Assessing the Performance of the Swiss Panel LIVES Calendar: Evidence from a Pilot Study

Davide Morselli*[1], Dario Spini[1], Jean-Marie Le Goff[1], Jacques-Antoine Gauthier[1], Karen Brändle[1], Estelle Mugnari[1], Nora Dasoki[1], Caroline Roberts[1], Laura Bernardi[1], Felix Bühlmann[1], Valérie-Anne Ryser[2], Robin Tillmann[2], Judith Kühr[3], and Alexandra Bumbaru[4]

[1]Swiss Centre of Competence in Research LIVES - University of Lausanne
[3]Department of Health and Social Action of the Vaud Canton, Switzerland
[2]Swiss Centre of Expertise in the Social Sciences (FORS), Switzerland
[4]University of Lausanne, Switzerland

Abstract

The use of life-history calendar methods in survey resaerch has been growing during the last couple of decades. This study presents a series of tests on the Swiss Panel LIVES Calendar (SPLC) aiming to understand the quality of the tool in collecting the expected data. The SPLC is a paper and pencil tool specifically designed for the first wave of the third sample of Swiss Household Panel to study life trajectories in relation to the concept of vulnerability. Thus, the SPLC is designed to be moderately standardized and relatively quick to complete, providing sufficient information on respondent's life trajectories in the domains of residential moves, cohabitation, intimate relationship, family, occupation, and health. The performance of the tool has been assessed using coherence and completeness indicators and by comparing the data collected through different modes, different sub-samples of the populations, and a previous survey (Family tiMes) which used the calendar method. Results showed a good performance of the SPLC both in terms of internal coherence and completeness. SPLC data are robust across modes and comparable to life-history calendar data collected with interactive face-to-face interviews.

*davide.morselli@unil.ch

Contents

1 Introduction

One of the methodological challenges of life course research is data collection (Levy, Ghisletta, Le Goff, Spini, & Widmer, 2005; Scott & Alwyn, 1998). As a matter of fact there are not many ways to assess the life history of individuals. Researchers can either prospectively observe respondents' life using repeated measures techniques (e.g., panels, cohort studies and other longitudinal designs), or asking respondents to recollect their past experiences and tell their personal history. Prospective studies have the advantage of observing the respondents' status minimizing their interference. In retrospective methods, researchers have to totally rely on the respondents' willingness and capability of disclosing their past, which could vary from person to person and therefore produce biased data (e.g., events are absent in some individuals' data not because they did not occur but because the respondents did not recall or not report). For this reason retrospective methods are sometimes considered less efficient than longitudinal design (Bidard, 2010). However equally important drawbacks apply to prospective designs, which would make the study of vulnerability across the whole life extremely difficult if not, in certain cases, impossible. For instance, for studying paths of migrations, researchers should estimate at the very beginning what respondents are more likely to migrate and being able to follow them across the different locations. Thus, prospective studies are generally more costly in terms of funding but also time.

Retrospective methods have the advantage of producing data in a time efficient manner and at relatively low costs. However, we have to trust respondents' capability of recollecting and reporting information on their past while several factors may influence memory, possibly leading to omission or misreporting of life events (Auriat, 1996; Reimer, 2001). Building on these considerations, several retrospective methods have been developed in social research to minimize memory bias in data, including decomposition (e.g., braking a class if events into subclasses), the use of reversed chronological order, the use of landmarks (e.g., transition points), incrementing the time to answer survey questions (see Tourangeau, 2000). Since the late 80s the life history calendar (LHC) method has been developed as a possible solution to many of the retrospective method's drawbacks (Caspi et al., 1996; Freedman, Thornton, Camburn, Alwin, & Young-DeMarcco, 1988).

The present paper discusses the validity of a LHC questionnaire designed to be implemented in the third sample of the Swiss Household Panel (SHP). Results from a pilot study conducted between mid-October 2012 and mid-March 2013 on a general population sample are presented.

1.1 The Performance of Life-History Calendars and Procedures of Validations

The LHC is substantially a two-way grid, with the temporal dimension on the one side, and different life domains (e.g., residence, family, employment) on the other. Respondents are asked to report events for each life domain, relating them to what happened across other domains or in reference to time landmarks. While filling the LHC, respondents can visualize their life trajectory, linking what happened to when, where and for how long it happened. It

has been argued that the calendar tools facilitate respondents to place events into a temporal context by relating them to other synchronic (parallel) or diachronic (hierarchical) events and episodes that occurred in the life span (Belli, 1998). In other words, the LHC's performance would be related to the efficacy of the tool to enhance several aspects of memory retrieving (Caspi et al., 1996). According to Conway (1992, 1996), autobiographical memory is indeed a non-linear process that works through different association mechanisms. Events can be recalled through their hierarchy (e.g., from more important to least important), through their sequence (e.g., in chronological order) or in relationship to other events and episodes. In addition, different individuals may be more comfortable or used to recollecting events in different ways.

Belli, Shay, and Stafford (2001) argues that the calendar facilitates the use of all three memory retrieval mechanisms as well as facilitating a flexible interaction between the respondent and the interviewer to clarify intended meanings and to reconstruct the past. In line with these assumptions, Glasner (2011) compared the number of edits (i.e., corrections) done by respondents who answered the same questions in an on-line conventional questionnaire as compared to an on-line LHC. Congruently with Belli's rationale, Glasner found that the number of edits in the calendar mode were as twice as high as those of the conventional questionnaire. In other words, biographical calendar data are more likely to be corrected by respondents than conventional questionnaire data.

To assess the reliability of LHCs two pathways are usually used: comparisons of LHC data with similar data collected with more traditional questionnaires, and comparison between data collected with LHC and external data source with no measurement error (medical records, purchase records, population registers). Concerning the latter, Rosenberg et al. (1983) performed a record check study using medical records as validation measures to test data on use of oral contraceptives. They found an agreement of 90% between the calendar data and the records. The agreement between medical records and self-reports decreased when brand and dose of contraceptive were also considered. Similar results were found in small-scale medical studies on the use of timelines (Searles, Helzer, & Walter, 2000). Auriat (1993) controlled the validity of dates of residential moves collected with the calendar by means of a comparison with the population registers. She found that misdating at one point of the interview increased the likelihood of misreporting subsequent events. Thus, Auriat argues "the life-history calendar or the time line techniques may be useful for decreasing the thread of errors that seems to exist in free series recall" (Auriat, 1993, p. 185). A comparison between retrospective data collected with calendar methods on respondents in early old age and archive material recorded 50 years previously showed satisfactory accuracy of the long distance dating of socio-demographic events (Berney & Blane, 1997).

When external sources of validation are not available, researchers have been using test-retest techniques and/or experimental designs that contrasted the LHC with other retrospective questionnaires. For instance, Engel, Keifer, and Zahm (2001) showed by means of a test-retest, that the calendar method had a higher reliability than conventional questionnaires, concerning agreement over eight to fourteen months span in reporting life event anchors such as marriages, or immigration. Similarly Freedman et al. (1988) compared respondents' self-reports from two waves of a longitudinal study, the second of which implemented a LHC. They found between 72% and 87% agreement between the first prospective wave and the second with LHC on respectively job and educational careers. If the memory bias is not completely neutralized with the use of LHC, it seems in general the calendar method has beneficial effects on the precision of data.

In addition to these indirect evaluations, some experimental comparisons between calendar instruments and conventional questionnaires showed that adding a time-line to the questionnaire enhanced data quality in comparison to the regular questionnaire procedure (Van der Vaart, 2004). Similarly, Belli et al. (2001) showed that the calendar produced fewer reporting errors than conventional question lists concerning the number of residential moves, income, weeks unemployed, and illness, but it over-reported history of cohabitation

and number of jobs.

Several indicators have been used to assess data quality with LHCs. In particular, Becker and Sosa (1992) used report consistency (i.e., less superposition of mutually exclusive behaviours) to show that the calendar was more internally coherent than the conventional questionnaire. Goldman, Moreno, and Westoff (1989), Yoshihama, Gillespie, Hammock, Belli, and Tolman (2005) and Engel et al. (2001) used instead the completeness of the calendar to asses data quality. Those authors argued that with the calendar methods reporting specific events are more complete (e.g., more episodes of contraceptive use, domestic violence victimization). Indeed, the visual nature of the calendar makes it easier for the interviewee and the interviewer to spot incongruous answers in the data, reducing the amount of time in the respondent's life course unaccounted for.

1.2 Life History Calendars for Studying Vulnerability

Aside practical implications the LHC method, it also worth noting its potential for studying vulnerability. Vulnerability is indeed a multidimensional and multifaceted concept which covers at the same time different aspects of life and their duration in time. It can be defined as the lack of resources that places individuals or groups at major risk of experiencing negative outcomes, incrementing the inability to cope effectively with those outcomes, and the recovery from their occurrence is either delayed or impossible (Spini, Hanappi, Bernardi, Oris, & Bickel, 2013).

According to this definition, vulnerability is not ascribed to specific segments of the society but it can be a rather wide spread phenomenon with different levels of intensity and chronicity. In addition, the same risk factors may have diversified effects on different individuals (Ranci, 2010). In the contemporary society, the confidence in the ability to keep social risk under control has been indeed replaced by the idea that risks are not fully predictable or controllable. Such change brought to a greater exposure to risk factors that put individuals in the conditions of experiencing negative social and economical outcomes, such as falling below the poverty threshold (Beck, 1992). New social risks emerge from the interaction between labour markets, families, and welfare systems, producing a broad spectrum of potential negative outcomes that are not fully overcame by the compensation of the outcome itself (Ranci, 2010). For instance, the compensations to losses, such as reintegration in the labour market after having lost a job, are not always effective in reducing social and psychological vulnerability and they perform differently for different segments of the society. Negative outcomes indeed need to be framed into an analytical strategy that includes individual interpretations of the events in relation to their social and historical context on the one side and the personal and family history on the other. The interaction between the social context and the individual life course defines to a great extent the meaning and expectations individuals attach to life events.

