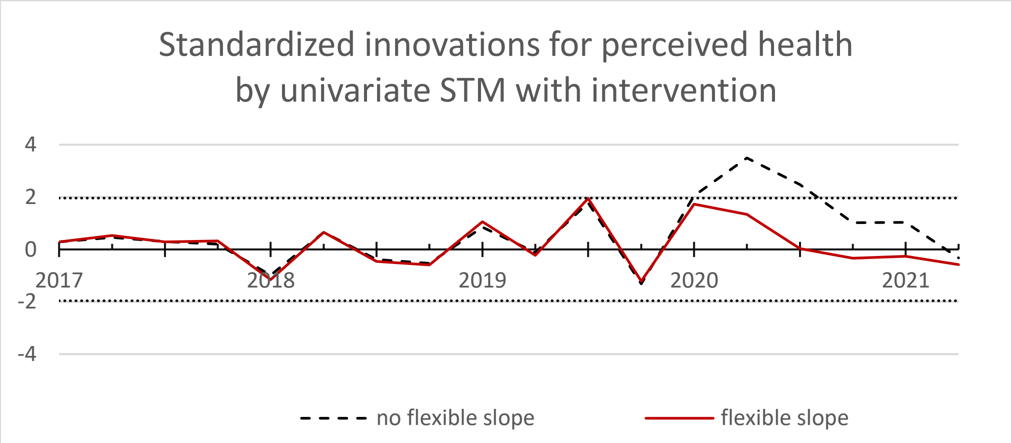


Figure 4.2: standardized innovations for perceived health estimated by univariate STM with intervention, given by (3.2)

The figures B.1 – B.16 in Appendix B show the results of the estimates for the eight variables and all three models. The displayed series start in the first quarter of 2017. Since a diffuse initialisation of the Kalman filter is applied, the model predictions for the first three years obtained with the STM are ignored. For all variables there are two figures that compare the direct estimates $\hat{y}_t^{\mathrm{C}}$ (dir compl) and $\hat{y}_t^{\mathrm{W}}$ (dir web) with the model-based estimates $\hat{L}_t + \hat{S}_t$ based on the bivariate STM (STM biv) and one of the univariate models with (STM univ intv) and without intervention (STM univ). Two other figures show the estimated standard errors of the quarterly estimates. Also plots are shown with the estimation of the intervention coefficient β (intervention univariate model) of model (3.2) and the systematic difference λ_t (bias bivariate model) of model (3.3) together with the corresponding 95% confidence intervals.

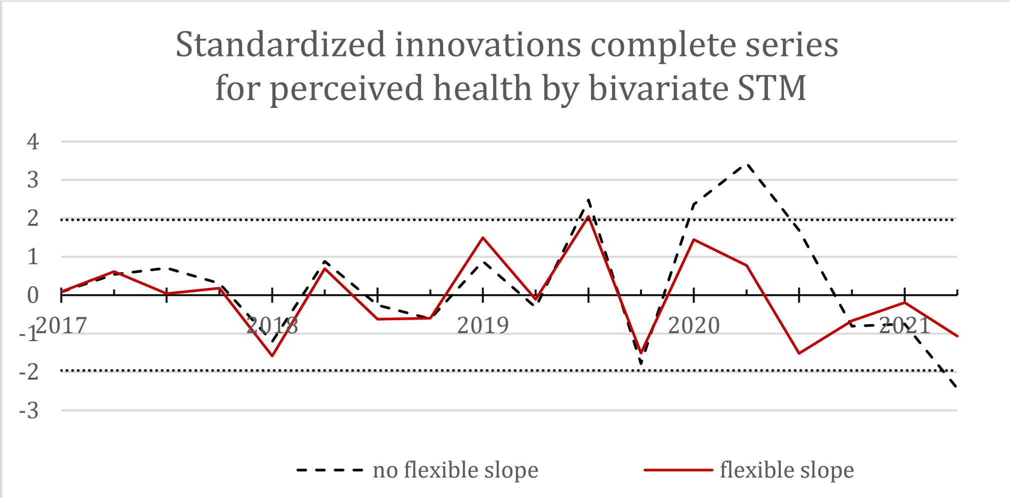


Figure 4.3: standardized innovations complete series for perceived health estimated by bivariate STM, given by (3.3)

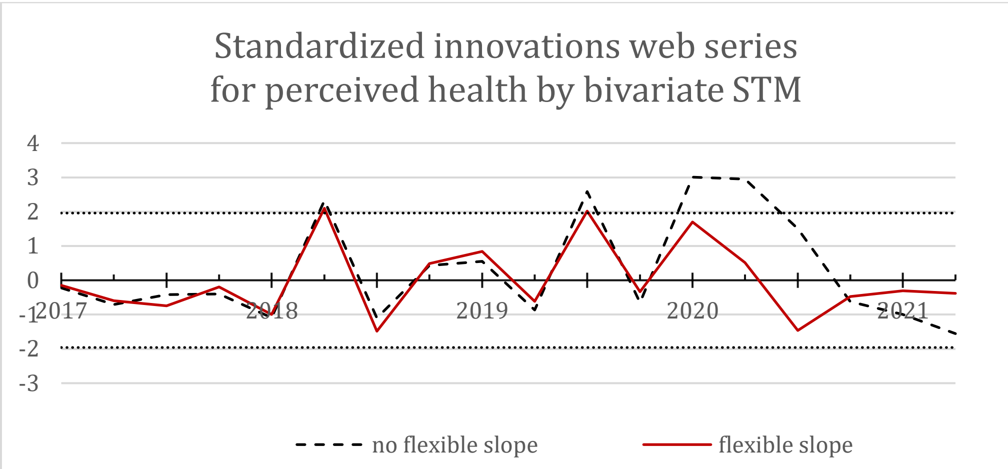


Figure 4.4: standardized innovations web series for perceived health estimated by bivariate STM, given by (3.3)

By comparing the direct estimates and by analysing the estimates of the systematic difference it follows that for most variables there is a clear mode effect between the CAPI and CAWI response. For perceived health, mentally unhealthy and overweight the differences between CAPI and CAWI are relatively small. For the care use variables GP consult, specialist consult and dental visit CAWI respondents score higher than CAPI respondents and the systematic difference λ_t varies between 1.5% and 2%. For daily smoking and excessive alcohol consumption it is just the other way around. Now CAPI scores higher than CAWI and for daily smoking the difference is the largest with a systematic difference λ_t around -4%.

Until 2020 there was no loss of CAPI and the STM estimates based on the univariate and bivariate models are very similar. During the Covid-19 pandemic that started in 2020 there are more clear differences between the STM estimates. Especially in quarters where CAPI is missing and for variables with a clear mode effect the univariate STM without intervention produces estimates at the level of the web series while the STM estimates by the bivariate model are at the level of the complete series. That is for example the case in the first quarters of 2020 for perceived health and dental visit and in the first quarter of 2021 for daily smoking, GP consult and mentally unhealthy. For the other variables similar effects are found in 2020 and 2021, but to a lesser extent. The univariate STM without intervention produces, as expected, biased estimates during the Covid-19 pandemic in quarters where CAPI is partially or completely missing.

The univariate STM with intervention also leads to biased estimates in quarters where CAPI is partially or completely missing during one of the lockdowns. This is because the model incorrectly interprets a part of the sudden changes in the real quarterly developments as differences in measurement bias and selection effects. This can result in a large estimates of the intervention coefficient β. The effect can be seen for all variables, and is small for mentally unhealthy (Figure B.3) and larger for the care use variables. For dental visit (Figure B.7) the resulting bias is the largest in quarter 2 of 2020, when dentists in the Netherlands were only open for emergency treatments.

The bivariate STM avoids that sudden changes in the developments of the population parameter are interpreted as differences in measurement and selection bias, because nowcasts are obtained for the missing estimates based on the complete response by means of the systematic difference λ_t in the model observed in the period before the lockdown. Estimates based on the bivariate STM are at the level of the complete series and are therefore used as the official quarterly DHS figures, since they provide the most plausible correction for the loss of the CAPI respondents.

For most variables the standard errors of the STM estimators are smaller than those of the direct estimators and the standard errors of the estimates based on the univariate models are generally smaller than those based on the bivariate model. In quarters where the flexibility parameter $f_t > 1$, the models assign more weight to the direct estimates and less strength is borrowed from the past. This results in larger standard errors that sometimes exceed the standard errors of the direct estimates. For the univariate STM with intervention this effect is large in the second quarter of 2020.

5. Official publications based on DHS

Official quarterly figures have been published for the eight selected DHS variables based on the bivariate STM (3.3). The first quarterly series were published in August 2020. These series ran from the first quarter of 2017 up to the second quarter in 2020. Subsequently, new estimates were published every quarter. The last quarter to be published in this series will probably be in the fourth quarter of 2021. The quarterly figures are computed in real time and will not be revised after publication.

Based on the quarterly figures also quarterly and annual developments are published. Quarterly developments are defined as the difference between two consecutive quarters and the annual developments as the difference between the same quarters in two consecutive years. The developments can be directly derived from the published quarterly figures. Standard errors for the quarterly developments are obtained by calculating the linear combination $\Delta_t^{\mathrm{Q}} = L_{t|t} - L_{t-1|t} + S_{t|t} - S_{t-1|t}$ via the Kalman filter recursion in (3.5) and (3.6). For the annual developments the standard errors are computed by calculating the linear combination of trends $\Delta_t^{\mathrm{A}} = L_{t|t} - L_{t-4|t}$. Here the linear combination of signals $L_{t|t} - L_{t-4|t} + S_{t|t} - S_{t-4|t}$ has not been used, because in that case many extra state variables should be specified in the vector $\boldsymbol{\alpha}_t$ in order to compute the seasonal components $S_{t-4|t}$. This may lead to unstable estimates.

The annual DHS figures for 2020 have been benchmarked with the quarterly figures by extending the regular weighting model with the quarterly STM estimates for the eight selected variables. In this way numerical consistency is achieved between the annual and quarterly publications. There is also a correction for the loss of CAPI for more detailed breakdowns of the eight variables. And finally a best possible correction is realized for the loss of CAPI for other related variables for which no model-based quarterly estimates are developed. Annual figures for 2021 will be benchmarked in a similar way. Quarterly and annual publications for 2017, 2018 and 2019 have not been made consistent with each other. Instead an explanatory note is added to these publications.

The effect of extending the weighting model of the annual figures for 2020 follows from Table 5.1. This table shows the summary statistics of the final weights for both the regular annual weighting and the extended annual weighting. The weights in the table are rounded to integers. It follows that the dispersion of the weights slightly increases by extending the weighting model. The standard deviation increases from 679 to 815. This is mainly because the minimum weight decreases from 457 to 67. The increase of the maximum weight is relative small.

	Regular weighting	Extended weighting
Minimum weight	457	67
First quantile	1493	1391
Median	1876	1859
Third quartile	2328	2433
Maximum weight	5507	6032
Standard deviation	679	815
Number of negative weights	0	0

Table 5.1: summary statistics final weights annual weighting 2020

In table 5.2 the results of the annual DHS figures are shown for 2020. In addition to the eight variables for which quarterly figures have been estimated, the variables cancer (ever had) and bronchitis (past 12 months) are displayed. The estimates in the table are percentages and the corresponding standard errors are given in parentheses. The corrections to the annual figures for the eight variables are in line with the previous results discussed in section 4. For daily smoking, overweight and excessive alcohol consumption there is a positive correction for the loss of CAPI in 2020. From section 4 we know that CAWI respondents score lower than CAPI respondents for these variables. For the other variables the correction on the annual figures is negative, while CAWI scores higher than CAPI. One would expect that cancer is related to the lifestyle variables, but this variable is negatively corrected from 6.47 (regular weighting) to 6.44 (extended weighting). For this variable the correction by means of the model-based quarterly figures does not work very well. On the other hand, this variable concerns all types of cancer and the relationship may be less strong. For bronchitis, where a strong relation is expected with daily smoking, the correction is indeed in the same direction as for daily smoking.

Variable	Regular weighting	Extended weighting
Perceived health	81.70 (0.45)	81.46 (0.46)
Mentally unhealthy	12.22 (0.42)	11.88 (0.42)
GP consult	25.26 (0.50)	24.65 (0.50)
Dental Visit	16.83 (0.42)	16.08 (0.42)
Specialist consult	14.94 (0.40)	14.41 (0.40)
Daily smoking	13.61 (0.45)	14.87 (0.49)
Overweight	49.96 (0.63)	50.02 (0.65)
Excessive alcohol consumption	6.43 (0.30)	6.93 (0.33)
Cancer	6.47 (0.26)	6.44 (0.26)
Bronchitis	4.28 (0.23)	4.33 (0.23)

Table 5.2: results annual figures DHS 2020. Estimates are in percentages and standard errors in parentheses.

6. Discussion

Based on the Dutch Health Survey (DHS), until 2020 only annual figures on health, care use and lifestyle were published by Statistics Netherlands. As a result of the corona crisis and the associated lockdown it was decided in June 2020 to publish a series of quarterly figures based on a structural time series model for a selection of eight DHS key variables. This serves multiple purposes. Firstly, with quarterly figures the period of the corona crisis can be better defined, so that possible effects of the crisis on the health figures can be measured better. Secondly, quarterly figures are more timely available, namely already during the statistical year and not only after the end of the reference year. This clearly increases the relevance of the health figures. Because the sample size of the DHS is too small to produce sufficiently precise quarterly figures with a direct estimator, the time series modelling approach is used as a form of small area estimation to improve the precision of the quarterly figures with sample information from preceding reference periods. And finally, it is possible to use the time series model to correct for the loss of face-to-face observation during the lockdown.

