Victimisation survey by three modes

Laaksonen, Seppo
University of Helsinki, Social Research
Box 68
University of Helsinki (00014), Finland
E-mail: Seppo.Laaksonen@Helsinki.Fi

Web surveys are becoming more common, not only due to their low cost. Unfortunately, the quality of web surveys may be problematic since coverage problems and non-response may be severe. Such problems do not automatically lead to considerably biased results, since the representativity of the data is most important. It is clear that a high response rate is still a good indicator for the representativity too. On the other hand, a response rate is rarely high in face-to-face or telephone surveys and consequently the representativity may be poor in these surveys as well. The paper compares the success of those three data collection modes as reliably as possible. Hence, our survey designs are as equal as possible. Our application area, crime victimization, is demanding also because the sensitiveness of questions varies. Our conclusions are not for these reasons straightforward.

Key words: Face-to-face interviewing, non-response vs. measurement error, probit regression, telephone interviewing, victimization indicators, web survey,

1. Introduction

Survey designs have been remarkably changing during some decades. As far as household surveys are concerned, face-to-face mode has been still considered to ensure the best quality, since all households can be in principle reached and interviewed. A good example is the European Social Survey (ESS) that only has allowed this mode for the core questionnaire during the four rounds from 2002 to 2009 (see EuropeanSocialSurvey.Org). Other alternatives are however under testing. Many other surveys, including recent Finnish national crime victim surveys, use computer-assisted telephone interviewing (CATI) although all households cannot be contacted via telephone. The reason is that CATI is relatively cheap, but cheaper techniques would be more preferable. Web surveys are much cheaper and hence more tempting. They are not however solely used in nation-wide cross-sectional demanding surveys due to quality doubts; the main reasons are obviously that the web coverage is not high and people are unaccustomed to use it in surveys.

Our aim is to make an experiment so that the results based on three data collection modes (face-to-face, phone and web) can be well compared. This is not our final target, since we do not yet believe that a pure web survey would be a good strategy in a near future. Hence, our experiment helps us to develop an optimal mixed-mode strategy that could give opportunity to survey reliably the whole target population. We do not here consider the problem of some surveys that some population groups are factually excluded from the sample (e.g. households whose telephone numbers are not found in pure CATI). So our study tries to estimate all results for the desired target population, nevertheless that we know that the sampling frame is not complete.

We anticipated in advance that the response rate will not be very high for our web survey compared to phone and face-to-face alternatives. A question is whether a survey with low response rate is not at all representative, and such estimates should not be used, consequently. This question has been recently

examined by some researchers. One is the RISQ project (see e.g. Shlomo et al 2009) and the other presented by Lundström and Särndal (2008). Both approaches take advantage of the representativity via auxiliary variables. This means, roughly saying, that it is good if the estimation methodology is such that it gives as the correct estimates for aggregate auxiliary variables as possible, it may work reasonably for proper survey estimates as well. Our solution for this is to adjust for the weights using a large and good pattern of auxiliary variables. This strategy does not correct for measurement problems due to different modes. This is our second big issue in the paper.

However, it is clear that representativeness increases but not monotonically with increasing response rate. But, representativeness depends on many other factors too, naturally including coverage that is often one of the key problems in web survey (see e.g. Lozar Manfreda and Vehovar 2008).

2. Modes and their effects

Jaeckle et al (2009, 18) say that there are literally hundreds of studies that have tested the comparability of data collected with different modes. It is still difficult to devise mode comparisons such that any differences in responses can clearly be attributed to the effect of mode on measurement. They mention that mode effects are often confounded with differences in sample composition due to differences in coverage, sampling error and non-response. Instead, coverage and non-response remain to be critical. Moreover, and not less importantly, these authors and many others (e.g. Dillman 2000, de Leeuw 2005&2008, de Leeuw and Dillman and Hox 2008) analyze differences in answering. For instance, they observe that a given difference in responses to survey questions may cause biases in some types of estimates, but not in others.

This study exploits and compares the three modes, face-to-face, telephone and web. Their characteristics are not similar. De Leeuw (2008, 122-123) presents a summary of these. In her view, face-to-face interviews are the most flexible form of data collection method. Visual and auditory stimuli may be used, all channels of communication are available for information transmission and feedback, and an interviewer is present as intermediary between respondent and researcher. It is possible to use a large variety of measurement instruments. Furthermore, highly complex questionnaires can be implemented as a trained interviewer takes care of navigation through the questionnaire. Currently, computer assisted personal interview (CAPI) is used generally, as in our study. However, most sensitive questions were answered anonymously using computer assisted self-interview technique (CASI).