Thus, in a context of uncertainty, stability of the life course is replaced by a constant exposure to multidimensional risk factors. As we defined it, vulnerability is not only the possibility for an individual of experiencing negative outcomes or significant damages as consequence of specific risk factors, but also as the different chances that individuals or social groups have of activating protective and compensatory factors at different levels, e.g., material, social, psychological. In other words, vulnerability is not a static but a dynamic condition, one definition of which being the time needed to recover from a negative outcome. Vulnerability is also multidimensional as it concerns simultaneously different aspects and dimensions of life, their interactions in time and the interpretation given to them (Ranci & Magliavacca, 2010). For these reasons, the study of vulnerability benefits from adopting a life course approach which accounts for multidimensional aspects of life and their interpretation. In particular the research on vulnerability and the life course paradigm overlap on three interrelated processes: diffusion, accumulation and interpretation. All those three processes can be easily studied using LHC data.

The process of diffusion is based on Schutz (1962)'s definition of life-spheres according to which the individual life is categorized into different domains. e.g., family, work, social relations, political life. Of course the definition of each sphere is socially constructed and the way each sphere is defined in the design of questionnaires or LHCs is charged with meanings. Nevertheless, each sphere can be the object of specific disciplinary investigations, given that the notion of sphere itself presumes that life-spheres are independent of each other to some extent. However, they are not completely independent, both positive and negative spillovers may happen. For instance, events that happen within the family life can influence the work sphere or social relations. Experiencing a divorce can strongly affect the network of interpersonal relations of the two ex-partners (Widmer, 2010). By accounting simultaneously for different life domains, spill-over effects can easily be studied with LHCs.

In addition, as life-spheres are interdependent, interdisciplinary research is needed. The LHC tool is particularly suitable for multi-disciplinary research because it produces a relatively large amount of data which can be analysed from different perspectives (demography, sociology, psychology, social-psychology...). Besides diffusion processes, the interaction among events that happen within each life-sphere is also fundamental for understanding vulnerability. The accumulation of disadvantages, including the temporal distance between negative and positive outcomes as well their sequence and timing, can put people in a severe condition of risk, affecting the reconstruction of resources needed to exit from vulnerability (Dannefer, 2003, 2009). The temporal structure of LHC data is particularly suitable for accounting for these aspects and can be used for performing sequence analysis, event-history analysis, or other longitudinal models (Axinn & Barber, 2001; Axinn, Pearce, & Ghimire, 1999).

Last but not least, the interpretation of what happened is at the same time a dependent and an independent variable in the study of vulnerability. The same event can assume very different meanings and affect peoples' live in very different ways. For instance, losing a job or being fired may not be experienced in the same way by members of different social classes, nor by members of the same class. Thus, the event that for someone can start a dangerous domino effect for another may be a minor problem. In addition, the meaning attached to events is culturally and socially embedded. Losing a job in a period of economic crisis may be experienced differently than losing it in a period of economic growth. On the other hand, in a period of crisis the loss of the job becomes sadly common while in other historical periods it could represent a less likely, and for this reason more stigmatizing, experience (Oesch & Lipps, 2012). The interpretation of events and their relationship with the social-historical context are tightly related and both can be addressed in two ways with LHCs.

Data in LHCs are connected to a temporal – i.e., the year in which events occurred – and a spatial dimension – where the respondent was located at the time the event occurred. This makes possible to relate events to places, and consequently to social and historical contexts. In addition, a retrospective evaluation of events can be surveyed using LHCs, collecting information on individual interpretations about life trajectories. This is probably one of the less developed uses of LHCs although in our understanding it is quite central to the study of vulnerability.

1.3 The Swiss Panel LIVES Calendar

Inspired by the French Ageven tradition (Vivier, 2006) and building on the previous experience on the use of LHCs in the Swiss context (see Morselli et al., in press), the Swiss Panel LIVES Calendar was designed to be inserted in the third sample of the SHP. In line with the life course paradigm (e.g. Mortimer & Shanahan, 2003), the SPLC focused on individual trajectories as produced by an events-related logic rather than by an underlying generative processes. SPLC aimed at explaining life-trajectories through the chronological order of events and the relative social statuses linked to each event.

The SPLC was printed on a paper sheet of 42 cm width, folded into three in order to

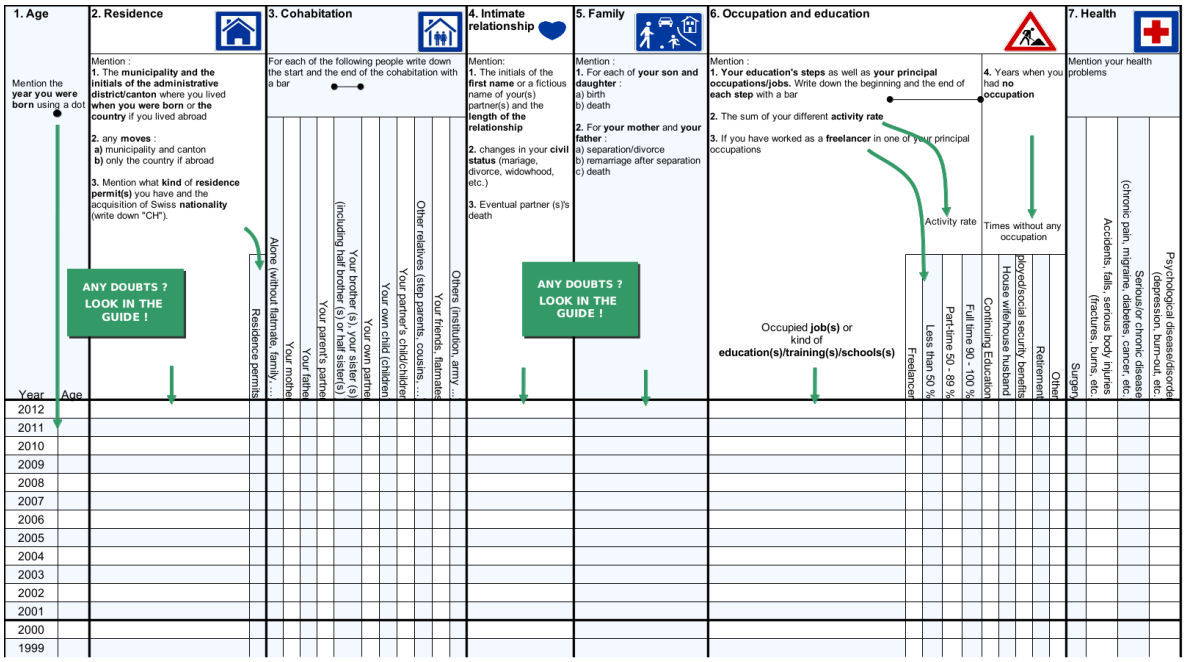

Figure 1: The Swiss Panel LIVES Calendar

create an agreeable 14×14 square, and directly filled by the respondent with a pencil. The layout was largely inspired on the LHCs used in two previous survey: *Vivre / Leben / Vivere* and *Family tiMes*[1]. The tool tapped six life domains: residential trajectory, cohabitation, intimate partnership, family history, activities and education, and health. It looked like a vertical grid with the life domains in column and the years in row, starting from the present (Figure 1). This LHC was designed to be self-administered, thus specific instructions for each column were printed in the top section. Respondents were also given a booklet with additional instructions to be consulted if needed, including an example of ready filled calendar. Icons and graphical elements were also added to facilitate the understanding of the tool.

Given that the Swiss Household Panel is a relatively complex survey with over 6000 respondents the SPLC had to respect some constraints in terms of time of administration and standardization of the tool. To facilitate the management of the survey, the tool was not personalized for each respondent, the actual year was put at the top of the page and respondents had to mark their birth year by scrolling down on the sheet. Although it is not uncommon to have calendars with the birth year on the top, pre-test cognitive interviews did not underline this "reversed" layout as a problem.

Slightly different ways of filling were requested across the tool. For residential moves, partnership, family and activities respondents were asked to write the required informations. For single events respondents were asked to draw a point on the row corresponding to the year when it occurred; for periods longer than one year they were asked to draw a line to indicate the beginning and the end. For cohabitation, health, job specifications (whether self-employed, and total employment rate), and periods without job, respondents were asked to give the information by drawing a line from the beginning to the end of each period. This format allowed to sensibly reduce the administration time which for this survey needed to be about 30-45 minutes.

[1]Swiss National Science Foundation, grants no 100017_130343/1

To further reduce the administration time, information about family and partnership had to be severely limited. Concerning partnership respondents were asked to report important relationship and any civil status change. About family history, respondent were asked birth and death of children, parents' separation divorce, and remarriage, and parents' death.

As discussed in section 1.2, the interpretation of life events, along side their occurrence or absence, is pivotal to the study of vulnerability. Not taking in consideration this aspect may lead to assume that similar events have the same impact on different people. According to our theoretical framework certain events (e.g. losing the job) may be perceived as normal for certain individuals or groups, but disruptive for others, and having very different effects in terms of vulnerability. For this reason the SPLC introduced subjective measures, with the aim of analysing the perception and interpretations of events. In particular it focused on the relationship between interpretation of past events and present well-being, and spillover effects.

