International Journal of Public Health (2020) 65:5–16
https://doi.org/10.1007/s00038-019-01327-9

ORIGINAL ARTICLE

Comparing web-based versus face-to-face and paper-and-pencil questionnaire data collected through two Belgian health surveys

Elise Braekman[1,2] · Rana Charafeddine[1] · Stefaan Demarest[1] · Sabine Drieskens[1] · Finaba Berete[1] · Lydia Gisle[1] · Johan Van der Heyden[1] · Guido Van Hal[2]

Received: 3 June 2019 / Revised: 10 December 2019 / Accepted: 30 December 2019 / Published online: 29 January 2020

Abstract

Objectives Using the European Health Interview Survey (EHIS) questionnaire, a web-based survey was organized alongside a face-to-face (F2F) survey including a paper-and-pencil (P&P) questionnaire for sensitive topics. Associated with these different modes, other design features varied too (e.g., recruitment, incentives, sampling). We assessed whether these whole data collection systems developed around the modes produced equivalent health estimates.

Methods Data were obtained from two population-based surveys: the EHISWEB (web-administered, n = 1010) and the Belgian Health Interview Survey 2018 (BHIS2018) (interviewer-administered, n = 2748). Logistic regression analyses were used to assess mode system differences while adjusting for socio-demographic differences in the net samples.

Results For the P&P mode of the BHIS, significant mode system differences were detected for 2 of the 9 health indicators. Among the indicators collected via the F2F mode, 9 of the 18 indicators showed significant differences.

Conclusions Indicators collected via the web-based and P&P self-administered modes were generally more comparable than indicators collected via the web-based and F2F mode. Furthermore, fewer differences were detected for indicators based on simple and factual questions compared to indicators based on subjective or complex questions.

Keywords Health surveys · Data collection · Web-based surveys · Face-to-face surveys · Mode systems · Data comparability

Introduction

General population health surveys serve as an important data source for monitoring population health and policy-making. The European Health Interview Survey (EHIS) takes on this role by periodically providing harmonized statistics on the health status, health care use and health determinants of the European Union and European Economic Area populations (Eurostat 2018; Santourian and Kitromilidou 2018). Two EHIS data collection waves have already been implemented, and the third wave was organized in 2019 (European Union 2018). The data collection is organized in each member state by conducting a stand-alone EHIS or integrating the EHIS within another survey. The member states are free to choose the data collection mode, but since the EHIS model questionnaire is long and contains complex questions, face-to-face (F2F) interviewing is recommended (Eurostat 2018; Santourian and Kitromilidou 2018). During the second EHIS wave, organized between 2013 and 2015, F2F interviewing with a paper-and-pencil (P&P) self-administered questionnaire for the sensitive questions was still a commonly used approach (Santourian and Kitromilidou 2018). However, for logistical and financial reasons, member states increasingly started using mixed-mode designs, including a web-based mode.

The comparability of data collected across different member states and at different time points is crucial for adequate policymaking (Verschuuren et al. 2013). Nonetheless, if different modes are used, the comparability of health estimates collected through these modes should

✉ Elise Braekman
elise.braekman@sciensano.be

[1] Department Epidemiology and Public Health, Sciensano, Juliette Wytsmanstraat 14, 1050 Brussels, Belgium

[2] Unit of Epidemiology and Social Medicine, University of Antwerp, Universiteitsplein 1, 2610 Antwerp, Belgium

be assessed. Indeed, the mode characteristics may affect how the respondent understands the question, retrieves relevant information, uses it to decide on an adequate response and finally chooses the answer (Tourangeau et al. 2000).

A web-based mode and a F2F mode (administered via computer-assisted personal interviewing—CAPI) differ in several ways. Firstly, in a web-based mode the questions and answer categories are presented visually, whereas in a F2F mode they are posed orally (except for questions and answer categories presented via show cards) (Tourangeau et al. 2000). The presentation may affect the order in which respondents consider answer options. Visual modes are more prone to primacy effects (selection of the first reasonable answer category in a list), whereas auditory modes are more susceptible to recency effects (selection of the last reasonable answer category) (Krosnick 1991; Schwarz and Hippler 1991). Auditory modes eliminate the need for literacy but in case of long and complex questions, they may overtax the respondents' listening and comprehension ability (Tourangeau et al. 2000). Visual modes may, however, impose a lower cognitive burden as respondents can consider the questions, instructions and answer categories more thoroughly (Tourangeau et al. 2013). Furthermore, web respondents can reread the questions at their own pace.

Secondly, a web-based mode is self-administered, whereas a F2F mode is interviewer-administered. The presence of an interviewer may result in higher levels of social desirability bias, e.g., an underreporting of undesirable health behavior and sensitive health problems (Bowling 2005; Burkill et al. 2016). Furthermore, interviewers can distract respondents and vary in the way they pose questions. Interviewer characteristics (gender, age, ethnicity, etc.) may also affect the responses (Tourangeau et al. 2000). On the other hand, interviewers may improve the data quality since they can clarify complex questions, probe for responses and keep respondents motivated through long questionnaires (Bowling 2005; Tourangeau et al. 2000).

Web-based and P&P modes do not differ in terms of these characteristics (both visual and self-administered). Nevertheless, in contrast to a P&P questionnaire, a web-based questionnaire is computer-assisted and can present the advantage of automatic data entry, integrated warning messages in case of missing, inconsistent and out-of-range answers, and automatic branching logic. Furthermore, a web-based mode can generate more honest responses than a P&P mode regarding (highly) sensitive topics (Gnambs and Kaspar 2015). Web respondents can be transported into a virtual world, wherein they forget their immediate surrounding and therefore it can create an illusion of privacy (Gnambs and Kaspar 2015, Weisband and Kiesler 1996).

