L I V E S
W O R K I N G
P A P E R
2 0 1 6 / 48

TITLE

Mixing modes of data collection in Swiss social surveys: Methodological Report of the LIVES-FORS Mixed Mode Experiment

Research report

Authors

Caroline Roberts

Dominique Joye

Michèle Ernst Stähli

http://dx.doi.org/10.12682/lives.2296-1658.2016.48
ISSN 2296-1658

FNSNF
Swiss National Science Foundation

The National Centres of Competence in Research (NCCR) are a research instrument of the Swiss National Science Foundation.

LIVES
Swiss National Centre of Competence in Research

Authors

Roberts, C. (1)

Joye, C. (1)

Stähli, M.E. (2)

Abstract

Survey-based data collection makes a fundamental contribution to social science research in Switzerland, and survey data form the empirical base of much of the research into vulnerability across the life course being carried out by the Swiss National Centre for Competence in Research, 'LIVES'. Because different features of the design of a survey can have implications for the quality of the data collected, optimising the survey design is key to ensuring the accuracy of the conclusions drawn from analyses of the data, and hence for the validity of both theoretical and policy developments derived from these. In this paper we present a methodological report of an experiment designed to provide evidence about which survey designs work best in the Swiss context, to maximise the quality of future quantitative research. The purpose of the study was to investigate the relation between survey design – in particular, the mode or modes of data collection used – and different sources of error affecting data quality, including coverage, nonresponse and measurement error. A key motivation for the research was the growing popularity of 'mixed mode' data collection, which offers a number of promising solutions to challenges facing traditional surveys (like under-coverage, declining response rates and rising costs), but carries the disadvantage of confounded measurement and selection errors that can compromise the comparability of the data collected in different modes. In this study, we compare single mode surveys (mail, CATI and web) and sequential mixed mode surveys (CATI plus mail, and web plus mail plus CATI/CAPI) with respect to response rates and the representativeness of the responding sample. The results lend support to the conclusion that mixing modes sequentially can help to increase response rates and improve sample representativeness, though differences are observed as a function of the availability of telephone numbers for sample members. Furthermore, these apparent benefits of mixing modes may not justify the additional burden such designs place on resources. Future research should, therefore, be directed at evaluating the cost-error trade-offs involved in different ways of combining modes from a total survey error perspective.

Keywords

Life course | Vulnerability

Authors' affiliations

(1) Institute of Social Sciences and NCCR LIVES
(2) FORS, Swiss Centre of Expertise in the Social Sciences

Correspondence to

caroline.roberts@unil.ch

1. Introduction

Survey data play a central role in the social sciences, with more and more researchers carrying out secondary analysis of large-scale datasets available through national data archives. In Switzerland, these include, among others, the Swiss Labour Force Survey (SLFS), the Survey of Income and Living Conditions (SILC), the Swiss Household Panel Survey (SHP), the Survey of Health, Aging and Retirement in Europe (SHARE), the European Social Survey (ESS), the International Social Survey Programme (MOSAiCH), and the European Values Survey (EVS). In the Swiss National Science Foundation-funded National Centre for Competence in Research (NCCR) 'LIVES', these major national studies of the general population have been supplemented by several purpose-designed quantitative surveys focused on different aspects of vulnerability and resilience among special subpopulations living in Switzerland. These include (at the time of writing) i) a longitudinal survey following a cohort sample of young adults (LIVES Cohort survey), including an over-sample of second generation immigrants from Albanian-speaking countries in the former Yugoslavia, being surveyed alongside the third sample of the SHP (Gomensoro and Bolzman, 2015), and an associated pilot survey that was designed to field test a suitable sampling strategy (carried out by the LIVES Individual Projects (IP)1 and IP2), ii) two studies investigating pathways out of unemployment (IP4), including a survey looking at the impact of mass redundancy among ex-employees of 5 firms that closed down between 2009 and 2010 (Baumann, Lipps, Oesch, and Vandenplas, 2016), and a survey of the newly unemployed in the canton of Vaud (Turtschi and von Ow, 2014); iii) a 7-wave longitudinal survey of workers and the unemployed looking at the impact of individual characteristics and resources on professional trajectories (IP7) (Maggiori, Rossier, Krings, Johnston, and Massoudi, 2016); iv) a 2-wave extension of an existing panel survey of married and unmarried couples looking at changing family configurations in response to critical events (IP8 – Widmer, Aeby, Sapin, 2013); v) a 2-wave panel study of the role of the couple relationship as a source of support for women facing breast cancer (IP11 – Charvoz, Favez, Notari, Panes-Ruedin, and Delaloye, 2016); vi) a longitudinal survey of divorcees and widows investigating the effects of losing an intimate partner in the second half of life (IP12 – Perrig-Chiello, Hutchison, and Knöpfli, 2016); and vii) a survey of

older adults aged 60 and over investigating inequalities during old age (IP13 – Oris et al., 2016).

Survey research is, thus, at the centre of the LIVES project, with over one hundred collaborators drawing on these (and other) data sources to develop new knowledge about the phenomenon of vulnerability and the resources people draw upon to overcome it, with a view to contributing to the development of innovative social policy measures informed by the findings of their research. The quality of the data collected is essential to the reliability and validity of these conclusions, and the effectiveness of any recommendations derived from them. Yet the quality of all survey data is inevitably compromised by trade-offs made in the survey design process – trade-offs that are becoming increasingly problematic as a result of growing challenges associated with carrying out surveys using traditional methods of data collection.

In this paper, we focus on the impact on data quality of a specific type of trade-off in survey design relevant to a number of the LIVES surveys, as well as to a growing number of larger-scale national and international surveys: the decision to use a mix of methods (or 'modes') of data collection (e.g. telephone or face-to-face interviews, paper or web self-administered questionnaires) to survey different sample members. We present a methodological report of a 'mixed mode experiment' – a collaboration between one of the methodological projects within LIVES (IP15) and the Swiss Centre of Expertise in the Social Sciences (FORS) – designed to investigate the effect of combining modes in different ways on various aspects of data quality in a survey on vulnerability in the general population. Before describing in more detail the aims and the design of the study, we first discuss the challenges currently facing survey research in Switzerland, the motivations behind the decision to mix data collection modes, and the reasons for undertaking methodological research on the effects of mixing modes in the context of a substantive research programme like LIVES.

1.1 Challenges to data quality in survey research

The LIVES studies face a unique set of challenges associated with sampling, and achieving an adequate representation of their chosen populations, many of which can be described as either 'hard-to-reach' or 'hard-to-survey', either because there are no suitable listings available for sampling purposes, or because they are notoriously hard to contact

or reluctant to participate in surveys (see Tourangeau et al,. 2014). Added to this, the substantive focus of the LIVES research poses quite particular, and often complex measurement challenges, such as how to ask about subjective phenomena likely to be perceived as sensitive by respondents, how best to capture life event histories or the configuration of social networks, and how to obtain accurate retrospective evaluations of wellbeing at different points across the life course. To complicate matters, many of the LIVES surveys have a longitudinal design, incurring an additional threat to data quality due to the risk of selective sample attrition. Each of these challenges and the way in which they are managed in the survey design process has implications for the amount of error affecting estimates derived from the data.

Though we do not consider the unique challenges facing the LIVES surveys in further detail here, they help to illustrate the main sources of error that affect the accuracy of survey estimates (Groves, 1989). These include: *coverage error* (associated with the failure to provide all eligible population members with a known and non-zero probability of being selected to participate in the survey); *nonresponse error* (resulting from nonparticipation among particular subgroups, and differences between the responding and nonresponding samples); and *measurement error* (resulting e.g. from problems with the design of the questionnaire, or the way in which respondents formulate their answers to the questions). The quality of any given survey estimate depends on how much it is affected by error from these various different sources. Other sources of error may be present (e.g. data input errors, coding errors, processing errors – see Groves (1989) for a detailed discussion), but we focus on the principal sources here.

While the unique challenges faced by the LIVES surveys and their implications for survey quality warrant detailed attention (see Oris, Roberts, Joye, and Ernst Stähli, 2016), the study reported here was motivated by more general challenges to data quality shared by the LIVES projects, as well as by many existing large-scale national and international surveys. In particular, these concern difficulties associated with the 'survey-taking climate' (Lyberg and Dean, 1992), which in Switzerland (in common with other western European countries and North America), is posing new constraints on how surveys can be carried out. This climate is characterised by public perceptions of being over-burdened by marketing calls and invitations to participate in surveys and pseudo-surveys, which has contributed to increasing non-contact and refusal rates (De Leeuw and De Heer, 2002).

As a consequence, survey costs have risen because of the need for additional response enhancement strategies (such as incentives, or refusal conversion efforts) to minimise the impact on response rates and the risk of nonresponse bias. Meanwhile, telephone interviewing, which has long been the mode of choice of the Swiss survey industry (and the Swiss Federal Statistical Office), has become increasingly problematic because of a substantial decline in recent years in the number of households with fixed-line telephones and/or registered telephone numbers. This has led to a growing problem of under-coverage in surveys conducted by phone (or surveys based on samples drawn from the telephone directory, which previously offered high levels of coverage) and a threat of bias due to differences in the characteristics of those who can and cannot now be surveyed by telephone (Ernst Stähli, 2012; Joye, Pollien, Sapin, and Ernst Stähli, 2012; Roberts, Lipps and Kissau, 2013; Lipps, Pekari, and Roberts, 2015).

This combination of factors has meant that survey organisations have had to adapt their traditional data collection methods to ensure adequate coverage for general population surveys, and offer clients affordable solutions. In both respects, single mode alternatives to telephone interviewing are unattractive. Face-to-face interviewing - often considered a 'gold standard' in many other countries with respect to gaining respondent cooperation - is rarely an option for most academically led surveys in Switzerland because it is too costly. However, cheaper, self-administered modes of data collection (namely, web and mailed paper questionnaires) do not necessarily offer reliable alternatives for complex, large-scale population surveys. Both mail surveys and web-surveys are unsuitable for those with literacy challenges, and particularly for anybody unable to read or write in the Swiss national languages. Web surveys offer advantages over mail surveys, in that they can handle more complicated questionnaire designs (e.g. with routing and filters), but suffer from more substantial coverage problems. The number of households with Internet access in Switzerland is currently over 85%, but this varies as a function of income (FSO, 2015), and almost certainly over-estimates the number of individual residents who would actually be able and willing to participate in a web survey.

1.2. Mixed mode data collection: Promises and problems

So-called 'mixed mode' data collection is increasingly being adopted, or considered, as a solution to these various constraints on traditional survey practice. For one, using a combination of data collection modes could provide an answer to the problem of non-coverage in telephone surveys (Atkeson et al., 2014; Groves, et al., 1988). Assuming the non-covered part of the population can be identified, an alternative mode can be used to collect data where telephone numbers are unavailable (Wagner et al., 2014). Another assumed benefit is that mixed mode survey designs that make use of cheaper modes (web and mail) are likely to offer substantial cost savings compared with pure face-to-face or telephone surveys (ibid.). A further hope is that using a combination of modes in sequence may help to reduce nonresponse bias, because using different modes to follow up nonrespondents may help to attract a more varied selection of respondents than would be possible in a single mode survey (Vannieuwenhuyze, 2014; Dillman, 2000).