This direction of research was inspired by analyses showing that the current well-being state is more related to recent daily hassles than critical past events. However, these results are usually obtained in associating reported past life events without a measure of the subjective importance of these events. Few studies have explored the effects on and of subjective interpretations. In particular, Lelièvre, Roubaud, Tichit, and Vivier (2006) added to life calendars the post-hoc evaluation of life periods on a 5-points Likert scale. A slightly different method was used by Clausen (1995) and, more recently, Perren, Keller, Passardi, and Scholz (2010), asking respondents to graph their life satisfaction as a curve chart. While Clausen used this method with a qualitative approach, Perren and colleagues translated curves into scales to be analysed with statistical tools. However, Perren's study was limited to specific life transitions rather than the whole life trajectory.

At the end of the SPLC interview, respondents were asked to cut their trajectory of three life domains (partnership, family life, and activity) into periods of different length. Respondents were informed that they could freely choose the length of the periods, and that periods and events should not necessarily overlap. For instance one get married once (event) but the marriage can have different phases (periods) with up and downs. Once respondents had completed the task of sub-setting the three life domains into periods they were asked to express an evaluation for each period on a 5-point bipolar scale ranging from -2 = "very negative" to $+2$ = "very positive" (see Figure 2).

2 The Swiss Household Panel III Pilot Study

2.1 Data

Data from the SHP pilot study was used to explore the validity of the SPLC. The study consisted in a four-sections questionnaire. The first two sections, namely the grid and household questionnaires are usual parts of the Swiss Household Panel interview. The grid was answered by the contact respondent and collected several socio-demographic information on the respondent's current situation and the household composition. In most cases the same respondent answered also the household questionnaire, which aimed to collect information on the household socio-economic condition. The two other sections were addressed to individuals instead of households and were answered by all available members of the household; 1.64 interviews per household were done on average. In this part of the interview respondents were asked to fill the SPLC (third section). In the last section, the interviewer used a standardized check-list to find incoherence, misreported or missing data[2] . Subsequently, respondents were asked a few question on their life satisfaction and to express an evaluation on their trajectories of three life domains: partnership, family and activity. The interview concluded asking whether respondents would be willing to participate again to the study.

[2]In the face-to-face mode part of the check was done during the interview

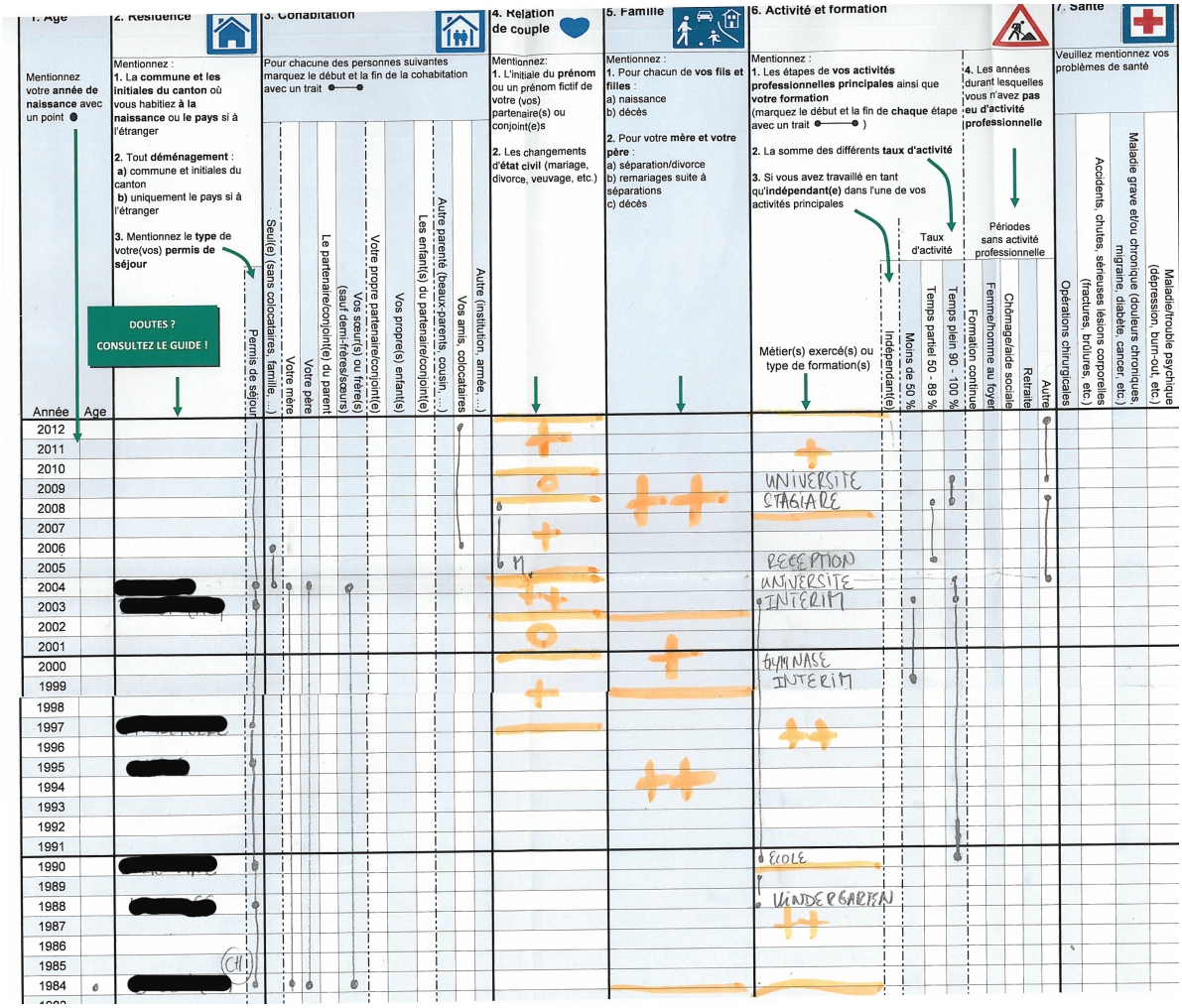

Figure 2: Example of Filled SPLC with Subjective Evaluation of Life Domains

420 households were interviewed between October 2012 and March 2013 with different contact and interviewing modes. Among members of those households 505 individuals completed the SPLC. 287 households were contacted by telephone and filled grid and household questionnaire by CATI. Among members of those households, for 165 individuals the SPLC was sent by mail and completed by the respondents without the interviewer, while for 138 people the SPLC was administered in a face-to-face (F2F) interview. Alongside those contacted by telephone, 133 households were contacted in F2F and filled the first two sections in CAPI. From those households, 202 individuals filled the SPLC in F2F mode. For all respondents the fourth section was in F2F.

The overall response rate per household was 37.5.4%, and 57.3% of the eligible individuals within the households responded to the SPLC. In 18.9% of the households all the eligible individuals filled the calendar. Concerning difference between contact mode, the response rate in the CATI contact was 40,1%. Half of those households received the SPLC by post, while the other half filled the questionnaire in F2F interview. Among those receiving the questionnaire by post, the response rate was 62.5%, and in 22.3% of the households all the eligible members answered the SPLC Concerning the F2F mode for the third interview section, the response rate was 42.5%, with 15.4% of the whole household filling the questionnaire.

Aside CATI contact, the F2F contact produced 31.1% of responses, of which a 79.5% filled the SPLC in F2F. 19.1% of the households the calendar was filled by all eligible members. All analyses were performed on raw (i.e., not cleaned and not weighted) data.

2.2 Description of the Sample

The SPLC was thus filled by 505 respondents. Gender distribution was 48.12% men and 50.3% women. Mean age was $M = 51.32$ with a standard deviation of 18.34. The oldest respondents were born in 1918 and 1922, while the five youngest were born in 1998. A distribution by cohort is presented in Table 1, while Table 2 and 3 present respectively civil status and level of education. All information are extracted from the SHP grid questionnaire. The missing values that appear in many tables correspond to a few cases for which the grid was wrongly administered after the calendar and the respondents refused to continue the interview, thus the grid was not filled.

Concerning education (Table 3), in order to simplify the analyses, the level of education was recoded into three categories: achieved or non-achieved primary education or post-compulsory orientation (24.03%); secondary education (58.86%); tertiary education (17.11%).

	Frequency	Percent
>1988	50.00	9.90
1979-88	59.00	11.68
1969-78	60.00	11.88
1959-68	111.00	21.98
1949-58	96.00	19.01
<1949	114.00	22.57
Missing	15.00	2.97
Total	505.00	100.00

Table 1: Birth Cohorts Distribution

	Frequency	Percent
Missing	8.00	1.58
Single / Never Married	120.00	23.76
Married	305.00	60.40
Separated	10.00	1.98
Divorced	36.00	7.13
Widow	24.00	4.75
Registered partnership (same sex partnership)	2.00	0.40
Total	505.00	100.00

Table 2: Civil Status Distribution

2.3 Questionnaire Design

The SPLC was designed on the calendar-based questionnaires elaborated for the *Vivre / Leben / Vivere* and the *Family tiMes* surveys. The *Family tiMes* LHC was administered and filled by an interviewer and asked specific questions on moves, family, job, couple relationship and cohabitation histories. The *Vivre / Leben / Vivere* was instead self-administered and had a lower level of standardization. The SPLC tried to achieve standardized data with a self-administered mode. As shown in Figure 1, the SPLC was composed in six columns, one for each life domain. The columns "cohabitation", "health", and part of the "activity" were to be filled marking lines or points to indicates the listed periods and events. In the other columns respondents were asked to handwrite the requested information. To maximize between-mode comparability, the SPLC was designed to be self-administered by the respondent also in F2F interviews. F2F interviewers had to follow an interview protocol aimed to guide the respondent in filling the questionnaire. This procedure had, in other words, the aim of minimizing the interviewer's effect which can be serious in LHC surveys (Brändle, Morselli, & Spini, 2013). The protocol followed the left-to-right direction of the questionnaire, but respondents were informed that at any moment they could either jump forward to fill other life domains first or step back to re-edit the information they had already indicated.

