Do Respondents Answer Differently in Web Survey than in Face-to-Face Interview: Field Work Experiment from the European Health Interview Survey (EHIS)

M. Kralj[1], M. Zaletel[2], D. Lavtar[3], T. Zupanič[4]

[1] *National Institute of Public Health, Ljubljana, Slovenia; marcel.kralj@nijz.si*
[2] *National Institute of Public Health, Ljubljana, Slovenia; metka.zaletel@nijz.si*
[3] *National Institute of Public Health, Ljubljana, Slovenia; darja.lavtar@nijz.si*
[4] *National Institute of Public Health, Ljubljana, Slovenia; tina.zupanic@nijz.si*

Abstract
Expansion of web surveys at the turn of 20th century brought new opportunities and new challenges at the same time. Web surveys enable lower costs and time savings for national data providers, and offer respondents a choice to complete the questionnaire at the time and with the pace they wish to. Mixed-mode designs have become a leading way of collecting survey data. Many organisations have experienced that the results from web surveys differ from other modes. There are many reasons for this, maybe the most important one is that many respondents provide different answers when completing the survey by themselves via web. How different are actually web surveys compared to other modes of data collection, and even more importantly, do they provide the same results? During the field work of European Health Interview Survey (2014), National Institute of Public Health carried out an experiment with re-interviews in order to test whether the respondents answer differently in web survey than they do in a face-to-face interview and to explore the magnitude of differences for different survey estimates. The analysis showed that, for some items, respondents gave different answers depending on the mode of data collection, which consequently can affect the final survey estimates.

Keywords: mixed-mode design, experiment, re-interview

1 Background

With the considerable growth of the internet use during the last decade (for example, in Europe 78% of individuals were using internet in 2015 (ITU 2015)) web surveys are becoming more and more important and widespread means of data collection, also in large (multi)national surveys. Higher speed, lower cost, greater flexibility for the respondents (they can answer the survey whenever and from wherever they choose), interactivity, and opportunity for including advanced functionalities are some of the key advantages of web surveys (Callegaro, Lozar Manfreda and Vehovar 2015; Berzelak 2016). However, because of the substantially lower response rate compared to the traditional data collection modes, such as telephone and face-to-face surveys, and for achieving better sample composition, web

surveys are increasingly used in combination with other data collection modes in so-called mixed mode design. But how different are web surveys compared to other modes of data collection, and do they provide the same results? Studies have shown that different data collection modes have different influences on the response behaviour, which results in respondents answering differently on the same questions in different data collection modes. One of the most important and frequently described factor is the social desirability effect. Social desirability bias occurs when individuals provide different responses in the presence of an interviewer in order to appear in a favourable light. Respondents exhibit this bias when they overreport socially approved behaviours and underreport socially disapproved behaviours (Groves et al. 2004). This issue appears especially in the case of questions about sensitive topics, such as questions about health, alcohol consumption, drug use, sexual behaviour etc. Many studies have found that web survey respondents are less susceptible to social desirability bias than respondents in face-to-face and telephone surveys because of the absence of an interviewer (Tourangeau, Conrad and Couper 2013; Duffy et al. 2005).

On the other hand, the interviewer's presence may importantly contribute to the reduction of respondent's burden and opportunities to take (cognitive) shortcuts while answering to the survey's questions[1]. Interviewer can additionally motivate the respondent, redirect the attention to specific topics, emphasize the instructions, and resolve any ambiguities or errors that occur during answering, which can consequently affect the data accuracy.

Heerwegh and Loosveldt (2008) confirmed the hypotheses that web surveys elicited more 'don't know' responses, more non-differentiation on rating scales and a higher item nonresponse rate. Some previous studies also suggest that in visual modes (such as a web survey) respondents for categorical questions tend to select answers that are shown at the beginning of a list, while in oral modes respondents tend to choose the last categories (Krosnick and Alwin 1987). Moreover, some researchers found out that telephone respondents are more likely to give positive answers to scale questions than web survey respondents (Christian, Dillman and Smyth 2005; Ye, Fulton and Tourangeau 2011). Another interesting results were obtained by Martin and Lynn (2011) who found suggestive evidence that web survey respondents give more pessimistic answers than respondents in a face-to-face interview.