Telephone interviews are less flexible according to de Leeuw (2008, 122). Their major drawback is the absence of visual cues during the interview. No show cards with lists of answer categories are available. The interviewer reads the question out aloud with the available response categories and the respondent has to rely solely on memory. In general, questions must be short and easily understandable over the phone. In telephone surveys the interviewer can assist respondents in understanding questions, administer questionnaires with a large number of screening questions, control the questionnaire sequence, and probe for answers on open questions. These options were important in our study, too. A big drawback is however, that the interviewer gets to know all answers, that is, anonymity is not possible.

Web surveys share the advantage of mail surveys regarding visual aids. Also, just as in mail surveys, the respondent is in charge and the situation may offer more privacy; sensitive questions are not a problem, at all. Because an interview program determines the order of the questions, more complex questionnaires can be used than in a paper mail survey. In this sense a web survey is equivalent to an interview survey. But

it also has a drawback, it is more perfunctory medium and people often just pay a flying visit.

Coverage and non-response are not easy to comparably analyze by modes. Naturally, everything can be made well, if the target population is identical and the sampling frame as well. This is the case in this study, but the survey can be something else in practice. The major problem is maybe the fact that some target population groups can be factually either completely or partially out of the sampling frame. A typical example is concerned telephone surveys in which case the full telephone number list is impossible to obtain in all countries. Consequently, some target population groups are avoidably out. It is of course possible to cut the target population to those who have access to telephone (or who's telephone numbers are known) but this is not correct. It follows that all such people should be considered as non-respondents as we do in this study.

There are also different options for this exclusion. Revilla (2010) argues:

Concerning the decision of participation, one element to take into account is the respondents' access to each mode: not all sampling units have a telephone or web access allowing them to complete a survey in that mode. A low coverage of the population of interest in one mode can be a barrier to the participation of some populations.

As far as Statistics Finland's practice in 2011 is concerned, two thirds of telephone numbers can be found automatically and with low cost. With an additional cost, about 80 per cent of numbers are found, and using more resources maximally around 90 per cent. This means, the coverage depends on the budget of the survey. In our study, we had medium resources and the telephone coverage was 83 per cent, consequently.

In the case of web surveys, analogical discussion can be made, since it is hard for non-web people to respond to such questionnaires. We have not found this discussion with face-to-face surveys since all can be reached in principle. Hence we here illustrate this aspect only for web surveys and for telephone surveys, respectively. Figure 1 shows the response rates by age so that the denominator include those who's telephone number has been available or who have used regularly Internet. This latter figure is not from individuals but estimated at aggregates from the fourth round of the Finnish European Social Survey (ESS).

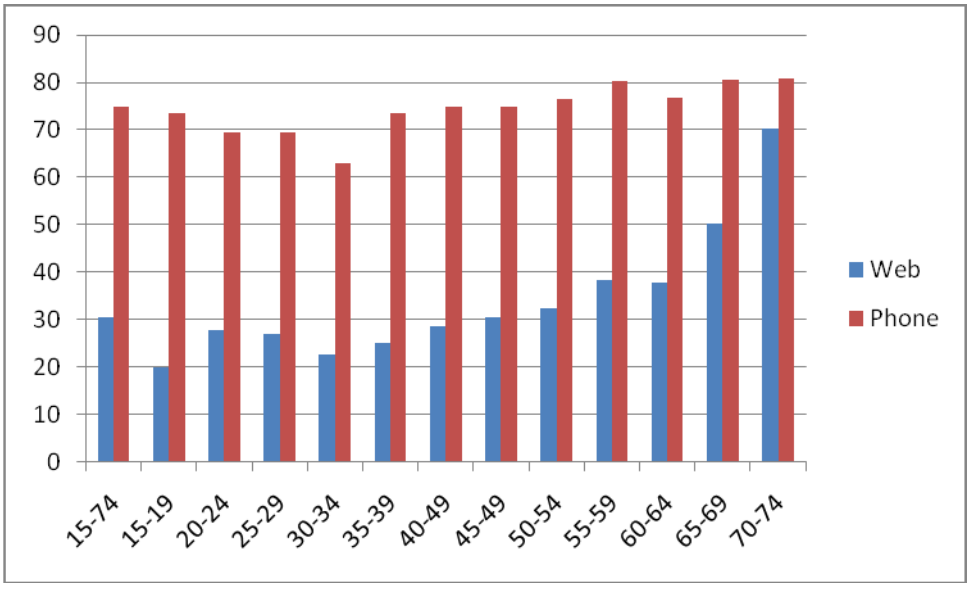

Fig. 1. 'Conditional' response rates by age group for our web survey and the telephone survey, respectively. Sources: our data and the ESS micro data of round four.