Previous mode comparison studies in the public health domain found evidence that differences are less likely to occur between different types of self-administered modes than between self- and interviewer-administered modes (Braekman et al. 2018; Hoebel et al. 2014). There is evidence that some types of health indicators are less prone to mode differences than others. Indicators based on factual questions are less affected by the mode than indicators based on questions requiring a subjective self-assessment (Braekman et al. 2018; Christensen et al. 2014; Feveile et al. 2007; Hoebel et al. 2014; Kralj et al. 2015; Tagseth et al. 2019). A factual question deals with objective issues, e.g., the presence of a disease or chronic condition, the use of medicines and health services or the occurrence of accidents and injuries. A subjective question requires a judgment of the own personal feelings when responding. Questions dealing with the level of pain experienced, self-rated general, mental and social health or socially undesirable health topics, such as substance abuse, could be examples of this. A distinction can further be made between easy-to-answer questions, i.e., questions with simple answer categories and more cognitively demanding questions that require attention to detail. The latter type of questions is more likely to differ between an interviewer-administered F2F mode and a self-administered P&P mode (Tipping et al. 2010).

In practice, a switch from a F2F EHIS, including a P&P self-administered questionnaire for sensitive topics, to a web-based EHIS, is more than a mode change as other features may also vary (sampling, recruitment strategy, incentives, questionnaire content, etc.) (Biemer and Lyberg 2003). Consequently, studies comparing whole systems of data collection developed around different modes—"mode system effects" instead of "pure mode effects"—are of particular interest for survey practitioners (Biemer and Lyberg 2003; Struminskaya et al. 2015). Our objective is to assess the differences in health estimates collected via two different mode systems; one developed for a web-based survey versus one developed around a F2F survey including a P&P self-administered questionnaire. We expect health indicators collected via the web-based versus P&P mode to yield more comparable results than indicators collected via the web-based versus F2F mode. Furthermore, fewer differences are expected for health indicators based upon simple factual questions than for indicators requiring a subjective assessment, or for those based on complex questions.

Methods

Study design

Two parallel cross-sectional population studies were organized: the pilot European Health Interview Survey by web (EHISWEB), a web-based survey, and the Belgian Health Interview Survey 2018 (BHIS2018), a F2F survey including a P&P self-administered questionnaire.

Pilot European Health Interview Survey by web

The data collection took place from April to November 2018, with a break during July and August. Individuals aged 16 to 85 years were selected from the National Register according to a multistage clustered sampling procedure. People belonging to collective or institutionalized households and those living in East Belgium (German speaking community, $< 1\%$ of the population) were excluded from the sampling frame. In a first step, a regional and provincial stratification was performed. In each stratum, municipalities were selected according to a "probability proportional to size" sampling. Next, a systematic sampling was applied to select individuals within municipalities. Sample substitution was applied during data collection: non-participating individuals were replaced, if necessary several times, by similar individuals matched on statistical sector (i.e., a subdivision of municipality), sex and age. The target sample size was 1000 individuals.

All selected individuals received a postal invitation letter including an URL, an ID-code and a password to access the questionnaire. Information about the purpose, content and voluntary character of the study, the incentive and the access period was also covered in this letter. The conditional incentive was a €10—gift voucher that could be spent in different types of stores. A reminder letter was sent to non-respondents 7 days after the invitation.

The EHISWEB questionnaire covered all the EHIS wave 3 variables requested by the European Statistical Office in three public health domains, i.e., health status, health care use and health determinants (Eurostat 2018). The survey addresses 21 health-related topics (i.e., EHIS sub-modules) with a coherent set of one or more questions within these domains. More details about the EHISWEB study design are described elsewhere (Braekman et al. 2019).

Belgian Health Interview Survey 2018

The BHIS is a household survey organized periodically since 1997 and designed to be representative for the Belgian population (Demarest et al. 2013). It supports the variables requested in the framework of the various EHIS waves. BHIS2018 data were used in this comparison study.

The sampling strategies applied in both studies were comparable, with the exception that the BHIS2018 is a household survey (max. 4 individuals per household were interviewed) with households selected according to a multistage clustered sampling procedure. A household instead of an individual level sampling is routinely applied in F2F surveys for logistical and financial reasons. Non-participating households were substituted with similar households matched on statistical sector, household size and age of the reference person. In contrast to EHISWEB, people living in elderly care facilities and in East Belgium were included in the sampling frame and there was no age limit. The target sample size was 11,300 individuals.

Selected households received an advance letter and were later contacted by an interviewer. No incentive for participation was foreseen. Proxy interviewing was conducted for selected individuals younger than 15 or for those not capable of responding themselves.

The BHIS2018 extended the set of EHIS variables with additional questions for national purposes and was therefore longer than the EHISWEB questionnaire. A F2F mode has greater potential in terms of questionnaire length than a web-based mode (De Leeuw 2008). The bulk of the questions were administered via CAPI, but a P&P self-administered questionnaire including sensitive topics also had to be completed during the interview session. Details concerning the BHIS2018 methodology can be found elsewhere (Demarest et al. 2019).

Both studies were based on mutually exclusive samples. In order to make the net samples of the BHIS2018 and EHISWEB comparable, only the BHIS data of respondents aged 16–85 years, not living in an institutionalized environment or in East Belgium and not having a proxy interview, were considered. Furthermore, to coincide with the EHISWEB data collection period, only respondents who completed the BHIS2018 between April and November 2018 (without the holiday months) were kept. Based on the region, sex and age distribution of the BHIS2018 net sample, post-stratification weights were calculated and assigned to the EHISWEB participants in order to compensate for the differential composition of the net samples of the two studies. For both studies, an authorization was received from the Ethics Committee of the University Hospital of Ghent and from the Sectoral Committee of the National Register.

Health outcomes

This study compared the outcomes of binary health indicators that are based on questions with a similar wording in both studies. One or more indicators are presented for each

EHIS sub-module. They were selected according to the following criteria: (1) indicators calculated using all or multiple questions from the sub-module; (2) (if possible) indicators applicable to all the respondents and not based upon follow-up questions depending on the conditional branching; (3) indicators with a prevalence of at least 5.0%; and (4) if different questions from the same sub-module could produce multiple indicators, these indicator outcomes were compared in terms of mode system differences. In this case, we present indicator(s) with an outcome in line with most other outcomes from the sub-module. Table 1 provides an overview of the selected indicators.