[1] To make question answering easier and to satisfy the survey request, the respondent may take various shortcuts: respondent does not answer the question or he/she responds with 'I don't know' or 'I don't want to answer'; he puts focus on the answer choices that are available at the beginning or at the end of a list; he responds with the same or almost the same answers; he tends to agree with the questions, answers positively; he selects the extreme or middle answers on the scale variables; or he chooses answers randomly (Berzelak 2016).

Because of such effects it is clear that final survey estimates may differ between data collection modes. In the health domain, findings of a study, performed by Hoebel et al. (2014), indicate that prevalence rates obtained from health interview surveys can vary with the mode of data collection, primarily between interviewer- and self-administered modes. They report that indicators, such as questions on prevalent diseases, may be less affected than the indicators of psychosocial and mental health, or health behaviours. Another study on mode effects in health survey revealed that in comparing face-to-face and self-administered modes no significant mode effects were observed for indicators related to the use of health services, but significant mode effects were observed for indicators related to the self-reported health-related quality of life, health behaviour, social relations and morbidity (Christensen et al. 2013). Also Bowling's (2005) results of the literature review suggest that the estimates of positive health status, health related quality of life, engaging in desirable behaviours and activities, appear likely to be exaggerated when obtained with face-to-face or telephone interviews, rather than with a self-administered mode. Socially undesirable behaviours (e.g. smoking) and sensitive health problems (e.g. prostatic disease, urinary symptoms) can also be underreported in face-to-face or telephone interviews, compared with self-administered questionnaires. However, she adds that some studies have reported no differences in the response between interviewer- and self-administered modes.

2 Experiment

2.1 EHIS 2014

The European Health Interview Survey (EHIS)[2] was carried out in autumn 2014 in Slovenia. It was conducted as a mixed-mode survey: web survey and face-to-face interviews. In November 2014, we made a test-retest experiment in order to examine whether the answers from web survey respondents differ from answers given in face-to-face interviewing. We invited 200 individuals who primarily responded to EHIS web survey to participate in our experiment. An incentive was offered for this additional participation. The length of the questionnaire in the experiment was about a third of the original EHIS questionnaire. We added some questions in order to detect the differences in the answers and to verify certain patterns of respondents' behaviour. 83 individuals volunteered to participate in the re-interviewing, out of them 69 were eligible to be taken into analysis.

[2] Detailed information on EHIS is available at: http://ec.europa.eu/eurostat/web/microdata/european-health-interview-survey.

2.2 Analysis method

For the evaluation of the differences in respondents' answers between both data collection modes, we carried out two sets of analyses. First, we analysed differences at the respondent level, where we observed for each variable the number of differences that occurred and the direction of difference as well. Here, the difference is defined as every pair of respondent's answers that are not completely identical in both data collection modes. In addition to this definition and approach for determining the number of differences, we used another definition for variables with 5- or more-point scales. In case of theses variables, the difference is defined as every pair of respondent's answers that differ in both data collection modes for more than one point regarding the scale measurement. The latter approach was taken under the assumption that the difference between two neighbouring categories/values in scales with higher number of categories can often be unrecognized or considered as irrelevant, especially among respondents. For testing whether the individuals provided different answers to web survey than they do in face-to-face interview, we used McNemar's test for nominal variables and Wilcoxon Signed-Ranks Test (a nonparametric equivalent to Paired-Samples T-Test) for ordinal, interval and ratio variables.

In second set of analyses, we compared data at the aggregated level and assessed differences in estimates between data collected via web survey and face-to-face interview. In this case, absolute and relative differences were calculated, namely for the mean and percentage estimates.

We did an analysis for 37 variables of different type and of diverse topics such as health status, accidents and injuries, use of day care, physical activity, consumption of vegetables, smoking, alcohol consumption, social support, etc. The list covers all key variables from the point of view of European Health Interview Survey.

3 Results

3.1 Individuals that responded to both data collection modes

Out of all 182 individuals (sample size was 200, but 18 individuals were excluded indirectly due to unforeseen absence of one interviewer), 83 (46%) participated in the experiment and gave answers to face-to-face interview. After the data cleaning process, 81 units were considered as eligible for the analysis. Table 1 shows demographic characteristics of

individuals who participated in the experiment, meaning they responded to both data collection modes – web survey and face-to-face interview.