2.3.1 Cognitive Interviews

Two sets of ten cognitive interviews were conducted in autumn 2011, with "think aloud" technique and probing questions (Willis, 2005). The aim was to improve the format of the calendar to facilitate self-administration and to observe potentialities and backdrops of this method. To fulfil these goals the SPLC was administered to respondents with high probability of having difficulties in filling the calendar without interviewer's intervention. Those were adult respondents, with low level of education and not native speaking. Twenty respondents evenly split by gender, aged 35-66 and medium-low level of education were selected. During the cognitive interview, respondents were informed that the questionnaire had to be filled without the help of the interviewer.

The interviews showed that respondents had difficulties to handle the large format of the tool. Some of the original instructions were vague or not clear, some other were too specific. The average time for filling the calendar was 60 minutes. Following these indications the outlook of tool was improved, as well as instructions were clarified and a manual on how to properly fill the questionnaire was created, including also a "Frequently Asked Questions" section.

The format of some sections of the calendar were restructured to speed up the filling process. In particular the sections of health and job specifications were redesigned to allow respondent to mark pre-listed events.

Concerning the way respondents filled the calendar, observations of the cognitive interviews confirmed that the LHC format allowed respondent to better rember and date some events by using others life domains as time landmarks. According to Belli et al. (2001)'s rationale for the use of the calendar and in line with Glasner (2011)' results showing that the number of edit (i.e., the corrections) in LHCs were as twice as much than those of the conventional questionnaires, the SPLC cognitive interviews showed different styles of filling the questionnaire, some of which benefited of cross-referencing between life domains.

During the interviews an external observer recorded the sequence used by the respondents to fill the different columns of the calendar. Three out of ten respondents filled the the calendar in linear order, starting from the first column on the left (residence), then passing to second (cohabitation), and so on until they reached the last column on the right (column 6, health). Another group of respondents completed the calendar in sequential order except for the first two columns (residence and cohabitation). They used in turn residence and cohabitation as anchoring for recollecting the changes in both domains. Once these two columns were properly completed, respondents went linearly through the rest of the calendar in a linear manner from columns 3 to 6. The third group of respondents (four out of ten) completed the task in a less sequential order, using the information of certain columns to step back and make changes in the previous data. While the first group respondents did not re-edit the information they had reported, the graphical and conceptual structure of the calendar helped the other respondents in retrieving and reporting autobiographical memories. Without the calendar we can imagine that respondents in the second group would have found more difficulties in accurately recalling their residence and cohabitation history and those in the third group would not have gone back and corrected the information they had previously reported.

These observation confirmed that the use of SPLC facilitate the report of chronological events, minimizing errors and misdating by allowing the respondents to link their answers on with another and re-edit them in case of mistakes.

2.3.2 Subjective Assessment of Life Trajectories

A series of pre-tests were conducted with undergraduate students to evaluate three methods of subjective assessment of life calendars: Lelièvre's method with Likert scales, curve graphing, and a hybrid method between the two. Questionnaires were administered by LIVES researchers, which recorded the administration and the final interview about respondents' reactions in order to test the instructions and the different methods of evaluation. The interviews suggested that the hybrid method was to be excluded, and that evaluation of specific life domains (e.g., work, family, partner relationship, health) were more reliable than a general evaluation of well-being across the life course. They also gave important information on how to set the instrument, the instructions to give and possible problems of data coding.

In a second set of interviews 60 undergraduate students were asked to fill the life-calendar and evaluate each domain of their life using both Likert type and the curve graphing methods. Questionnaires were self-administered and the order between the two methods of evaluation was randomized. After filling the calendar and the subjective evaluation, respondents were asked to fill a questionnaire about their reactions to the instruments. To code graphical data an ad hoc procedure was set up for estimating the interpolation function of the graph. Results showed that the two methods generally produced similar data (Spearman's $\rho > .60$), but not for all life domains. More incongruence between the two methods of evaluation were found for activity and partner relationship Brändle (2012). In addition, respondents considered the graph method more difficult to understand than the Likert, even if it was also decalred to be more satisfactory and complete.

Similarly, in the cognitive interviews respondents were asked to fill both the Likert and the graph methods to evaluate partnership, family, activities, and health separately. The results were fairly similar to the pre-test with students, although some aged respondents

were more at ease with the Likert method and some others complained that graphing was cognitively more demanding because focused the attention more on single years than periods. For these reasons the Likert was preferred over the graph method and the final version asked respondents to divide their trajectories into periods and express an evaluation on a 5-point bipolar scale (ranging from -2 = "very negative" to $+2$ = "very positive").

2.4 Indicators of Performance

Given the absence of an external source of exact data (i.e., without measurement error), the present study focused on two main parameters used in previous studies for assessing the quality of the SPLC: completeness (see Engel et al., 2001) and coherence (see Becker & Sosa, 1992).

In this study, completeness was estimated by computing five indicators: number of residential moves; number of intimate partners; number of children; number of employments; number of unemployment periods. Those indicators were therefore used to compare the performance of the tool across different modes of data collection and in comparison to the life-history calendar used in a earlier cross-sectional survey (*Family tiMes*), that was conducted in Switzerland between 2010 and 2011.

In addition to completeness, coherence was assessed in three ways: analysis of interviewers' corrections of the calendar during the check section of the interview; internal coherence between mutually exclusive dimensions of the SPLC; coherence between data collected in the first section of the interview (i.e., the grid) and data collected with the calendar.

Given that the SPLC was designed to study vulenrability, the tests of coherence were also performed comparing different subsamples. In particular a subsample of respondents defined as "vulnerable" (i.e., low economic and cultural capital) were compared to the general population.

3 Descriptive Analysis of the SPLC

The residential trajectory is one of less controversial parts of the calendar to be filled. As suggested by some studies (e.g., Auriat, 1993; Berney & Blane, 1997; Blane, 1996), and confirmed by our cognitive interviews, the residential moves are important anchoring landmarks that facilitate memory recollection of other events. Respondents reported on average 4.93 residential moves along their life course (SD =2.99), with the largest group of respondents reporting 3 moves (see Figure 3).

Concerning cohabitation, Figure 4 shows the frequency of the cohabitation trajectories. As it might be expected, the majority of respondents reported of having lived with their parents. A very few respondents (5.94%) declared of having shared the house with the parent's partner. A certain caution should be used in interpreting this result as from debriefings with the interviewers revealed that some respondents had sometimes misread the label and indicated their own partner instead of the parent's one. In principle this information should have been checked by the interviewers in both the F2F mode and mail modes, but a certain degree of errors cannot be completely excluded. Few periods of cohabitation were also recorded for partner's children (4.36%), non-listed relatives (10.69%), friends (16.44%), and other (10.69%).

Respondents reported on average 1.43 intimate relationships (SD =1.11, $Mode$ =3). In relation to those, 371 marriages reported, as well as 2 registered partnership, 27 separations, 64 divorces, and 26 widowhoods.

However, it should be noticed that sometimes the civil status change is incoherent or not clear. In particular, the 1.19% of the participants reported separation (either separation, divorce, or widowhood) as a first civil status change, which is impossible (Table 4). This may depend on the fact that non-married long term couples may have reported separations