To date, however, research evidence testifying to the purported benefits of combining modes of data collection has yielded mixed results. In particular, the apparent cost-saving advantages have been questioned, because of the increased fixed costs involved in mounting a survey in multiple modes (Vannieuwenhuyze, 2014). Furthermore, researchers have questioned whether reductions in single sources of survey error – e.g. coverage or nonresponse bias – guarantee a reduction in *total* survey error (the combined influence of different error sources on survey estimates) (Buelens and van den Brakel, 2015). This is because different modes of data collection have unique measurement properties, and influence respondents' answers to survey questions in varied ways (Dillman, et al., 2009), meaning measurement error is likely to be increased when modes are mixed. An increase in measurement error may, therefore, outweigh the benefits of a reduction in coverage and/or nonresponse error, resulting in a net increase in total error for certain survey estimates (Vannieuwenhuzye, 2014; de Leeuw, 2005).

As well as the net impact of mixing modes on total survey error, a further data quality concern relates to the confounding of different error sources. Given that each mode has its own implications for the amount of coverage, nonresponse and measurement error in survey statistics, estimates based on mixed mode data will not be strictly equivalent, making comparisons between subgroups responding in different modes

problematic (a particular concern for researchers interested in comparing estimates over time, or across countries). Disentangling the combined effects of different modes on survey estimates invariably requires complex multivariate modelling techniques (e.g. Vannieuwenhuyze, Loosveldt, and Molenberghs, 2014; Lugtig, Lensvelt-Mulders, Frerichs and Greven, 2011; Vannieuwenhuyze and Loosveldt, 2012), rendering mixed mode data less intuitive for data users. Efforts to develop methods of correcting for mode effects at the analysis stage are still in their infancy (although, see Martin and Lynn, 2011; Vannieuwenhuyze and Lynn, 2014), so while savings may be possible at the data collection stage for the survey sponsor, it is likely that the costs may be passed on to data users having to handle the increased complexity in their analyses. The most likely result is that data users simply use the data, unaware of, or turning a blind eye to, the implications confounded survey errors may have for the accuracy of their conclusions.

These various concerns highlight some of the trade-offs inherent in the decision to mix modes of data collection between different sources of survey error, and particularly, between survey errors and costs (De Leeuw, 2005). Though literature in this field is burgeoning, there is still a lack of concrete guidance available to researchers considering mixing modes about how these trade-off decisions should be approached, and about their implications for the accuracy of survey estimates. A number of studies have begun to fill this gap (e.g. Vannieuwenhuyze, 2014; Wagner et al., 2014; Klausch et al., 2015), but the need to replicate research across different national contexts is essential and to date, relatively few methodological studies have focused on the advantages and disadvantages of mixing modes in the Swiss context. The study reported here was designed to address these research needs.

1.3. The Present Study

The LIVES-FORS mixed mode experiment was designed to address a number of practical and theoretical concerns relating to the use of different modes in the Swiss context to collect data relating to topics of interest to the NCCR LIVES. The practical concerns related to the feasibility and relative costs of using different fieldwork designs to survey the general population, while the theoretical concerns related to the effects of mixing modes on different aspects of data quality. The reason for addressing more general, practical considerations was to evaluate the costs and benefits of different ways

of conducting surveys in Switzerland, with a view to developing recommendations about optimal survey designs for the future (including the possibility of tailoring fieldwork methods for different population subgroups). We were particularly interested, for example, in how best to ensure adequate coverage of people without fixed-line telephone numbers, who are at risk of being under-represented in traditional single mode survey designs, and who differ along various dimensions from those with fixed-line telephone numbers. The theoretical focus of the research was on the impact of different modes on different types of survey error (coverage, nonresponse and measurement error), and on how to evaluate the unique and combined effects of these errors on survey estimates derived from single and mixed mode data.

We do not attempt to address these multiple concerns in the present report. Rather, our main aim is to describe in detail the methodology of the study undertaken, and to present an overview of results relating to one of the proposed benefits of mixing modes that were discussed in the Introduction. Specifically, we address the questions of whether mixing modes of data collection can help to improve survey response rates compared with single mode data collection designs, and reduce the risk of nonresponse error associated with differential participation across population subgroups. We do not present results relating to how mixing modes influences coverage and measurement errors, as we address these elsewhere (see Roberts, Ernst Stähli, and Joye, 2013; and Sanchez Tome, Roberts, Ernst Stähli, and Joye, 2014). Instead, we present the final outcomes of cases assigned to each of the mode treatment groups, and calculate response and refusal rates for each survey design. We then present the composition of the samples responding in each group. In the following section, we describe in detail the research design and the methods used.

2. Methods

2.1. Sample

The population for this study was adults aged 15 years and older, resident in Switzerland. The reason for this focus was to survey the same population as the European Social Survey (ESS), which as a high quality face-to-face survey provided an important benchmark for the experiment[1]. Due to budgetary constraints, however, it was necessary to limit the scale of the study by focusing on a single linguistic region in Switzerland, and

for pragmatic reasons, we chose the French-speaking districts, referred to locally as 'Suisse romande'. Suisse romande is made up of around 1.6 million French speakers, representing about 20% of the total Swiss population. Though not precisely defined, it generally includes the Cantons of Fribourg, Geneva, Jura, Neuchâtel, western Valais, Vaud, and the northern part of Berne.

The sample for the experiment was supplied by the Swiss Federal Statistical Office (SFSO), and was drawn from their sampling frame of individuals (Stichprobenrahmen für Personen- und Haushaltserhebungen - SRPH), which is based on population registers maintained by municipalities. The SFSO makes register-based samples available for surveys funded by the Swiss Science Foundation, which are considered to be of national importance, form part of major international projects, or that have a repeated element. In this case, because the experiment was linked to the Swiss European Social Survey, the research team was able to benefit from this arrangement.

The SFSO only provides telephone numbers for sample members that are listed publicly in the telephone directory maintained by Swisscom (for its own surveys, it has access to unlisted telephone numbers also), but because of the dramatic decline in landline coverage, and because registration of telephone numbers in the directory is not compulsory (as it was until the early 1990s), telephone numbers are not available for all sample members (Ernst Stähli, 2012; Joye et al., 2012). As mentioned, an important feature of the design of our study was to assess the problem of under-coverage posed by the non-availability of unlisted and mobile telephone numbers, and to investigate which mode, or combination of modes, would provide the optimal way of surveying the part of the population that cannot easily be contacted for survey research by telephone. For this reason, our experimental design (described in detail below) treats sample members with and without publicly listed telephone numbers separately.

We made a rough estimation of required sample sizes in each treatment group of the experimental design (described in detail below) based on expected response rates in each mode of data collection. Based on these calculations, we established a sample size requirement of 3600 (2100 with telephone numbers, and 1500 without). As the SRPH consists of individuals, it was possible for the SFSO to draw a simple random sample of residents in French-speaking municipalities. In practice, this involves sampling from two bilingual (German- and French-speaking) municipalities (Fribourg and Bienne), meaning

some members (German speakers) of the gross sample would not be eligible to participate. To obtain a sufficient number of eligible cases (defined as adults aged 15 and over, resident in French-speaking municipalities, or in bilingual communes for whom the preferred language (recorded on the register) was French) to assign to the treatment groups in the experimental design, as well as to ensure a sufficient number of sample members with no listed fixed-line telephone number, we requested a gross sample from the SFSO of 4000 cases.

In the end, we were supplied 4002 names and addresses, from which it was necessary to make an additional selection to obtain a gross sample of French-speakers only. This was done by selecting only those sample members from the bilingual communes of Fribourg and Bienne for whom the preferred language was French. Eighty-three cases were German speakers living in Fribourg and Bienne, and these were subsequently dropped from the sample, leaving a total of 3919 cases. Of these, 2306 (58.8%) had one or more registered fixed-line telephone numbers. We selected a random sample of 2100 from this group, and a random sample of 1500 from the remaining 1613 (41.2%) cases without a telephone number. These cases were then randomly assigned to the treatment groups, as described in the following section. The remaining 319 cases that were not sampled at this stage were retained as a 'reserve sample', which we used as a control group to assess some of the confounding factors affecting the interpretation of results from a shortened follow-up questionnaire, which was sent to nonrespondents following the main data collection phase (described further below).

2.2. Experimental design and fieldwork procedures

The experimental design consisted of three principal mode treatments, each of which can be considered as independent survey designs: 1) a pure **mail** survey with a paper questionnaire; 2) a mixed mode sequential survey starting with a **web** survey, with one follow-up by paper questionnaire, and then a follow-up either by telephone (for those with a telephone number) or by face-to-face (computer-assisted personal interviewing – **CAPI**) interview (for those without a telephone number); and 3) a computer-assisted telephone **(CATI)** survey (for sample members with a telephone number only), with a follow-up by paper questionnaire for nonrespondents. All three mode treatments culminated in a non-response follow-up at the end of the main fieldwork period,

consisting of a considerably reduced length version of the paper questionnaire (24 items only). As the mode assignments differed depending on whether a sample member had a telephone number or not, the resulting design included five separate treatment groups, shown in Table 1. An important implication of the design is that respondents had both an *assigned* mode (their mode treatment group), and a *response* mode (the mode they finally responded in). Based on this design, we are able to compare three single mode survey designs (mail, web and CATI) by looking at respondents who answered in their assigned modes, and two mixed mode designs, involving CATI plus mail, and a web plus mail plus CATI/CAPI. In all the comparisons reported here, we further distinguish sample members with and without publicly listed telephone numbers, to ensure comparability with the CATI group, and to investigate the extent of noncoverage error in samples drawn telephone directories.

All data collection was carried out by the fieldwork agency, M.I.S. Trend SA. Fieldwork started on the 22nd November 2012, and was completed by 8th March 2013. Table 1 also shows the timetable for the data collection period, which consisted of a series of predetermined contacts, spaced apart by a set number of days, the timing of which was based on recommendations derived from Dillman, Smyth and Christian's (2009) 'Tailored Design Method'. These recommendations are based on empirically-established best practice methods for mail surveys (see Dillman (1978), but were adapted here to accommodate the sequential mixed mode design employed in the web and CATI groups. The Tailored Design Method rests on the idea that both response rates and achieved sample representativeness can be improved by developing survey procedures that establish trust and increase the perceived benefits of participation, while decreasing the expected costs of taking part (Dillman et al., 2009; p.38). Key features of this methodology include providing a token of appreciation in advance (in this case, an unconditional cash incentive of 10.- Swiss Francs), and varying contact strategies (e.g. by using different stationery for different mailings) to ensure repeated requests to participate are not ignored, and (ideally) attract different types of respondents. Details of all the mailings used in the study are provided in the appendix to this report.