In this paper three time series methods are considered, two based on a univariate structural time series model (STM) and one based on a bivariate STM. Estimates based on the first univariate STM assume that there are no mode effects between web response and face-to-face response. In quarters where face-to-face response is missing due to lockdown measures this method does not correct for the loss of face-to-face observation. The second univariate STM attempt to model the change in measurement and selection bias with a level intervention variable. This is also a less optimal solution, since the lockdown also has a strong effect on the population parameters. A part of the real evolution of the population parameters is incorrectly absorbed in the level intervention, resulting in biased model predictions for the population parameters of interest. The third approach combines two series of direct estimates, a series based on complete response and a series based on web response, in a bivariate STM. This model uses the observed differences between the complete series and the series based on web response to make a time-varying estimate for the sudden change in measurement and selection effects due to the loss of CAPI during the lockdown period. In quarters where face-to-face response is missing, there are no estimates available based on the complete response, and nowcasts are obtained through the bivariate STM. With this approach, sudden changes in the period-to-period change of the population parameter are not interpreted as differences in measurement and selection bias. It is however assumed that the observed differences between the two input series in the period before the lockdown, do not change during the lockdown. For the selected DHS variables there are clearly mode effects implying that the first univariate STM produces biased estimates in quarters during the lockdown when there is no or less face-to-face observation possible. The second univariate STM incorrectly interprets a part of the sudden changes in the real period-to-period developments as differences in measurement bias and selection effects. For these reasons the univariate models are unsuitable for estimating quarterly figures during the Covid-19 pandemic.

Based on the bivariate STM official quarterly figures are published for the eight selected DHS variables. The published series starts in the first quarter of 2017 and will probably continue until the last quarter of 2021, because after that the Covid-19 pandemic will probably be over. This paper concerns the estimation of the series until the second quarter of 2021. Figures for the first and fourth quarter in 2020 are treated as if face-to-face observation is completely available. The loss of these observations in the last two weeks of March and December is ignored, which is a reasonable assumption. In the third and fourth quarter of 2020 and in the second quarter of 2021 the CAPI response is lower than before the corona crisis. In the second quarter of 2020 and the first quarter of 2021 no face-to-face observation was available and in the third quarter of 2020 face-to-face observation was only available in the months August and September. For these quarters a correction has been applied through the bivariate STM. This correction relies on two assumptions. The first assumption is that the composition of the web response does not change during the corona crisis. This assumption has been evaluated and confirmed through a response analysis. The second assumption is that the difference between the web and face-to-face response does not change due to the corona crisis. The latter assumption cannot be evaluated. During the pandemic in 2020 and 2021 people selected for face-to-face interviewing were given the opportunity to respond via the internet, which leads to a third response mode. Since it does not yield major differences in the estimates whether the web respondents that are selected for a face-to-face interview are treated as web or face-to-face respondents, these respondents are considered as web response.

The corrections for the loss of face-to-face observation have been incorporated in the annual figures of 2020 by including in the weighting model of the annual response a table with the corrected model-based quarterly figures for the eight selected DHS variables. In this way the best possible correction is realized for the loss of face-to-face response for other related variables for which no model-based quarterly estimates are developed.

An essential advantage of using the STM is that model-based estimates are more accurate than direct estimates. In particular, period-by-period developments can be estimated much more accurate thanks to the positive correlation between trend estimates and consecutive periods. For some variables the corona crisis seems to have had a major effect on the development. In order to account for the sudden increase in the dynamics of these figures in the time series model, it is necessary to make the trend component more flexible during the Covid-19 pandemic. This has been done by increasing the variance of the disturbance terms of the trend component during the Covid-19 pandemic. A consequence is that the standard errors of the model-based estimates increase for these quarters and are in some cases larger than the standard errors of the direct estimates.

The Covid-19 crisis increased the awareness that variance is not the only quality concept for official statistics, but that other quality dimensions such as timeliness and comparability over time are at least as important. As a result of this awareness, Statistics Netherlands extended the traditional design-based inference approach for the annual publications of the DHS, with a model-based inference method as a form of small area estimation to produce more timely figures. At the same time, the proposed method compensates for the bias that occurs as a result of the temporal loss of CAPI responses to maintain comparability over time and avoid a sudden increased MSE.

References

Binder, D.A., and J.P. Dick (1990). A method for the analysis of seasonal ARIMA models. Survey Methodology, 16, pp. 239-253.

Boonstra, H.J. (2019). Weging Gezondheidsenquête 2014. Discussion paper CBS, Statistics Netherlands, Heerlen. URL: https://www.cbs.nl/nl-nl/achtergrond/2019/18/weging-gezondheidsenquete-2014.

Buelens, B. and J. A. van den Brakel (2015). Measurement error calibration in mixed-mode surveys. Sociological Methods & Research, 44, pp. 391-426.

Doornik, J.A. (2009). An Object-oriented Matrix Programming Language Ox 6. London: Timberlake Consultants Press.

Durbin, J. and S. Koopman (2012). Time Series Analysis by State Space Methods (second edition). Oxford University Press, Oxford.

Kish, L. (1965). Survey Sampling, New York: John Wiley and Sons.

Koopman, S.J., Shephard, N. and Doornik, J.A. (2008). SsfPack 3.0: Statistical Algorithms for Models in State Space Form. London: Timberlake Consultants Press.

Lumley, T. (2014). 'survey': analysis of complex survey samples. R package version 3.30.
R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/.

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

Rao, J.N.K. and I. Molina (2015). Small Area Estimation, 2nd edition, New York: John Wiley and Sons.

Van den Brakel, J.A., M. Souren and S. Krieg (2021).

Appendix A. Components of applied time series models

The trend component L_t in models (3.1) and (3.3) is modelled by a smooth trend model with a flexible slope, given by

$$L_t = L_{t-1} + R_{t-1} \qquad \text{(A.1)}$$
$$R_t = R_{t-1} + \eta_t^R, \text{ where}$$

$$\eta_t^R \sim \mathcal{N}(0, f_t \sigma_{\mathrm{R}}^2)$$
$$\mathrm{Cov}(\eta_t^R, \eta_{t'}^R) = 0, \text{ for } t \neq t', f_t \geq 1 \text{ and } t \geq 0.$$

The seasonal component S_t in models (3.1) and (3.3) is modelled by a trigonometric seasonal model, given by

$$S_t = \gamma_{1,t} + \cdots + \gamma_{J/2,t}, \text{ where} \qquad \text{(A.2)}$$

$$\gamma_{j,t} = \gamma_{j,t-1} \cos\left(\frac{\pi j}{J/2}\right) + \gamma_{j,t-1}^* \sin\left(\frac{\pi j}{J/2}\right) + \omega_{j,t}$$
$$\gamma_{j,t}^* = \gamma_{j,t-1}^* \cos\left(\frac{\pi j}{J/2}\right) - \gamma_{j,t-1} \sin\left(\frac{\pi j}{J/2}\right) + \omega_{j,t}^* \text{ for } j = 1, \ldots, J/2 \text{ and } J \text{ even.}$$

For odd values of J, $J/2$ is replaced by $(J-1)/2$ in equation (A.2). For quarters $J = 4$ it holds that $\gamma_{j,t}^* = 0$, because $\sin(\pi) = 0$. So the trigonometric model reduces to

$$S_t = \gamma_{1,t} + \gamma_{2,t}, \text{ where}$$

$$\gamma_{1,t} = \gamma_{1,t-1}^* + \omega_{1,t}$$
$$\gamma_{1,t}^* = -\gamma_{1,t-1} + \omega_{1,t}^*$$
$$\gamma_{2,t} = -\gamma_{2,t-1} + \omega_{2,t}, \text{ and}$$

$$\omega_{1,t} \sim \mathcal{N}(0, \sigma_\omega^2)$$
$$\omega_{1,t}^* \sim \mathcal{N}(0, \sigma_\omega^2)$$
$$\omega_{2,t} \sim \mathcal{N}(0, \sigma_\omega^2), \text{ and}$$

$$\mathrm{Cov}(\omega_{j,t}, \omega_{j,t'}) = 0, \text{ for } t \neq t' \text{ and } j = 1, 2,$$
$$\mathrm{Cov}(\omega_{1,t}^*, \omega_{1,t'}^*) = 0, \text{ for } t \neq t',$$
$$\mathrm{Cov}(\omega_{j,t}, \omega_{1,t}^*) = 0, \text{ for all } t \text{ and } j = 1, 2,$$
$$\mathrm{Cov}(\omega_{1,t}, \omega_{2,t}) = 0, \text{ for all } t.$$

The systematic difference λ_t in model (3.3) between the regular series and the web series is modelled as a random walk, given by

$$\lambda_t = \lambda_{t-1} + \eta_{\lambda,t}, \text{ where} \qquad \text{(A.3)}$$

$$\eta_{\lambda,t} \sim \mathcal{N}(0, \sigma_\lambda^2)$$

$\text{Cov}(\eta_{\lambda,t}, \eta_{\lambda,t'}) = 0$, for $t \neq t'$.

The measurement errors e_t^{A} in model (3.1) and e_t^{C} and e_t^{W} in model (3.3) are modelled by the following measurement error model (Binder and Dick, 1990):

$$e_t^j = \sqrt{\widehat{\mathrm{V}}(\hat{y}_t^j)}\, \tilde{e}_t^j \text{ for } j \in \{\mathrm{A}, \mathrm{C}, \mathrm{W}\}, \text{ where} \tag{A.4}$$

$$\tilde{e}_t^j \sim \mathcal{N}(0, \sigma_{\mathrm{e},j}^2)$$
$$\text{Cov}(\tilde{e}_t^j, \tilde{e}_{t'}^j) = 0, \text{ for } t \neq t'$$

The measurement errors of the input series of the bivariate STM are correlated, due to the overlap in the response on which the direct estimates $\hat{y}_t^{\text{C}}$ and $\hat{y}_t^{\text{W}}$ are based. The covariance between the measurement errors are given by (3.4). This expression is obtained as follows. Following Kish (1965), the correlation between two variables observed in two partial overlapping samples is given by

$$\mathrm{Cor}(z_1, z_2) = \rho \frac{n_{1\cap 2}}{\sqrt{n_1}\sqrt{n_2}}, \text{ where}$$

- z_1 the variable observed in sample s_1 of size n_1,
- z_2 the variable observed in sample s_2 of size n_2,
- $n_{1\cap 2}$ the size of the sample overlap between s_1 and s_2,
- ρ the correlation between z_1 and z_2 based on the $n_{1\cap 2}$ respondents that are included in s_1 and s_2.

In this application, sample s_1 is the sample with CAWI respondents and s_2 the sample of complete response. This implies that $z_1 = \hat{y}_t^{\text{W}}$, $z_2 = \hat{y}_t^{\text{C}}$, $n_1 = n_t^{\text{W}}$, and $n_2 = n_t^{\text{C}}$, with n_t^{W} is the size of the web response in quarter t and n_t^{C} the size of the complete response in quarter t. In this case the sample overlap is also the sample with CAWI respondents. Therefore we have $n_{1\cap 2} = n_t^{\text{W}}$ and $\rho = 1$. From this it follows that

$$\mathrm{Cor}(\hat{y}_t^{\text{W}}, \hat{y}_t^{\text{C}}) = \rho \frac{n_t^{\text{W}}}{\sqrt{n_t^{\text{W}}}\sqrt{n_t^{C}}} = \frac{\sqrt{n_t^{\text{W}}}}{\sqrt{n_t^{C}}}$$

and

$$\text{Cov}(\hat{y}_t^{\text{W}}, \hat{y}_t^{\text{C}}) = \frac{\sqrt{n_t^{\text{W}}}}{\sqrt{n_t^{C}}} \sqrt{\widehat{\mathrm{V}}(\hat{y}_t^{\text{C}})}\sqrt{\widehat{\mathrm{V}}(\hat{y}_t^{\text{W}})}$$

The latter expression is used as an estimator for the covariance between the measurement errors of the input series of the bivariate STM in (3.3).