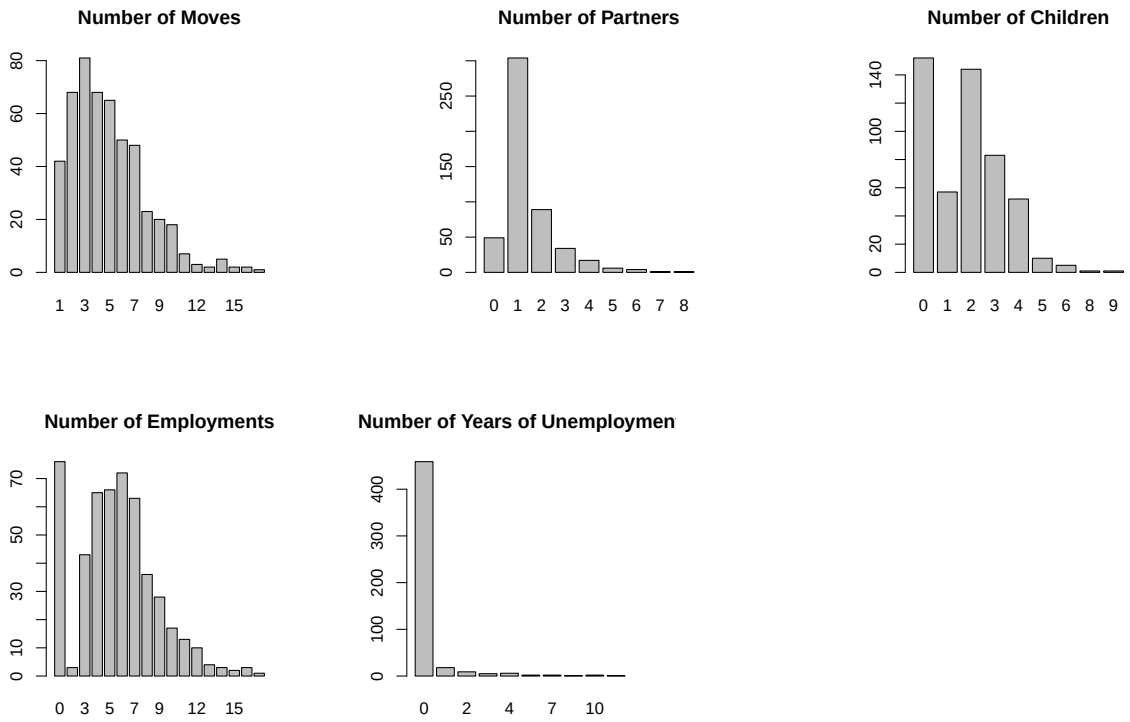

Figure 3: Frequencies of the Indicators of Completeness

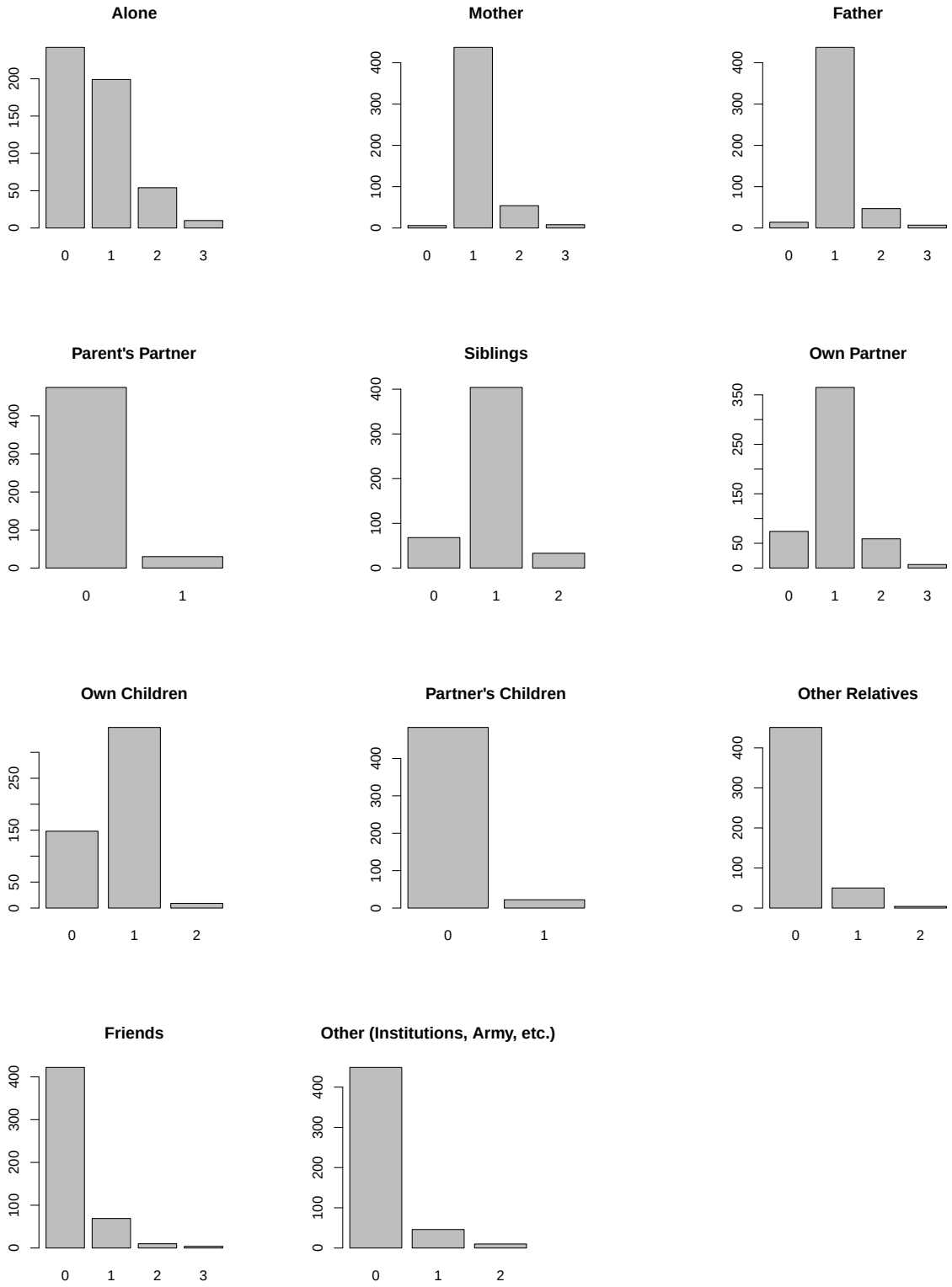

Figure 4: Frequencies of Cohabitation Responses

from the partner in terms of life event instead civil status change. Experience of separation or bereavement are indeed important life markers even if they are not mirrored in a change of civil status. On the other hand, they could also be the symptoms of missing or mis-reported data.

Table 5 shows the frequency of family events across the sample. The question on second marriage of parents as listed on the calendar was probably not fully clear. Three episodes of second marriage of the parents did not specify to which parent were referred to. In total, recorded parent's second marriage were 32. The variable "other" (n =20) refer mainly to birth of siblings. Information on siblings were not directly asked in the column family, but were nevertheless reported by some respondents.

Concerning the number of children, figure 3 show a zero-inflated distribution. Among respondents with children the average was 1.78 children, SD =1.53, which is in line with other statistics of the Swiss population (Jeanneret & Goebel, 2012).

The number of jobs per respondent is displayed in figure 3, while figure 5 reports job specifications and additional information about periods without employment. The number of respondents reporting unemployed periods was 9.11% of the sample, however the length of the periods differed substantially from respondent to respondent, ranging from 0 to 12 years (M =0.28, SD =1.19). If the rate of respondents reporting unemployment seems low, it may be instead plausible once framed within the unemployment rate of Switzerland, which historically is lower than other European countries and only increasing in recent years (Champion, 2011).

Concerning employment, the calendar asked respondents to report the changes in job rates, by indicating the sum of the activity rate at any change. The majority of the reported jobs were indicated at full-time rate. However, the there is a certain discrepancy between the overall number of reported jobs (n = 2331) and the overall number of activity rate changes (n = 1888). This is discrepancy is given by the SPLC asking the changes in the overall activity rate and not the rate of each activity.

The last column was dedicated to health problems. It is known that health-related events are keen to be under reported in retrospective data collection (Means, Nigam, Zarrow, Loftus, & Donaldson, 1989). Following observations and respondent reaction during the cognitive interviewing, the health events were grouped in four main categories. This choice was made to reduce the sensitivity of the response and social desirability on the one side, and to reduce the time needed to fill the questionnaire on the other. Figure 6 reports frequencies of for each category. 19.8% of the sample reported at least one serious illness, and 6.53% reported at least one psychological problem such as burn-out or depression. Thus, if on the one side some information on specific illnesses is lost with this method, on the other it seems to encourage response.

4 Between-Mode Comparison

The SHP pilot study was structured into two contact modes (telephone and F2F). In the telephone contact, respondents answered to the grid questionnaire via CATI and then were randomly assigned to fill the life calendar by post or with a F2F interview. In the case they received the questionnaire by post, an interviewer was sent to collect the completed calendar, to check whether it was properly filled (i.e., without incongruence), and in half of the cases to administer an additional part of the interview with the subjective evaluation of the life trajectory. In both F2F administrations of the calendar (CATI+F2F and F2F+F2F) the calendar was completed an checked in one single time, along side with the last part of the interview.

In all the administration modes the SPLC was self-completed on paper by the respondent. The interviewer followed indeed a protocol to help the respondent understanding the task, but did not completed himself the calendar as in a classical CAPI interview. In other words,