The different contact strategies used in this study (summarised in Table 1) were as follows: all sample members assigned to web and mail mode were sent a pre-notification letter informing them that they would shortly be receiving an invitation to participate in a

survey, and that they would receive a small thank-you gift. Sample members assigned to the CATI group received a pre-notification letter directly with the cash incentive informing them they would soon be contacted by telephone for an interview (telephone contact attempts and interviews began a few days later). For the web and mail groups, the pre-notification letter was followed three days later by a second letter with the cash incentive, plus the paper questionnaire and a return envelope (for the mail group), or a link to the web questionnaire and log-in details (for the web group). Letters contained details of a toll-free hotline telephone number people could use to contact the fieldwork agency if they did not wish to be contacted further, or (in the case of those assigned to the web group) if they wanted to request a paper version of the questionnaire. One week after the initial mailing, all sample members assigned to the web and mail groups were sent a reminder postcard, to thank those who had already participated, and to motivate those who had not already done so, to complete the questionnaire. Three weeks later, sample members in all mode groups who had not responded so far in their assigned mode were mailed a paper questionnaire and return envelope (a replacement questionnaire was sent to the mail group). This mailing was the last contact in 2012.

The final contact for the mail and CATI groups took place at the start of 2013, and consisted of a letter plus the nonresponse follow-up questionnaire (together with a pen with the LIVES logo on it as a small incentive, and a return envelope). Because responses to the nonresponse follow-up questionnaire are likely to be affected (e.g. see Vandenplas, et al., 2015) by a) context effects associated with reducing the length and changing the order of questions in the questionnaire (Schwarz and Sudman, 1992); and b) possible timing of fieldwork effects (answering questions about well-being in the weeks following the Christmas vacation may lead to different answers to those given in the week preceding the holidays), we decided to make use of our small reserve sample to provide a control group for the main sample. The reserve sample was, therefore, mailed the short questionnaire and a pen at the same time as the mail and CATI nonrespondents. At this point, nonrespondents in the web group without telephone numbers were sent a letter informing them that an interviewer would visit them to attempt to complete an interview at their home. Web nonrespondents with telephone numbers immediately started being contacted by telephone interviewers. (Nonrespondents in the reserve sample were sent a

reminder postcard one week later, and a replacement questionnaire three weeks following the initial mailing to try to boost response.)

For both the CATI group and the web-CATI group, CATI interviewers were instructed to make up to 50 contact attempts before assigning 'non-contact' as the final outcome. CAPI interviewers were instructed to make a minimum of 5 face-to-face visits before a case could be defined as a non-contact. The only deviation from the timetable shown in Table 1 concerned the date of the final contact with the web groups. Because the face-to-face fieldwork did not progress as quickly as hoped, the mailing of the non-response questionnaire for the web-CAPI sub-group was postponed for one week to allow time for the final contact attempts to be made (the mailing for the web-CATI sub-group was sent on the 8th February, as scheduled). As mentioned, all letters and postcards used in the fieldwork can be found in the Appendix.

Table 1: Experimental Design

Samples		Assigned Mode	N	1st Contact Day 1	2nd Contact Day 4	3rd Contact Day 11	4th Contact Day 26	5th Contact Day 50	6th Contact Day 85
With listed telephone number (n=2100)	1	**CATI**	600	Pre-notification letter with incentive	Telephone call attempts for a CATI interview		Letter + paper questionnaire	Letter + NR paper questionnaire	End
	2	**Mail**	500	Pre-notification letter	Letter + paper questionnaire + incentive	Postcard reminder	Letter + paper questionnaire	Letter + NR paper questionnaire	End
	3	**Web**	1000	Pre-notification letter	Letter with link to web survey + incentive	Postcard reminder (with web link)	Letter (+ web link) + paper questionnaire	Telephone call attempts for a CATI interview	Letter + NR paper questionnaire
Without listed telephone number (n=1500)	4	**Mail**	500	Pre-notification letter	Letter + paper questionnaire + incentive	Postcard reminder	Letter + paper questionnaire	Letter + NR paper questionnaire	End
	5	**Web**	1000	Preotification letter	Letter with link to web survey + incentive	Postcard reminder (with web link)	Letter (+ web link) + paper questionnaire	Letter announcing face-to-face visits for CAPI interview	Letter + NR paper questionnaire

2.3. Questionnaires

Cover story

Participants in the study were not made aware of the fact that they were participating in a methodological experiment. The 'cover story' for the experiment was a survey on the well-being of residents in Suisse romande (and indeed, the data may be used for this purpose, if they are handled appropriately given their mixed mode nature). The survey was called 'Bien-être et Mal-être en Suisse romande' (translated literally as, 'Well-being and Unease in French-speaking Switzerland'), and all the letters sent to respondents emphasised this as the survey topic, and the fact that the study formed part of the research activities of the NCCR 'LIVES'. This was a deliberate tactic, as our research questions were focused on how a potentially sensitive survey topic central to the study of vulnerability would impact on decisions to participate across different subgroups, as well as on responses to survey questions, and how this, in turn, might vary across data collection mode. A news article on the LIVES website was published soon after the start of the data collection period, which reinforced this cover story: http://www.lives-nccr.ch/en/actualites/201211.

Source questions

The questionnaire was made up of questions from two principal sources. The first source included questionnaires developed by other LIVES Individual Projects[2] for the purposes of their own surveys. We parsed all the available LIVES questionnaires to first identify common themes relevant to the measurement of well-being, and selected candidate questions related to each theme. The second source was the Round 6 ESS questionnaire, and in particular, its module on Personal and Social Well-being. In general, we selected questions on themes common to LIVES surveys and the ESS module, and favoured ESS versions of questions common to both, because they had previously been validated across multiple countries when the same module was fielded in Round 3 of the ESS. The motivation for using pre-existing LIVES questions was to investigate the possible impact of mode on key measures of interest to the NCCR (some of which have been fielded in mixed mode surveys). The motivation for selecting items from the ESS was to be able to make comparisons with a high quality face-to-face benchmark survey (fieldwork for Round 6 was being

carried out during the same period as the mixed mode experiment). The source of all questions included in the questionnaire is provided in the 'Question Mapping' table in the appendix.

Questionnaire length and content

The questionnaire was designed to take around 25-30 minutes to complete, and consisted of a total of ≈125 items, of which 44 were measures of the respondents' socio-economic background (the ESS and MOSAiCH were the principal source surveys for these questions), and 41 items were measures of well-being. The well-being related items included measures of health and health-related behaviours (smoking and drinking), measures of happiness and life satisfaction, negative and positive affect, stress, social support and experienced life events. In addition to substantive measures of well-being, we included some more general questions (around 20 in total) on society in general, including measures of social trust, fear of crime, political interest and participation, perceptions of inequality, and attitudes towards immigration, some of which were of interest as they have been found elsewhere to be correlates of survey participation. Finally, we included around 20 items designed to address methodological questions relating to survey attitudes, mode preferences, and social desirability bias (the latter being particularly relevant to evaluating measurement effects on sensitive survey items).

Adaptation to different modes

We simultaneously developed four versions of the questionnaire, making small adaptations to the questions as needed to make them suitable for administering in each of the different modes. As questions from the ESS made up the large majority of the questions in the questionnaire, our primary mode was face-to-face interviewing. Face-to-face questions from the ESS are typically administered with the aid of showcards on which respondents can read the available response alternatives. To adapt showcard questions to make them suitable for telephone administration, several changes were necessary depending on the nature of the question and response alternatives. For example, for questions with response scales, it was necessary to provide a description of the response scale and its labels in the question stem for interviewers to read out. For more complex items, such as questions asking about income, and main activity,

more substantial changes were needed, including asking the question as an open-ended question (in the case of income) and breaking down the question into one or more separate items (in the case of main activity). More substantial changes such as these may well have implications for comparability across the modes, and should be investigated before proceeding with analyses.

To adapt the questionnaire to make it suitable for web administration, we followed specifications and guidelines developed by the Core Scientific Team of the ESS for participants in their mixed mode research programme (available on request). These guidelines included recommendations regarding the introduction to the survey; the survey's URL, usernames, passwords, and options for pausing and resuming questionnaire completion; response formats; validity checks for numerical responses and open-ended questions; the visual design of the questionnaire (e.g. the use of bold and regular fonts, on-screen formatting of questions and response options); paging (the questionnaire used a screen-by-screen format with each item on a separate screen, except for questions with a branching format and batteries of items sharing a response scale); and importantly, the treatment of item nonresponse (see below). Details of important differences between the questionnaires in each mode are included in the Question Mapping table in the appendix.

Having followed the ESS guidelines for adapting the face-to-face questionnaire for web administration, we were able to produce a paper questionnaire based on the design of the web instrument. Once again following recommendations by Dillman et al. (2009), our aim was to create a common visual stimulus for both the web and paper questionnaires, to minimise the likelihood of mode effects. The paper questionnaire was formatted so that its visual design was similar to that of the web instrument (e.g. using radio buttons for response options instead of check boxes), and the layout was designed to be both visually attractive and user-friendly. The homepage for the web questionnaire and the cover of the paper questionnaire both showed the title of the survey 'Bien-être et Mal-être en Suisse romande', with the sub-heading 'Une initiative pour comprendre comment la vie se passe pour les habitants de notre région' ('An initiative to understand how life is going for the residents of our region'), the University of Lausanne and LIVES logos, and a red 'tangram' of a figure running across the page, which is part of the LIVES visual identity (see screenshots in the appendix).

A few additional specific adaptations were necessary for the paper version of the questionnaire, including notably, some routing instructions to guide respondents to applicable questions. Simple instructions saying e.g. 'Go to Q18' were provided next to the relevant response alternatives from which the respondent was routed. We deliberately kept routing to a minimum in the questionnaire to avoid over-complicating the paper questionnaire.

Treatment of item nonresponse

The ESS questionnaire does not (except in specific instances) permit explicit 'Don't Know' response options, but the interviewer is allowed to code spontaneous 'Don't Know' answers given by the respondent. This poses a problem for how to collect equivalent data in self-administered modes. Offering respondents an explicit 'Don't Know' response would change the distribution of substantive answers to the question (Krosnick, et al., 2002). On the other hand, not offering a 'Don't Know' option could lead respondents to skip the question, and the reasons for the item nonresponse would be unclear to the researcher. The ESS-recommended solution that we adopted in our web questionnaire was to show an error message when a respondent attempted to skip a question, offering them the opportunity to answer 'Don't Know' or 'Prefer not to answer' if these responses were more appropriate, or to skip the question altogether. In the paper questionnaire, we were not able to resolve this problem, but instead elected to offer explicit 'Don't Know' response alternatives for a selection of questions where it was genuinely possible to not know the answer. For all other items skipped by respondents, we will not be able to determine the reasons for item non-response as is possible in the other modes.

In practice, these differences in the treatment of item nonresponse across the modes translated into differences in the rates of missing answers in each of the three modes. Overall, item nonresponse rates were lowest in the web survey (affecting a mean of 0.5 items), and highest in the mail survey (affecting a mean of 3.2 items), and the rate for CATI was between these two (affecting a mean of 1.4 items). However, there was variation depending on the type of nonresponse. Explicit refusals to respond to specific items were rare and similar in number across the three modes. No answers were virtually non-existent in web and CATI modes, but affected a mean of 2.5 items in the mail questionnaire. Meanwhile, don’t knows were highest in the telephone

survey, affecting a mean of 1.2 items. These mostly affected sensitive attitudinal measures (e.g. on immigration and discrimination), and some more complex methodological items in which respondents were asked to assess the sensitivity of the questions (for the purpose of evaluating the risk of social desirability bias on different topics). In the paper questionnaire, questions presented in grids were more likely to suffer from missing answers than other items. In common with other studies, the income measures were particularly prone to being left unanswered (although considerably less so by the web respondents). For example, for the item measuring *household* income, the treatment groups varied as follows: mail (31% respondents refused to answer this item), CATI (22%), and web (7%)[3]. These results compare with a missing rate of 17% on the household income measure in the main ESS face-to-face survey.