The average accessibilities are very close to each other both for web and telephone (around 80 %). This means that from this point of view the high response rate is not impossible in web either. Almost 100 per cent of young agegroups are using regularly internet but older people much less. Interestingly, older people seem to be more motivated to participate given that they are using internet.

3. Victimisation surveys and their special nature

Victimisation surveys have a 40 year long history in Finland as well as in other pioneer countries. The basic idea of the victimisation survey is to ask of a population sample whether they have fallen victims for certain crimes during a defined time period (Heiskanen 2002). Traditionally, crime data have been received from judicial institutions, such as police, courts and prison. These data sources, although they might describe exactly the activities of the authorities cannot be used as a valid measure of criminality and its development, because crimes are more or less comprehensively reported to the authorities, and the propensity to report different crimes differ, as may also do reporting of similar crimes in time.

What makes criminal victimisation a very specific topic is the nature of crime: by essence crime is concealed, because it is forbidden in the penal code and the consequences of the crime are often painful and sensitive to the victims? Especially the sensitivity connected with victimisation is of importance. For instance, details of violence in close relations may be an issue which the victim is not willing to reveal to anyone.

The secrecy of responding may differ across the modes: in mail and web surveys the respondent discusses with the somewhere existing, unknown researcher. The privacy protection of the respondent is more secure in mail and web surveys than in modes that use an interviewer for conducting the data collection. Telephone surveys are sometimes assessed as a convenient mode for victimisation surveys: because the respondent does not meet personally the interviewer, there is a higher degree of anonymity compared to the face-to-face method. On the other hand, the interviewer can in the face-to-face interview in many ways facilitate the interviewee in the interview situation, and thus encourage the respondent for giving the reliable rapport on victimisation. A natural strategy is to give the interviewee anonymously to give his/her answer. In this study, we used two strategies for face-to-face interviewing, (i) self-administered answers for most sensitive questions, CASI (e.g. violence, sexual harassment) and (ii) completely interviewer-administrated (e.g. background variables, theft and other ordinary crimes). The detailed description of our example variables is given in the output table, Table 5.

Our standpoint is that 1) the mode that secures a better secrecy for the respondent will produce a higher victimisation prevalence; 2) the differences between modes are larger when more sensitive questions are asked.

4. Our survey

We thus made our best to organize the fieldwork fairly for each mode. The invitation for the survey was mad by mail to all and basically similar instructions were given to all. The questionnaires are as similar as possible. Web survey is self-administrated and responses are anonymous whereas face-to-face and telephone surveys are administrated by an interviewer. Anonymity was not possible to arrange for telephone surveys, but it was partially possible for face-to-face surveys due self-administered questionnaire for most sensitive questions. Naturally, motivation to surveys can be made better for telephone and face-to-face surveys since

the sample persons were contacted many times by personal contacts. This could be possible for web-survey responders too, but we only sent one reminder letter to all who had not answered in two weeks time.

The fieldwork period lasted about two and half months, but two weeks more for telephone interviewing. A reason was that the same interviewers were used for face-to-face survey and they wanted to finish this job before the holiday period, but continued by phone after that. This is obviously a reason for a relatively low response rate in face-to-face survey. Note that telephone interviewing was not done in a CATi center but at homes of the interviewers.

Web survey was our special pilot since it is expected to use increasingly in future. This was the first country-level web study in Statistics Finland and many statisticians did not believe that it will be successful. For example, response rates were guessed to be quite low, less than 10 percent, but our optimistic rate 25 percent was about realised. Naturally, we compare the success of the web mode against the two other survey modes, that is, interviewing by telephone and face-to-face. We thus arranged an as fair 'competition' as possible between these modes. Hence we use the same sampling design and compute the similar sampling weights for each data. Some steps however differ. The sample people for the web survey are requested to go to fill in the questionnaire on a specific website and they will receive the unique password for this purpose. In the cases of the other two modes, the interviewers administered the fieldwork. The contact efforts cannot be considered to be exactly equal for each mode, maybe efforts for web were least whereas these were hardest for phone.