Appendix B. Results STM

	Perceived health					
	2019 Q2	2020 Q2	2020 Q3	2020 Q4	2021 Q1	2021 Q2
	Univariate STM without intervention					
$\hat{\sigma}_R$	<0.001	0.001	<0.001	0.001	0.001	0.001
$\hat{\sigma}_\omega$	<0.001	0.002	0.001	0.002	0.002	0.001
$\hat{\sigma}_{e,A}$	0.007	0.007	0.008	0.007	0.007	0.007
	Univariate STM with intervention					
$\hat{\sigma}_R$	<0.001	<0.001	<0.001	<0.001	0.001	<0.001
$\hat{\sigma}_\omega$	<0.001	0.002	0.002	0.002	0.002	0.001
$\hat{\sigma}_{e,A}$	0.007	0.006	0.006	0.007	0.008	0.008
	Bivariate STM					
$\hat{\sigma}_R$	<0.001	0.001	<0.001	<0.001	<0.001	0.001
$\hat{\sigma}_\omega$	<0.001	<0.001	0.002	0.002	0.002	0.002
$\hat{\sigma}_\lambda$	<0.001	<0.001	0.002	<0.001	<0.001	0.004
$\hat{\sigma}_{e,C}$	0.957	1.120	0.979	0.934	0.928	0.862
$\hat{\sigma}_{e,W}$	1.310	1.340	1.180	1.240	1.240	1.040
	Mentally unhealthy					
	Univariate STM without intervention					
$\hat{\sigma}_R$	<0.001	<0.001	<0.001	<0.001	0.002	0.001
$\hat{\sigma}_\omega$	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
$\hat{\sigma}_{e,A}$	0.006	0.005	0.005	0.005	0.005	0.006
	Univariate STM with intervention					
$\hat{\sigma}_R$	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
$\hat{\sigma}_\omega$	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
$\hat{\sigma}_{e,A}$	0.006	0.006	0.005	0.006	0.006	0.006
	Bivariate STM					
$\hat{\sigma}_R$	<0.001	<0.001	<0.001	<0.001	0.001	<0.001
$\hat{\sigma}_\omega$	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
$\hat{\sigma}_\lambda$	0.032	0.003	0.003	<0.001	0.003	0.003
$\hat{\sigma}_{e,C}$	1.030	0.964	0.918	1.030	0.987	0.927
$\hat{\sigma}_{e,W}$	0.818	0.783	0.784	1.110	0.815	0.880

Table B.1: Concurrent ML estimates hyperparameters STM for perceived health and mentally unhealthy problems

	GP consult					
	2019 Q2	2020 Q2	2020 Q3	2020 Q4	2021 Q1	2021 Q2
	Univariate STM without intervention					
$\hat{\sigma}_{\mathrm{R}}$	<0.001	0.001	<0.001	<0.001	<0.001	<0.001
$\hat{\sigma}_{\omega}$	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
$\hat{\sigma}_{\mathrm{e,A}}$	0.008	0.007	0.008	0.008	0.008	0.007
	Univariate STM with intervention					
$\hat{\sigma}_{\mathrm{R}}$	<0.001	0.001	<0.001	<0.001	<0.001	<0.001
$\hat{\sigma}_{\omega}$	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
$\hat{\sigma}_{\mathrm{e,A}}$	0.008	0.007	0.007	0.007	0.008	0.007
	Bivariate STM					
$\hat{\sigma}_{\mathrm{R}}$	<0.001	0.002	0.001	<0.001	0.001	0.002
$\hat{\sigma}_{\omega}$	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
$\hat{\sigma}_{\lambda}$	0.002	<0.001	0.002	0.002	0.002	0.002
$\hat{\sigma}_{\mathrm{e,C}}$	1.040	1.030	1.150	1.160	1.230	1.180
$\hat{\sigma}_{\mathrm{e,W}}$	0.901	1.210	0.936	0.943	0.901	0.906
	Dental visit					
	Univariate STM without intervention					
$\hat{\sigma}_{\mathrm{R}}$	<0.001	0.003	0.005	<0.001	0.002	0.003
$\hat{\sigma}_{\omega}$	<0.001	<0.001	<0.001	0.004	0.003	0.002
$\hat{\sigma}_{\mathrm{e,A}}$	0.009	0.008	0.008	0.011	0.009	0.010
	Univariate STM with intervention					
$\hat{\sigma}_{\mathrm{R}}$	<0.001	0.003	0.005	<0.001	0.002	0.003
$\hat{\sigma}_{\omega}$	<0.001	<0.001	<0.001	0.004	0.003	0.002
$\hat{\sigma}_{\mathrm{e,A}}$	0.009	0.008	0.008	0.011	0.009	0.010
	Bivariate STM					
$\hat{\sigma}_{\mathrm{R}}$	<0.001	0.003	0.006	0.004	0.003	0.006
$\hat{\sigma}_{\omega}$	<0.001	<0.001	<0.001	0.004	0.004	0.001
$\hat{\sigma}_{\lambda}$	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
$\hat{\sigma}_{\mathrm{e,C}}$	1.080	1.130	1.100	0.713	0.624	0.954
$\hat{\sigma}_{\mathrm{e,W}}$	1.070	1.080	1.140	1.050	1.010	1.190

Table B.2: Concurrent ML estimates hyperparameters STM for GP consult and dental visit

	Specialist consult					
	2019 Q2	2020 Q2	2020 Q3	2020 Q4	2021 Q1	2021 Q2
	Univariate STM without intervention					
$\hat{\sigma}_{\mathrm{R}}$	<0.001	0.002	0.001	0.001	0.001	<0.001
$\hat{\sigma}_{\omega}$	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
$\hat{\sigma}_{\mathrm{e,A}}$	0.008	0.008	0.009	0.009	0.009	0.008
	Univariate STM with intervention					
$\hat{\sigma}_{\mathrm{R}}$	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
$\hat{\sigma}_{\omega}$	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
$\hat{\sigma}_{\mathrm{e,A}}$	0.008	0.008	0.008	0.008	0.009	0.009
	Bivariate STM					
$\hat{\sigma}_{\mathrm{R}}$	<0.001	0.002	0.002	0.002	0.001	<0.001
$\hat{\sigma}_{\omega}$	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
$\hat{\sigma}_{\lambda}$	<0.001	<0.001	<0.001	<0.001	<0.001	0.002
$\hat{\sigma}_{\mathrm{e,C}}$	1.020	1.060	1.090	1.100	1.120	1.120
$\hat{\sigma}_{\mathrm{e,W}}$	0.973	1.040	1.110	1.130	1.120	1.140
	Daily smoking					
	Univariate STM without intervention					
$\hat{\sigma}_{\mathrm{R}}$	<0.001	<0.001	<0.001	<0.001	<0.001	0.002
$\hat{\sigma}_{\omega}$	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
$\hat{\sigma}_{\mathrm{e,A}}$	0.010	0.011	0.010	0.01	0.010	0.010
	Univariate STM with intervention					
$\hat{\sigma}_{\mathrm{R}}$	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
$\hat{\sigma}_{\omega}$	<0.001	<0.001	<0.001	0.001	<0.001	0.002
$\hat{\sigma}_{\mathrm{e,A}}$	0.010	0.009	0.010	0.009	0.009	0.009
	Bivariate STM					
$\hat{\sigma}_{\mathrm{R}}$	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
$\hat{\sigma}_{\omega}$	<0.001	<0.001	<0.001	0.001	0.003	0.003
$\hat{\sigma}_{\lambda}$	0.002	0.003	0.007	0.01	0.002	0.002
$\hat{\sigma}_{\mathrm{e,C}}$	1.310	1.220	1.200	1.250	1.710	1.710
$\hat{\sigma}_{\mathrm{e,W}}$	1.280	1.210	1.100	0.653	0.718	0.760

Table B.3: Concurrent ML estimates hyperparameters STM for specialist consult and daily smoking

	Overweight					
	2019 Q2	2020 Q2	2020 Q3	2020 Q4	2021 Q1	2021 Q2
	Univariate STM without intervention					
$\hat{\sigma}_{\mathrm{R}}$	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
$\hat{\sigma}_{\omega}$	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
$\hat{\sigma}_{\mathrm{e,A}}$	0.014	0.013	0.013	0.013	0.013	0.013
	Univariate STM with intervention					
$\hat{\sigma}_{\mathrm{R}}$	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
$\hat{\sigma}_{\omega}$	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
$\hat{\sigma}_{\mathrm{e,A}}$	0.014	0.013	0.013	0.012	0.010	0.012
	Bivariate STM					
$\hat{\sigma}_{\mathrm{R}}$	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
$\hat{\sigma}_{\omega}$	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
$\hat{\sigma}_{\lambda}$	<0.001	<0.001	0.004	0.004	0.004	0.004
$\hat{\sigma}_{\mathrm{e,C}}$	1.130	1.080	1.040	1.020	1.020	1.060
$\hat{\sigma}_{\mathrm{e,W}}$	1.070	1.030	1.020	1.090	1.090	1.030
	Excessive alcohol consumption					
	Univariate STM without intervention					
$\hat{\sigma}_{\mathrm{R}}$	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
$\hat{\sigma}_{\omega}$	0.002	0.001	0.001	0.001	0.001	<0.001
$\hat{\sigma}_{\mathrm{e,A}}$	0.008	0.010	0.009	0.009	0.009	0.010
	Univariate STM with intervention					
$\hat{\sigma}_{\mathrm{R}}$	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
$\hat{\sigma}_{\omega}$	0.002	<0.001	<0.001	0.001	0.001	<0.001
$\hat{\sigma}_{\mathrm{e,A}}$	0.008	0.010	0.010	0.010	0.009	0.009
	Bivariate STM					
$\hat{\sigma}_{\mathrm{R}}$	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
$\hat{\sigma}_{\omega}$	0.002	0.004	<0.001	<0.001	<0.001	<0.001
$\hat{\sigma}_{\lambda}$	0.003	0.002	0.002	0.002	0.003	0.002
$\hat{\sigma}_{\mathrm{e,C}}$	1.360	1.460	1.450	1.480	1.490	1.470
$\hat{\sigma}_{\mathrm{e,W}}$	0.822	1.070	1.050	1.040	1.040	1.050

Table B.4: Concurrent ML estimates hyperparameters STM for overweight and excessive alcohol consumption