Nonresponse follow-up questionnaire

As mentioned, the final step in our contact strategy was to send nonrespondents a reduced length questionnaire (see Appendix). The aim of the nonresponse questionnaire was to obtain information from as many nonrespondents as possible about key survey variables likely to be at risk of nonresponse bias. We selected 24 items from the main questionnaire on the grounds that they seemed likely to correlate with both, or either, the decision to participate in the survey and/or different aspects of vulnerability.

3. Results

To recap, in this report we present key results from the experiment to show how different single and mixed mode data collection designs affect survey response rates in the Swiss context. We also look at how mixing modes of data collection can affect the composition of the responding sample, and with it, the potential for nonresponse bias related to the under- or over-representation of specific population subgroups. To address these twin concerns, we firstly present the final dispositions of cases assigned to each of mode treatment groups, and response and refusal rates for each of the survey designs, taking into consideration the impact of subsequent fieldwork efforts. Secondly, we present the composition of the samples responding in each group by looking at frequency distributions across socio-demographic variables present on the sampling frame.

3.1 Composition of the gross sample

Before presenting the survey outcome rates and results relating to the impact of non-response on quality, we first provide preliminary information about the socio-demographic composition of the samples, using data from the sampling register. The purpose of this exercise is to summarise how people with listed telephone numbers differ from people without, and to highlight the possible implications this may have for surveys interested in studying vulnerable populations.

Table 2 shows frequency distributions across a selection of socio-demographic variables for the gross sample (experimental sample plus the reserve sample), with significant differences between the sub-samples with and without telephone numbers highlighted in column 4, and differences between the sample with telephone numbers and the gross sample highlighted in column 5. The latter in particular provide information about the extent of coverage error that would be present in surveys relying on samples based on the public telephone directory, or surveys using telephone interviewing as the sole mode of data collection (based on the assumption that no efforts are made to supplement the available telephone numbers, which can go some way to reduce error – see Lipps, Pekari, and Roberts, 2015).

We use Chi-square Tests of Independence to compare the subsamples with and without telephone numbers. Statistically significant differences are evident on all variables. Compared with people without listed telephone numbers, people with listed telephone numbers are more likely to be female, to be married or have been married, and to live in multi-person households (4 people plus). They are also older on average (aged 45 and older), and more likely to have been born in Switzerland, and/or to have Swiss nationality. They are also more likely to live in suburban or rural areas and less likely to live in city/town centres.

To compare the sample with listed numbers to the gross sample we use non-parametric one-sample chi-square tests (see Parke, 2013), which assesses how well the distribution of the responding sample across a number of socio-demographic variables reflects the distribution for all sample members (ibid., p.12). The results indicate that the CATI sample is significantly different to the gross sample on age (X^2 (df = 3) = 81.19, p < .001), with fewer individuals aged 44 and under, and more individuals aged 65+); marital status (X^2 (df = 3) = 20.25, p < .001), with more married and fewer single individuals; nationality and country of birth (X^2 (df = 3) = 82.12, p < .001), with more Swiss nationals and fewer citizens from non-bordering

countries; household size (X^2 (*df* = 5) = 29.16, $p < .001$), with fewer individuals living in single-person households; and on regional indicators – e.g. with fewer individuals living in city/town centres (X^2 (*df* = 3) = 10.86, $p < .05$). These differences would, therefore, result in an under-representation in surveys of the young, unmarried, foreign, and urban population(s).

3.2 Outcome rates across treatment groups

To present the outcome rates from the experiment, we follow the recommendations of the American Association for Public Opinion Research (AAPOR, 2011). Our analysis is based on administrative data provided by the fieldwork organisation, which provides a detailed record of all contact attempts (in all modes of data collection) made with sampled cases, as well as contacts initiated by respondents and non-respondents with either the fieldwork organisation, or the University of Lausanne[4]. These data provide information about the timing and outcome of contacts, and enable us to examine in detail what happened to each sample member during data collection, and what their final status was at the end of the survey fieldwork. Contact records in a mixed mode survey context are inherently complex, and can be analysed in a variety of ways. In the following, we only provide results of a preliminary analysis aimed at identifying final dispositions for all sample members, using standard definitions proposed by AAPOR to enable the calculation of response rates.

Table 2: Socio-demographic characteristics of the gross sample, those with listed telephone numbers and those without

Auxiliary Variables	(1) Gross sample supplied by the OFS[1] (n=3919)		(2) With listed telephone numbers (n=2306)		(3) Without listed telephone numbers (n=1613)		(4) Difference b/w telephone and no telephone samples (3)-(2)		(5) Difference b/w telephone and gross sample (2)-(1)	
	%	Std. Err.	%	Std. Err.	%	Std. Err.	%	Sig.[5]	%	X^2 [6]
Male	49.5	(0.8)	47.9	(1.0)	51.8	(1.2)	3.9	*	-1.6	
Age (mean in years)	47.1	(0.3)	50.2	(0.4)	42.6	(0.4)	-7.58	***	3.1	
Age group										81.2***
<30	21.1	(0.7)	18.9	(0.8)	24.1	(1.1)	5.2	***	-2.2	
30-44	26.7	(0.7)	20.8	(0.9)	35.1	(1.2)	14.3	***	-5.9	
45-64	31.8	(0.7)	33.6	(1.0)	29.2	(1.1)	-4.4	**	1.8	
65+	20.5	(0.7)	26.8	(0.9)	11.6	(0.8)	-15.2	***	6.3	
Marital Status										20.3***
Single	32.4	(0.8)	29.1	(1.0)	37.0	(1.2)	7.9	***	-3.3	
Married	52.8	(0.8)	56.2	(1.0)	47.9	(1.2)	-8.3	***	3.4	
Widowed	5.0	(0.4)	6.0	(0.5)	3.7	(0.5)	-2.3	**	1.0	
Divorced	9.6	(0.5)	8.4	(0.6)	11.3	(0.8)	2.9	**	-1.2	
Legal partnership[2]	0.2	(0.1)	0.3	(0.1)	0.2	(0.1)	-0.1		0.1	
Nationality										82.1***
Swiss	70.7	(0.7)	78.9	(0.9)	59.0	(1.2)	-19.9	***	8.2	
Bordering country	10.0	(0.5)	8.7	(0.6)	11.9	(0.8)	3.2	**	-1.3	
Other	19.3	(0.6)	12.4	(0.7)	29.1	(1.1)	16.7	***	-6.9	
Country of birth[3]										74.7***
Switzerland	61.4	(0.8)	69.9	(1.0)	49.3	(1.3)	-20.6	***	8.5	
Bordering country	12.5	(0.5)	11.2	(0.7)	14.2	(0.9)	3.0	**	-1.3	
Other	26.1	(0.7)	18.9	(0.8)	36.3	(1.2)	17.4	***	-7.2	
Residence Permit										94.1***
CH, no permit needed	70.7	(0.7)	78.9	(0.9)	59.0	(1.3)	-19.9	***	8.2	
C – settlement permit[4]	18.1	(0.6)	15.9	(0.8)	22.7	(1.0)	6.9	***	-2.9	
B – residence permit	10.1	(0.5)	5.1	(0.5)	17.2	(0.9)	12.1	***	-5.0	
L – short term residence permit	0.5	-	0.0	-	1.1	-	-		-0.5	
Household size (mean persons)	2.8	(.02)	2.9	(.03)	2.6	(.03)	-0.32	***	0.1	29.2***[7]
Urbanisation										10.9*
City/town centre	31.3	(0.7)	28.1	(0.9)	35.8	(1.2)	7.7	***	-3.2	
City/town suburbs	41.1	(0.8)	42.6	(1.0)	39.0	(1.2)	-3.6	*	1.5	
Isolated town	1.0	(1.6)	1.1	(2.2)	0.8	(2.0)	-0.3		0.1	
Rural community	26.6	(0.7)	28.1	(0.9)	24.4	(1.1)	-3.7	**	1.5	

Notes: [1]Not including 84 residents of the bilingual communes Bienne and Fribourg whose preferred language was listed as German. [2]Includes current and dissolved legal partnerships. [3]28 cases have no country of birth data recorded. [4]Includes family members of intergovernmental organisations. [5]Chi-square tests of independence: *** p<.001, **p<.01, *p<.05. [6] Non-parametric one-sample chi-square tests of goodness of fit: *** p<.001, **p<.01, *p<.05. [7]Chi-square test on distribution across household sizes of 1, 2, 3, 4, 5, and 6+.

The reporting of response rates, refusal rates and contact rates typically depends on the possibility of identifying four types of final disposition: a) interviews/ completed questionnaires; b) eligible cases that are not interviewed (non-respondents); c) cases of unknown eligibility; and d) cases that are not eligible (AAPOR, 2011; p.7). This requires a clear definition of eligibility, which will vary depending on the population for the study and how and when the sample was drawn. In the present study, because we had access to the population register-based sampling frame and the possibility of a sample of named individuals, being eligible to participate was defined as being aged 15 or over, and resident in private households in French-speaking municipalities, on the day that the sample was drawn. The SFSO regularly updates the SRPH. For the present study, the sample was drawn from a version of the SRPH last updated on the 31st August 2012. As fieldwork did not get underway until November 2012, inevitably during the intervening period some sampled individuals had moved home (making them more difficult, or impossible to contact), died, or moved out of the region of interest or Switzerland altogether (making them no longer eligible to participate) by the time the data collection started. Such changes may equally have occurred later during the data collection period itself, affecting the eligibility of sampled cases as fieldwork progressed. As we cannot determine the dates when the eligibility status of cases may have changed, we assume that all named individuals sampled for the study were eligible to participate, so there is no need to identify not-eligible cases or cases of unknown eligibility (nor estimate the likely proportion of eligible cases among cases of unknown eligibility – see AAPOR, 2011).

Final dispositions may reflect the results of the last contact attempt made (e.g. if the contact resulted in a completed interview/questionnaire, a so-called ‘hard’ refusal or identified some other status preventing further participation in the study (e.g. being deceased, having a long-standing illness or disability, being unable to communicate in French, or an incorrect address listing). In other cases, the last contact attempt made to a sampled case may result in an outcome that provides less information about the ability or willingness of a sample member to participate in the study than a previous contact. For this reason, we base our analysis on all available contact information, to determine final disposition (as required) from prior contact attempts providing the most information about a case. Again, we follow the AAPOR guidelines, which propose a hierarchy of outcomes giving precedence to attempts involving human

contact (i.e. prior refusals supersede noncontacts, as does information establishing ineligibility, or in the case of the present study, a sampled person's inability to participate during the designated fieldwork period) (*ibid.,* p.11).