Table 1: Demographic characteristics of experiment participants

	n	%
Sex		
Male	39	48.1
Female	41	51.9
Age		
15-34	24	29.6
35-54	26	32.1
55-74	29	35.8
≥75	2	2.5
Education		
Primary or lower	7	8.6
Lower secondary	2	2.5
Upper secondary	38	46.9
Tertiary	34	42.0
Labour status		
Carries out a job or profession	37	45.7
Self-employed	4	4.9
Unemployed	6	7.4
Pupil, student	10	12.3
Retired	22	27.2
Other	2	2.5
Total	81	100.0

3.2 Differences in respondents' answers between data collection modes

Before we looked into the differences in respondents' answers, additional 12 units were screened out from the analysis as they reported that some changes, which could affect the difference in their answers, happened to them in the period between both surveys[3]. The analysis was therefore performed on a sample of 69 units. It is worth mentioning that the average time between the first (web survey) and second (face-to-face interview) participation was 29 days.

For majority of variables no significant differences in respondents' answers were found. However, for 11 out of 37 variables, statistically significant differences in respondents' answers between both data collection modes were observed. In addition, we could identify relatively large number of changes in respondents' answers when asked by a different survey mode. Table 2 shows the results of statistical significance as well as the number and type of differences that occurred for each variable.

[3] This screening question was asked at the end of the interview.

The respondents rated their general health status significantly better in face-to-face interview than in web survey. They also rated their physical pain at a higher level in web survey compared to face-to-face interview, and the level of how this pain interferes with their normal work. In this case, for example, 33 individuals rated their physical pain at higher level in web survey, while much less respondents (11) rated it at higher level in face-to-face mode. Thus, the number of respondents who changed their answers with data collection mode was 44 (65%). When we looked for answers that differed for more than one point according to measuring scale, the number of respondents with different answers was 18 (26%). On the other hand, their answers did not change significantly with data collection mode when they reported about having longstanding illness/health problem, limitation because of a health problem, disease or chronic conditions such as diabetes and depression, or when they reported on the use of inpatient and day care. Respondents did not report differently about their previous accidents that resulted in injury. For example, only three respondents reported their road traffic accident exclusively in web survey, while none of them reported it exclusively in face-to-face interview.

In almost all variables (7 out of 8) that are measuring mental health, we could find more individuals who rated their mental well-being worse in web survey than later in face-to-face interview. However, statistically significant differences in answers were observed for three variables. These three variables relate to problems such as sleeping problems, poor appetite or overeating, and despondency or disappointment.

When we looked at the indicators for health-risk behaviours, we found out that answers on daily physical activity did not change with survey mode. Only five differences in respondents' answers were identified in this case. Respondents' answers were also not statistically different between data collection modes when they reported on the frequency of vegetable consumption, although we can notice that much more individuals rated higher frequency of eating vegetables in a face-to-face interview rather than in web survey. Reporting on smoking habits did not change as well. However, we could observe that three individuals reported they tried to stop smoking and reported this exclusively in the web survey. This number is not large but we need to take into account that the entire sample includes only eight smokers. Significant differences in answers occurred in some questions about the alcohol consumption, but generally, respondents reported higher frequency of alcohol consumption in every question on the web survey.

In face-to-face interview, participants tended to answer more towards greater social support they possessed. Significant difference (at level of 0.1) in respondents' answers between both data collection modes was found regarding the question on how easy is to get practical help from neighbours. Regardless of the direction of difference, we can notice that high number of differences in respondents' answers occurred in variables for social support. Approximately 50% of all respondents have changed their answers with survey mode in all three variables. Mode effect could also be observed for a variable that measured overall life satisfaction, where respondents more frequently assessed higher satisfaction level in a face-to-face interview. In this experiment, for example, 31 individuals rated their life satisfaction at higher level in face-to-face interview, while much less respondents (9) rated it at higher level in web survey.