The overall target for both for telephone and face-to-face survey was about 65 percent. By telephone mode the goal was almost achieved (62%) but face-to-face interviewing did not succeed well (50%).

5. Sampling design and auxiliary variables

The target population covers 15-74 years old residents in Finland who's mother tongue is not Swedish (6 % of Finnish population are Swedish speaking). This choice was due to our limited budget; so we did not need e.g. Swedish questionnaires. The size of the population was about 3.8 million at the time when the sampling frame was created.

Our sampling design can be called *stratified cluster two-stage sampling* but stratification was made twice. First, we used four large regions as pre-strata. Within each region small-area clusters, together 100, were picked up by pps with replacement where size is the corresponding frame population of the region. The clusters were constructed using municipality codes and postal codes. In the second stage, we formed new strata by gender and four age groups. So, we have finally 32 pre-strata (4 x 2 x 4). Respectively, the people to the gross sample were drawn at random within each sampled cluster. The sample allocation into the 32 strata is not equal for two reasons. First, the anticipated response rates were used to some extent so that if an anticipated response rate is lower, the sampling fraction is higher, and vice versa. Secondly, our client desired to analyze male victimisation in more detail and hence the sampling fractions are essentially higher for males in the web survey and in the telephone survey, but only little in the face-to-face survey (see Aromaa et al 2010). We do not present here all details, but only illustrate the gross sampling weights in Table 1.

Table 1: Some figures of the overall design weights

Count	Mean	Sum	Minimum	Maximum	Coeff. of variation, %
7737	494	3 825 000	36	1397	32

In fact, we created the three gross samples from the overall sample of those 7737 people, one for each 'mode survey.' This sampling was completely random, that is, now created the three independent samples. The allocation was based on the pre-determined sizes so that our budget allowed a rather small size for the face-to-face survey (735) but much higher for the telephone survey (3009) and for the web survey (3993), respectively. The number is highest for the web survey since we did not believe to get a very high response rate for this group but we desired to get enough respondents in absolute measures.

The population register (FPR) includes a number of auxiliary variables and these were exploited as completely as possible. In addition, we downloaded auxiliary variables from the education register (EDR) that is based on completed exams in public Finnish schools and other education institutions. Thus, the exams received in other countries are not covered, and hence we had to put such people at the lowest education levels. Moreover, we took some variables from the employment register (EMR) and the tax register (TR). Table 2 gives the potential list of our auxiliary variables. For the sampling design, the first three ones were used as explained above. These and all the others are at least attempted when analysing non-response and adjusting the weights, see the following chapters.

Table 2. Auxiliary variables of the sampling file

Variable	**Source**	**Used for** **S= sampling** **U = unit non-response** **W= weighting**
Age in years	FPR	S U W
Gender	FPR	S U W
Regional variables including municipality, postal codes, coordinates of the home	FPR	S U W
Marital status with different options, marriage year, number of marriages	FPR	U W
Mother tongue and citizenship	FPR	U W
Rough occupation or socioeconomic status (not for all)	FPR	U
Household composition including number of children	FPR	U W
House variables such as size, number of rooms and type of kitchen	FPR	U W
Education level and field	EDR	U W
Unemployed, number of months as unemployed	EMR	U
Taxable income	TR	U

5. Response rates and response modelling

The overall response rate is rather low on the average (44%). As expected, the rate is lowest for the web survey (almost 25%). We are not satisfied with the response rate of the face-to-face survey (50%) since the rate has been 10-15 percent higher in recent surveys of Statistics Finland. One reason for that was obviously a high workload of our interviewers who at the same time collected the telephone survey data. This mode was more successful with 62 percent response rate.

In this section, we first present one simple rate by education level but go forward using multidimensional analysis, since this is needed for the next steps.

Figure 2 is thus one-dimensional. The profiles of the response rates are quite different by modes. The rate increases almost linearly by education level for web, but not for phone and face-to-face. The middle-educated are best respondents in phone. Interestingly, there is no clear line for face-to-face respondents by education level, but their profile is most even.