Table 3 provides a summary of the final disposition of cases in each of the experimental treatment groups, broken down by whether or not respondents had listed telephone numbers available from the sampling register or not. Because all cases included in the study are assumed to be eligible, response rates can simply be calculated as the total number of completed interviews divided by the total number cases in each group (see row highlighted in grey). We provide two response rates for each group – one based on the total number of complete interviews/questionnaires obtained, and the other based on the addition of completed NRFU questionnaires. Overall, the highest response rates for complete interviews were obtained in the CATI and the web groups (70.6% and 70.2% respectively), and the mail group was 5 percentage points lower at 65.4%. Differences are evident based on whether or not the sample member had a listed telephone number or not. Among those with telephone numbers, each of the survey designs produced similar response rates. However, response rates were significantly lower among sample members with no telephone number assigned to mail mode (60.6% compared with 70.2% for those with telephone numbers assigned to this mode). They were only slightly lower, however, among sample members with no telephone number assigned to web mode (69.7% compared with 71.4%). These findings suggest that the use of additional modes to follow up non-respondents is an effective way of reducing nonresponse rates, particularly among the subpopulation with no listed telephone numbers (the CAPI phase producing more additional interviews than the CATI phase used for those with telephone numbers).

As with response rates, we can also calculate refusal rates for the survey by simply dividing the number of refusals by the total number of cases assigned to each group. The refusal rate was highest in the web group (13.6%), particularly among those re-contacted at the CATI and CAPI phase, and lowest among sample members assigned to the mail group (7.9%). The refusal rate for the CATI group was 11.0%. Again, there are noteworthy differences among those with and without listed telephone numbers assigned to the mail group. Note, however, that comparisons across modes on this metric are problematic because the refusal rate includes only

explicit refusals for the mail and web phases[5]. In mail surveys, an implicit way of refusing to participate is to simply ignore the mailings and never respond – an outcome that we do not include in the refusal rate here. Instead, people who never responded to mailings are included in the non-contact (for other reasons) rate. Similarly, because of the difficulty of knowing whether contact has actually been made with the target respondent in a mail survey, we are not able to compare contact rates across the mode treatment groups. Information about the ability of sample members to participate was obtained from a mix of contacts with the target respondent and members of his/her household and returns from the post office. For this reason, it is not always clear what qualifies as a successful contact in each of the different modes, so we do not attempt to analyse contact rates further. Not surprisingly, we can conclude that the use of interviewer-administered modes decreases non-contact rates and allows more information to be gained about the ability and motivation of sample members to participate in the survey.

We can further address the question of whether mixing modes helps to reduce overall nonresponse rates more effectively than single mode follow-ups by comparing the number of completed interviews across different phases of data collection between the groups assigned to mail and the groups assigned to web and CATI. To do this, we break down the total fieldwork period into the following 5 phases:

> *Phase 1* – 19 – 29th November, 2012: Week 1 of fieldwork from the mailing of the advance letter, to the mailing of the postcard reminder in the mail and web groups.
>
> *Phase 2* – 30th November – 14th December, 2012: Weeks 2 and 3 of fieldwork, from the mailing of the postcard reminder (web and mail) to the mailing of the reminder letter and the paper questionnaire (all groups).

Table 3 : Final dispositions by treatment group

	CATI		Mail						Web					
			With Listed Numbers		Without Listed Numbers		Total		With Listed Numbers		Without Listed Numbers		Total	
	n	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**
Gross sample	600	100	500	100	500	100	1000	100	1000	100	1000	100	2000	100
Completed interviews														
CATI	364	60.7	-	-	-	-			42	4.2	-	-	42	2.1
Web	-	-	-	-	-	-			457	45.7	432	43.2	889	44.5
Mail	57	9.5	351	70.2	303	60.6	654	65.4	215	21.5	184	18.4	399	20.0
CAPI	-		-	-	-	-					81	8.1	81	4.1
Nonresponse Follow-up	36	6.0	22	4.4	28	5.6	50	5.0	37	3.7	24	2.4	61	3.1
Total Interviews	421	70.2	351	70.2	303	60.6	654	65.4	714	71.4	697	69.7	1411	70.6
Total including NRFU	457	76.2	373	74.6	331	66.2	704	70.4	751	75.1	721	72.1	1472	73.6
Non-response														
Refusal	66	11.0	51	10.2	28	5.6	79	7.9	151	15.1	121	12.1	272	13.6
Deceased	1	0.2	-	-	2	0.4	2	0.2	2	0.2	2	0.2	4	0.2
Language problem	19	3.2	1	0.2	3	0.6	4	0.4	16	1.6	6	0.6	22	1.1
Disabled or long-term illness	13	2.2	5	1.0	1	0.2	6	0.6	27	2.7	4	0.4	31	1.6
Absent during fieldwork period	11	1.8	1	0.2	1	0.2	2	0.2	12	1.2	4	0.4	16	0.8
Moved abroad	-	-	-	-	2	0.4	2	0.2	-	-	4	0.4	4	0.2
No longer at address	7	1.2	3	-	18	3.6	21	2.1	8	0.8	56	5.6	64	3.2
Non-contact[1] (other reasons)	26	4.3	66	13.2	114	22.8	180	18.0	33	3.3	82	8.2	115	5.8

Notes: [1]Includes cases who never responded to questionnaire mailings.

Phase 3 – 15th December 2012 – 4th/8th January 2013: Weeks 5 to 7 of fieldwork, from the mailing of the paper questionnaire (all groups) to the mailing of a letter announcing the visit of the face-to-face interviewer/ start of telephone call attempts (web group only), or to the mailing of the NRFU questionnaire (mail and CATI groups).

Phase 4 – 5th/9th January – 11th/18th February 2013: Weeks 8 to 13/14 of fieldwork, from the mailing of the NRFU questionnaire to the mail and CATI groups to the mailing of the NRFU to the web groups[6].

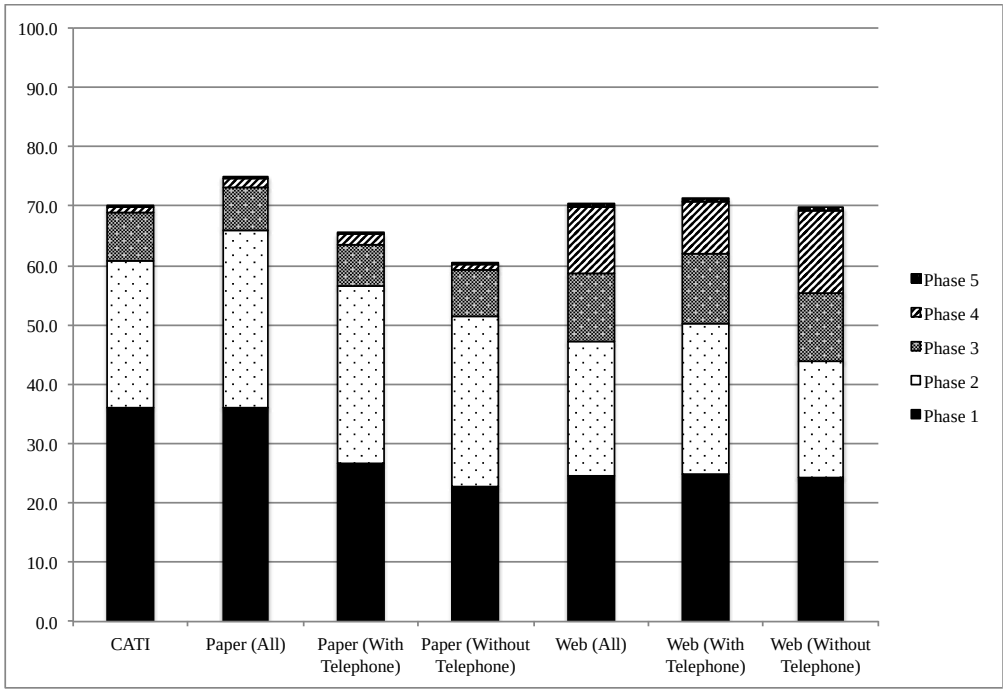
Phase 5 - 12th/19th February 2013 - 31st May 2013: Weeks 13/14 to 21 of fieldwork, from the mailing of the NRFU questionnaire to the web groups to the end of the fieldwork period.

Figure 1 shows the evolution in the response rate across the five fieldwork phases, for the groups assigned to the different survey designs. The 'control' groups are those assigned to mail, a single mode survey. The comparison groups are those assigned to web and CATI, which both involved mixed mode follow-ups of nonrespondents.

For sample members with telephone numbers, the single-mode mail survey design and the mixed mode CATI plus mail design were equally effective in the first two phases of the survey, achieving response rates of 60% after only three weeks of fieldwork. By contrast, response rates for the web group were only 50% at the end of these phases. The (re-)mailing of the paper questionnaire in phase 3 made it possible to boost response rates in both the CATI and the mail group by a further ≈10% by the end of the fieldwork period to reach a total of 70% in both groups. In the web (with telephone numbers) group, however, the final response rate slightly exceeded that for the CATI and mail groups, thanks to the sequencing of modes, firstly with the paper questionnaire, and then by CATI. The paper questionnaire follow-up for the web group added just over 10% to the response rate, but an additional ≈10% was obtained with the CATI follow-ups in phase 4. While the mixed mode strategy of chasing nonrespondents proved effective in this study, for sample members with telephone numbers, a single mode mail-only survey design represented the most cost-efficient survey design (Roberts, Ernst Stähli, and Joye, 2013).

For the sample members without telephone numbers the progression of fieldwork was somewhat different. As mentioned, the response rate in the mail group was already lower than that for the group with telephone numbers, but this difference emerged early on in the fieldwork. In the group with phone numbers, the response rate after the first phase was 30.6%, while in the group without phone numbers, the response rate was only 22.8% after phase 1. Following the mailing of the reminder postcard response rates in both groups more than doubled, but at this phase there was a 10 percent point difference between the two subsamples. Following the mailing of the reminder questionnaire and the NRFU questionnaire, response rate gains were slightly higher in the no-telephone group, but this progress did not succeed in closing the gap between the groups. In the web sample with no phone numbers, the response rates following each phase of the fieldwork were more similar to, but again, lower than those for the group with phone numbers. However, the CAPI phase proved to be more effective than the CATI phase for the latter group, and this helped to reduce the gap in the final response rate between the two groups.

Figure 1: Response rates by assigned mode and fieldwork phase (main questionnaire only)



3.3 Composition of the responding samples

Next, we present information about the composition of the responding samples in each of the experimental treatment groups, in order to find out whether the improvements in response rates resulted in a reduction in nonresponse bias using auxiliary data from the register-based sampling frame. To enable a general comparison between single and mixed mode survey designs, our assessment of nonresponse bias is based on the difference between the gross sample and a) the sample responding in their assigned mode, and b) the sample responding after all follow-ups in alternative modes at different phases of the survey. Again, we treat the sample with listed telephone numbers separately from the sample without.