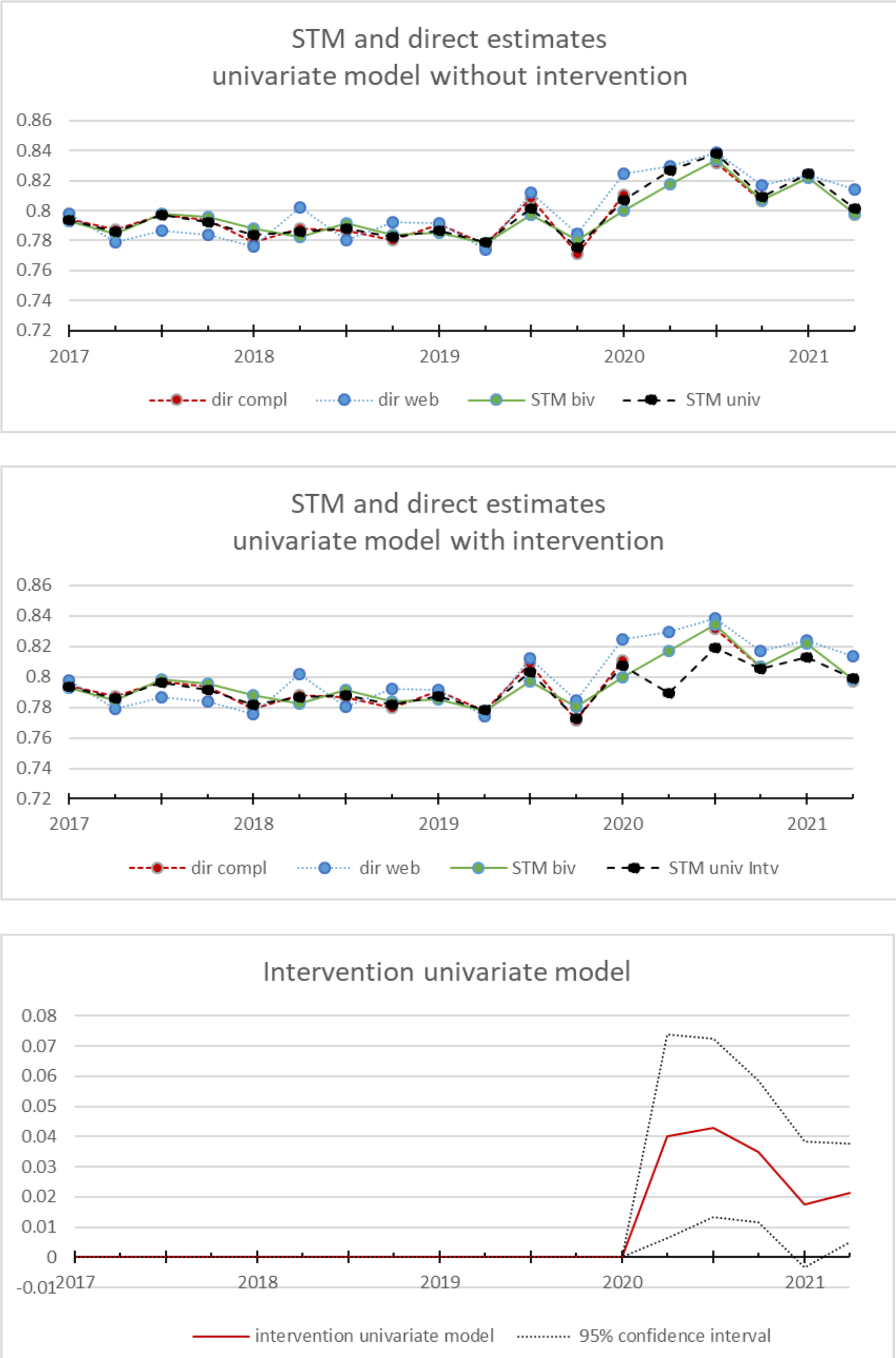


Figure B.1: results STM for perceived health

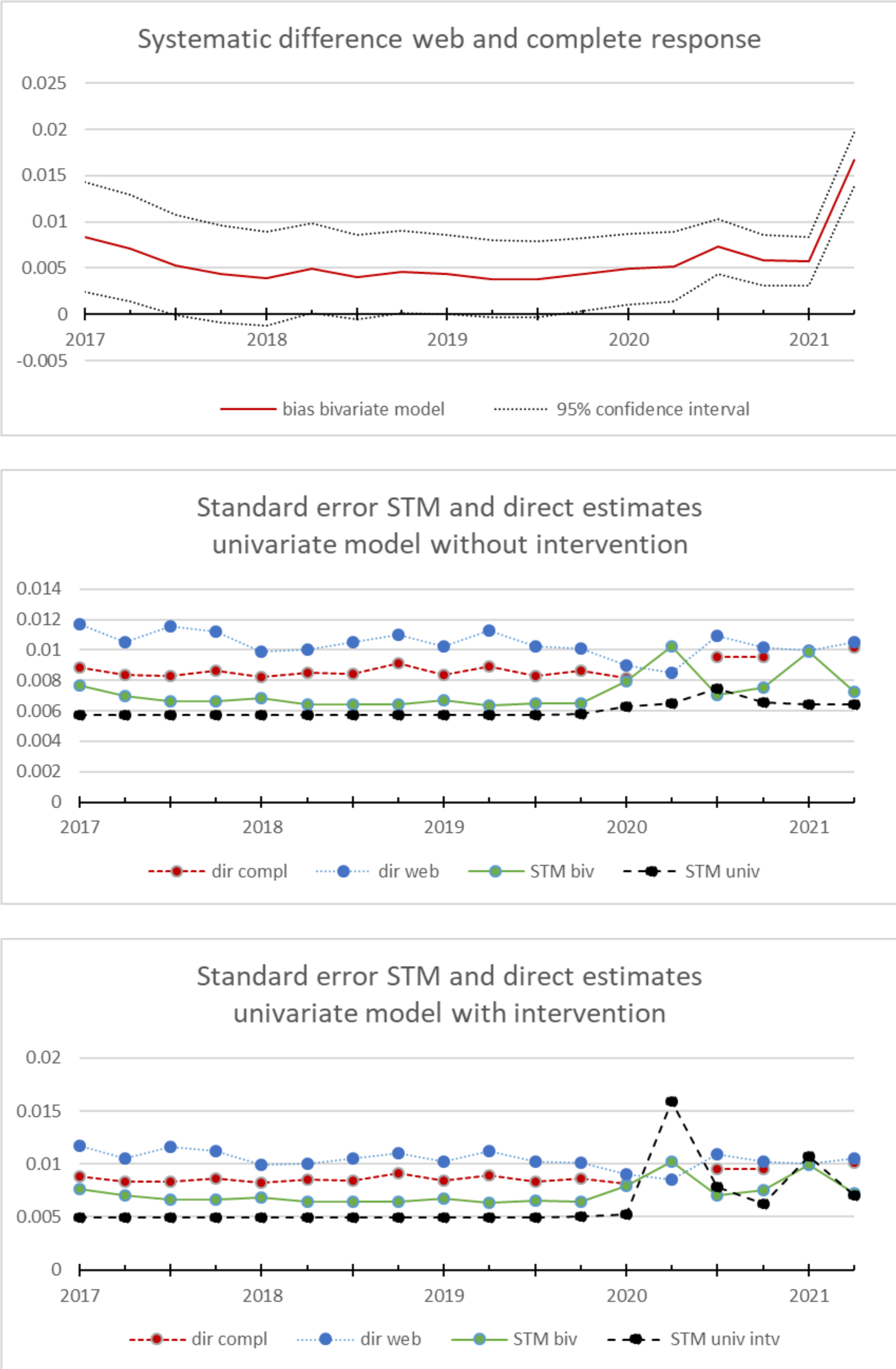


Figure B.2: results STM for perceived health

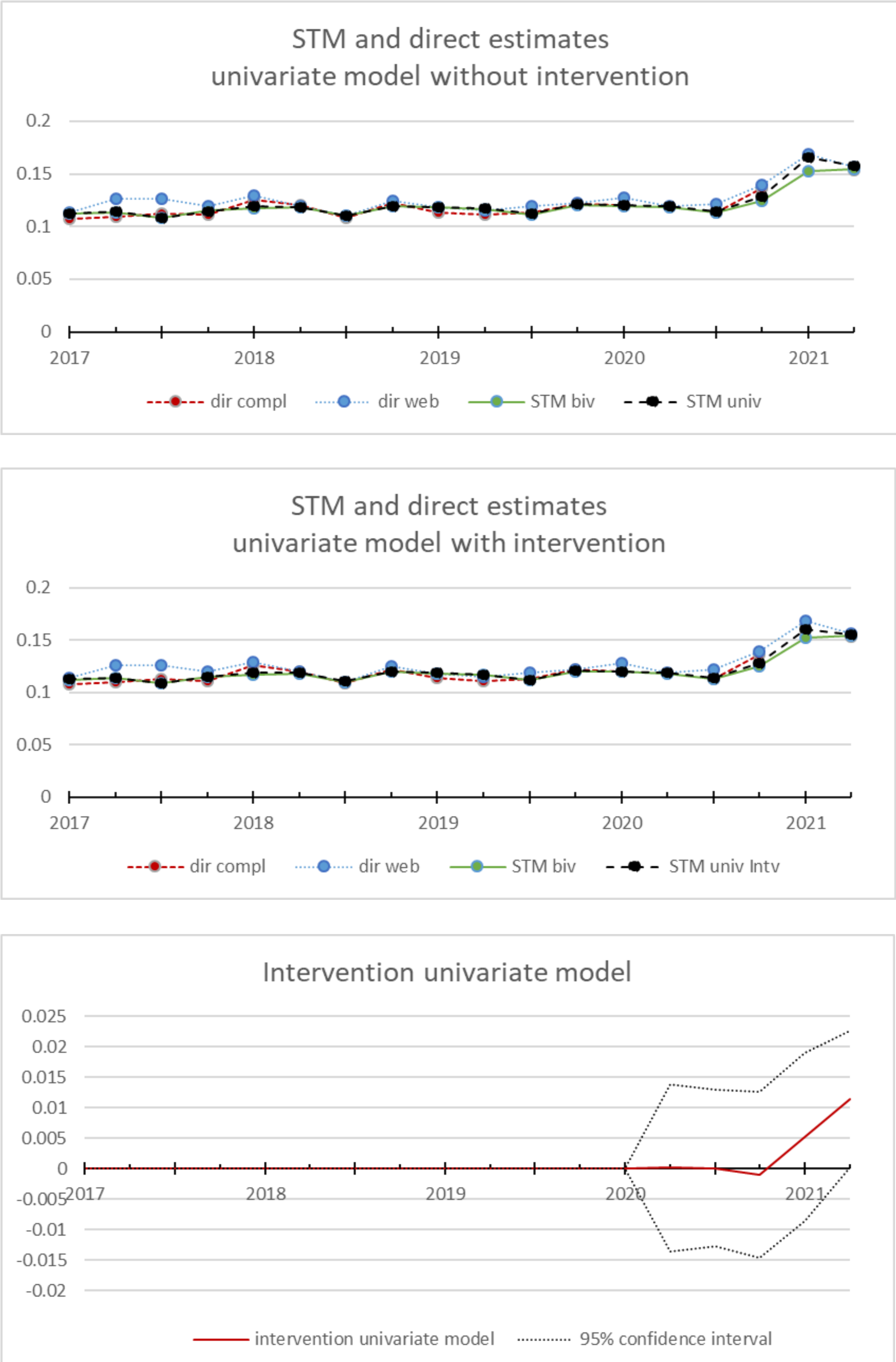


Figure B.3: results STM for mentally unhealthy

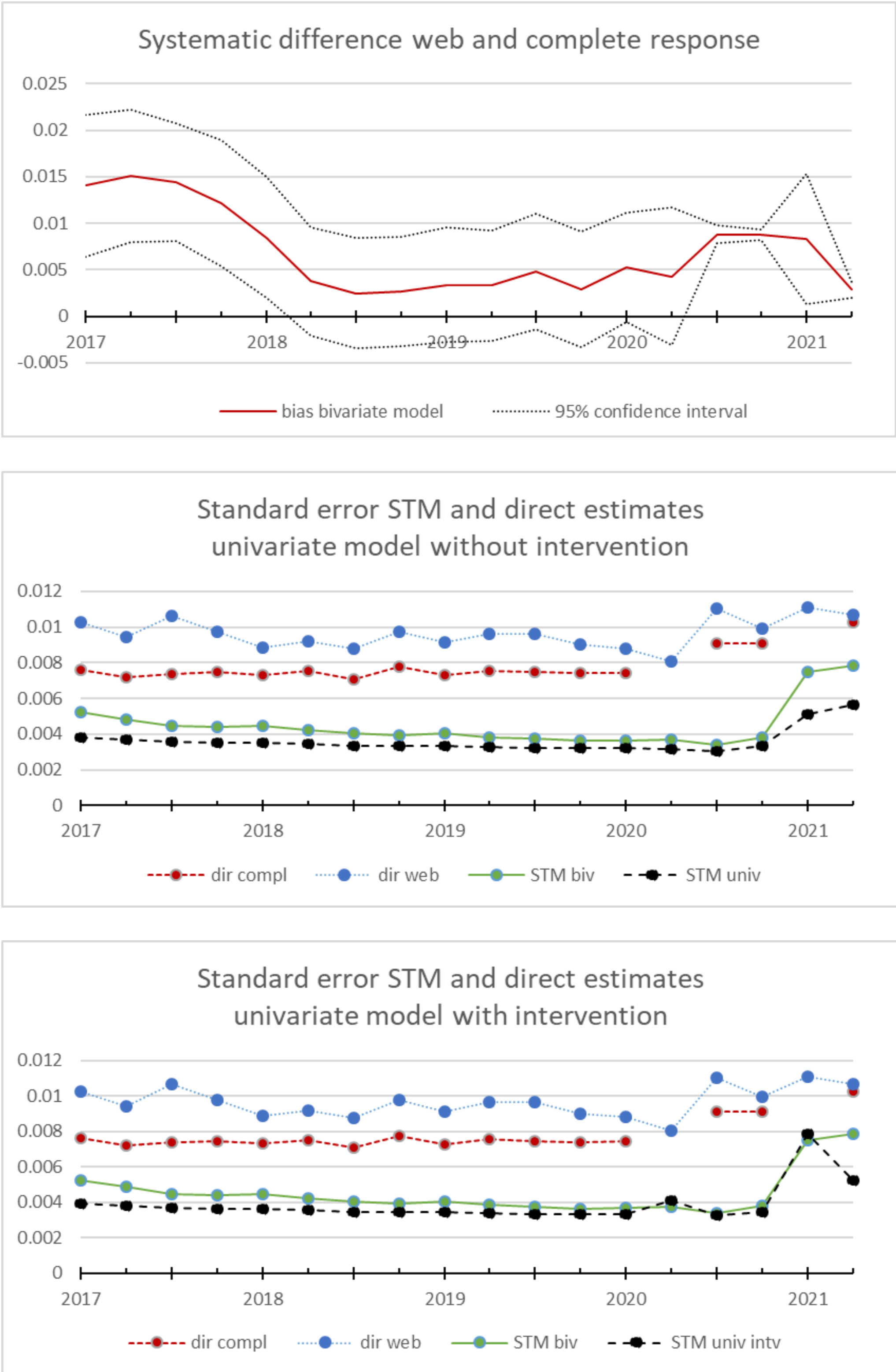