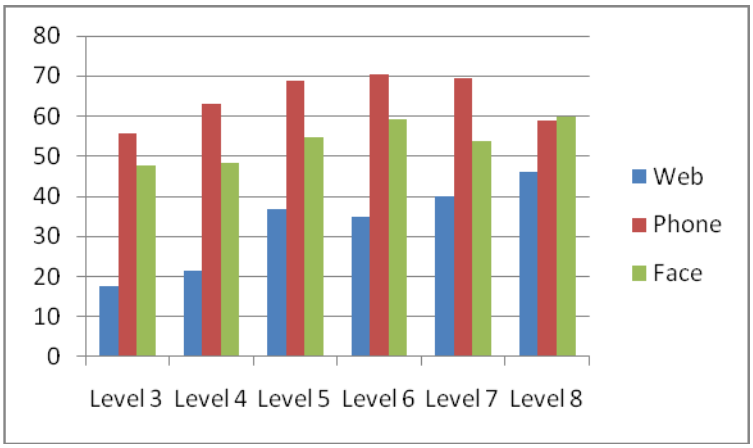

Fig. 2. Response rates by education level for each mode, %

There are at least four acceptable alternatives to analyse responding multi-dimensionally (Laaksonen 2006). In all these cases, the dependent variable is a binary response indicator (1=respondent, 0=no) but the link function can be chosen differently. The results do not differ dramatically, fortunately (Laaksonen 2006). We here use a probit link and the outputs for the following step are thus probits. These are easy to interpret since they are probabilities and illustrate differences in these indicators.

We tried with a number of the combinations of available auxiliary variables and also so that we created new auxiliary variables from the initial ones. The purpose in this is naturally to improve the fitness of the model (and predictability). But what this means? Since we are going to adjust the initial weights, our one criterion is that the same specification is required for each mode survey. This ensures the best comparable weights and consequently the the estimates, and the weights adjust for representativity to some extent.

The second question is: Should the explanatory variables in each model be statistically significant? Our answer is: 'No.' The reason is that even an insignificant variable may improve the adjusted weights although not substantially. Of course, we have to be careful and to avoid implausible (often extreme) weights. Nevertheless, we took the significance account to some extent so that our models include only such variables that are significant at least in one mode model except that we always included gender due to its general interest. It should be noted that gender was significant as the only auxiliary variable for web, and almost

for phone. In both these cases, the probit was lower of males, that is, they responded somewhat worse.

We expected that unemployed people will respond worse than the others but this is not the case. Maybe the reason is the subject of the survey that gives opportunity to tell about unpleasant experiences. Also, we expected that the taxable income is significant. We tried to scale it taking into account the household size with the same result. Very often, poor and very rich people are less willing to participate (e.g. Laaksonen 1991) but this survey is different, maybe for the similar reason as for unemployed. This study is thus interesting since these variables are not significant.

The main results on the probit response propensity models are given in Table 3. There are some special transformed variables that need to be explained better. One is partnership that we constructed un-usually in order to include more aspects within the variable like the usual marital status offers. Our partnership is done so that first the widowed are distinguished, then the singles and then those who have married two or more times and finally the other marriages were classified into three groups. Fresh marriage means the year 2000 or later, medium between 1980 and 2000 whereas old marriages were made this agreement before 1980.

We see that singles, widowed and those who have married twice or more have been responded relatively worst whereas old and medium marriages best. The differences for web surveys are relatively slightest. Freshly marriages are surprisingly bad responders in phone surveys. This is much due to non-availability of a found phone number (that is above 20% for them and for singles, and almost 20% for many marriages people). These fresh marriages are, on the other hand, relatively best respondents in the web survey.

Table 3. Probit estimates by mode and for the whole data. The estimates for occupation, type of kitchen and education field are not included in the table but are used in models.

Auxiliary variable	*Web*	*Phone*	*Face-to-face*	*Overall*
Face-to-face vs web				0.7343
Phone vs web				1.0779
Male vs female	-0.0565	-0.0256	0.1038	-0.0266
Mother tongue				
Finnish vs Russia	0.2599	0.2319	-0.1071	0.2743
Other vs Russia	-0.1135	-0.1085	-0.3186	-0.1011
Education levels				
Level 2 vs level 8	-0.8251	0.01111	-0.3798	-0.5620
Level 3 vs level 8	-0.6313	0.3858	-0.2512	-0.3142
Level 5 vs level 8	-0.1875	0.3874	-0.0285	-0.0749
Level 6 vs level 8	-0.2137	0.6398	0.0408	0.0227
Level 7 vs level 8	-0.1046	0.6062	0.0860	0.0954
Large regions				
Southern towns vs Helsinki	-0.0869	-0.0227	0.1125	-0.0268
Other towns vs Helsinki	-0.1294	-0.0060	0.0336	-0.0650