Overall, the results are encouraging, with relatively few differences observed between the responding samples and the gross sample on the register variables. For the sample with listed telephone numbers (see Table 4), the sample responding to the single-mode mail survey differed significantly from the gross sample on two variables: marital status (X^2 (*df* = 3) = 12.81, $p < .01$), and the proportion living in the Geneva-Lausanne metropolitan area (X^2 (*df* = 1) = 4.18, $p < .05$). Single people and people living in the Geneva-Lausanne area are under-represented, while married people and people outside of the metropolitan area are over-represented. This latter finding was indicative of the fact that mail respondents were slightly more likely (approaching significance) to live in rural areas compared to urban ones (X^2 (*df* = 3) = 7.04, $p < .1$). The sample with telephone numbers responding by web differed significantly from the gross sample on age (X^2 (*df* = 3) = 38.93, $p < .001$), marital status (X^2 (*df* = 3) = 10.04, $p < .05$), and household size (X^2 (*df* = 3) = 17.17, $p < .01$). The web survey over-represents people aged less than 30 and those aged between 45 and 64, while it under-represents those aged 65 or older. Single people are over-represented, as are those living in multi-person (4+) households, while those living in single person households are underrepresented. However, the use of additional modes (mail and CATI) for the web group with phone numbers helped to eliminate all bias on the register variables. Unlike respondents in the mail and web surveys, the CATI respondents were significantly more likely to be of Swiss nationality (X^2 (*df* = 2) = 8.73, $p < .05$) and also differed from the gross sample on age (X^2 (*df* = 3) = 8.07, $p < .05$). The addition of the mail follow-up failed to eliminate these biases (X^2 (*df* = 2) = 9.39, $p < .01$; and X^2 (*df* = 3) = 7.61, $p < .1$ respectively for nationality and age). The

final responding sample in the CATI group slightly over-represented the under 30s, and under-represented the over 65s, as well as Swiss citizens.

For the sample without listed numbers, we can compare those responding by web to respondents to the mail survey returning the questionnaire either in phases 1 or 2 (see Table 5). The web survey significantly over-represents people with Swiss nationality (X^2 (df = 2) = 29.45, $p < .001$). By contrast, the initial response to the mail survey slightly underrepresents men (X^2 (df = 1) = 3.55, $p < .1$) and people living in city centres ((X^2 (df = 2) = 6.08, $p < .05$) and consequently, the metropolitan Geneva-Lausanne region (X^2 (df = 1) = 2.88, $p < .1$)). Though these differences are only significant at the 5-10% level, given the comparatively smaller sample size for this treatment group, it is possible that they would translate into larger biases in surveys with a larger sample size. The differences persist after phase 3 following the mail-out of the reminder questionnaire, and the CAPI follow-ups for the web group. Specifically, the mail survey continues to under-represent the urban population (X^2 (df = 2) = 6.52, $p < .05$), and the web plus mail, and web plus mail plus CAPI response groups continue to over-represent Swiss people ((X^2 (df = 2) = 26.70, $p < .001$) and (X^2 (df = 2) = 10.57, $p < .01$) respectively). Thus, unlike for the sample with telephone numbers, the mixing of modes for the web group makes little difference to the composition of the responding sample on the observed variables.

Table 4: Composition of responding samples by treatment group and samples responding in primary assigned mode (sample members with telephone numbers)

	(1) Gross sample with telephone numbers (n=2100)		(2) CATI Respondents by CATI (n=364)			(3) CATI All respondents in treatment group (n=421)			(4) Mail All respondents in treatment group (n=351)			(5) Web Respondents by Web (n=457)			(6) Web All respondents in treatment group (n=714)		
Auxiliary Variables	**%**	**Std. Err.**	**%**	**Std. Err.**	*p*	**%**	**Std. Err.**	*p*	**%**	**Std. Err.**	*p*	**%**	**Std. Err.**	*p*	**%**	**Std. Err.**	*p*
Male	47.4	(1.1)	47.0	(2.6)		48.7	(2.4)		45.9	(2.7)		49.0	(2.3)		47.9	(1.9)	
Age (mean in years)	50.3	(0.4)	48.5	(1.0)		47.6	(0.9)		50.3	(1.0)		45.1	(0.8)		49.2	(0.7)	
Age group					*			*						***			
<30	18.5	(0.9)	21.2	(2.1)		21.1	(2.0)		16.2	(2.0)		24.5	(2.0)		21.0	(1.5)	
30-44	20.8	(0.9)	18.7	(2.1)		21.6	(2.0)		23.4	(2.3)		22.3	(2.0)		19.5	(1.5)	
45-64	33.9	(1.0)	37.6	(2.5)		36.3	(2.4)		35.0	(2.6)		38.7	(2.3)		34.3	(1.8)	
65+	26.9	(1.0)	22.5	(2.2)		20.9	(2.0)		25.4	(2.3)		14.4	(1.7)		25.2	(1.6)	
Marital Status											**			*			
Single	28.9	(1.0)	28.8	(2.4)		30.2	(2.2)		24.8	(2.3)		34.1	(2.2)		30.1	(1.7)	
Married	56.9	(1.1)	58.0	(2.6)		56.3	(2.4)		64.4	(2.6)		54.3	(2.3)		55.6	(1.9)	
Nationality					*			*									
Swiss	79.2	(0.9)	86.3	(1.8)		84.6	(1.8)		81.5	(2.1)		81.6	(1.8)		81.1	(1.5)	
Bordering country	8.5	(0.6)	6.6	(1.3)		7.6	(1.3)		7.7	(1.4)		9.2	(1.4)		8.7	(1.1)	
Other	12.3	(0.4)	7.1	(1.4)		7.8	(0.8)		10.8	(1.7)		9.2	(1.4)		10.2	(1.1)	
Household size														**			
1	16.0	(0.8)	14.6	(1.9)		15.4	(1.8)		14.2	(1.9)		10.7	(1.5)		13.6	(1.3)	
2	30.6	(1.0)	27.7	(2.4)		26.8	(2.2)		30.8	(2.5)		28.0	(2.1)		31.4	(1.7)	
3	18.6	(0.9)	20.6	(2.1)		20.0	(2.0)		19.7	(2.1)		18.6	(1.8)		17.9	(1.4)	
4+	34.8	(1.0)	37.1	(2.5)		37.8	(2.4)		35.3	(2.6)		42.7	(2.3)		37.1	(1.8)	
NUTS region																	
Région Lémanique	71.0	(1.0)	69.5	(2.4)		69.6	(2.3)		68.9	(2.5)		69.1	(2.2)		70.3	(1.7)	
Espace Mitteland	29.0	(1.0)	30.5	(2.4)		30.4	(2.3)		31.1	(2.5)		30.9	(2.2)		29.7	(1.7)	
Urbanisation											†						
City/town centre	28.4	(1.0)	26.9	(2.3)		27.3	(2.2)		23.1	(2.3)		26.3	(2.1)		27.7	(1.7)	
City/town suburbs	42.8	(1.1)	42.6	(2.6)		42.3	(2.4)		43.3	(2.7)		45.3	(2.3)		43.4	(1.9)	
Isolated town	1.1	(0.2)	1.6	(0.7)		1.9	(0.7)		0.9	(0.5)		0.7	(0.4)		0.6	(0.3)	
Rural community	27.7	(1.0)	28.8	(2.4)		28.5	(2.2)		32.8	(2.5)		27.8	(2.1)		28.3	(1.7)	
Metropolitan Area Geneva-Lausanne	48.5	(1.1)	45.6	(2.6)		45.6	(2.4)		43.0	(2.7)	*	49.0	(2.3)		48.9	(1.9)	

Notes: Non-parametric one-sample chi-square tests of goodness of fit: *** p<.001, **p<.01, *p<.05, †p<.

Table 5: Composition of responding samples by treatment group and samples responding in primary assigned mode (sample members without telephone numbers)

	(1) Gross sample without telephone numbers (n=1500)		(2) Mail (Contacts 1-3) (n=257)			(3) Mail All respondents in treatment group (n=303)			(4) Web Respondents by Web (n=432)			(5) Web Web + Paper (n=616)			(6) Web All respondents in treatment group (n=697)		
Auxiliary Variables	**%**	**Std. Err.**	**%**	**Std. Err.**	***p***	**%**	**Std. Err.**	***p***	**%**	**Std. Err.**	***p***	**%**	**Std. Err.**	***p***	**%**	**Std. Err.**	***p***
Male	51.4	(1.3)	45.5	(3.0)	†	46.2	(2.9)	†	54.4	(2.4)		52.7	(2.0)		53.9	(1.9)	
Age (mean in years)	42.5	(4.2)	43.4	(1.1)		41.9	(1.0)		41.2	(0.7)		43.4	(0.7)		43.1	(0.6)	
Age group																	
<30	24.0	(1.1)	22.6	(2.6)		25.7	(2.5)		24.5	(2.1)		22.9	(1.7)		23.1	(1.6)	
30-44	35.1	(1.2)	34.6	(3.0)		35.0	(2.7)		37.9	(2.3)		33.5	(1.9)		33.6	(1.8)	
45-64	29.5	(1.2)	28.8	(2.8)		27.1	(2.6)		30.1	(2.2)		31.5	(1.9)		31.8	(1.8)	
65+	11.4	(0.8)	14.0	(2.2)		12.2	(1.9)		8.3	(1.3)		12.0	(1.3)		11.5	(1.2)	
Marital Status											†						
Single	37.2	(1.3)	36.6	(3.0)		38.9	(2.8)		38.4	(2.3)		35.1	(1.9)		35.1	(1.8)	
Married	48.5	(1.3)	48.2	(3.1)		46.9	(2.9)		51.6	(2.4)		51.4	(2.0)		51.4	(1.9)	
Nationality											***			***			**
Swiss	58.9	(1.3)	64.2	(3.0)		63.4	(2.8)		70.1	(2.2)		68.3	(1.9)		64.7	(1.8)	
Bordering country	11.6	(0.8)	8.6	(1.8)		9.6	(1.7)		12.0	(1.6)		11.2	(1.3)		10.9	(1.2)	
Other	29.5	(1.2)	27.2	(2.8)		27.1	(2.6)		17.8	(1.8)		20.5	(1.6)		24.4	(1.6)	
Household size																	
1	25.4	(1.1)	21.8	(2.6)		21.8	(2.4)		22.2	(2.0)		23.6	(1.7)		23.3	(1.6)	
2	29.4	(1.2)	31.1	(2.9)		30.0	(2.6)		29.6	(2.2)		30.6	(1.9)		29.9	(1.7)	
3	19.1	(1.0)	21.0	(2.6)		21.5	(2.4)		18.3	(1.9)		18.4	(1.6)		18.7	(1.5)	
4+	26.1	(1.1)	26.1	(2.7)		26.7	(2.6)		29.9	(2.2)		27.5	(1.8)		28.2	(1.7)	
NUTS region																	
Région Lémanique	76.7	(1.1)	73.5	(2.8)		73.6	(2.5)		76.4	(2.1)		76.6	(1.7)		76.4	(1.6)	
Espace Mitteland	23.3	(1.1)	26.5	(2.8)		26.4	(2.5)		23.6	(2.1)		23.4	(1.7)		23.6	(1.6)	
Urbanisation					*			*									
City/town centre	35.3	(1.2)	29.2	(2.8)		28.7	(2.6)		32.2	(2.3)		35.1	(1.9)		35.2	(1.8)	
City/town suburbs	39.5	(1.3)	40.9	(3.1)		42.9	(2.9)		44.0	(2.4)		40.0	(2.0)		39.8	(1.9)	
Isolated town	0.9	(0.2)	-	-		-	-		1.2	(0.5)		1.3	(0.5)		1.6	(0.5)	
Rural community	24.3	(1.1)	30.0	(2.9)		28.4	(2.6)		22.7	(2.0)		23.6	(1.7)		23.4	(1.6)	
Metropolitan Area Geneva-Lausanne	55.1	(1.3)	49.8	(3.1)	†	51.5	(2.9)		56.0	(2.4)		55.3	(2.0)		54.5	(1.9)	

Notes: Non-parametric one-sample chi-square tests of goodness of fit: *** p<.001, **p<.01, *p<.05, †p<.