Statistical analyses

The (unweighted) socio-demographic characteristics of the EHISWEB net sample were compared with those of the BHIS2018 net sample.

We calculated differences between the weighted prevalence rates of the EHISWEB and the BHIS2018 study by subtracting the EHISWEB estimate from the BHIS2018 estimate for each health indicator. Corresponding confidence intervals (CI) were assessed using the Delta method (Oehlert 1992). Logistic regression analyses were conducted to assess possible mode system differences (crude analyses). Subsequently, logistic regression analyses adjusting for sex, age, region, marital status, household size, educational attainment and country of birth were conducted to observe possible mode system differences when accounting for the differential composition in the net samples (adjusted analyses). All analyses took into account the multistage stratified sampling design of both surveys by using specific survey analysis techniques offered by our software package that included a "weights," "strata" and "cluster" statement. Statistical analyses were performed using SAS® Enterprise Guide 7.1 and Stata/SE® 13.

Results

Net sample characteristics

Compared to the BHIS2018 respondents, EHISWEB respondents were more likely to be younger, live in the Brussels Capital Region and be more educated (Table 2). They were less likely to be divorced, separated or widowed, live alone or be born outside Europe.

Data comparability

Health status

Both the crude and adjusted analyses showed no significant differences between three chronic health conditions collected via web versus F2F: high cholesterol level, low back disorder and asthma (Table 3). In addition, no significant differences were detected for restrictions in performing household activities. However, EHISWEB respondents were more likely to report a leisure activity injury, absence from work and having trouble remembering and concentrating. They reported suffering from bodily pain during the past month less often than BHIS2018 respondents. These four differences found in the crude analyses remained statistically significant after adjustment for socio-demographic characteristics. Higher levels of restrictions in performing daily activities were reported in EHISWEB compared to BHIS2018 respondents, but this difference only became significant after adjusting for socio-demographic characteristics.

Good self-rated health was reported more often among EHISWEB than BHIS2018 P&P respondents, but this difference was no longer significant after adjusting for socio-demographic characteristics. No significant differences were detected between the web and P&P mode system for experiencing a chronic health problem, daily activity limitations or a depressive disorder.

Health care

The crude and adjusted analyses showed no significant differences by mode system for three out of five health care indicators: "contact with a specialist," "blood pressure measured within the past 5 years" and "postponement of medical consumption" (Table 4). EHISWEB respondents were more likely to report an inpatient hospitalization in the past 12 months compared to BHIS2018 respondents, but this difference only became (borderline) significant in the adjusted analysis. Both the crude and adjusted analyses indicated that significantly more EHISWEB than BHIS2018 respondents reported using non-prescribed medicines in the past 2 weeks.

Health determinants

For the indicators collected via web versus F2F, the crude analysis showed lower rates of obesity among EHISWEB respondents (Table 5). After adjustment for socio-demographic characteristics, this difference was no longer significant. In addition, no significant difference was found for the indicator "eating at least 2 portions of fruit daily." According to the crude and adjusted analyses, EHISWEB

Table 1 Overview of the selected health indicators (Pilot European Health Interview Survey by web, Belgian Health Interview Survey 2018, Belgium 2018)

Sub-module	Indicator	# items/ questions	Instrument[a]	Comparison[b]
European health status module				
Health status	Good self-rated health	1	MEHM	Web vs. P&P
	Chronic health problem	1	MEHM	Web vs. P&P
	Daily activity limitations	1	MEHM	Web vs. P&P
Diseases and chronic health conditions	High cholesterol level in the past 12 months	1		Web vs. F2F
	Low back disorder in the past 12 months	1		Web vs. F2F
	Asthma in the past 12 months	1		Web vs. F2F
Accidents and injuries	Leisure accident resulting in injury in the past 12 months	1		Web vs. F2F
Absence from work (due to health problems)	Absence from work due to a health problem in the past 12 months[c]	2		Web vs. F2F
Functional limitations	Difficulties in remembering and concentrating	1	BI-M2	Web vs. F2F
Personal care activities	Restrictions in performing daily activities in people aged 55 and older[d,e]	6	ADL	Web vs. F2F
Household activities	Restrictions in performing household activities in people aged 55 and older[d]	7	IADL	Web vs. F2F
Pain	Moderate to very severe bodily pain in the past 4 weeks[d]	1	SF-36v2®	Web vs. F2F
Mental Health	Depressive disorder[f]	9	PHQ-9	Web vs. P&P
European health care module				
Use of inpatient and day care	Inpatient hospitalizations in the past 12 months	2		Web vs. F2F
Use of ambulatory and home care	Contact with a specialist in the past 12 months	1		Web vs. F2F
Medicine use	Usage of non-prescribed medicines in the past 2 weeks	1		Web vs. F2F
Preventive services	Blood pressure measured within the past 5 years	2		Web vs. F2F
Unmet needs for health care	Postponement of medical consumption due to financial reasons[g]	5		Web vs. F2F
European health determinants module				
Weight and height	Obesity (body mass index $\geq$ 30) in adults (people aged 18 and older)	3		Web vs. F2F
Physical activity/exercise	Being sufficiently physically active in total[d]	6	EHIS-PAQ	Web vs. F2F
Dietary habits	Eating at least 2 portions of fruit daily[d]	2		Web vs. F2F
Smoking	Current smokers	3		Web vs. P&P
	Ever daily smokers	4		Web vs. P&P
Alcohol consumption	Daily drinkers	1		Web vs. P&P
	Weekly risky single occasion alcohol drinking (6 + drinks/occasion)	2		Web vs. P&P
Social Support	Poor social support	3	OSS-3	Web vs. P&P
Provision of informal care or assistance	Providing informal care	1		Web vs. F2F