Figure B.4: results STM for mentally unhealthy

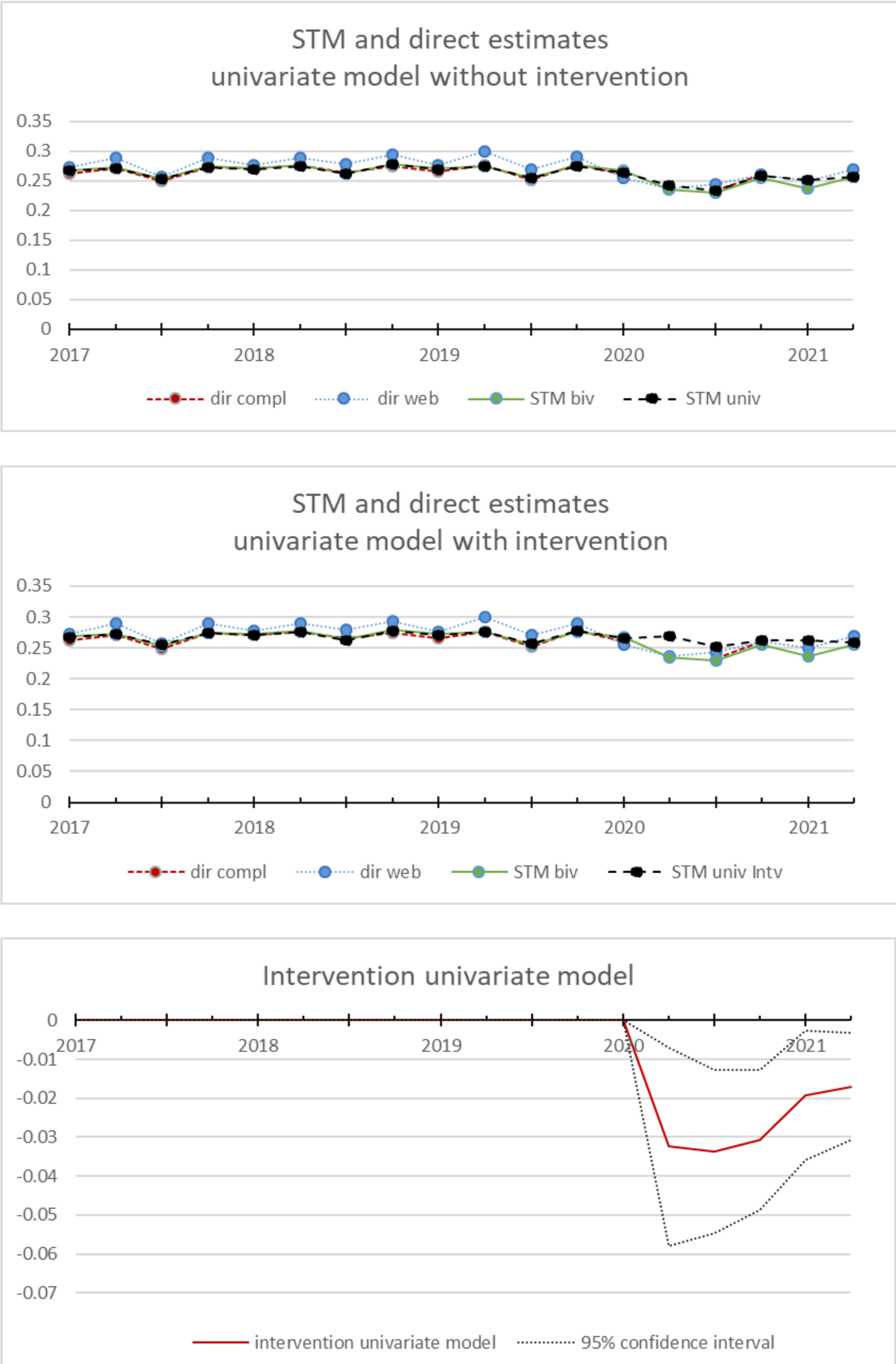
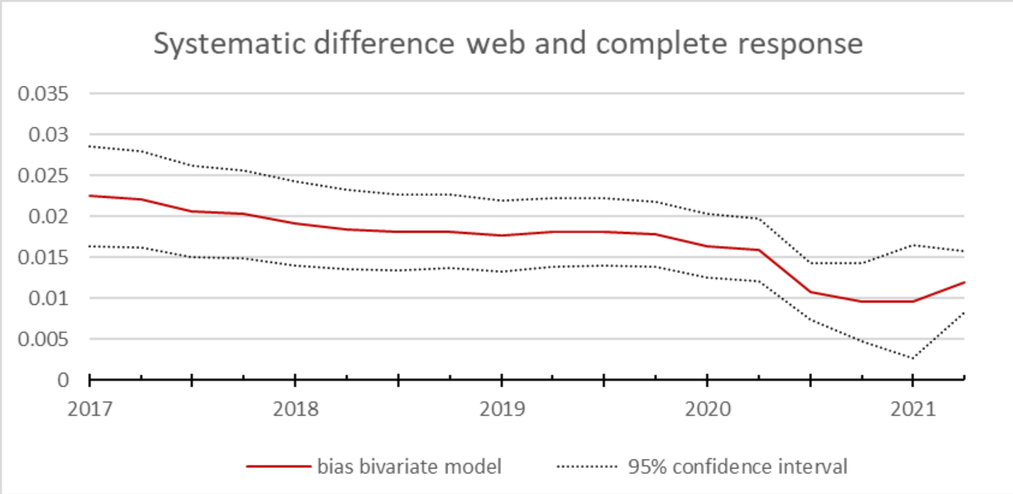
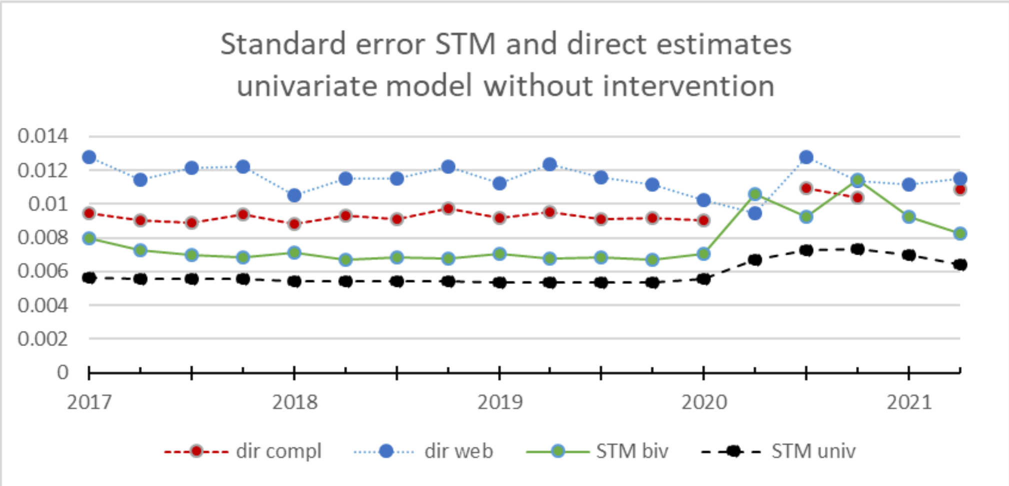


Figure B.5: results STM for GP consult





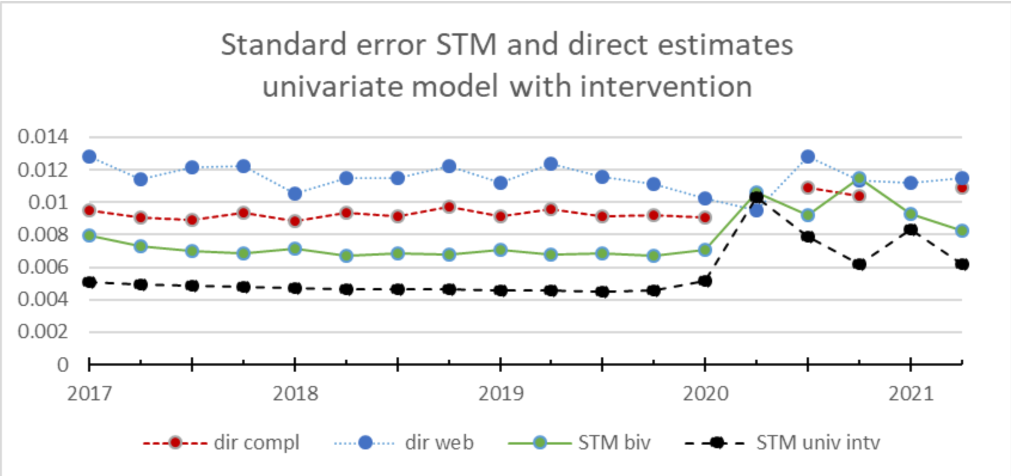
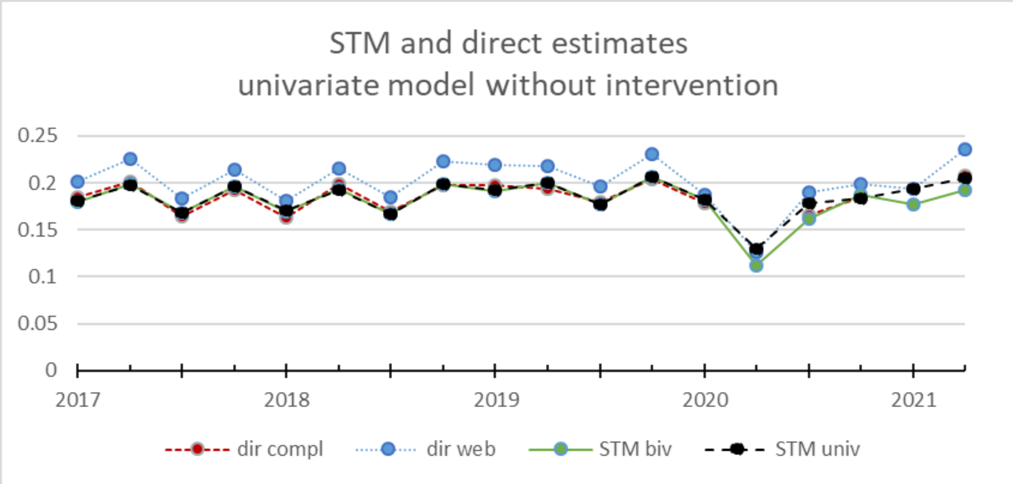
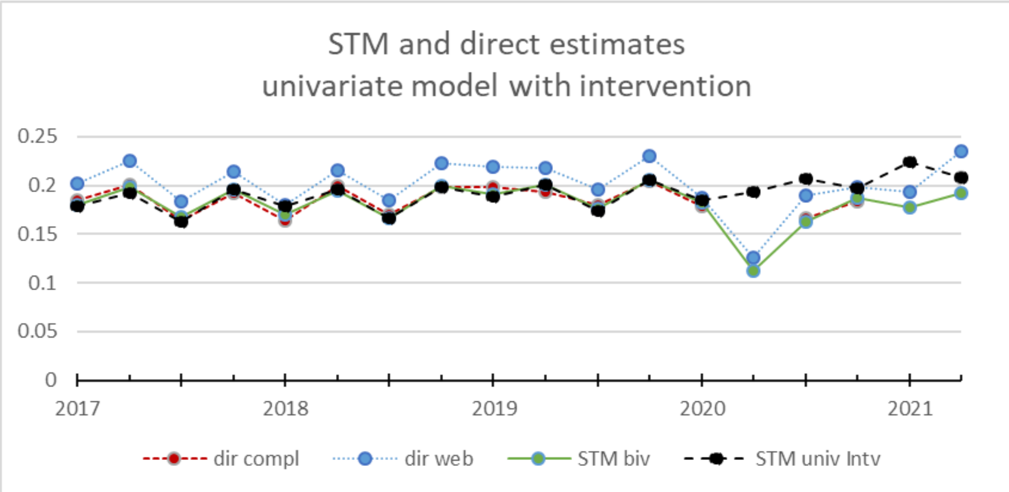