4. Discussion and Conclusion

In this paper, we present the methodological report for an experiment carried out by the LIVES IP15 in collaboration with FORS, designed to investigate advantages and disadvantages of different types of survey design and their impact on estimates of different aspects of vulnerability. The report describes in detail the design of the study and the questionnaires, and presents an overview of key outcomes relating to survey quality for each of the designs under investigation. Specifically, we present response rates and an analysis of the composition of the responding samples, using data from the sampling frame (which was based on population registers), allowing us to assess the extent of nonresponse bias on a number of socio-demographic measures not typically available in general population surveys. These auxiliary data, together with an experimental design that permits comparisons across multiple single and mixed mode survey designs, provides a rich and unique empirical base on which to optimise the design of Swiss surveys in this and other fields in the future, as well as to contribute to the international literature on mixed mode survey methodology. Additional features of the study's design offer other research opportunities of interest to the scientific community. In particular, the sample design for the experiment provides the possibility of comparing people with and without listed telephone numbers, which will give important insights into the extent of coverage error in telephone surveys, as well as into the various challenges associated with conducting surveys in different modes with different subpopulations. In the following, we present a summary of the findings presented in the report and discuss some of the limitations of the study and directions for future research.

4.1 Summary of findings

Overall, response rates to the experiment were higher than anticipated, and higher than those typically obtained in social science surveys in Switzerland. This unexpected result no doubt relates to the general interest of the survey topic (subjective wellbeing), but it also reflects the variety of efforts that were made to optimise the contact procedures and follow-up of nonrespondents. In particular, reliance on Dillman and his colleagues' (2009) tailored design method allowed us to schedule an effective protocol for encouraging the prompt participation of sample members, and this combined with a cash incentive, the legitimacy afforded by the (local) University of Lausanne, and the visually appealing LIVES logo and tangram (shown in the appendix), proved to be an effective

mix of ingredients for attracting different types of respondent. This is evident not only in the good response rates, but also in the composition of the responding samples across the different treatment groups, which generally represented the population well. A number of important variations across survey designs and across the with- and without-telephone subpopulations are worth highlighting here.

Final response rates for the mixed mode web and CATI surveys were higher than that for the single mode mail survey, confirming that mixing modes sequentially can help to boost response. This finding adds to the existing somewhat mixed findings on the benefits of sequencing modes in surveys (see Wagner, Arrieta, Guyer and Ofstedal (2014) for a review) but suggests that in Switzerland, at least, mixed mode designs can improve on single mode designs with respect to response. Comparing the single mode designs (i.e., mail with CATI and web without the follow-up phases), web fared poorest with a response rate of 44.5%, followed by mail, which at the equivalent phase of the survey (i.e., following phase 2) had a response rate of 56.5%, followed by CATI, which after 3 weeks had obtained a response rate of 60.7%. By contrast, the mail response rate following the final reminder was 65.4%, the mixed mode web survey was 70.2%, and the mixed mode CATI survey was 70.6%. Given the additional steps involved in the web survey, the CATI and mail designs were considerably more cost and time-efficient. Without the CATI/CAPI follow-up, the response rate for the web plus mail design was equivalent to the mail-only design at 64.4%

Differences in response rates between the modes were moderated by whether or not a sample member had a listed telephone number. People without listed telephone numbers were less likely to participate in the survey overall, particularly in the single-mode mail survey. In the mixed mode web survey, the CAPI phase helped to improve participation in this group, but failed to fully close the gap between those with and without telephone numbers. By contrast, for people with listed telephone numbers, response rates were equally high for the mail and CATI surveys, while the web survey achieved the highest response rate, but again, only at the expense of extra resources.

Overall, all the surveys achieved reasonably good representation of the population, with relatively few statistically significant differences observed on the auxiliary variables available for all sample members. The different modes attracted different types of people, however, and again, this was moderated by whether or not the sample had listed telephone numbers or not. For example, in the single mode designs, for the sample with telephone numbers, the mail survey overrepresented single people and people living in

city centres, the CATI survey overrepresented the Swiss, and the web survey overrepresented younger members of the population, single people and people living in larger households. Mixing modes eliminated bias in the web survey for this sample, but failed to do so in the CATI survey (young people and Swiss people were overrepresented). For the sample without phone numbers, the mail survey slightly underrepresented men and people living in city centres (differences were not significant), and the web survey overrepresented the Swiss – a bias that was not eliminated by the sequencing of modes for nonrespondents.

Thus, the findings of the experiment not only provide new evidence that mixing modes can improve response rates, but also that differential selection effects between modes can be exploited when combining modes to reduce overall selection bias. However, this picture is complicated by whether or not sample members have a listed telephone number or not – a variable that has been found in other studies to be correlated with survey participation (e.g. Cobben and Bethlehem, 2005; Roberts, Vandenplas, and Ernst Stähli, 2014) partly due to its implications for a person's contactability, but also due to its correlation with other variables linked to willingness to take part in surveys. In our analysis of the sampling frame data, we find that people with telephone numbers are significantly older, more likely to be married, more likely to be Swiss and more likely to be living outside of city centres. In future research, we will be investigating in more detail how these characteristics correlate with other questionnaire variables linked to response propensity, to gain a more detailed picture of the mechanisms underlying nonresponse bias.

4.2 Limitations and directions for future research

In this paper, we focus only on error associated with nonresponse in different types of single and mixed mode survey designs. In our analysis we take advantage of the auxiliary variables available on the sampling register, which provide a rich source of information about nonrespondents to the survey, not typically available to other researchers. However, it is noteworthy that these mostly socio-demographic variables may ultimately be ineffective for predicting the risk of nonresponse bias on the key substantive variables of interest in the survey (Peytcheva and Groves, 2009). For this reason, in other research we are making use of data from the nonresponse follow-up survey to shed light on differences between respondents and nonrespondents in each of the survey designs to learn more about the variables most likely to be affected by

nonresponse bias (Roberts, Ernst Stähli, Joye and Sanchez Tome, 2015). The correlation between having a listed telephone number and survey participation suggests a potential for confounding between errors from noncoverage and nonresponse in survey designs that exclude some or all population members without a publicly listed telephone number (Cobben and Bethlehem, 2005). We do not attempt to assess the interplay between multiple sources of survey error simultaneously here, but given the inherent trade-offs involved in decisions to mix modes in surveys, a more sophisticated analysis of the relation between different sources of error is likely to be helpful for improving future survey designs. For example, it is of interest to know whether efforts to reduce coverage errors by introducing additional modes are offset or compounded by the introduction of differential nonresponse bias.

While the results reported here appear to lend support to arguments in favour of mixing modes, a proper evaluation of mixed mode survey designs alongside single mode designs would be incomplete without an assessment of the extent and nature of measurement error in each. Different modes of data collection have different measurement properties, which mean that the answers respondents give in those modes, and consequently the estimates derived from the survey, may differ (e.g. de Leeuw, 2005). For example, it is well established that self-administered modes obtain more honest answers from respondents on sensitive measures than interviewer-administered surveys (Tourangeau and Yan, 2007). Further differences in measurement may arise due to the increased cognitive burden associated with different modes (Holbrook, Green and Krosnick, 2003; Roberts, 2007). This makes it difficult to compare estimates from surveys conducted in different modes. However, the situation is complicated further by the fact that, as we have seen, different modes attract different types of respondent, who might genuinely differ on the variables affected by differential measurement error. Thus, selection and measurement effects are confounded in a way that limits the comparability of data collected by different modes (either in different surveys, or in the same mixed-mode survey), and this once again, calls for a more sophisticated analysis of the interplay between different sources of survey error.

The confounding between selection and measurement effects represents one of the principal disadvantages of mixing modes and a complex challenge for methodologists and analysts handling the data. A proper treatment of mixed mode data involves first assessing the extent and nature of selection effects in each mode (often thwarted by a lack of data about nonrespondents to a survey), and then assessing the extent of differential

measurement error between modes, while controlling for the observed selection effects. Once the measurement error has been estimated, efforts should ideally be made to correct for the error, in order to permit comparisons between modes, or more accurate estimation of statistics based on the mixed mode survey. Different approaches to the problem of disentangling selection and measurement errors in mixed mode surveys are available (see, e.g., Vannieuwenhuyze and Loosveldt, 2012), but to date, there remains a lack of clear guidance for researchers about the procedures that should be followed when assessing the extent of mode differences in measurement prior to data analysis. Due to the complexity of the problem, we do not address differential measurement error in this paper (aside from a brief discussion of variations observed across modes in rates of item nonresponse). We address this problem elsewhere (e.g., Sanchez Tome, Roberts, Joye and Ernst Stähli, 2014), but it is of key importance that the results of such an evaluation be taken into consideration alongside the results reported here in order to draw correct conclusions about the relative costs and benefits of the different survey designs tested in this experiment.

Once again the need to properly evaluate the extent of differential measurement error alongside both errors of nonresponse and coverage concerns the general question of cost and error trade-offs implicit in mixed modes surveys (de Leeuw, 2005). Increased response rates, and a reduction in sampling error and the risk of nonresponse bias may be a key priority for many survey designers, but if these benefits come at the cost of an inflated fieldwork budget and data collection period, increased measurement error, and greater burden on data users, then it is important to make empirically-informed decisions about which survey designs are better or worse. Such an appraisal can only be made via an assessment of the total survey error (Biemer, 2010) affecting survey estimates (in this case of vulnerability) alongside an analysis of survey costs (Groves, 1989), and in future research we will be turning our attention to this challenge using data from the experiment reported here.

5. Acknowledgements

This work was supported by the Swiss National Science Foundation, as part of the Swiss National Centre of Competence in Research (NCCR) 'LIVES'. We are grateful to the Swiss Federal Statistical Office for supplying a sample from their frame based on population registers maintained by municipalities. We would particularly like to thank Emilie Bornet and Mathias Humery at MIS Trend SA., for their careful management of the fieldwork.