[a]MEHM = Minimum European Health Module (Cox et al. 2009), BI-M2 = Budapest Initiative-Mark 2 (Washington group 2012), ADL = Activities of Daily Living, IADL = Instrumental Activities of Daily Living, SF-36v2® = 36-item Short Form Health Questionnaire, 2nd edition (domain: Bodily pain) (Ware and Sherbourne 1992), PHQ-9 = Patient Health Questionnaire, 9-item depression screener (Kroenke et al. 2001), EHIS-PAQ = the European Health Interview Survey-Physical Activity Questionnaire (Finger et al. 2015), OSS-3 = Oslo 3-items Social Support Scale (Meltzer 2003)

[b]The EHISWEB study included a web-based mode, and the BHIS2018 included a face-to-face (F2F) mode with a supplementary paper-and-pencil (P&P) questionnaire

[c]This indicator was only calculated for currently employed respondents

[d]For items/questions of these indicators, BHIS2018 interviewers had to refer to show cards. These cards included the response categories, which respondents had to read before choosing the appropriate category

[e]Eurostat requests including 5 personal care activities for which respondents may experience restrictions in performing them. In the EHISWEB and BHIS2018, one of these requested items "Getting in and out of a bed or chair" was split in 2 items: "Getting in and out of a bed" and "Getting in and out of a chair"

[f]Eurostat recommends the use of the PHQ-8 due to the high sensitivity of the 9th item. However, the PHQ-9 was used in the EHISWEB and BHIS2018

[g]Questions related to this indicator were asked to only one household member (reference person or partner) in the BHIS2018 (n = 1672). Eurostat requests including 4 types of health service that may not be affordable because of financial reasons. In the EHISWEB and BHIS2018, a fifth item was included "Eye glasses or contact lenses." Contrary to the recommendation of Eurostat, no answer category "No need for…" was foreseen

Table 2 Socio-demographic characteristics of the Pilot European Health Interview Survey by web and the Belgian Health Interview Survey 2018 study population (Belgium 2018)

	EHISWEB (n = 1010)		BHIS2018 (n = 2748)	
	n	% (95% CI)	n	% (95% CI)
Characteristics used for the EHISWEB sampling design				
Sex				
Male	501	49.6 (46.5–52.7)	1293	47.1 (45.2–48.9)
Female	509	50.4 (47.3–53.5)	1455	52.9 (51.1–54.8)
Age				
16–24	142	14.1 (11.9–16.2)	233	8.5 (7.4–9.5)
25–44	309	30.6 (27.8–33.4)	847	30.8 (29.1–32.6)
45–64	380	37.6 (34.6–40.6)	1041	37.9 (36.1–39.7)
65–85	179	17.7 (15.4–20.1)	627	22.8 (21.2–24.4)
Region				
Flemish	355	35.1 (32.2–38.1)	1091	39.7 (37.9–41.5)
Brussels Capital	330	32.7 (29.8–35.6)	746	27.1 (25.5–28.8)
Walloon	325	32.2 (29.3–35.1)	911	33.2 (31.4–34.9)
Latent characteristics				
Marital status				
Never married or in a registered partnership	308	30.5 (27.7–33.3)	766	27.9 (26.2–29.6)
Married or in registered partnership	573	56.7 (53.7–59.8)	1528	55.6 (53.7–57.5)
Divorced, separated or widowed	124	12.3 (10.3–14.3)	454	16.5 (15.1–17.9)
Missing	5	0.5 (0.1–0.9)	0	
Household size				
1	153	15.1 (12.9–17.4)	601	21.9 (20.3–23.4)
2	351	34.8 (31.8–37.7)	915	33.3 (31.5–35.1)
3	168	16.6 (14.3–18.9)	453	16.5 (15.1–17.9)
3+	332	32.9 (30.0–35.8)	779	28.3 (26.7–30.0)
Missing	6	0.6 (0.1–1.1)	0	
Educational attainment				
Low	194	19.2 (16.8–21.6)	684	24.9 (23.3–26.5)
Intermediate	318	31.5 (28.6–34.4)	910	33.1 (31.4–34.9)
High	458	45.3 (42.3–48.4)	1088	39.6 (37.8–41.4)
Missing	40	4.0 (2.8–5.2)	66	2.4 (1.8–3.0)
Country of birth				
Belgium	817	80.9 (78.5–83.3)	2094	76.2 (74.6–77.8)
European country	114	11.3 (9.3–13.2)	252	9.2 (8.1–10.2)
Non-European country	76	7.5 (5.9–9.2)	397	14.4 (13.1–15.8)
Missing	3	0.3 (0.0–0.6)	5	0.2 (0.0–0.3)

EHISWEB Pilot European Health Interview Survey by web, *BHIS2018* Belgian Health Interview Survey 2018, *CI* confidence interval

respondents were significantly more likely to be sufficiently physically active and to provide informal care compared to BHIS2018 respondents.

For the alcohol indicators ("daily drinkers" and "weekly risky single occasion alcohol drinking") and the smoking indicator "ever daily smoker," collected via web versus P&P, both the crude and adjusted analyses showed no significant differences. However, according to the crude and adjusted analyses, EHISWEB respondents were less likely to smoke currently. Higher levels of receiving poor social support were reported in EHISWEB compared to BHIS2018 respondents, but this difference only became significant in the adjusted analysis.