Rural vs Helsinki	-0.1209	0.2477	0.3718	0.0943
Partnership				
Widowed vs old marriage	-0.2224	0.1060	-0.1071	-0.0407
Single vs old marriage	-0.3414	-0.2130	-0.3115	-0.2660
Many vs old marriage	-0.1236	-0.2956	-0.1181	-0.1691
Fresh vs old marriage	-0.3011	-0.0946	-0.2245	-0.1718
Medium vs old marriage	-0.1526	-0.1472	0.1363	-0.1018
No kids vs 1+ kids	0.1835	-0.1360	- 0.0015	0.0227
1 room vs 4+ rooms	-0.0427	-0.4151	-0.1458	-0.2633
2-3 rooms vs 4+ rooms	-0.0374	-0.1151	0.0062	-0.0657

Although the variable 'age' is always exiting, it is not included in Table 3. The reason is that we used the four age variables (at least one of these was significant in each case). One is age in years and three others with its second, third and fourth power (i.e. polynom transformation). This strategy gives the smoothed profiles for response propensities in probits. To facilitate the interpretation: The values are scaled so that the probit = 0 corresponds to the estimated response rate equal = 0.5.

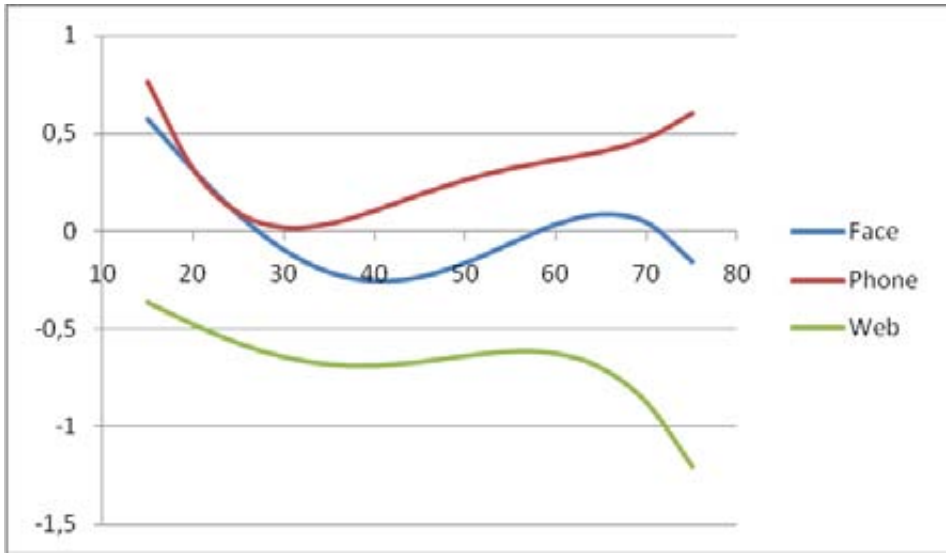

Fig. 4. Response premia by age for each mode based on the probit model.

There are several interesting features in this figure. Older people are participating quite poorly in the web (cf. conditional rates in Figure 1) whereas they respond best in the phone mode. This is also due to the fact their phone number was well available for the survey institute. Another general feature seems to be a rather low participation of 30-40 year olds. Is it maybe so that this generation has not time enough for surveys and possibly their attitudes are against surveys, or against this particular sensitive survey.

6. Harmonised weights

Multivariate unit response analysis is more illustrative and was done by probit regression. The same model was also used for adjusting weights. The weighting methodology follows the same procedure as presented in Laaksonen and Chambers (2006) and Laaksonen (2007). That is, the weighted response propensity (probit) model is used to estimate the response propensities to each respondent by multiplying the basic weights of

the respondents by their inverse, and then the sums of the weights were benchmarked to the 32 known strata population.

The modelling and weighting were done for each mode survey separately, and the respective adjusted weights are used in estimations. The basic weights were thus adjusted further by exploiting a number of auxiliary variables. This operation naturally improves as well the quality of estimates, their representativity and makes them more comparable. It is however clear that some quality problems remain. We assess these questions in Section 6.