[1] The study was also designed to coincide with other mixed mode experiments being conducted alongside Round 6 of the European Social Survey by a number of participating countries.
[2] These included IP1, IP4, IP7, IP8, IP11, IP12, and IP13.
[3] Rates are provided for respondents answering in their assigned mode only.
[4] Mailings for the survey were sent on LIVES/UNIL headed stationery, giving the office address for Professor Dominique Joye, so a large proportion of returned mail from non-respondents and the post office was handled at the University of Lausanne. The remainder was addressed directly to MIS Trend, SA., either by mail or via the hotline telephone number and email address provided in the mailings. Records were kept of all such communications (including reasons given for nonresponse).
[5] Implicit refusals are included for the CATI and CAPI mode. Implicit refusals included cases who made and broke appointments for interviews, or who stated that they were ill but would be available for an appointment at a later stage. The decision to code these as refusals is also based on AAPOR (2011).
[6] As mentioned, due to delays with the face-to-face fieldwork, the mailing of the NRFU for the web, no-phone group was postponed by one week. For the purposes of comparison, phase 4 is calculated for all other mode groups from the 11th February.

6. References

American Association for Public Opinion Research. (2011). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. 7th edition. AAPOR.

Atkeson, L.R., Adams, A.N., and Alvarez, R.M. (2014). Nonresponse and mode effects in self- and interviewer-administered surveys. *Political Analysis*, 22: 304-320.

Baumann, I., Lipps, O., Oesch, D., and Vandenplas, C. (2016). How to survey displaced workers in Switzerland? Sources of bias and ways around them. In M. Oris, C. Roberts, D. Joye, and M. Ernst Stähli (Eds.). *Surveying human vulnerabilities across the life course*. New York: Springer.

Biemer, P. (2010). Total survey error: Design, implementation, and evaluation. *Public Opinion Quarterly*, 74(5): 817-848.

Buelens, B., and van den Brakel, J.A. (2015). Measurement error calibration in mixed mode sample surveys. *Sociological Methods and Research*, 44(3): 391-426.

Cobben, F., and Bethlehem, J.G. (2005). *Adjusting undercoverage and nonresponse bias in telephone surveys*. Discussion paper 05006, Statistics Netherlands.

Charvoz, L., Favez, N., Notari, S.C., Panes-Ruedin, B., and Delaloye, J-F. (2016). A survey of couples facing breast cancer in women. In M. Oris, C. Roberts, D. Joye, and M. Ernst Stähli (Eds.). *Surveying human vulnerabilities across the life course*. New York: Springer.

De Leeuw, E. (2005). To mix or not to mix? Data collection modes in surveys. *Journal of Official Statistics, 21*(2), 1-23.

De Leeuw, E., and De Heer, W. (2002). Trends in Household Survey Nonresponse: A Longitudinal and International Comparison. In R. Groves, D. Dillman, J. Eltinge, & R. J. A. Little (Eds.), *Survey nonresponse* (p. 41-54). New York: Wiley.

Dillman, D. A. (1978). *Mail and telephone surveys: The total design method*. New York: Wiley-Interscience.

Dillman, D.A. (2000). *Mail and internet surveys: The tailored design method.* 2nd ed. New York: Wiley.

Dillman, D.A., Smyth, J.D., and Christian, L.M. (2009). *Internet, mail, and mixed-mode surveys. The Tailored Design Method (3rd Edition)*. Hoboken, NJ.: John Wiley and Sons.

Ernst Stähli, M. (2012). Spotlight on Switzerland. In M. Häder, S. Häder, M. Kühne (eds), *Telephone Surveys in Europe: Research and Practice*, pp. 25–36. Berlin Heidelberg: Springer.

Gomensoro, A., and Bolzman, C. (2015). The effect of socio-economic status of ethnic groups on educational inequalities in Switzerland: which 'hidden' mechanisms? *Italian journal of sociology of education*, (Feb.).

Groves, R.M. (1989). *Survey errors and survey costs*. New York: John Wiley and Sons.

Groves, R.M., Biemer, P.P., Lyberg, L.E., Massey, J.T., Nicholls II, W.L., and Waksberg, J. (1988). *Telephone survey methodology. Wiley Series in Survey Methodology.* New York: John Wiley and Sons.

Holbrook, Green and Krosnick, 2003; Krosnick, J.A., Holbrook, A.L., Berent, M.K., Carson, R.T., Hanemann, W.M., Kopp, R.J., Mitchell, R.C., Presser, S., Ruud, P.A., Smith, V.K., Moody, W.R., Green, M.C., and Conaway, M. (2002). The impact of 'no opinion' response options on data quality. *Public Opinion Quarterly*, 66: 371-403.

Joye, D., Pollien, A., Sapin, M., and Ernst Stähli, M. (2012). Who can be contacted by phone? Lessons from Switzerland. In M. Häder, S. Häder, M. Kühne (eds), *Telephone Surveys in Europe: Research and Practice*, pp. 85–102. Berlin Heidelberg: Springer.

Klausch, T., Hox, J., and Schouten, B. (2015). Selection error in single- and mixed mode surveys of the Dutch general population. *Journal of the Royal Statistical Society, Series A,* 178(4): 945-961.

Lipps, O., Pekari, N., and Roberts, C. (2015) Undercoverage and nonresponse in a list-sampled telephone election survey. *Survey Research Methods,* 9(2): 71-82.

Lugtig, P., Lensvelt-Mulders, G.J.L.M, Frerichs, R., and Greven, A. (2011). Estimating nonresponse bias and mode effects in a mixed mode survey. *International Journal of Market Research*, 53(5): 669-686.

Lyberg, L. and Dean, P. (1992). *Methods for reducing nonresponse rates: A review*. Paper presented at the annual meeting of the American Association for Public Opinion Research. St. Petersburg, Fl.

Maggiori, C., Rossier, J., Krings, F., Johnston, C., and Massoudi, K. (2016). A longitudinal research on professional trajectories: Preliminary results from the 1st wave of data collection on a representative sample. In M. Oris, C. Roberts, D.

Joye, and M. Ernst Stähli (Eds.). *Surveying human vulnerabilities across the life course*. New York: Springer.

Martin, P., and Lynn, P. (2011) The effect of mixed mode survey designs on simple and complex analyses. *Centre for Comparative Social Surveys Working Paper Series, 2011-04.* City University London.

Office fédéral de la statistique. (2015). *L'accès des ménages à internet et son utilisation par les individus en Suisse. Enquête sur les technologies de l'information et de la communication 2014 auprès des ménages.* Neuchâtel: OFS.

Oris, M., Guichard, E., Nicolet, M., Gabriel, R., Tholomier, A., Monnot, C., Fagot, D., and Joye, D. (2016). Representation of vulnerability and the elderly. A total survey error perspective on the VLV survey. In M. Oris, C. Roberts, D. Joye, and M. Ernst Stähli (Eds.). *Surveying human vulnerabilities across the life course*. New York: Springer.

Oris, M., Roberts, C., Joye, D., and Ernst Stähli, M. (2016). *Surveying human vulnerabilities across the life course*. New York: Springer.

Parke, C. (2013). *Essential first steps to data analysis. Scenario-based examples using SPSS.* Thousand Oaks, CA.: Sage Publications.

Perrig-Chiello, P., Hutchison, S., and Knöpfli, B. (2016). Vulnerability following a critical life event: temporary crisis or chronic distress? A psychological controversy, methodological considerations, and empirical evidence. In M. Oris, C. Roberts, D. Joye, and M. Ernst Stähli (Eds.). *Surveying human vulnerabilities across the life course*. New York: Springer.

Peytcheva, E., and Groves, R.M. (2009). Using variation in response rates of demographic subgroups as evidence of non-response bias in survey estimates. *Journal of Official Statistics* 25: 193-201.

Roberts, C. (2007). Mixing modes of data collection in surveys: A methodological review. *NCRM Methods Review Papers*, 008.

Roberts, C., Ernst Staehli, M., and Joye, D. (2013). *Mixed mode solutions to coverage and nonresponse error: Evaluation of the cost-error trade-off.* Paper presented at the 5th Conference of the European Survey Research Association, Ljubljana, Slovenia - 15-19/7/2013.

Roberts, C., Lipps, O. and Kissau, K. (2013) Improving Data Quality in Swiss Social Surveys - Using the Swiss population register for methodological research. *FORS*

Working Paper Series, 2013-01.

Roberts, C., Ernst Stähli, M., and Joye, D. (Under revision). Reducing coverage error in telephone surveys: an experimental comparison of mixed mode alternatives. Submitted to *Journal of Social Statistics and Methodology*.

Roberts, C., Ernst Stähli, M., Joye, D., and Sanchez Tome, R. (2015). *Combining web and mail sequentially – error and cost advantages in the Swiss context.* Paper presented at the 6th Conference of the European Survey Research Association, Reykjavik, Iceland - 13-17/7/2015.

Roberts, C., Vandenplas, C., & Ernst Stähli, M. (2014). Using register data to assess the impact of response enhancement methods on the risk of nonresponse bias. *Survey Research Methods*, 8(2), 67–80.

Sanchez Tome, R., Roberts, C., Ernst Staehli, M., and Joye, D. (2014). *Mode effects in measures of subjective well-being – Implications for cross-survey comparisons.* Paper presented at the 67th Annual Conference of the World Association for Public Opinion Research, Nice, France - 4-6/9/2014

Schwarz, N, and Sudman, S. (1992). *Context effects in social and psychological research.* New York, NY.: Springer-Verlag.

Tourangeau, R., Edwards, B., Johnson, T.P., Wolter, K.M., and Bates, N. (2014) *Hard-to-survey populations*. Cambridge: Cambridge University Press.

Tourangeau, R., and Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, 133(5): 859-883.

Turtschi, N., and von Ow, A. (2014). Réseau social et inégalités de réinsertion. *Panorama*, 6: 26-27.

Vandenplas, C., Joye D., Ernst Stähli, M., and Pollien, A (2015). Identifying pertinent variables for nonresponse follow-up surveys. Lessons learned from 4 cases in Switzerland. *Survey Research Methods, 9(3): 141-158.*

Vannieuwenhuyze, J.T.A. (2014). On the relative advantages of mixed-mode versus single-mode surveys. *Survey Research Methods*, 8(1):31-42.

Vannieuwenhuyze, J.T.A. and Loosveldt, G. (2012). Evaluating relative mode effects in mixed-mode surveys: Three methods to disentangle selection and measurement effects. *Sociological Methods and Research*, 42(1): 82-104.

Vannieuwenhuyze, J.T.A., and Lynn, P. (2014). Measurement effects between CAPI and Web questionnaires in the UK Household Longitudinal Study. *Understanding Society Working Paper Series*, 2014-01.

Vannieuwenhuyze, J.T.A. Loosveldt, G., and Molenberghs, G. (2014). Evaluating mode effects in mixed-mode survey data using covariate adjustment models. *Journal of Official Statistics*, 30(1): 1-21.

Wagner, J., Arrieta, J., Guyer, H., and Ofstedal, M.B. (2014). Does sequence matter in multimode surveys: Results from an experiment. *Field Methods*, 26(2): 141-155.

Widmer, E., Aeby, G., Sapin, M. (2013). Collecting family network data. *International Review of Sociology*, 23 (1), 27-46.

7. Appendix

https://www.lives-nccr.ch/sites/default/files/pdf/publication/48_lives_wp_appendix.pdf