Table 2: Differences in respondents' answers between data collection modes (CAPI and CAWI), for 37 variables

Type of variable	Number of differences	Types of differences, $\|f_1$ (f_2)	p-value	n
HS1 - How is your health in general?				
Ordinal (5 pts)	18 (26.1%)	Better status reported in CAWI |3 (0) Better status reported in CAPI |15 (2)	.004	69|69
HS2 - Do you have any longstanding illness or longstanding health problem?				
Nominal (binary)	7 (10.1%)	Illness/health problem reported only in CAWI |5 Illness/health problem reported only in CAPI |2	.453	69|69
HS3 - For at least the past 6 months, to what extent have you been limited because of a health problem in activities people usually do?				
Ordinal (3 pts)	16 (23.2%)	Higher limitation reported in CAWI |11 Higher limitation reported in CAPI |5	.180	69|69
CD1J - During the past 12 months, have you had any of the following diseases or conditions – Diabetes?				
Nominal (binary)	11 (15.9%)	Nonresponse in CAWI, without a disease in CAPI |9 Nonresponse in CAPI, without a disease in CAWI |1 Nonresponse in CAWI, disease reported in CAPI |1	1.000	69|58
CD1O - During the past 12 months, have you had any of the following diseases or conditions – Depression?				
Nominal (binary)	10 (14.5%)	Nonresponse in CAWI, without a disease in CAPI |7 Nonresponse in CAPI, without a disease in CAWI |1 Nonresponse in CAWI, disease reported in CAPI |1 Nonresponse in CAPI, disease reported in CAWI |1	1.000	69|59
AC1A - In the past 12 months, have you had any of the following type of accidents resulting in injury – Road traffic accident?				
Nominal (binary)	3 (4.3%)	Accident reported only in CAWI |3	.250	69|69
AC1B - In the past 12 months, have you had any of the following type of accidents resulting in injury – Home accident?				
Nominal (binary)	12 (17.4%)	Accident reported only in CAWI |7 Accident reported only in CAPI |3 Nonresponse in CAWI, without accident in CAPI |2	.344	69|67
AC1C - In the past 12 months, have you had any of the following type of accidents resulting in injury – Leisure accident?				
Nominal (binary)	11 (15.9%)	Accident reported only in CAWI |6 Accident reported only in CAPI |2 Nonresponse in CAWI, without accident in CAPI |2 Nonresponse in CAWI, accident reported in CAPI |1	.289	69|66
PL6 - Do you have difficulty walking half a km on level ground without the use of any aid?				
Ordinal (4 pts)	7 (10.3%)	Higher difficulty reported in CAPI |4 Higher difficulty reported in CAWI |3	.726	68|68