The next section compares the results from those three independent surveys and uses the equally adjusted weights in estimation. Table 5 illustrates these weights roughly (compare these against the design weights in Table 1). The net sample for face-to-face interviewing is rather low and estimates for this are not very accurate consequently. On the other hand, the variation of these weights is moderate, highest being for the web survey. This implicates, for instance, that we have been able to distinguish many features on high non-response rates via response propensity modelling, thanks for good auxiliary variables. Thus, if we had poor auxiliary variables, the variability would not be much higher than obtained for the design weights.

Table 4: Some figures of the adjusted sampling weights

Mode	Count	Mean	Sum	Minimum	Maximum	Coeff. of variation, %
Web	971	3940	3 825 000	214	17095	58
Phone	1866	2052	3 825 000	168	8939	37
Face-to-face	366	10453	3 825 000	1002	41909	48

7. Victimization results by mode

Table 5 presents a number of selected results. We have included most common victimization indicators in the table. Very rare crimes are not included since their estimates do not make any difference between modes and they are in some cases even almost zeros. Some of the indicators are calculated from one initial variable but some from many. In the latter case, the crime means that at least one of the given list has been occurred for a respondent.

Table 5. Some estimates by data collection mode (standard errors in parenthesis). Highest rates are bolded.

Indicator *(from how many variables)*	***Face-to-face***	***Phone***	***Web***
Part 1: Fear	*Interviewer types*	*Interviewer types*	*Self-administered*
Feeling unsafe (1)	*19.9 (1.3)*	*19.8 (1.3)*	***25.5** (0.8)*
Burglary fear (1)	**24.4** (1.4)	21.0 (0.6)	**26.5** (0.8)
Assault fear (1)	20.0 (1.3)	*20.4 (0.6)*	***26.1** (0.8)*
Fear on family and friends (1)	*33.5 (1.6)*	*32.0 (0.7)*	***37.6** (0.9)*
Part 2: Ordinary crimes	*Interviewer types*	*Interviewer types*	*Self-administered*
Car stolen given owned (1)	1.7 (0.8)	*2.5 (0.4)*	***3.8** (0.7)*
Car damaged (1)	10.1 (1.7)	*10.1 (0.8)*	***15.6** (1.5)*
Bicycle stolen (1)	*17.3 (2.2)*	*17.3 (1.0)*	***23.5** (1.6)*
Cottage burglary (1)	*6.5 (2.3)*	***8.6** (1.3)*	***9.0** (1.6)*
Stolen using violence (1)	*1.5 (0.7)*	*2.6 (0.4)*	***3.5** (0.7)*
Stolen, other (1)	*9.9 (1.8)*	*9.4 (0.8)*	***12.1** (1.2)*
Home burglary (1)	2.2 (0.7)	2.7 (0.4)	**5.0** (0.8)
Cheating (1)	8.4 (1.5)	**9.6** (0.7)	**10.2** (1.0)
Part 3: Very sensitive crimes	*Anonymous, Self-administered*	*Interviewer*	*Self-administered*
Sexual violence (1)	*10.2 (1.7)*	*9.4 (0.8)*	***15.1** (1.3)*
Sexual harassment, ever (11)	**52.3** (2.8)	37.9 (1.3)	45.4. (1.9)
Sexual harassment, recent (11)	**21.8** (2.4)	10.1 (0.8)	**22.2** (1.6)
Violence by unknown (9)	**35.1** (2.7)	29.6 (1.1)	**34.5** (1.7)
Violence by partner (11)	**3.8** (1.5)	2.2 (0.6)	**3.7** (0.9)

Our one purpose was to compare the results between the three mode surveys. We thus used our best knowledge and methodology for getting the as ideal adjusted weights as possible. These weights thus harmonise the estimates so that the structural differences of these three data are controlled so as our probit models do. Basically, they adjust for the representativity of the data sets, too. Nevertheless, the response rates vary substantially, and especially a low response rate for the web survey rises a problem in representativity, since we have no such auxiliary variables that could adjust for the use of web.

We naturally see that the web users are more often in large towns (e.g. Helsinki area in Table 4). Since at the same time, many crimes are more common in the areas, it is expected that such indicators will be

overestimated. This is the main reason that most of the part 2 indicators are highest in web. Cheating and cottage burglary are exceptions, but these crimes are not very common in large towns. Estimates of the part 1 indicators on fear are also overestimated in the web survey but not dramatically. This reflects the same representativity problem as for part 2.