Figure B.6: results STM for GP consult





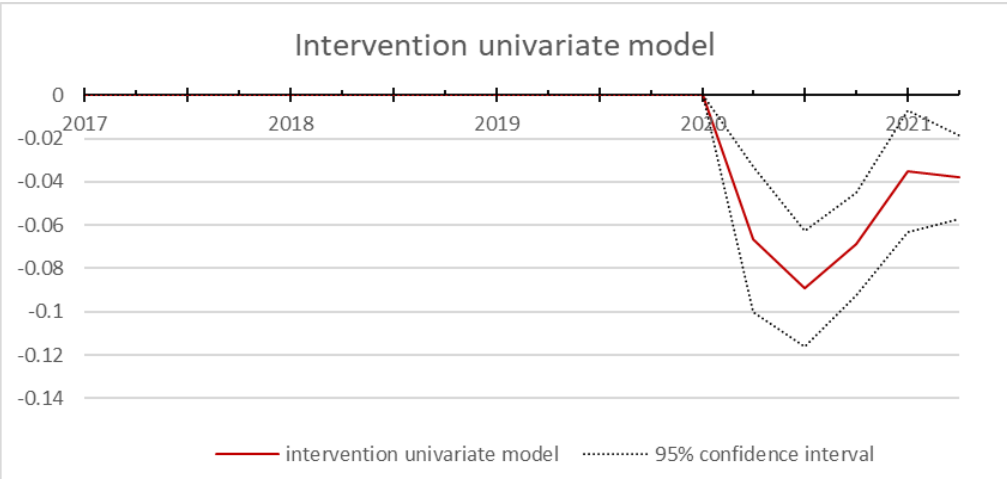


Figure B.7: results STM for dental visit

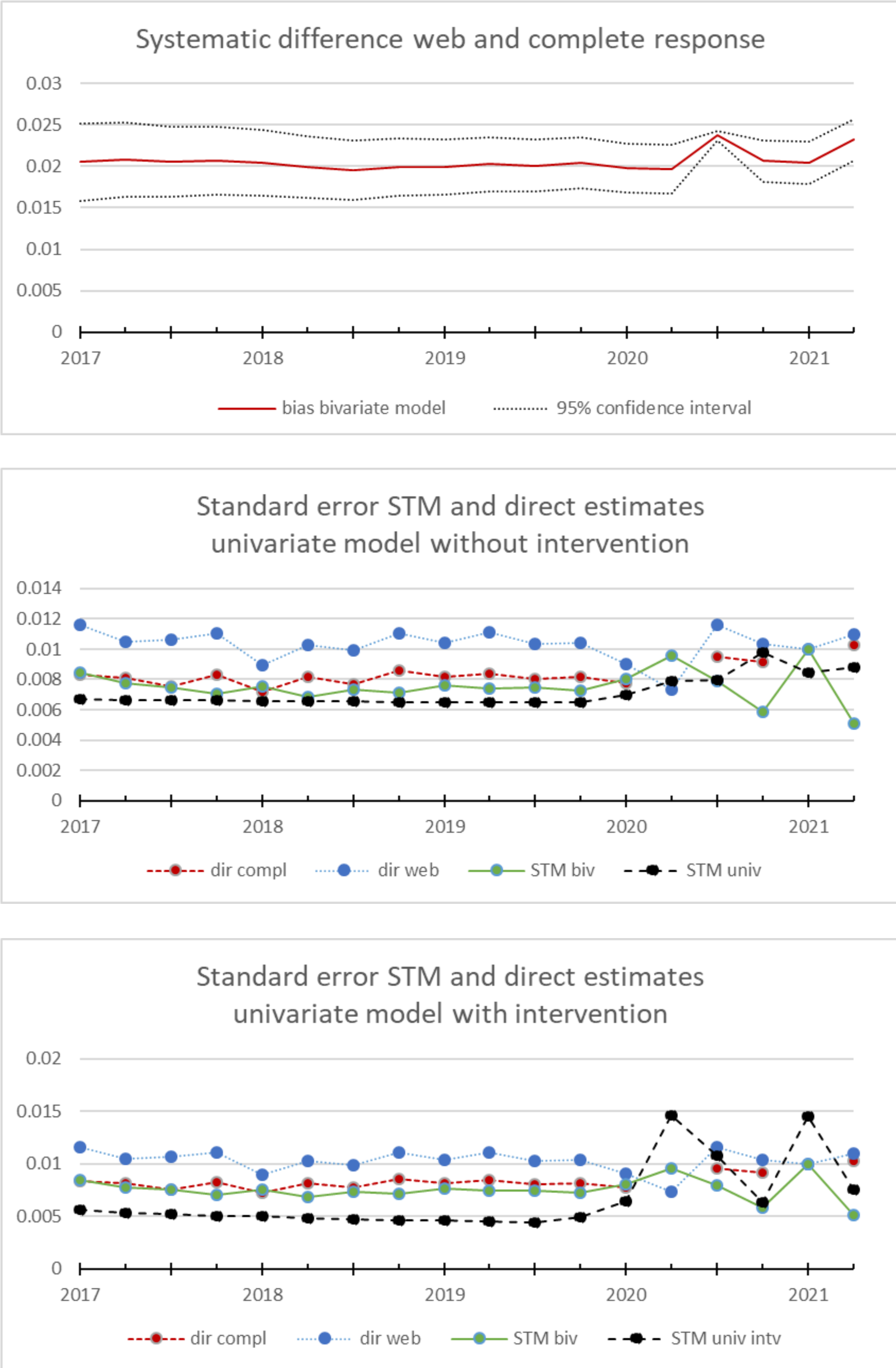
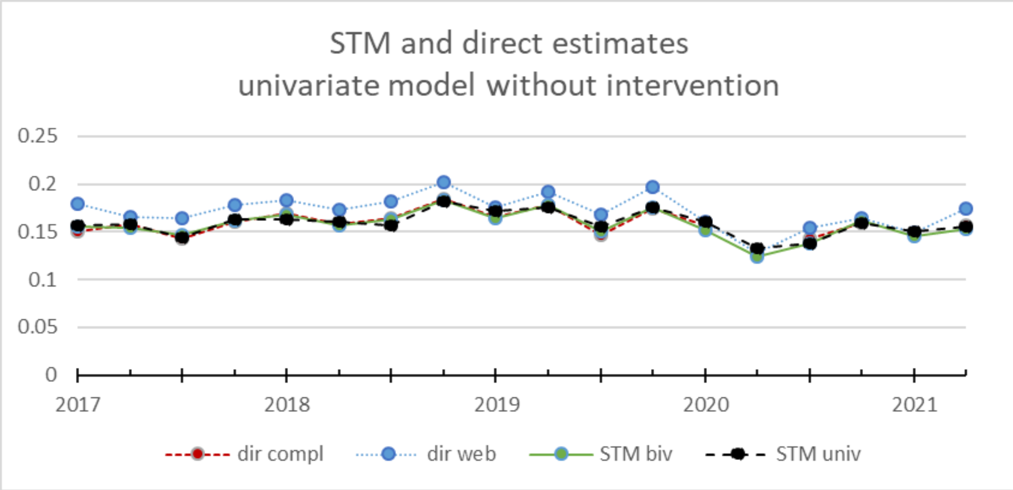
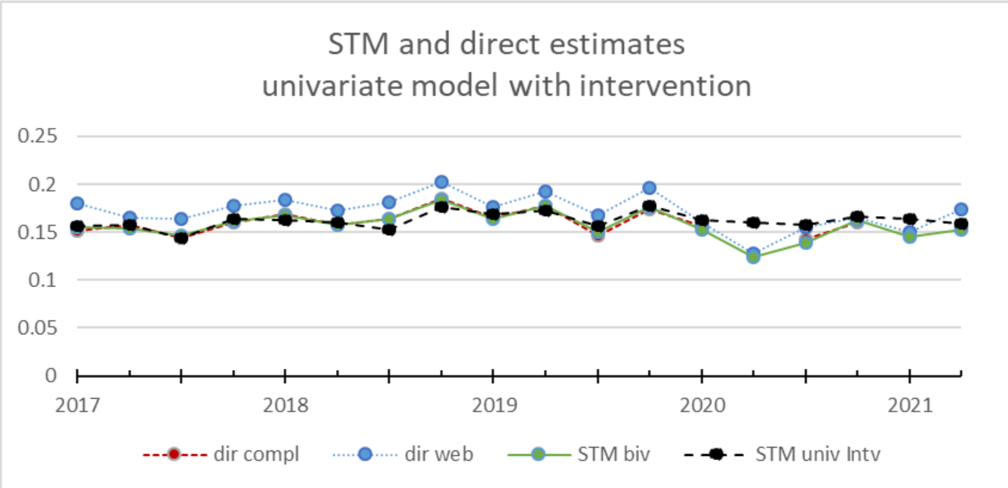


Figure B.8: results STM for dental visit





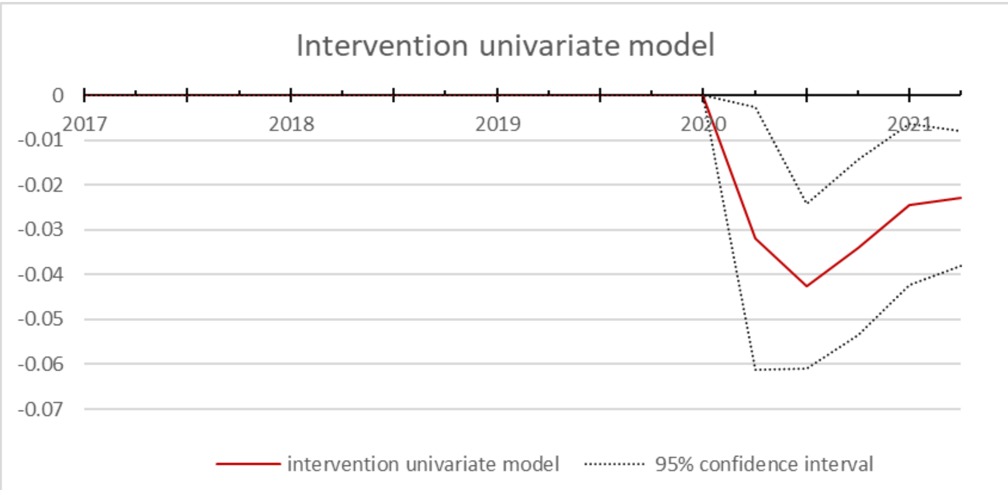
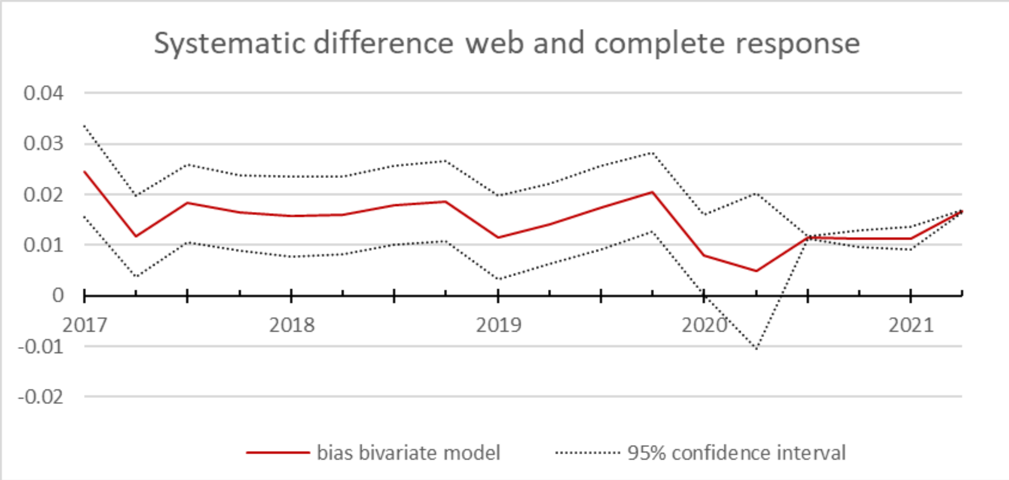
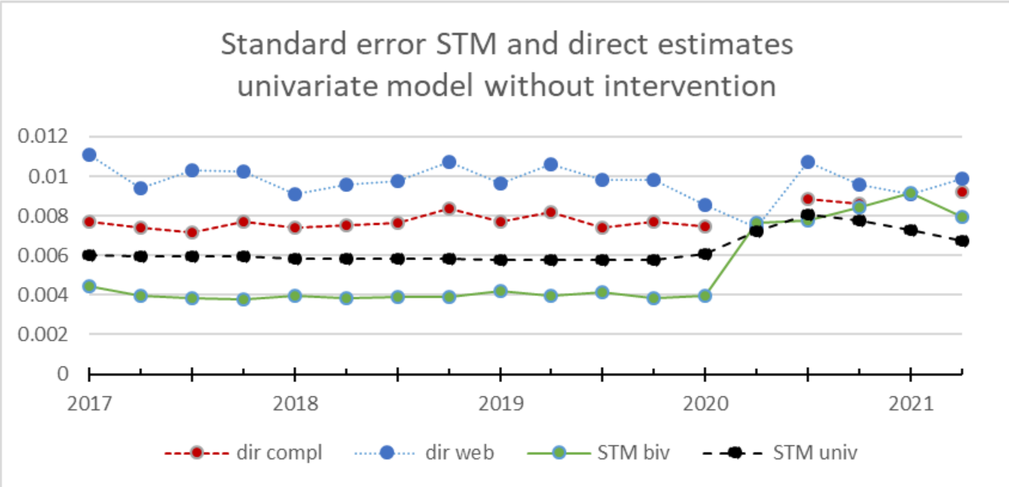


Figure B.9: results STM for specialist consult





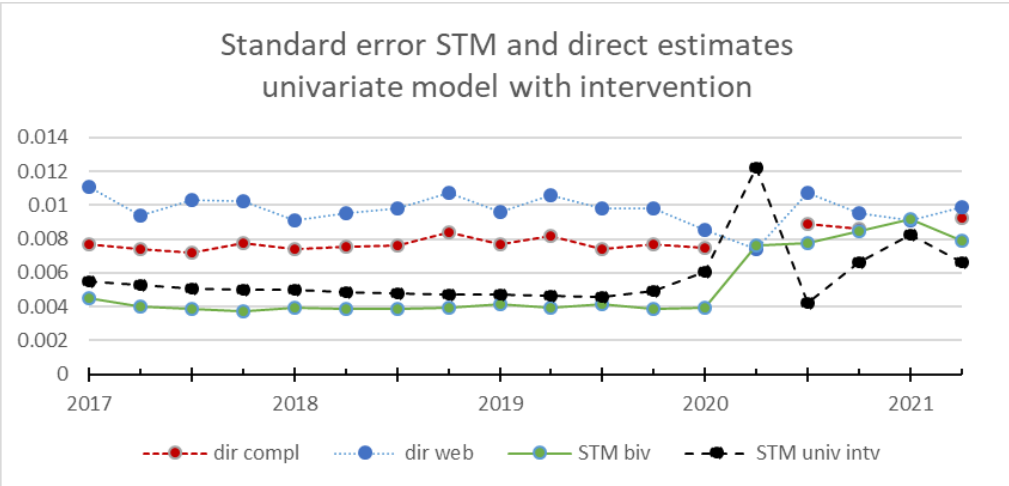
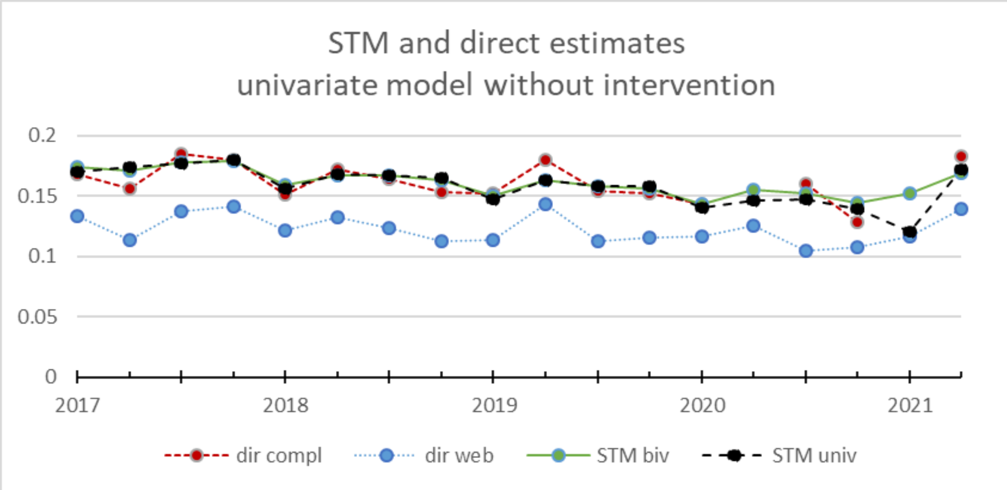
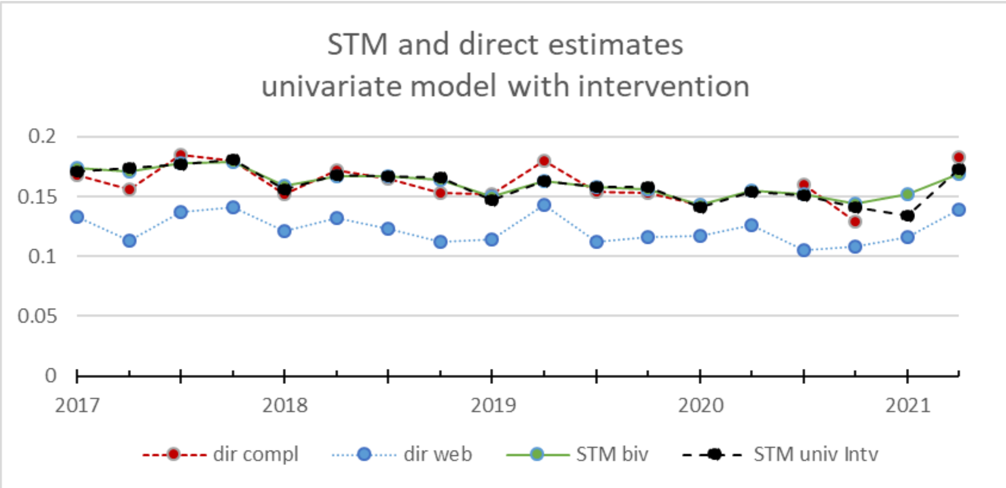