Discussion

Due to its logistical and financial advantages, web-based data collection is gaining popularity for large-scale health surveys. In this study, we assessed the comparability of

Table 3 Weighted prevalence rates of *health status* indicators and results from logistic regression analyses showing the association between mode system and health status indicators (Pilot European Health Interview Survey by web, Belgian Health Interview Survey 2018, Belgium 2018)

	Weighted prevalence			Crude OR (EHISWEB vs. BHIS2018)[b] (95% CI)	Adjusted OR (EHISWEB vs. BHIS2018)[c] (95% CI)
	EHISWEB %	BHIS2018 %	Diff.[a] % (95% CI)		
Web versus F2F					
High cholesterol level in the past 12 months	21.1	18.7	2.4 (− 0.7 to 5.5)	1.16 (0.96–1.41)	1.21 (0.98–1.49)
Low back disorder in the past 12 months	22.5	23.5	− 1.0 (− 4.2 to 2.2)	0.95 (0.79–1.13)	0.94 (0.78–1.14)
Asthma in the past 12 months	7.6	6.6	1.0 (− 1.0 to 2.9)	1.16 (0.87–1.55)	1.22 (0.91–1.64)
Leisure accident resulting in injury in the past 12 months	7.3	4.4	2.8 (1.0 to 4.6)	1.68 (1.24–2.28)	1.62 (1.17–2.24)
Absence from work due to a health problem in the past 12 months	52.9	44.4	8.5 (3.5 to 13.5)	1.41 (1.15–1.72)	1.43 (1.16–1.77)
Difficulties in remembering and concentrating	34.4	20.4	14.0 (10.5 to 17.5)	2.05 (1.73–2.42)	2.20 (1.84–2.63)
Restrictions in performing daily activities among people aged 55 and older	17.4	14.0	3.5 (− 1.0 to 7.9)	1.30 (0.94–1.80)	1.62 (1.14–2.32)
Restrictions in performing household activities among people aged 55 and older	26.2	27.4	− 1.2 (− 6.6 to 4.2)	0.94 (0.71–1.24)	1.19 (0.87–1.64)
Moderate to very severe bodily pain in the past 4 weeks	24.1	29.0	− 5.0 (− 8.3 to − 1.7)	0.77 (0.65–0.92)	0.80 (0.66–0.95)
Web versus P&P					
Good self-rated health	80.0	76.2	3.9 (0.7 to 7.0)	1.25 (1.04–1.52)	1.16 (0.94–1.42)
Chronic health problem	30.6	30.8	− 0.2 (− 3.8 to 3.3)	0.99 (0.84–1.17)	1.01 (0.85–1.21)
Daily activity limitations	20.8	23.6	− 2.8 (− 6.0 to 0.4)	0.85 (0.71–1.03)	0.90 (0.74–1.09)
Depressive disorder	15.0	16.4	− 1.5 (− 4.2 to 1.3)	0.90 (0.73–1.10)	0.91 (0.73–1.13)

All analyses were done taking into account the multistage stratified sampling design of both surveys (post-stratification weights, strata and clusters)

EHISWEB Pilot European Health Interview Survey by web, *BHIS2018* Belgian Health Interview Survey 2018, *OR* odds ratio, *CI* confidence interval, *F2F* face-to-face, *P&P* paper-and-pencil

[a]Diff. = difference in prevalence rates

[b]Result of a logistic regression analysis with only mode system as independent variable

[c]Result of a logistic regression analysis with mode system as independent variable and an adjustment for sex, age, region, marital status, household size, educational attainment and country of birth

health estimates collected by a mode system developed for a web-based EHIS versus by a mode system developed for a F2F EHIS including a P&P self-administered questionnaire for sensitive topics. For less than half of the presented health indicators, significant mode system differences were detected and indicators in the domains of health status, health care use and health determinants were equally prone to differences.

In line with the results of previous studies (Braekman et al. 2018; Hoebel et al. 2014), indicators collected via the two self-administered modes were less likely to differ than indicators collected via the self- and interviewer-administered mode. There were two exceptions: EHISWEB respondents reported lower perceived social support and lower rates of current smoking compared to BHIS2018 respondents. Other studies also found differences in the reporting of social support between self-administered modes (Hoebel et al. 2014; Tipping et al. 2010). Contrary to our findings, other studies usually found high mode equivalence for smoking status indicators (Braekman et al. 2018; Hoebel et al. 2014; Vergnaud et al. 2011). When considering all the analyzed smoking indicators, we found that indicators related to current smoking behavior ("current smoking" and "current daily smoking") were

Table 4 Weighted prevalence rates of *health care* indicators and results from logistic regression analyses showing the association between mode system and health care indicators (Pilot European Health Interview Survey by web, Belgian Health Interview Survey 2018, Belgium 2018)

	Weighted prevalence			Crude OR (EHISWEB vs. BHIS2018)[b] (95% CI)	Adjusted OR (EHISWEB vs. BHIS2018)[c] (95% CI)
	EHISWEB %	BHIS2018 %	Diff.[a] % (95% CI)		
Web versus F2F					
Inpatient hospitalizations in the past 12 months	12.1	10.1	2.0 (− 0.4 to 4.4)	1.22 (0.97–1.55)	1.28 (1.00–1.63)
Contact with a specialist in the past 12 months	63.1	60.2	2.9 (− 0.7 to 6.5)	1.13 (0.97–1.32)	1.11 (0.94–1.30)
Usage of non-prescribed medicines in the past 2 weeks	44.6	29.0	15.6 (11.9 to 19.3)	1.97 (1.68–2.30)	1.88 (1.60–2.22)
Blood pressure measured within the past 5 years	93.8	93.2	0.6 (− 1.2 to 2.3)	1.10 (0.82–1.47)	1.03 (0.74–1.42)
Postponement of medical consumption due to financial reasons in the past 12 months	13.4	12.4	1.0 (− 1.6 to 3.7)	1.10 (0.87–1.38)	1.19 (0.92–1.55)

All analyses were done taking into account the multistage stratified sampling design of both surveys (post-stratification weights, strata and clusters)

EHISWEB Pilot European Health Interview Survey by web, BHIS2018 Belgian Health Interview Survey 2018, *OR* odds ratio, *CI* confidence interval, *F2F* face-to-face

[a]Diff. = difference in prevalence rates

[b]Result of a logistic regression analysis with only mode system as independent variable

[c]Result of a logistic regression analysis with mode system as independent variable and an adjustment for sex, age, region, marital status, household size, educational attainment and country of birth

significantly higher among BHIS2018 compared to EHISWEB respondents. Nevertheless, indicators referring to ever smoking ("ever smoker," "ever daily smoker" and "ever tried electronic cigarettes") did not differ significantly.