PN1 - How much bodily pain have you had during the past 4 weeks?				
Ordinal (6 pts)	44 (64.7%)	Higher pain reported in CAPI |11 (2) Higher pain reported in CAWI |33 (16)	.000	68|68
PN2 - During the past 4 weeks, how much did pain interfere with your normal work (including both work outside the home and housework)?				
Ordinal (5 pts)	32 (47.1%)	Higher limitation reported in CAPI |6 (2) Higher limitation reported in CAWI |26 (9)	.001	68|68
MH1A - Over the last 2 weeks, how often have you been bothered by any of the following problems - Little interest or pleasure in doing things?				
Ordinal (4 pts)	18 (26.5%)	Higher frequency of a problem reported in CAPI |7 Higher frequency of a problem reported in CAWI |9 Nonresponse in CAWI, without a problem in CAPI |2	.491	68|66
MH1B - … - Feeling down, depressed or hopeless.				
Ordinal (4 pts)	20 (29.4%)	Higher frequency of a problem reported in CAPI |5 Higher frequency of a problem reported in CAWI |10 Nonresponse in CAWI, without a problem in CAPI |3 Nonresponse in CAWI, problem reported in CAPI |2	.197	68|63
MH1C - … - Trouble falling or staying asleep, or sleeping too much.				
Ordinal (4 pts)	31 (45.6%)	Higher frequency of a problem reported in CAPI |4 Higher frequency of a problem reported in CAWI |25 Nonresponse in CAWI, problem reported in CAPI |2	.000	68|66
MH1D - … - Feeling tired or having little energy.				
Ordinal (4 pts)	24 (35.3%)	Higher frequency of a problem reported in CAPI |8 Higher frequency of a problem reported in CAWI |13 Nonresponse in CAWI, without a problem in CAPI |2 Nonresponse in CAWI, problem reported in CAPI |1	.221	68|65
MH1E - … - Poor appetite or overeating.				
Ordinal (4 pts)	21 (30.9%)	Higher frequency of a problem reported in CAPI |3 Higher frequency of a problem reported in CAWI |16 Nonresponse in CAWI, without a problem in CAPI |2	.006	68|66
MH1F - … - Feeling bad about yourself or that you are a failure or have let yourself or your family down.				
Ordinal (4 pts)	16 (23.5%)	Higher frequency of a problem reported in CAPI |2 Higher frequency of a problem reported in CAWI |12 Nonresponse in CAWI, without a problem in CAPI |2	.007	68|66
MH1G - … - Trouble concentrating on things, such as reading the newspaper or watching television				
Ordinal (4 pts)	19 (27.3%)	Higher frequency of a problem reported in CAPI |6 Higher frequency of a problem reported in CAWI |10 Nonresponse in CAWI, without a problem in CAPI |2 Nonresponse in CAPI, problem reported in CAWI |1	.196	68|65
MH1H - … - Moving or speaking so slowly that other people could have noticed. Or the opposite—being so fidgety or restless that you have been moving around a lot more than usual.				
Ordinal (4 pts)	10 (14.7%)	Higher frequency of a problem reported in CAPI |4 Higher frequency of a problem reported in CAWI |4 Nonresponse in CAWI, without a problem in CAPI |2	1.000	68|66
HO1 - In the past 12 months have you been in hospital as an inpatient, that is overnight or longer?				
Nominal (binary)	5 (7.4%)	Hospitalisation reported only in CAWI |3 Hospitalisation reported only in CAPI |1 Nonresponse in CAWI, without hospitalisation in CAPI |1	.625	68|67
HO3 - In the past 12 months, have you been admitted to hospital as a day patient, that is admitted to hospital for diagnosis, treatment or other types of health care, but not required to remain overnight?				
Nominal (binary)	16 (23.5%)	Admission reported only in CAWI |7 Admission reported only in CAPI |9	.804	68|68
UN1A - Have you experienced delay in getting health care in the past 12 months because the time needed to obtain an appointment was too long?				
Nominal (3 pts)	26 (38.2%)	Delay reported only in CAWI |4 Delay reported only in CAPI |6	1.000	68|39
BM1 - How tall are you without shoes?				
Open numerical	23 (33.8%)	Higher number reported in CAPI |9 Higher number reported in CAWI |13 Nonresponse in CAWI, height reported in CAPI |1	.057	68|67
BM2 - How much do you weigh without clothes and shoes?				
Open	38 (55.9%)	Higher number reported in CAPI |16	.324	68|67