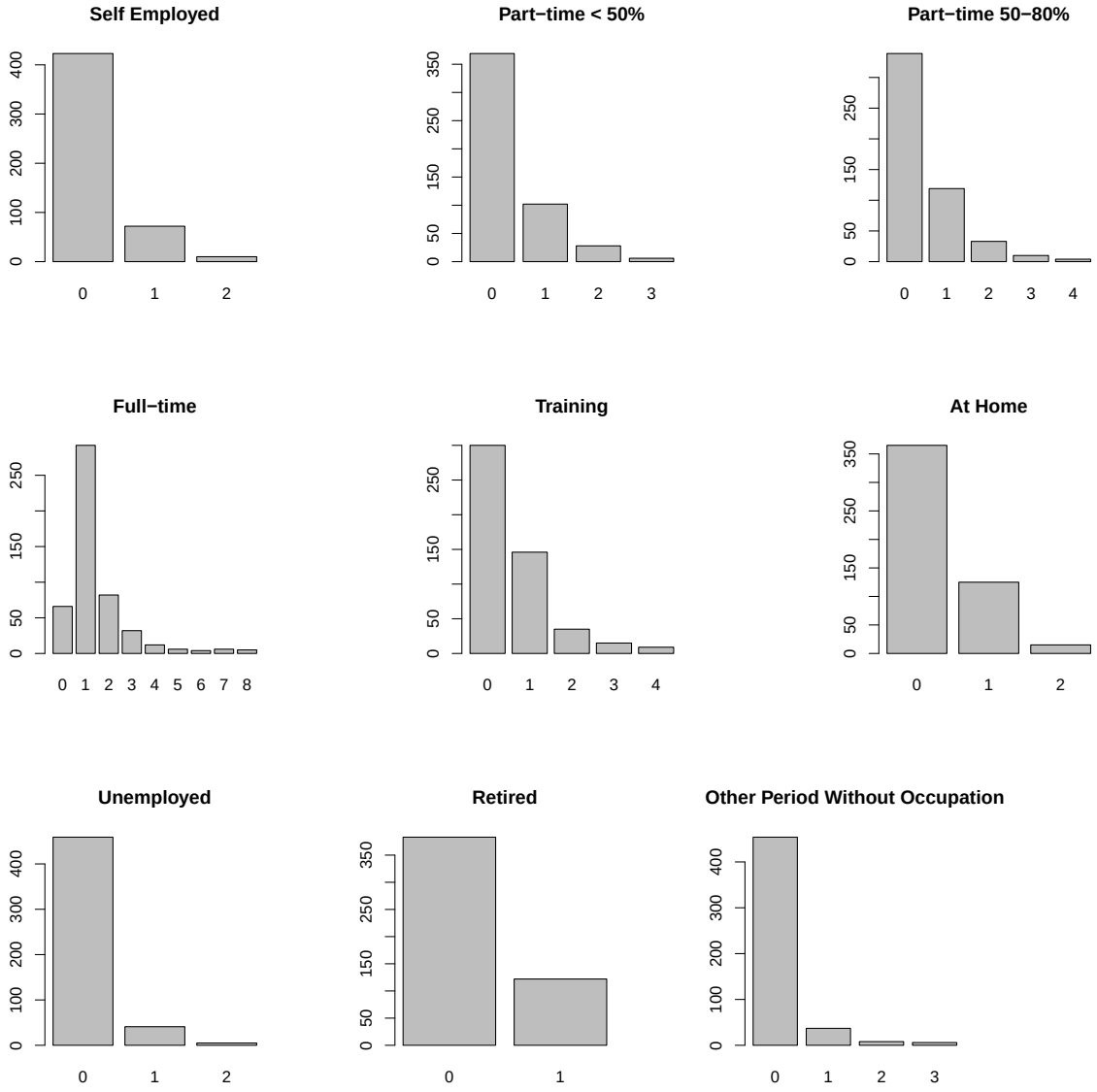

Figure 5: Frequencies of Job Specifications and Periods without Activities

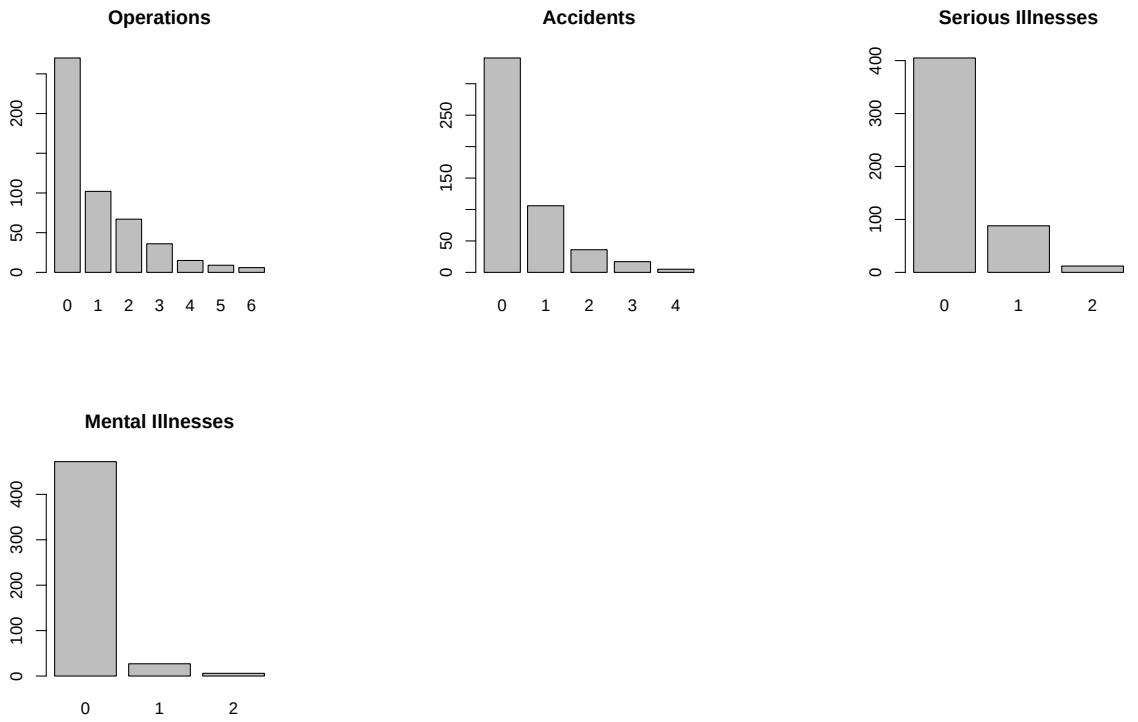

Figure 6: Frequencies of Health Problems

all the three modes were paper-and-pencil based, the interviewer was present to support the self-administration.

The average time of the interview in the F2F mode (about one third of the sample) was 45 minutes, to be split into 40 minutes for the short version of the interview (i.e., without the subjective evaluation) and 50 minutes for the version with the subjective evaluation. However, the adminstration time was collected only forthe F2F mode, thus the mean times of the interview were estimated on relatively small subsamples. The distribution between modes is reported in Table 6.

As shown in Table 7, there were not gender and education differences among respondents in the three modes. Concerning income, in the F2F+F2F mode there were higher proportions of the lower incomes categories, while the middle class was slightly more concentrated in the CATI+Mail mode. No differences concerned the higher classes. However the χ^2 test was not significant and the Cramer's V was very small ($V = 0.1$).

A different picture emerged when looking at the age distribution. Both Table 7 and Figure 7 show that the age was not evenly distributed across the three modes. In particular the CATI+F2F are slightly older, while the younger proportion of the sample is concentrated on the F2F+F2F mode. However, the Cramer's V was relatively small ($V = 0.19$) suggesting that the difference between cohort was indeed marginal. Nevertheless, given that respondents were randomly assigned to the three conditions, the between-mode difference on age suggest that the contact mode may have generated a bias at the very beginning of the data collection process. Thus, age difference was used as covariate in the analyses of the next section for disentangling the mode effect.

On the other hand, given that the SPLC was self-administered in all the modes and interviews were structured to minimize the interviewers effect, we expected to find none or little differences in the distributions of the five indicators across interview-modes. That is the mode had a marginal effect on the answers reported by respondents in the three conditions. Otherwise we would need to look to socio-demographics variables that might explain why under certain modes we might obtain different distributions of events. In case that also the explanatory variables failed to explain the between-mode variance we might conclude that the mode of data collection facilitated or inhibited reporting certain events in the calendar.

4.1 Between-Mode Difference on Completeness

The five indicators of completeness were used to investigate between-mode difference: number of moves, number of intimate partner relationship, number of children, number of experienced jobs, number of years of unemployment.

Figure 8 provides an overview of the differences between modes. The graphs represent the distribution of the maximum number of events per respondent among the three modes. In line with the expectations, little difference was found on all five indicators. However a little caution should taken in interpreting these results. As discussed earlier, the between mode difference might be a consequence of a slight selection bias and age difference. Thus, an analysis that can control for covariance between different variables is needed.

4.2 Modelling Over Representations of Zeros: The Hurdle Model

As showed in the previous section, the indicators of completeness had a non-normal distribution, and therefore violated the assumptions of linear regression modelling. In particular, the number of moves and the number of partners had a Poisson distribution while the other three presented an over representations of zeros, as well an over dispersed Poisson distribution. The inflation of zeros in life event data is not surprising. In particular, there may be two reasons for which zeros were over represented. The first one is because the respondents did not mention any event for a particular life domain or a relatively rare event type, for example unemployment. The second one consisted of declaring only some events for that domain,

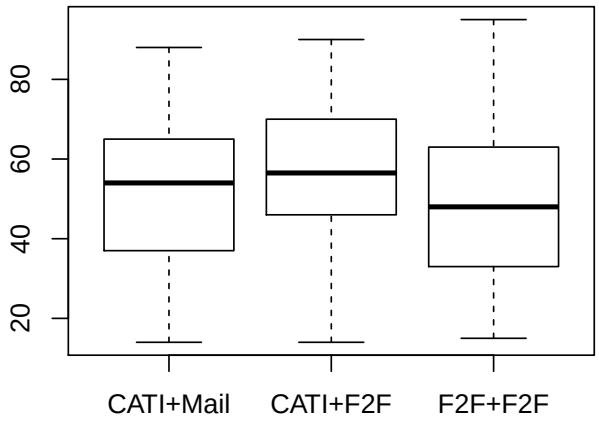

Figure 7: Differences between Modes on Age

but not all of them. For example, short periods of unemployment could be mis-reported or not reported.

Because we were interested here to analyse the effect of the mode of interviewing on each of these two aspects, we estimated for each of our indicators of completeness a hurdle model (Zeileis, Kleiber, & Jackman, 2008). Hurdle models are counting process models taking into account an over-representation of zero count and are divided into two components. The aim of the first component is to analyse cases with zero count versus cases with at least one count. In the present case, this first component considers whether the respondent had mentioned at least one event or none. It is a logistic model in which we estimate the effect of covariates on the odds to indicate at least one event versus none. The mode of interviewing is introduced in this logit model as covariate.