For indicators collected via web versus F2F, three reasons could potentially explain why some health indicators are less prone to mode system differences. Firstly, health indicators derived from factual questions, such as those about diseases or health care use, are less likely to differ between self- and interviewer-administered modes (Christensen et al. 2014; Hoebel et al. 2014; Kralj et al. 2015; Tipping et al. 2010). In agreement with this reasoning, no differences were detected for the presence of a chronic disease or condition (i.e., high cholesterol level, low back disorder, obesity and asthma) and for two health care indicators (i.e., "contact with a specialist" and "blood pressure measurement"). Nevertheless, significant differences emerged for two other health care indicators: "inpatient hospitalizations" and "use of non-prescribed medicines" were significantly higher among EHISWEB respondents. Potential alternative reasons for these differences are; (1) EHISWEB respondents could reflect longer about the questions and the most adequate response, allowing them to better recall hospitalizations or use of non-prescribed medicines; (2) BHIS2018 respondents were more likely to give socially desirable answers due to the presence of an interviewer (e.g., using non-prescribed medicines could be regarded as socially undesirable by some respondents).

Secondly, according to Tipping et al. (2010) health indicators based on questions that are more cognitively demanding and require attention to detail are more likely to differ between a self-administered P&P mode and an interviewer-administered F2F mode. Some elements make questions more cognitively demanding: the length, the complexity of the sentences and the words (i.e., the reading level), the inclusion of abstract concepts and the request for qualified judgments (i.e., questions with a qualified time frame or a qualified definition) (Holbrook et al. 2006). A mode system difference was detected for the physical activity indicator derived from the EHIS-PAQ instrument, which contains detailed questions on work-related, transport-related and leisure-time physical activity (Finger et al. 2015). These questions use complex wording, contain abstract concepts (e.g., "a typical week") and include multiple qualified time frames (e.g., "at least 10 min").

The health status indicator "leisure accident resulting in an injury" and the health determinants indicator "providing informal care" are based upon factual questions. Nevertheless, the prevalence of these indicators was significantly higher among EHISWEB versus BHIS2018

Table 5 Weighted prevalence rates of *health determinants* indicators and results from logistic regression analyses showing the association between mode system and health determinants indicators (Pilot European Health Interview Survey by web, Belgian Health Interview Survey 2018, Belgium 2018)

	Weighted prevalence			Crude OR (EHISWEB vs. BHIS2018)[b] (95% CI)	Adjusted OR (EHISWEB vs. BHIS2018)[c] (95% CI)
	EHISWEB %	BHIS2018 %	Diff.[a] % (95% CI)		
Web versus F2F					
Obesity in adults	14.0	17.0	− 3.0 (− 5.7 to − 0.3)	0.80 (0.64–0.99)	0.84 (0.67–1.05)
Being sufficiently physically active in total	62.9	52.2	10.7 (7.0 to 14.4)	1.55 (1.33–1.82)	1.56 (1.32–1.83)
Eating at least 2 portions of fruit daily	36.8	34.4	2.4 (− 1.3 to 6.1)	1.11 (0.95–1.30)	1.08 (0.92–1.28)
Providing informal care	23.6	14.1	9.5 (6.4 to 12.6)	1.89 (1.55–2.29)	1.80 (1.47–2.21)
Web versus P&P					
Current smokers	15.4	19.9	− 4.5 (− 7.4 to − 1.6)	0.73 (0.60–0.90)	0.76 (0.62–0.95)
Ever daily smokers	41.2	38.2	3.0 (− 0.8 to 6.9)	1.14 (0.97–1.33)	1.13 (0.96–1.34)
Daily drinkers	12.9	10.8	2.1 (− 0.5 to 4.7)	1.22 (0.96–1.55)	1.16 (0.90–1.51)
Weekly risky single occasion alcohol drinking	9.4	8.3	1.0 (− 1.2 to 3.2)	1.14 (0.87–1.49)	1.19 (0.90–1.58)
Poor social support	20.2	17.4	2.8 (− 0.2 to 5.9)	1.21 (0.99–1.47)	1.26 (1.03–1.55)

All analyses were done taking into account the multistage stratified sampling design of both surveys (post-stratification weights, strata and clusters)

EHISWEB Pilot European Health Interview Survey by web, *BHIS2018* Belgian Health Interview Survey 2018, *OR* odds ratio, *CI* confidence interval, *F2F* face-to-face, *P&P* paper-and-pencil

[a]Diff. = difference in prevalence rates

[b]Result of a logistic regression analysis with only mode system as independent variable

[c]Result of a logistic regression analysis with mode system as independent variable and an adjustment for sex, age, region, marital status, household size, educational attainment and country of birth

respondents. One possible explanation for this mode system difference could be that these indicators are also created using rather complex questions that contain some abstract concepts, e.g., defining a "leisure accident," an "injury," providing "care" and "help." When posing complex questions via a F2F interview, interviewers could give more explanation, clarify abstract concepts or emphasize the reference periods. This is not an option in web-based surveys, but web respondents do have the ability to reread the questions and instructions at their own pace and to think longer about the most adequate response.

Thirdly, there is evidence that indicators from questions assessing subjective information, such as general health, mental health or socially undesirable topics, have a higher chance of differing between interviewer- and self-administered modes (Bowling 2005; Christensen et al. 2014; Feveile et al. 2007; Hoebel et al. 2014; Kralj et al. 2015; Tipping et al. 2010). This might explain why a difference was found for the health status indicators "experiencing bodily pain," "difficulties in remembering and concentrating" and "restrictions in performing daily activities," which involve a self-assessment of subjective information. Furthermore, underreporting "being absent from work," "having memory problems," "facing problems while conducting basic daily activities such as dressing, washing and feeding yourself" or "using non-prescribed medicines" may be related to a social desirability bias, as people may be reluctant to report this to an interviewer. Nevertheless, there was certainly no exhaustive evidence of this increased social desirability bias found when using a F2F mode as no significantly higher rates of indicators typically sensitive to socially desirable answering [e.g., performing physical activity (Adams et al. 2005) or experiencing financial troubles paying for health care (Keeter et al. 2015)] were reported in the EHISWEB study.