numerical		Higher number reported in CAWI |21 Nonresponse in CAWI, weight reported in CAPI |1		
PE9 – Are you physically active at least 30 minutes per day or 150 minutes per week? Include all physical activities at work, daily tasks and leisure activities.				
Nominal (binary)	5 (7.4%)	Physically active only in CAWI |2 Physically active only in CAPI |1 Nonresponse in CAWI, physically active in CAPI |2	1.000	68|66
FV3 - How often do you eat vegetables or salad, excluding potatoes and juice made from concentrate?				
Ordinal (5 pts)	27 (39.7%)	Higher frequency of vegetable intake reported in CAWI |9 (1) Higher frequency of vegetable intake reported in CAPI |17 (3) Nonresponse in CAWI, frequent vegetable intake in CAPI |1 (1)	.105	68|67
SK1 - Do you smoke?				
Ordinal (3 pts)	4 (5.9%)	More regular smoking reported in CAPI |2 More regular smoking reported in CAWI |2	1.000	68|68
SK3 - On average, how many cigarettes do you smoke each day?				
Open numerical	1 (16.7%)	Higher number reported in CAWI |1	.317	6|6
SK7 – Did you try to stop smoking in last 12 months?				
Nominal (binary)	3 (37.5%)	Attempt to stop smoking reported only in CAWI |3	.250	8|8
AL1 - In the past 12 months, how often have you had an alcoholic drink of any kind [beer, wine, cider, spirits, cocktails, premixes, liquor, homemade alcohol…]?				
Ordinal (9 pts)	37 (54.4%)	More frequent alcohol consumption in CAWI |25 (8) More frequent alcohol consumption in CAPI |11 (7) Nonresponse in CAWI, infrequent alcohol consumption in CAPI |1 (1)	.095	68|67
AL4 - Thinking of Friday to Sunday, on how many of these 3 days do you usually drink alcohol?				
Ratio (4 pts)	10 (52.6%)	More days of drinking reported in CAWI |6 More days of drinking reported in CAPI |3 Nonresponse in CAWI, no drinking in CAPI |1	.248	19|18
AL6 - In the past 12 months, how often have you had 6 or more drinks containing alcohol on one occasion? For instance, during a party, a meal, an evening out with friends, alone at home, …				
Ordinal (9 pts)	31 (58.5%)	More frequent binge drinking in CAWI |23 (7) More frequent binge drinking in CAPI |8 (3)	.022	53|53
SS1 - How many people are so close to you that you can count on them if you have serious personal problems?				
Ordinal (4 pts)	31 (45.6%)	Higher number reported in CAPI |18 Higher number reported in CAWI |11 Nonresponse in CAWI, number reported in CAPI |2	.239	68|66
SS2 - How much concern do people show in what you are doing?				
Ordinal (5 pts)	37 (54.4%)	Higher concern and interest reported in CAWI |16 (2) Higher concern and interest reported in CAPI |18 (4) Nonresponse in CAWI, response in CAPI |3 (3)	.550	68|65
SS3 - How easy is it to get practical help from neighbours if you should need it?				
Ordinal (5 pts)	36 (52.9%)	Greater possibility of getting help reported in CAWI |10 (4) Greater possibility of getting help reported in CAPI |24 (4) Nonresponse in CAWI, response in CAPI |2 (2)	.087	68|66
DOH – What is the net monthly income of your household? (in EUR)				
Open numerical	34 (51.5%)	Higher income reported in CAPI |7 Higher income reported in CAWI |3 Income reported only in CAPI |9 Income reported only in CAWI |15	.682	66|15
ZAD – How much are you satisfied with your life as a whole?				
Interval (11 pts)	41 (62.1%)	Greater life satisfaction reported in CAPI |31 (13) Greater life satisfaction reported in CAWI |9 (4) Nonresponse in CAWI, life satisfaction reported in CAPI |1 (1)	.002	66|65

Note: f_1 = frequency for the individual type of difference; (f_2) = frequency for the individual type of difference, taking into account only respondents' answers that differ for more than one point regarding the scale measurement; n = number of cases (first figure shows all cases including cases with missing data, second figure shows number of cases that are included in statistical test – cases with no missing data)

3.3 Differences in survey estimates

In the following subchapter, we look into the differences in survey estimates obtained from the web survey and face-to-face interview, and explore whether the estimates were affected by the differences in data collection modes. Table 3 shows the absolute and relative differences in estimates between both survey modes. For the purpose of this paper, the results are presented for variables where significant differences were observed in the previous set of analyses.

It can be noticed that the relative differences between estimates often exceed 10%, which is a value that usually should not be negligible. The highest relative difference in estimates was found for the level/prevalence of bodily pain (22%), pain interference (25%) and mental health problems (15-21%). In all cases the estimates obtained from the web survey show higher prevalence of such problems. Relative differences reach around 10% also for the estimates that measured the general health status, social support and life satisfaction. Also for these three dimensions, the estimates that were obtained from the face-to-face interview reflect better status. Smaller relative differences were observed for estimates of alcohol consumption, but the values were still around 5%. However, if we check the differences in percentage estimates, we can get for some specific estimates (e.g. prevalence of having alcoholic drink 3-4 days a week or more in past 12 months) a relative change that is even 30% or more. Of course, we need to be aware that percentage estimates are generally more susceptible to the effects of data collection modes compared to the mean estimates (Jäckle, Roberts and Lynn 2006). Regardless to that, a review of the differences suggests that web survey produced estimates that show higher level of alcohol consumption compared to face-to-face interview. The smallest relative difference was observed for average body height (0.6%), concretely, this means the difference of one centimetre.