The second component of the model is a left truncated counting process model in order that estimates the number of counts for cases in which there is at least one count. In the present case, this second component aimed to analyse the number of events declared by respondents according to the mode. Given the tendency to over-dispersion of the empiric distribution of number of events, negative binomial function was used as parametric distribution for this second component rather than the Poisson. This hurdle model can be written:

$$\begin{cases} Log[E(y)] = a + bx & \text{with } var(y) = E(y) + \frac{E(y)^2}{\theta} \\ Log\left[\frac{P(y)=0}{P(y)>0}\right] = c + dx' \end{cases} \tag{1}$$

Where y is the number of events, $E(y)$ represents the expectancy of y and x and x' are independent covariates respectively introduced in the counting process and the logit sub-models. a and b in one side and c and d in the other sides are parameters to be estimated, while θ, also to be estimated, is a parameter related to the variance of the negative distribution. In the present case, covariates introduced in both sub-models are the same ($x = x'$).

Hurdle models were estimated using the `pscl` library for `R` (Jackman, 2012). The main

covariate that was used was the mode divided into three modalities: contact by mail (reference category), face to face with a contact by phone, and face to face with a direct contact with the interviewer. Given that the young people are as a matter of fact less likely to have experienced several births of their own children, unemployment periods, or changes of jobs, the youngest cohort (people with less than 25 years) was removed from the analyses.

4.3 Model Results

The baseline models with only the mode as predictor are displayed in Table 8 and 9. Table 8 shows the probability of passing from none to one event, coefficients should to be interpreted as in logistic regression model. On contrary, coefficients in Table 9 can be interpreted as in Poisson (count) regression and models the variance of events from one to infinity. No significant difference between the mail and F2F modes was found in the count model, that is the number of events once reported was similar across modes. By contrast a marginal effect was found in the zero model for the number of children and unemployment.

Gender, education and cohort were introduced in two steps to account for mode difference in number of reported children and number of episodes of unemployment (Table 10). The effect of mode became non significant once controlled for age (Models 2 and 4). Compared to F2F+F2F respondents, in the CATI+Mail mode respondents were on average 74.55% more likely to report children (*Odds Ratio* = 1.75) but 222.51% more likely of not reporting unemployment ($OR = 0.45$). Similarly CATI+F2F were 25.9% more likely to report children ($OR = 1.26$) and slightly less likely of reporting unemployment (OR =0.86). However, standard errors were relatively large and coefficients were not statistically significant.

Thus, we can conclude that the mode effect (once controlled for age) was negligible, and confirmed our expectation that using paper-and-pencil self-administration and non-intrusive interview protocol allow to produce highly comparable data. In the next section the SPLC data were tested for external validation, by comparing results from the SPLC and the *Family tiMes* survey, which used a similar calendar with high interactive F2F interviews.

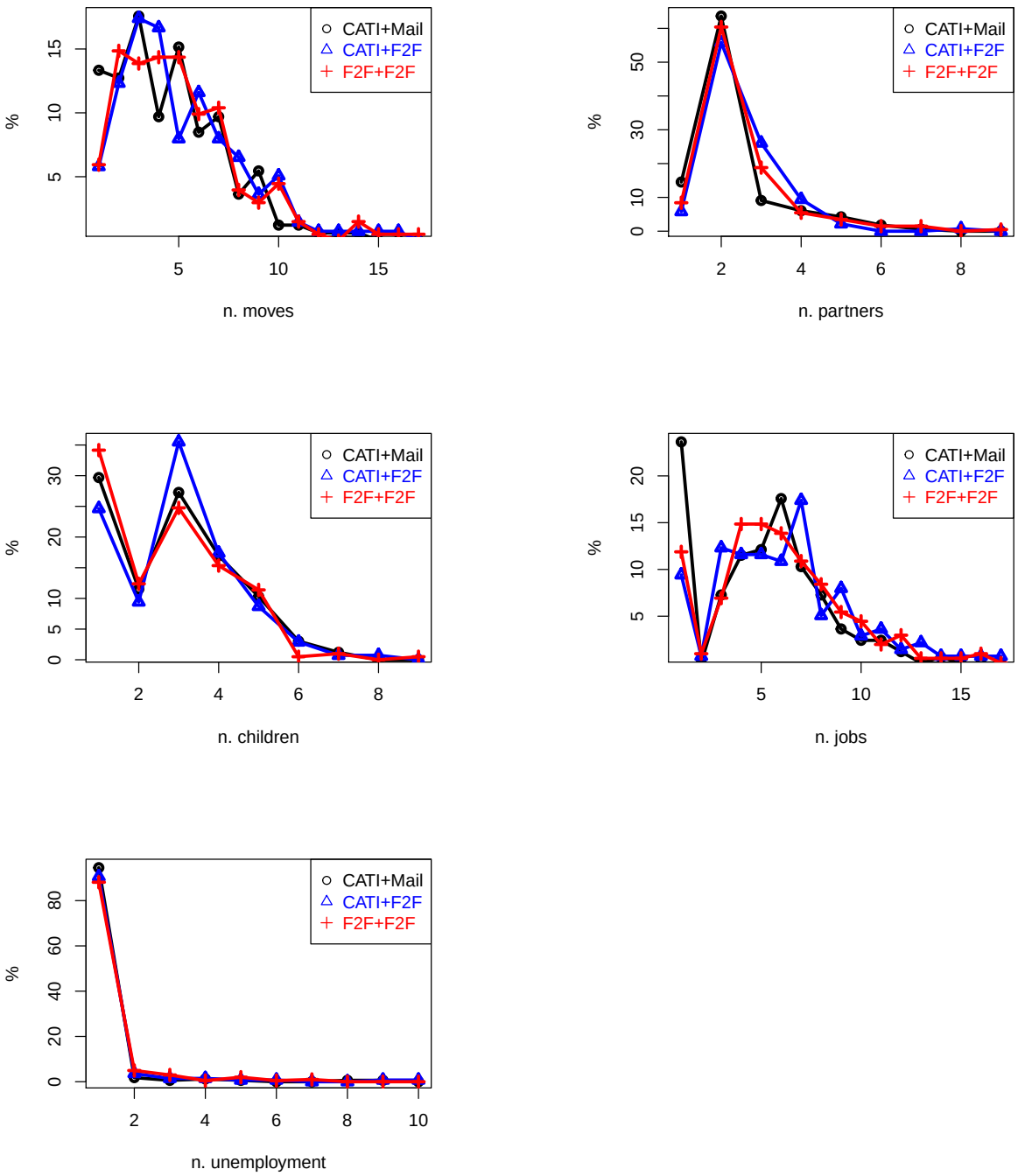

Figure 8: Mode Comparison

	Frequency	Percent
missing	8.00	1.58
pre-scolarite non-obligatoire	2.00	0.40
ecole obligatoire inachevee (ou jusqu a la 7eme annee incluse)	20.00	3.96
n a termine que l ecole obligatoire (secondaire i: 8eme ou 9eme annee, y inclus cycle d orientation, pregymnase, classe	62.00	12.28
2 ans: formation professionnelle duale (entrainement + ecole) avec attestation federale ou ecole professionnelle a plein	19.00	3.76
3 a 4 ans : apprentissage, formation professionnelle duale avec niveau cfc (sans maturite professionnelle)	184.00	36.44
3 a 4 ans : ecole professionnelle a plein temps, ecole superieur de commerce, ecole de metiers	16.00	3.17
maturite professionnelle (integree ou apres un apprentissage ou l ecole superieure de commerce)	10.00	1.98
2 a 3 ans: ecole de formation generale (ex. ecole d administration et des transports, ecole de culture generale), ecole	6.00	1.19
baccalaureat / maturite (gymnase)	20.00	3.96
1 an: preapprentissage, ecole commerciale, formation generale, 10eme annee scolaire, stage menager ou sejour linguistiqu	16.00	3.17
formation professionnelle superieure avec maitrise, brevet federal ou diplome federal ou formation equivalent (ex. profe	38.00	7.52
2 ans a temps plein ou 3 ans a temps partiel: ecole technique ou professionnelle, technicien et, ecole sup. de gestion	7.00	1.39
3 ans temps plein ou 4 ans a temps partiel: ecole professionnelle superieure, ets, escea, esaa, ies	13.00	2.57
universite, haute ecole universitaires, epf (hautes ecole universitaires cantonales, hautes ecoles polytechniques federa	55.00	10.89
haute ecole pedagogique hep (etudes pour l enseignement aux degres prescolaire et primaire, l enseignement pour le degre	4.00	0.79
haute ecole specialise hes, bachelor hes, master hes, diplome postgrade hes	16.00	3.17
ecole normale	9.00	1.78
Total	505.00	100.00

Table 3: Highest Level of Education

	Frequency	Percent
missing	157.00	31.09
married	329.00	65.15
registered partnership	2.00	0.40
separated from spouse	2.00	0.40
divorced	3.00	0.59
widowed	1.00	0.20
other	11.00	2.18
Total	505.00	100.00

Table 4: Calendar's First Civil Status Episode

Event	Frequency
Parents' separation	17
Parents' divorce	18
Parents' marriage	3
Mother's marriage	17
Mother's death	185
Father death	250
Father's marriage	12
Child's birth	769
Child's death	15
Other	20

Table 5: Frequency of Family Events in the Sample

	Frequency	Percent
CATI+Mail	165.00	32.67
CATI+F2F	138.00	27.33
F2F+F2F	202.00	40.00
Total	505.00	100.00

Table 6: Modes: Contact and Grid + Life Calendar.