Figure B.10: results STM for specialist consult





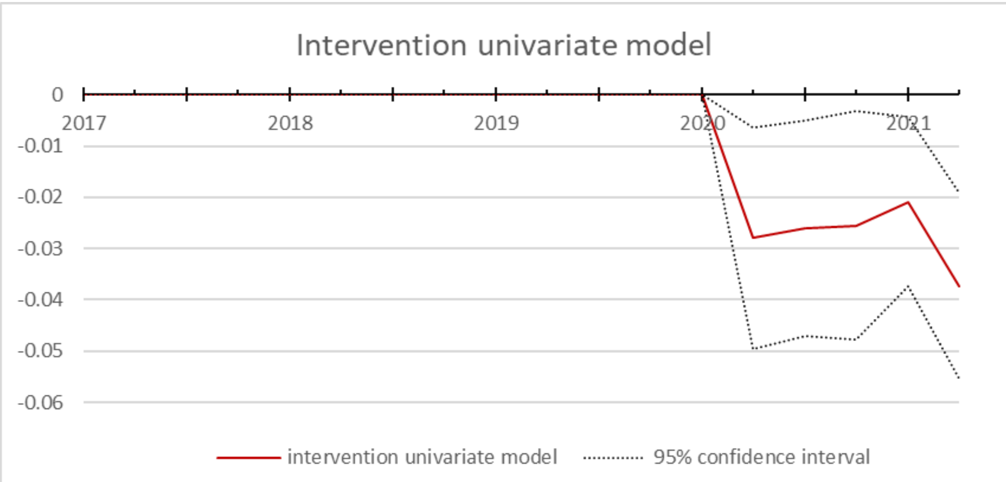


Figure B.11: results STM for daily smoking

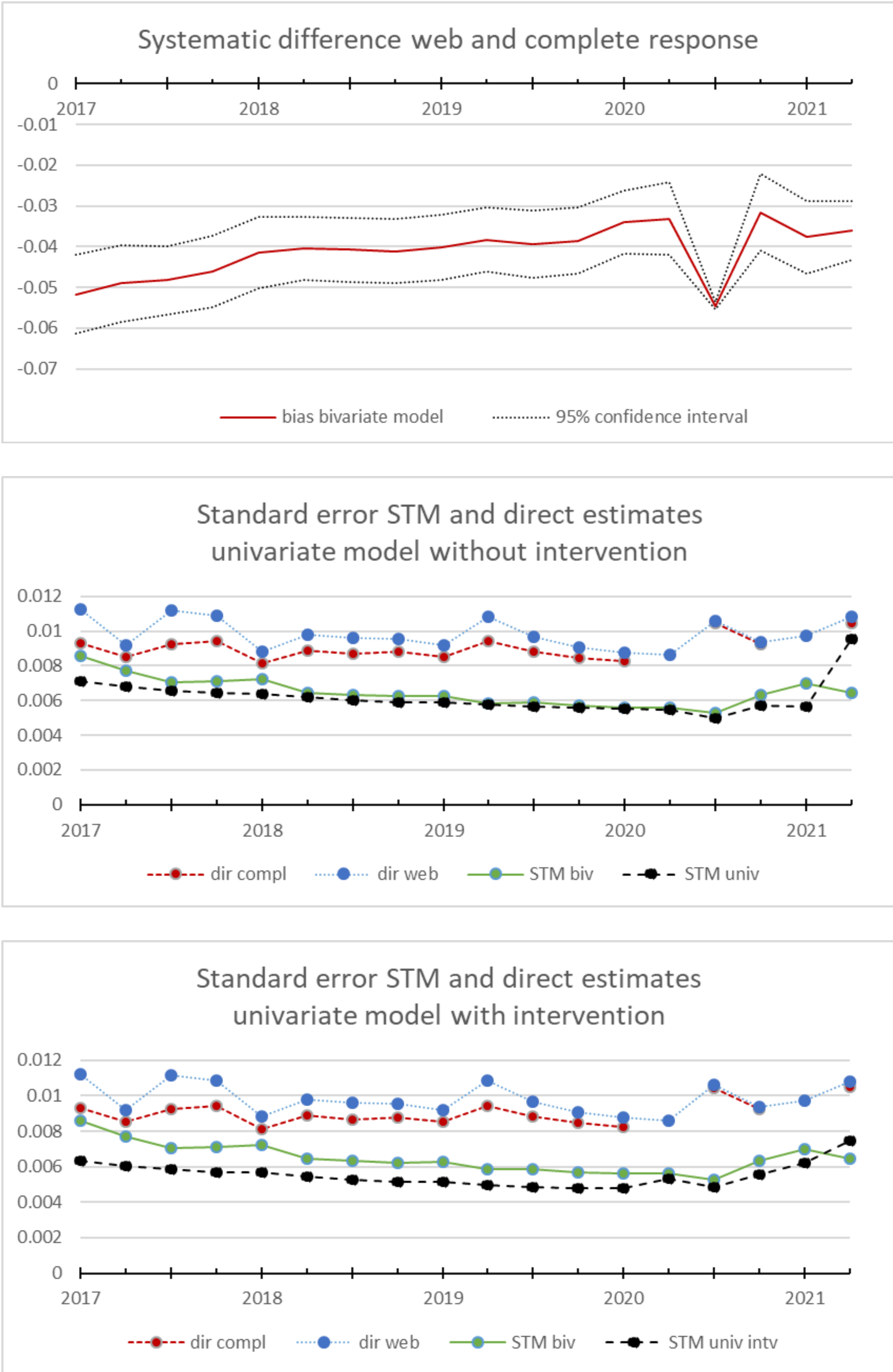
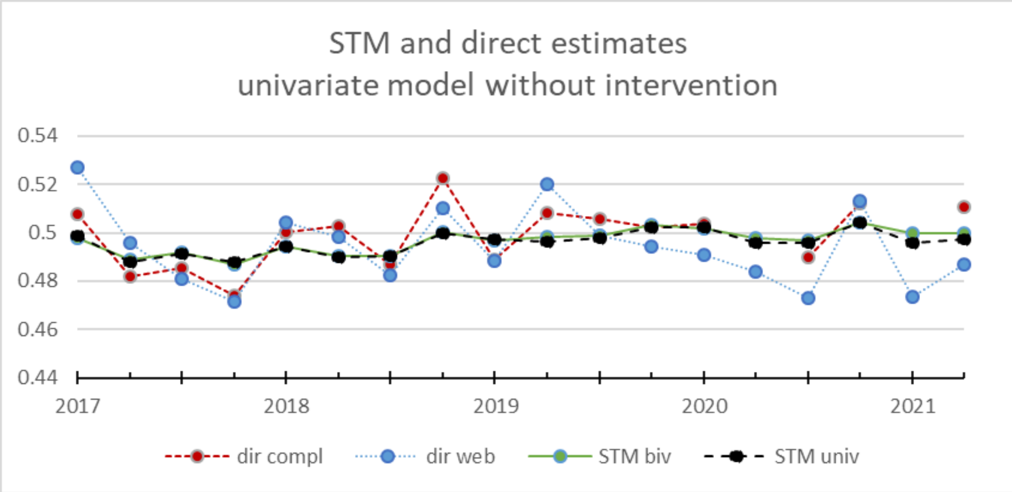
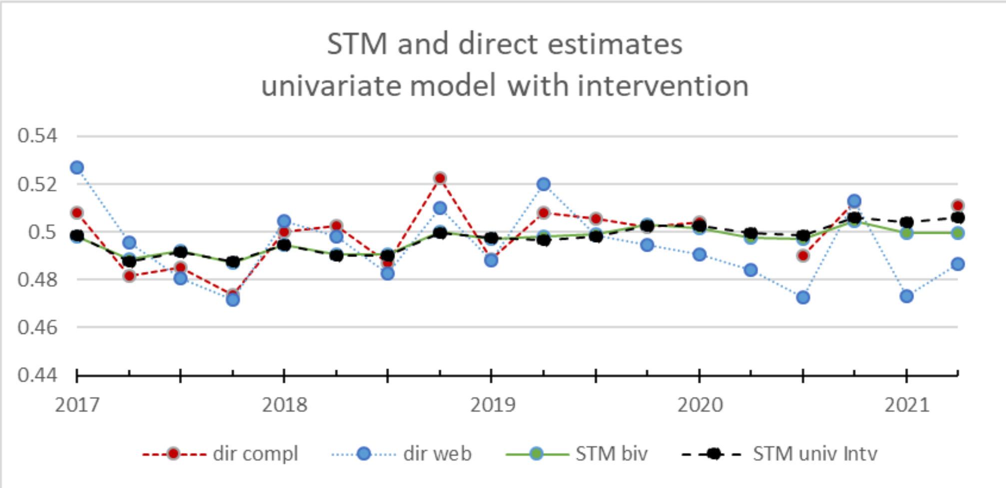


Figure B.12: results STM for daily smoking





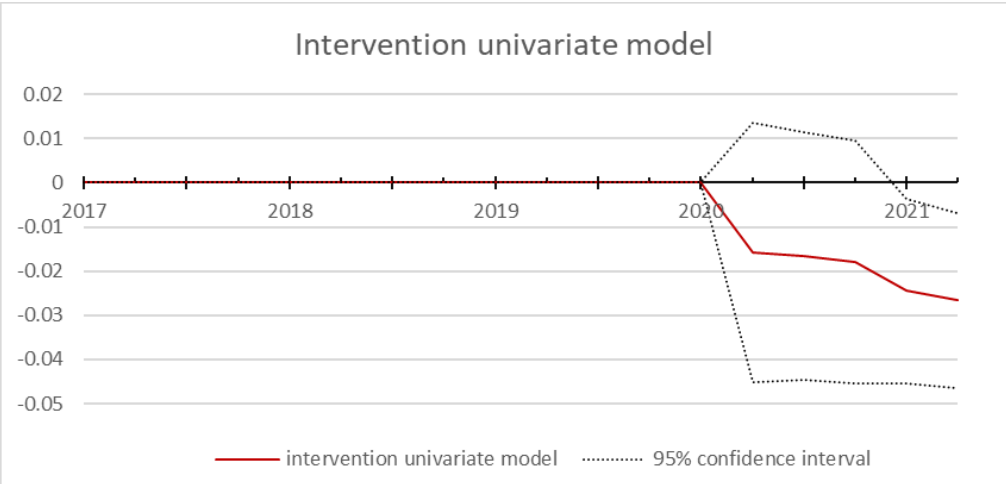
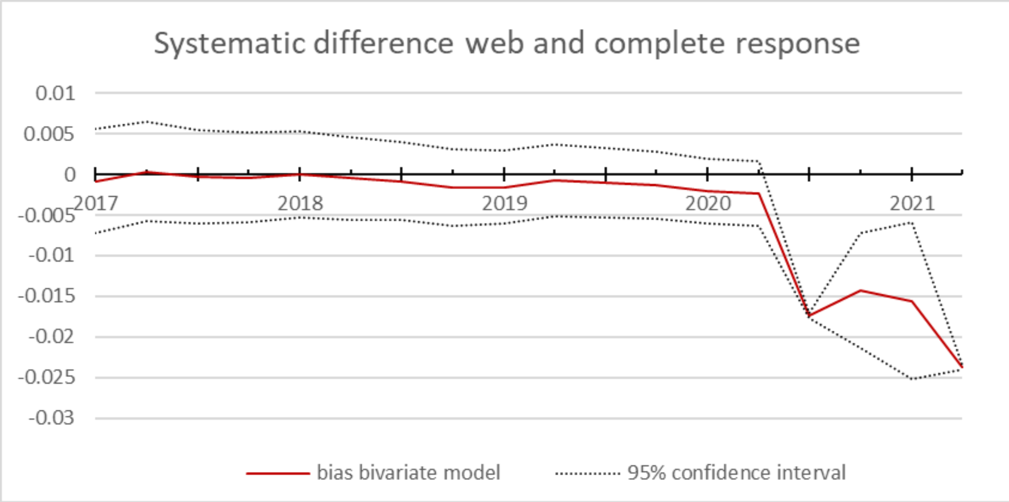
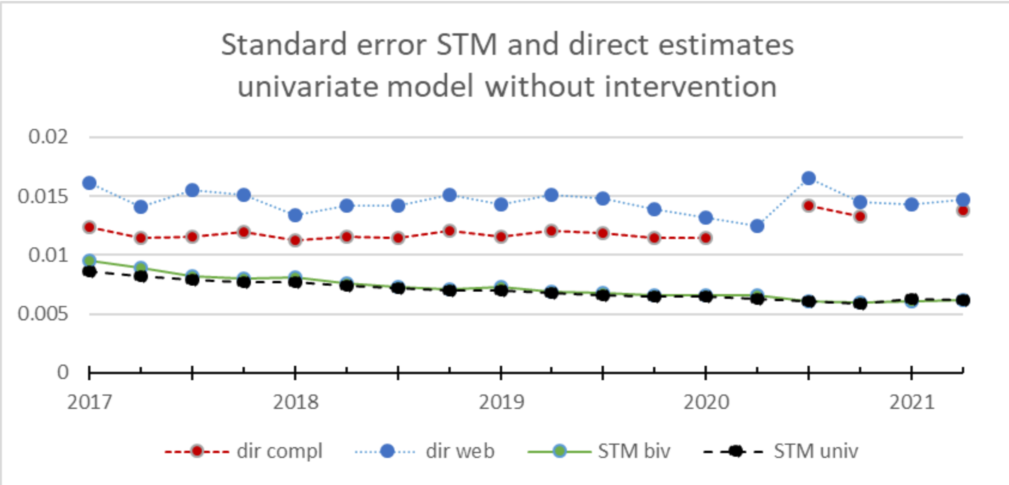