Table 3: Differences in survey estimates obtain from web survey (CAWI) and face-to-face interview (CAPI) mode

	(1) CAWI	(2) CAPI	Absolute diff. (2) - (1)	Relative diff. ((2)-(1))/(1)
HS1 - How is your health in general? (1 - Very good, 5 - Very bad; Mean)	2.1	1.8	-0.2	-11.2%
HS1 - How is your health in general? (Very good + good; %)	75.4	84.1	8.7	11.5%
PN1 - How much bodily pain have you had during the past 4 weeks? (1 - None, 6 - Very severe; Mean)	2.8	2.2	-0.6	-22.2%
PN2 - During the past 4 weeks, how much did pain interfere with your normal work (including both work outside the home and housework)? (1 - Not at all, 5 - Extremely; Mean)	1.9	1.4	-0.5	-25.1%
MH1C - Trouble falling or staying asleep, or sleeping too much. (1 - Not at all, 4 - Nearly every day; Mean)	1.8	1.5	-0.3	-18.7%

MH1E - Poor appetite or overeating. (1 - Not at all, 4 - Nearly every day; Mean)	1.4	1.1	-0.3	-20.8%
MH1F - Feeling bad about yourself or that you are a failure or have let yourself or your family down (1 - Not at all, 4 - Nearly every day; Mean)	1.3	1.1	-0.2	-15.3%
BM1 - How tall are you without shoes? (Mean)	172.8	171.8	-1.0	-0.6%
AL1 - In the past 12 months, how often have you had an alcoholic drink of any kind [beer, wine, cider, spirits, cocktails, premixes, liquor, homemade alcohol…]? (1 - Every day or almost every day, 9 - Never or only a few sips or trials, in my whole life; Mean)	5.3	5.6	0.3	5.3%
AL1 - In the past 12 months, how often have you had an alcoholic drink of any kind [beer, wine, cider, spirits, cocktails, premixes, liquor, homemade alcohol…]? (3-4 days a week or more, %)	16.4	11.6	-4.8	-29.4%
AL6 - In the past 12 months, how often have you had 6 or more drinks containing alcohol on one occasion? For instance, during a party, a meal, an evening out with friends, alone at home, … (1 - Every day or almost, 9 - Never in my whole life; Mean)	7.0	7.3	0.3	4.7%
AL6 - In the past 12 months, how often have you had 6 or more drinks containing alcohol on one occasion? For instance, during a party, a meal, an evening out with friends, alone at home … (Once a month, %)	15.5	10.3	-5.2	-33.3%
SS3 - How easy is it to get practical help from neighbours if you should need it? (1 - Very easy, 5 - Very difficult; Mean)	1.8	1.6	-0.2	-11.5%
ZAD – How much are you satisfied with your life as a whole? (0 - Very dissatisfied, 10 - Very satisfied; Mean)	7.6	8.2	0.6	8.2%

4 Conclusion

With this test-retest experiment, we can conclude that for the majority of variables differences are not significant, but for some items, respondents gave different answers depending on the mode of data collection, which consequently can affect the final survey estimates. Questions based on factual information (e.g. longstanding illness, disease/chronic conditions, accidents and injuries, use of health services) were less susceptible to mode effect than questions on subjective information (e.g. level of pain, mental health, general health, life satisfaction, social support) or some questions that focus on undesirable behaviours, such as alcohol consumption.

Differences in respondents' answers between both data collection modes (web survey and face-to-face interview) may be attributed to the reduced social desirability bias in web survey and to the fact that respondents give more positive responses in face-to-face interviews than in self-administered surveys, according to the results of other studies (Bowling 2005). As Ye, Fulton and Tourangeau (2011) write, tendency to give positive ratings appears to be related to the presence of an interviewer, and it may reflect respondents' reluctance to express bad news. However, the complexity of the interviewing process should not be overlooked. Many factors can have different influence on the respondent's behaviour and on the answers, which

he or she provides for a specific question. The respondent's characteristics (e.g. motivation, skills, disability), interviewer's experiences, interviewer-respondent relationship (e.g. level of perceived anonymity), characteristics and actual performance of the interview (e.g. use of technology, use of visual stimulus of a show card, disturbances during the interview), question type and its difficulty level (e.g. understandability, number of response categories) are just a few of them. Therefore, to understand the reasons for the differences that occur with the use of different data collection modes, more research is needed. However, such experiments are an important action for the identification and assessment of the potential size of the mode effect.

Nevertheless, we should not forget that the mode of data collection is only one component that contributes to the total error of survey estimate (another important component is, for instance, the nonresponse bias) (Groves and Lyberg 2010), and that a decision on which data collection mode or modes will be used is made based on the combination of target population under study, available resources, data quality and (geographical and longitudinal) comparability, especially in the field of official statistics.