Although this study found evidence for some mode system differences, future EHIS data are expected to be increasingly collected via web-based questionnaires due to the cost advantage. Therefore, we advise taking actions to minimize these differences. One strategy could be the development of an EHIS model questionnaire specifically designed for a web-based mode. The questions and

instructions could be altered in order to give the same perceived stimuli to respondents as questions posed through F2F interviewing (De Leeuw et al. 2008). This could be done by adding comprehensive definitions for abstract concepts such as "providing care" in the web-based mode. Furthermore, motivational phrases and feedback mechanisms that resemble an interviewer's interaction could be integrated. Showing warning messages if respondents give inconsistent answers or respond too quickly are examples of this. An extensive pre-testing of this model questionnaire using different devices (computers, tablets and smartphones) should be done in order to assess its equivalence with other modes. How mode system differences can be accounted for during analysis should also be thoroughly assessed.

Strengths and limitations

The main strength of this study lies in the assessment of mode system differences for a wide arrange of health indicators, which proves usefulness within the broader European context. The trend toward more web-based data collection is expected to continue in future EHIS data collection.

Using a different data collection mode was linked to applying a different sampling and recruitment strategy and caused a differential non-response in the EHISWEB and BHIS2018 studies. This led to differences in the socio-demographic composition of the net samples: EHISWEB respondents were generally younger, more often living in the Brussels Capital region and more highly educated than BHIS2018 respondents. As we aimed to assess mode system differences in a systematic way, post-stratification weights were applied and a statistical adjustment for socio-demographic characteristics was conducted in order to control for the differences in the socio-demographic compositions of the net samples. Nevertheless, it is still possible that some of the health outcomes are impacted by composition effects that could not be accounted for. Schnell et al. (2017) found, for example, that internet users have a better subjective health status than internet non-users and that weighting for socio-demographic characteristics does not eliminate this observed health difference. Moreover, in our study the adjusted odds ratios that take into account the socio-demographic characteristics of the net samples hardly differ from the crude odds ratios, which indicates that socio-demographic differences in the net sample compositions have a poor impact on the differences between the health estimates from the EHISWEB and the BHIS2018 data sources.

The BHIS2018 survey included more questions than the EHISWEB survey and this might also have influenced the outcomes. Furthermore, the context for completing the two self-administered modes (web and P&P) differed substantially (e.g., the presence of an interviewer) which probably also affected the responses.

Conclusion

Our findings suggest that the existence of mode system differences depends on the type of modes that are compared. Fewer differences were detected for health indicators collected via the web-based and P&P mode versus via the web-based and F2F mode. Furthermore, not all health indicators are equally prone to mode system differences. Indicators based on simple and factual questions, such as having an objective physical condition, are generally more comparable than indicators from questions that are rather complex or that contain abstract concepts. Finally, indicators derived from questions involving a certain amount of subjective assessment when responding are slightly more susceptible to mode system differences.

Acknowledgements We wish to thank the EHISWEB and BHIS2018 study participants.

Funding This research project was funded by the Belgian Science Policy (Belspo).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval All procedures performed in these studies were in accordance with the ethical standards of the Ethics Committee of the University Hospital of Ghent.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

Adams SA, Matthews CE, Ebbeling CB et al (2005) The effect of social desirability and social approval on self-reports of physical activity. Am J Epidemiol 161:389–398. https://doi.org/10.1093/aje/kwi054

Biemer PP, Lyberg LE (2003) Introduction to survey quality. Wiley, Hoboken

Bowling A (2005) Mode of questionnaire administration can have serious effects on data quality. J Public Health 27:281–291. https://doi.org/10.1093/pubmed/fdi031

Braekman E, Berete F, Charafeddine R et al (2018) Measurement agreement of the self-administered questionnaire of the Belgian Health Interview Survey: paper-and-pencil versus web-based mode. PLoS ONE 13:e0197434. https://doi.org/10.1371/journal.pone.0197434

Braekman E, Charafeddine R, Demarest S et al (2019) Is the European Health Interview Survey online yet? Response and net sample composition of a web-based data collection. Eur J Public Health. https://doi.org/10.1093/eurpub/ckz206

Burkill S, Copas A, Couper MP et al (2016) Using the web to collect data on sensitive behaviours: a study looking at mode effects on the British National Survey of Sexual Attitudes and Lifestyles. PLoS ONE 11:e0147983. https://doi.org/10.1371/journal.pone.0147983

Christensen AI, Ekholm O, Glümer C et al (2014) Effect of survey mode on response patterns: comparison of face-to-face and self-administered modes in health surveys. Eur J Public Health 24:327–332. https://doi.org/10.1093/eurpub/ckt067

Cox B, Van Oyen H, Cambois E et al (2009) The reliability of the minimum European health module. Int J Public Health 54:55–60. https://doi.org/10.1007/s00038-009-7104-y

De Leeuw ED (2008) Choosing the method of data collection. In: De Leeuw ED, Hox JJ, Dillman DA (eds) International handbook of survey methodology. Lawrence Erlbaum Associates, New York, pp 113–135

De Leeuw ED, Hox JJ, Dillman DA (2008) Mixed-mode surveys: when and why. In: De Leeuw ED, Hox JJ, Dillman DA (eds) International handbook of survey methodology. Lawrence Erlbaum Associates, New York, pp 299–316

Demarest S, Berete F, Charafeddine R, Van der Heyden J. Gezondheidsenquête 2018: Methodologie. Report number: D/2019/14.440/34. Sciensano, Brussels. https://www.gezondheidsenquete.be. Accessed 6 November 2019.

Demarest S, Van der Heyden J, Charafeddine R et al (2013) Methodological basics and evolution of the Belgian Health Interview Survey 1997–2008. Arch Public Health 71:24. https://doi.org/10.1186/0778-7367-71-24

European Union (2018) Commission Regulation (EU) 2018/255: Implementing Regulation (EC) No 1338/2008 of the European Parliament and of the Council as regards statistics based on the European Health Interview Survey (EHIS). Off J Eur Union 48/12:dd21-2-2018.