Figure B.13: results STM for overweight





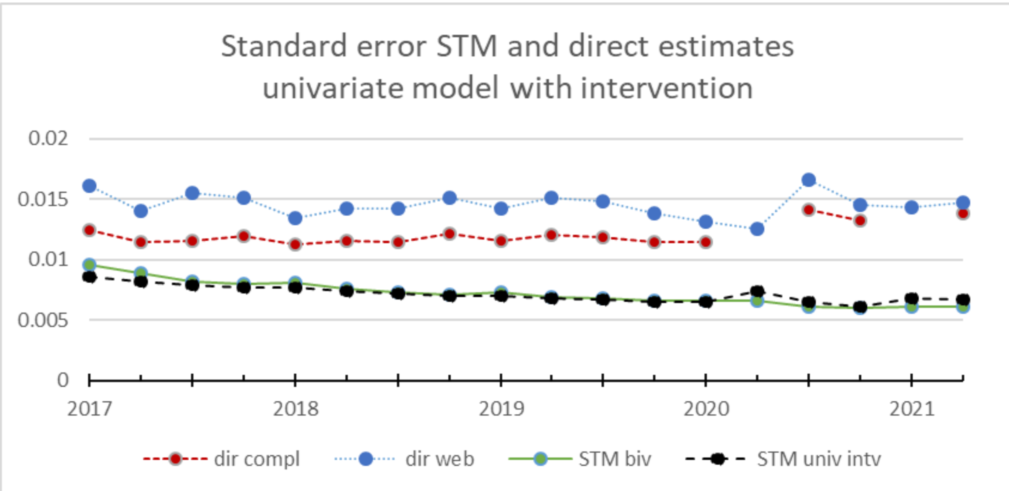


Figure B.14: results STM for overweight

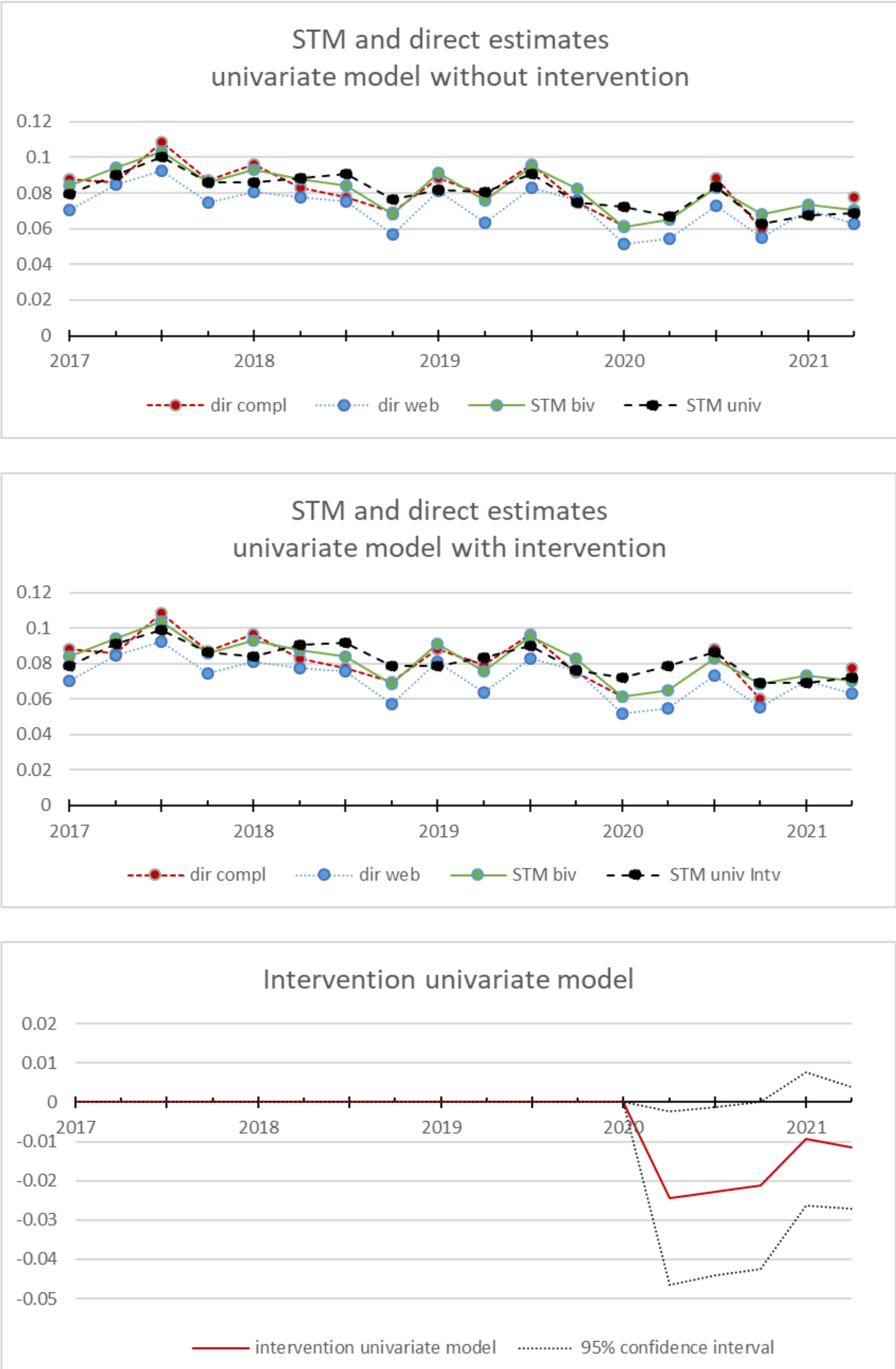
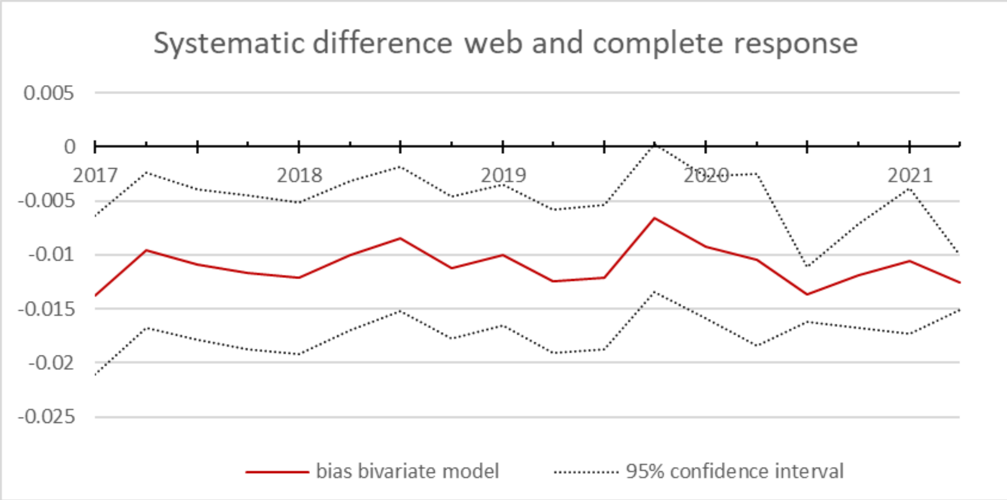
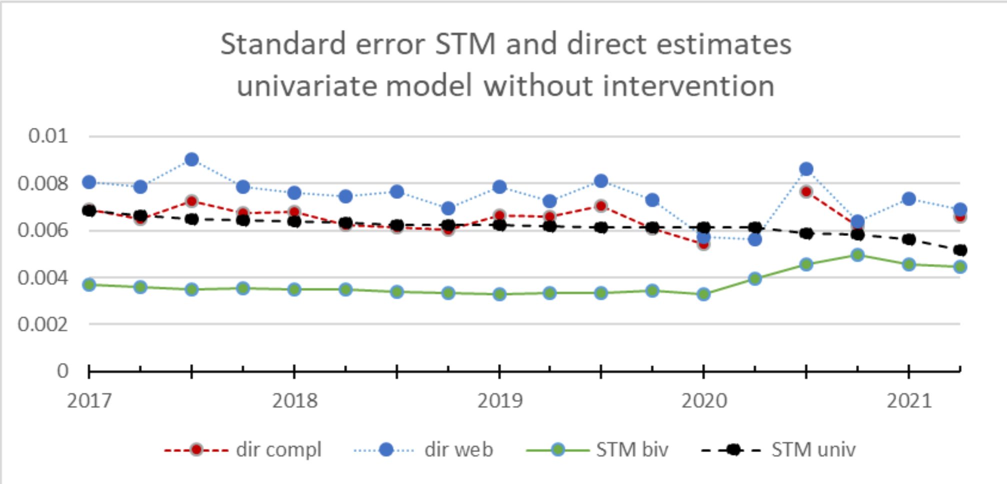


Figure B.15: results STM for excessive alcohol consumption





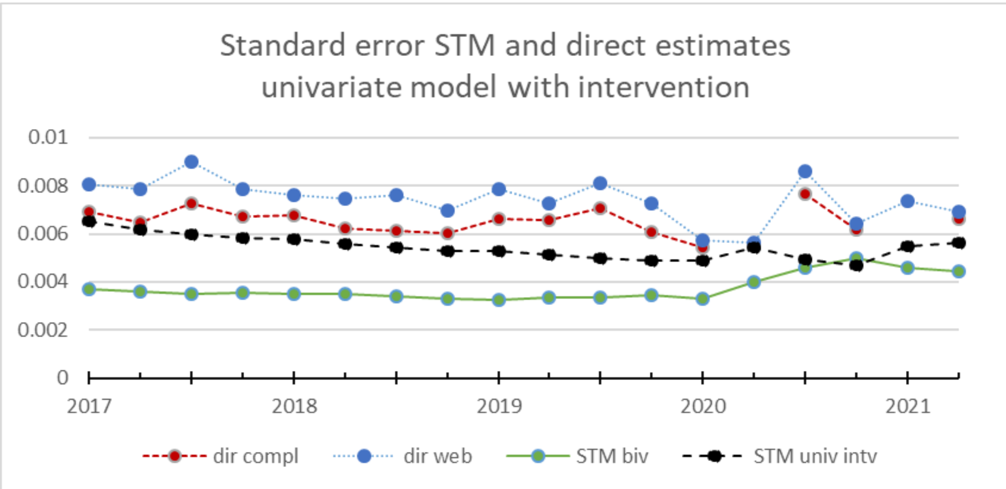


Figure B.16: results STM for excessive alcohol consumption

Explanation of symbols

Empty cell	Figure not applicable
.	Figure is unknown, insufficiently reliable or confidential
*	Provisional figure
**	Revised provisional figure
2017–2018	2017 to 2018 inclusive
2017/2018	Average for 2017 to 2018 inclusive
2017/’18	Crop year, financial year, school year, etc., beginning in 2017 and ending in 2018
2013/’14–2017/’18	Crop year, financial year, etc., 2015/’16 to 2017/’18 inclusive

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

Colophon

Publisher
Centraal Bureau voor de Statistiek
Henri Faasdreef 312, 2492 JP Den Haag
www.cbs.nl

Prepress
Statistics Netherlands, CCN Creation and visualisation

Design
Edenspiekermann

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