The indicators of part 3 are in most cases different. The highest estimates are for web and face-to-face surveys, and the lowest for phone. This is interesting but how to explain it? In any case, the representativity of the web survey is no more as bad as in the previous cases. Maybe these two surveys are quite well comparable. At contrast, the phone survey estimates are underestimated, although this data set is most representative in response rates. But this does not help if there are bias due to measurement errors, that in this case very obvious. The phone survey does not offer anonymity for the respondent like the web and the face-to-face with CASI. We expected some downward bias but not so big as we find for the indicators that are concerned several unpleasant crimes. The first indicator, a general level one, does not tell the same story.

We tried to find explanations for phone survey results also from the workload of the interviewers that was hard during the fieldwork but was stressed them in the end when they used much time for increasing the response rate for the phone survey. We made a survey for interviewers, too, asked from them about how pressing they feel their job. This shows that telephone interviewing was in general more pressing than face-to-face interviewing. This not usually observed but we interpret it to be due the topic of this survey. It is easier to communicate with respondents in face-to-face situation. On other hand, web is best in anonymity, the respondents tell the 'truth' without extra explanations.

References

Aromaa, K., Heiskanen, M., Laaksonen, S., Nikula, J., and Ruuskanen, E. (2010): Finnish results from the EU-pilot survey. The final report delivered to the European Commission 22.2.2010. 71 pp. Not published. Available from the authors.

de Leeuw, E. D. (2008): Choosing the method of data collection. In: *International Handbook of Survey Methodology* (Eds. de Leeuw, E.D., Hox, J.J. and Dillman, D.A). 299-316. Taylor & Francis Group, LLC. New York

de Leeuw, E. D., (2005): To Mix or Not to Mix Data Collection Modes in Surveys. *Journal of Official Statistics* 21. 2, 233-255.

de Leeuw, E. D and Dillman, D.A, Hox, J.J. (2008): Mixed mode surveys: When and why. In: *International Handbook of Survey Methodology* (Eds. de Leeuw, E.D., Hox, J.J., Dillman, D.A). 299-316. Taylor & Francis Group, LLC. New York.

Dillman, D. A. (2000): *Mail and Internet Surveys: The Tailored Design Method*. Second Edition. Wiley. New York.

Heiskanen, M. (2002) Väkivalta, pelko, turvattomuus. Surveytutkimusten näkökulmia suomalaisten turvallisuuteen. Tutkimuksia 236. Tilastokeskus. Helsinki.

Jäckle, A., Roberts, C. and Lynn, P. (2009). Assessing the Effect of Data Collection Mode on Measurement. *International Statistical Review* 78, 1, 3-20.

Laaksonen, S. (2007): Weighting for Two-phase Surveyed Data. *Survey Methodology* 33. 2. 121-130.

Laaksonen, S. and Chambers, R. (2006): Survey Estimation Under Informative Nonresponse with Follow-up.

Journal of Official Statistics 22, 1, 81-95.

Laaksonen S. (2006). Does the choice of link function matter in response propensity modelling? *Model Assisted Statistics and Applications, An International Journal*, Volume 1, Number 2 / 2005 / 2006. pp. 95 – 100. Publisher: IOS Press.

Laaksonen, S. (1991). Adjustment for Non-response in Two-year Panel Data: Applications to Problems of Household Income Distribution. *The Statistician* 40, 153-168.

Lozar Manfreda, K. and Vehovar, V. (2008). Internet surveys. In: *International Handbook of Survey Methodology* (Eds. de Leeuw, E.D., Hox, J.J. and Dillman, D.A). 264-284. Taylor & Francis Group, LLC. New York.

Revilla, M. (2010). Quality in Unimode and Mixed-Mode Designs: A Multitrait-Multimethod Approach. Survey Research Methods 4, 151-164. http://www.surveymethods.org

Schlomo, N., Skinner, C., Schouten, B., Bethlehem, J. and Zhang, L.C. (2009). Statistical Properties of R-Indicators. *Work Package 3. RISQ Deliverable 2.1*. Available www.r-indicator.eu.

Särndal, C-E. and Lundström, S. (2008). Assessing Auxiliary Vectors for Control of Nonresponse Bias in the Calibration Estimator. *Journal of Official Statistics* 24, 167-191.

Voogt, R.J.J. and Saris, W.E. (2005). Mixed Mode Designs: Finding the Balance Between Nonresponse Bias and Mode Effects. *Journal of Official Statistics* 21, 36