5 References

Berzelak, N. (2016). Razlike v odgovorih med spletnim in drugimi načini anketiranja. Bilten Statističnega društva Slovenije, Vol. 59/38.

Bowling, A. (2005). Mode of questionnaire administration can have serious effects on data quality. Journal of Public Health, VoI. 27, No. 3, pp. 281–291. Accessed: http://jpubhealth.oxfordjournals.org/content/27/3/281.long [23.4.2016].

Callegaro M., Lozar Manfreda, K., Vehovar, V. (2015). Web Survey Methodology. London: Sage.

Christensen, A.I., Ekholm, O., Glümer, C., Juel, K. (2013). Effect of survey mode on response patterns: comparison of face-to-face and self-administered modes in health surveys. The European Journal of Public Health. Accessed: http://eurpub.oxfordjournals.org/content/24/2/327 [23.4.2016].

Christian, L.M., Dillman, D., Smyth, J.D. (2005). The effects of mode and format on answers to scalar questions in telephone and web surveys. Accessed: https://www.sesrc.wsu.edu/dillman/papers/2006/theeffectsofmodeandformat.pdf [23.4.2016].

Duffy, B., Smith, K., Terhanian, G., Bremer, J. (2005). Comparing data from online and face-to-face surveys. Int. J. MarketRes.47 (6), 615–639. Accessed: https://www.ipsos-mori.com/Assets/Docs/Archive/Publications/comparing-data.pdf [23.4.2016].

Groves, R. M., Fowler jr, F. J., Couper, M. P., Lepkowski, J. M., Singer, E. and Tourangeau, R. (2004). Survey Methodology. Wiley.

Groves, R. M. and Lyberg, L. (2010). Total survey error - past, present, and future. Public Opinion Quarterly, Vol. 74, No. 5, 2010, pp. 849–879. Accessed: http://poq.oxfordjournals.org/content/74/5/849.full.pdf+html [23.4.2016].

Heerwegh, D., Loosveldt, G. (2008). Face-to-face versus Web surveying in a high-Internet-coverage population: differences in response quality. Public Opinion. Q.72, 836–846. Accessed: http://poq.oxfordjournals.org/content/72/5/836.full.pdf [23.4.2016].

Hoebel, J., von der Lippe, E., Lange, C., Ziese, T. (2014). Mode differences in a mixed-mode health interview survey among adults. Archives of Public Health 2014, 72:46. Accessed: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4373057/ [23.4.2016].

ITU. (2015). Key 2005-2015 ICT data for the world. Accessed: http://www.itu.int/en/ITU-D/Statistics/Documents/statistics/2015/ITU_Key_2005-2015_ICT_data.xls [23.4.2016].

Jäckle, A., Roberts, C. and Lynn, P. (2006). Telephone versus Face-to-Face Interviewing: Mode Effects on Data Quality and Likely Causes (ISER Working Paper No. 2006-41). Essex, UK: Institute for Social & Economic Research. Accessed: https://www.iser.essex.ac.uk/files/iser_working_papers/2006-41.pdf [23.4.2016].

Krosnick J. A., Alwin, D.F. (1987). An evaluation of a cognitive theory of response-order effects in survey measurement. Public Opinion. Q 1987, 51:201–219. Accessed: https://pprg.stanford.edu/wp-content/uploads/1987-an-evaluation-of-a-cognitive-theory-of-response-order-effect.pdf [23.4.2016].

Peter Martin, P. and Lynn P. (2011). The Effects of Mixed Mode Survey Designs on Simple and Complex Analyses. CCSS Working Paper, no. 04. Accessed: https://www.city.ac.uk/__data/assets/pdf_file/0017/125135/CCSS-Working-Paper-No-04.pdf [23.4.2016].

Tourangeau, R., Conrad, F. G. in Couper, M. (2013). The science of web surveys. Oxford, UK: Oxford University Press.

Ye C., Fulton, J., Tourangeau, R. (2011). More positive or more extreme? A meta-analysis of mode differences in response choice. Public Opinion Q 2011, 75:349–365. Accessed: http://poq.oxfordjournals.org/content/75/2/349.short [23.4.2016].