Eurostat (2018) European Health Interview Survey (EHIS wave 3) methodological manual. Publication Office of the European Union. https://ec.europa.eu/eurostat/documents/3859598/8762193/KS-02-18-240-EN-N.pdf/5fa53ed4-4367-41c4-b3f5-260ced9ff2f6. Accessed 30 May 2018

Feveile H, Olsen O, Hogh A (2007) A randomized trial of mailed questionnaires versus telephone interviews: response patterns in a survey. BMC Med Res Methodol 7:27. https://doi.org/10.1186/1471-2288-7-27

Finger JD, Tafforeau J, Gisle L et al (2015) Development of the European Health Interview Survey-Physical Activity Questionnaire (EHIS-PAQ) to monitor physical activity in the European Union. Arch Public Health 73:59. https://doi.org/10.1186/s13690-015-0110-z

Gnambs T, Kaspar K (2015) Disclosure of sensitive behaviors across self-administered survey modes: a meta-analysis. Behav Res Methods 47:1237–1259. https://doi.org/10.3758/s13428-014-0533-4

Hoebel J, von der Lippe E, Lange C et al (2014) Mode differences in a mixed-mode health interview survey among adults. Arch Public Health 72:46. https://doi.org/10.1186/2049-3258-72-46

Holbrook A, Cho YI, Johnson T (2006) The impact of question and respondent characteristics on comprehension and mapping difficulties. Public Opin Q 70:565–595. https://doi.org/10.1093/poq/nfl027

Keeter S, McGeeney K, Igielnik R et al (2015) From telephone to the web: the challenge of mode of interview effects in public opinion polls. Pew Research Center. https://www.pewresearch.org/wp-content/uploads/2018/08/2015-05-13_mode-study_REPORT.pdf. Accessed 30 May 2019

Kralj M, Zaletel M, Lavtar D et al (2015) Do respondents answer differently in web survey than in face-to-face interview: field work experiment from the European Health Interview Survey (EHIS). https://www.ine.es/q2016/docs/q2016Final00137.pdf. Accessed 30 May 2019

Kroenke K, Spitzer RL, Williams JB (2001) The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 16:606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Krosnick JA (1991) Response strategies for coping with the cognitive demands of attitude measures in surveys. Appl Cogn Psychol 5:213–236. https://doi.org/10.1002/acp.2350050305

Meltzer H (2003) Development of a common instrument for mental health. In: Nosikov A, Gudex C (eds) EUROHIS: developing common instruments for health surveys. IOS Press, Amsterdam, pp 35–60

Oehlert GW (1992) A note on the Delta method. Am Stat 46:27–29. https://doi.org/10.1080/00031305.1992.10475842

Santourian A, Kitromilidou S (2018) Quality report of the second wave of the European Health Interview Survey. Publication Office of the European Union. https://ec.europa.eu/eurostat/documents/7870049/8920155/KS-FT-18-003-EN-N.pdf/eb85522d-bd6d-460d-b830-4b2b49ac9b03. Accessed 30 May 2019

Schnell R, Noack M, Torregroza S (2017) Differences in general health of internet users and non-users and implications for the use of web surveys. Surv Res Methods 11:105–123. https://doi.org/10.18148/srm/2017.v11i2.6803

Schwarz N, Hippler H-J (1991) Response alternatives: the impact of their choice and presentation order. In: Biemer PP, Groves RM, Lyberg LE, Mathiowetz NA, Sudman S (eds) Measurement errors in surveys. Wiley, New York, pp 41–56

Struminskaya B, de Leeuw ED, Kaczmirek L (2015) Mode system effects in an online panel study: comparing a probability-based online panel with two face-to-face reference surveys. Methods Data Anal 9:3–56. https://doi.org/10.12758/mda.2015.001

Tagseth M, Sund ER, Hallman GT et al (2019) May telephone surveys provide reliable public health surveillance data for municipalities? Mode effects differ between categories of questions. The HUNT Study. Norway. Norsk Epidemiologi 28:105–116. https://doi.org/10.5324/nje.v28i1-2.3057

Tipping S, Hope S, Pickering K et al (2010) The effect of mode and context on survey results: analysis of data from the Health Survey for England 2006 and the Boost Survey for London. BMC Med Res Methodol 10:84. https://doi.org/10.1186/1471-2288-10-84

Tourangeau R, Conrad FG, Couper MP (2013) The science of web surveys. Oxford University Press, New York

Tourangeau R, Rips LJ, Rasinski K (2000) The psychology of survey response. Cambridge University Press, Cambridge

Vergnaud AC, Touvier M, Méjean C et al (2011) Agreement between web-based and paper versions of a socio-demographic questionnaire in the NutriNet-Santé study. Int J Public Health 56:407–417. https://doi.org/10.1007/s00038-011-0257-5

Verschuuren M, Gissler M, Kilpeläinen K et al (2013) Public health indicators for the EU: the joint action for ECHIM (European Community Health Indicators & Monitoring). Arch Public Health 71:12. https://doi.org/10.1186/0778-7367-71-12

Ware JE, Sherbourne CD (1992) The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection. Med care 30:473–483

Washington group (2012) Washington group on disability statistics (WG), Budapest Initiative (BI) & United Nations Economic &

Social Commission for Asia & the Pacific (UNESCAP). Development of disability measures for surveys: the extended set on functioning. https://www.washingtongroup-disability.com/wp-content/uploads/2016/01/Development_of_Disability_Measures_for_Surveys_The_Extended_Set_on_Functioning.pdf. Accessed 30 May 2019

Weisband S, Kiesler S (1996) Self disclosure on computer forms: meta-analysis and implications. In: Bilger R, Guest S, Tauber MJ (eds) CHI'96 Electronic Proceedings. Association for Computing Machinery Inc, New York, pp 1–14