Variable		CATI +Mail	%	CATI +F2F	%	F2F +F2F	%	χ^2	p
Sex	Men	77	0.47	61	0.44	102	0.44		
	Women	88	0.53	77	0.56	86	0.56	3.71	0.16
Education	Primary (acheived or not)	37	0.22	35	0.25	46	0.25		
	Secondary	96	0.58	85	0.62	108	0.62		
	Tertiary	32	0.19	18	0.13	34	0.13	2.49	0.65
Income per year	< 50000	19	0.14	28	0.23	39	0.23		
	50000-89000	52	0.37	47	0.38	58	0.38		
	89000-143000	44	0.32	28	0.23	43	0.23		
	>143000	24	0.17	20	0.16	20	0.16	7.90	0.25
Age	>1988	26	0.16	8	0.06	16	0.06		
	1979-88	12	0.07	8	0.06	39	0.06		
	1969-78	18	0.11	16	0.12	26	0.12		
	1959-68	34	0.21	34	0.25	43	0.25		
	1949-58	34	0.21	30	0.22	32	0.22		
	<1949	41	0.25	42	0.30	31	0.30	36.85	0.00

Table 7: Crosstabulation between Mode and Gender, Level of Education, Income, and Cohort

Table 8: Zero Hurdle Models for the Mode Comaprison (Passage from Zero to Non-Zero Event)

	Dependent variable:									
	Moves		Partners		Children		Jobs		Unemployment	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Intercept	2.80^{***}	3.60^{***}	3.60^{***}	3.40^{***}	1.60^{***}	0.82^{***}	2.20^{***}	2.60^{***}	-2.90^{***}	-1.90^{***}
	(0.39)	(0.51)	(0.58)	(0.46)	(0.25)	(0.18)	(0.30)	(0.33)	(0.42)	(0.24)
CATI+F2F (ref. CATI+Mail)	1.30		0.41		−0.13		0.93		0.66	
	(0.81)		(0.92)		(0.34)		(0.55)		(0.53)	
F2F+F2F	0.83		−0.28		-0.75^{*}		0.46		1.10^{*}	
	(0.64)		(0.74)		(0.30)		(0.45)		(0.48)	
CATI+F2F (ref. F2F+F2F)		0.46		0.69		0.62^{*}		0.47		−0.40
		(0.88)		(0.85)		(0.30)		(0.56)		(0.40)
CATI+Mail		−0.83		0.28		0.75^{*}		−0.46		-1.10^{*}
		(0.64)		(0.74)		(0.30)		(0.45)		(0.48)
Observations	378	378	378	378	378	378	378	378	378	378

Note: *p<0.05; **p<0.01; ***p<0.001

Table 9: Count Hurdle Models for the Mode Comparison (Events from One to Infinite)

	Dependent variable:									
	Moves		Partners		Children		Jobs		Unemployment	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Intercept	1.40***	1.40***	−1.00	−0.91	0.84***	0.82***	1.80***	1.90***	0.05	−0.43
	(0.07)	(0.06)	(0.61)	(0.60)	(0.07)	(0.07)	(0.04)	(0.04)	(1.40)	(1.30)
CATI+F2F (ref. CATI+Mail)	0.04		0.0000		−0.02		0.07		0.31	
	(0.10)		(0.24)		(0.11)		(0.06)		(0.83)	
F2F+F2F	0.01		0.12		−0.02		0.03		−0.48	
	(0.09)		(0.23)		(0.10)		(0.05)		(0.77)	
CATI+F2F (ref. F2F+F2F)		0.03		−0.12		0.004		0.04		0.79
		(0.09)		(0.23)		(0.10)		(0.05)		(0.63)
CATI+Mail		−0.01		−0.12		0.02		−0.03		0.48
		(0.09)		(0.23)		(0.10)		(0.05)		(0.77)
Observations	378	378	378	378	378	378	378	378	378	378
Log likelihood	−895.00	−895.00	−448.00	−448.00	−647.00	−647.00	−918.00	−918.00	−184.00	−184.00

Note: *p<0.05; **p<0.01; ***p<0.001

Table 10: Count Hurdle Models for Numebr of Children and Years of Unemployment

	Dependent variable:			
	Children		Unemployment	
	(1)	(2)	(3)	(4)
Intercept	1.70^{***}	0.28	-1.90^{***}	-1.20^{*}
	(0.41)	(0.49)	(0.47)	(0.57)
CATI+F2F (ref. F2F+F2F)	0.61^{*}	0.23	-0.43	-0.15
	(0.31)	(0.33)	(0.40)	(0.42)
CATI+Mail	0.84^{**}	0.56	-1.10^{*}	-0.80
	(0.31)	(0.34)	(0.49)	(0.50)
Women	-0.23	-0.29	0.29	0.25
	(0.26)	(0.27)	(0.36)	(0.37)
Secondary educ. (ref. Primary)	-0.67	-0.51	-0.17	-0.50
	(0.40)	(0.43)	(0.45)	(0.48)
Tertiary educ.	-1.60^{***}	-1.40^{**}	0.03	-0.41
	(0.44)	(0.48)	(0.54)	(0.57)
1969-78 (ref 1979-88)		1.70^{***}		-0.19
		(0.46)		(0.56)
1959-88		1.80^{***}		-0.05
		(0.39)		(0.48)
1949-58		1.80^{***}		-1.50^{*}
		(0.42)		(0.71)
$<$ 1949		2.20^{***}		-1.90^{*}
		(0.46)		(0.73)
Observations	378	378	378	378
Log likelihood	-637.00	-617.00	-174.00	-164.00
Note:	*p<0.05; **p<0.01; ***p<0.001			

5 Comparison with Interviewer-Respondent Interactive Calendar Interview

The LHC literature suggest that part of benefit of the LHCs depends on the interviewer-respondent interaction (e.g., Belli et al., 2001). In other words, the benefits of calendar's structure in parallel life domain is amplified by the interviewer's help to liaising the different domains. This effect has been often underlined as a key aspect of the LHC interview and equally important in different modes (e.g., Belli, Bilgen, & Al Baghal, 2013; Bilgen & Belli, 2010). To a certain extent, such interaction was also introduced in CAWI self-administered LHCs by Glasner (2011) . According to her procedure, the respondent is guided throughout the questionnaire by specific indications and questions prompted on the screen to help the respondents to fill the calendar.

On the contrary, the SPLC tried to limit the interviewer-respondent interaction to facilitate the comparison between different administration modes. As seen in the previous section, this procedure seems indeed to work and the mode effect is marginal. However, because of the absence of interviewer-respondent interaction the data of the SPLC could be biased in all modes, and be affected by severe under-reporting. In this section we computed the same indicators of completeness for a F2F interactive LHC interview to be used as parameter for evaluating the SPLC performance. In line with the literature we would expect the interactive interview being more complete than the self-administered mode.

5.1 Data

The SPLC shares several dimensions and part of the layout with the LHC used in the *Family tiMes* survey. However, the *Family tiMes* interview was interviewer-administered and the calendar was displayed to the respondent but actually filled by the interviewer during a semi-structured interview. As already explained, the SPLC was self-administered also in the F2F mode. In other words, while the *Family tiMes* LHC relied on the interaction between respondent and interview, in the SPLC the interaction was minimized.

The *Family tiMes* survey is part of a cross-sectional study conducted in 2010 on two cohorts of 400 individuals born in 1950-55 and 1970-75 drawn from a representative sample of the Swiss resident population. This survey aimed at capturing in a life course perspective the changing web of constraints and opportunities in which family and occupational trajectories of individual unfold.

Beside questions on values and opinion and an instrument measuring personal network of relationships (Widmer & La Farga, 2000) that were recorded in CAPI mode, the *Family tiMes* survey included a life-history calendar administered by mean of a paper questionnaire in a F2F interview setting. The calendar was inspired by the French Ageven tradition (Vivier, 2006) and brought together six life domains (cohabitation, residence, intimate relationships, occupational activities and critical life events), with semester as time unit. The *Family tiMes* calendar was filled by the interviewer during the interview, and code in a standardized format in a second moment.

Respondents were asked to mention all qualitative as well as quantitative changes that had occurred in each domain. Similarly to the SPLC, to indicate a specific period, the interviewer had to draw a line from the beginning to the ending semester of the corresponding life-spheres, whereas qualitative information was recorder by indicating a rank number (of child, of partner, of job, ...). Concerning residence, the calendar recorded the number of

moves as well as the corresponding destination area code (even for changes within the same area). For intimate relationships longer than three months, each partner was identified and reasons for separation were recorded. Number, type, status and activity rate were recorded to tap occupational careers.

5.2 Descriptive Analysis of the Two Surveys

Because *Family tiMes* targeted two specific cohorts (i.e.,1950-55 and 1970-75), the same age groups were selected in the SHP database. This left the SHP pilot study sample with a relatively small sample size compared to *Family tiMes* ($n_{SHP} = 89$; $n_{FT} = 803$) and this need to be kept in mind when interpreting the results. Given the small sample size and the marginal mode effect in the SHP pilot study, all three modes were pooled together.

As shown in Table 11, there are no gender differences between the two samples. Also the level of education is similar, with a slightly tendence to have higher educated respondents in the SHP data. On the contrary, χ^2 statistics suggest some difference in relation to income distribution and age. Concerning income, the significant χ^2 is most likely affected by the high number of categories and the low effectives in the SHP sub-sample. When the income categories are reduced to only four, the test is no longer significant ($\chi^2 = 4.39$, $p = 0.22$) Nevertheless, in the SHP there is a relativity high number of middle-income respondents, while *Family tiMes* is more evenly distributed among low-, middle- and high incomes.

There are also marginal differences in the age distributions, with the SHP having a slight higher rate of middle-aged respondents. That is not surprising, given that the SHP sample unlike *Family tiMes* was not stratified by age. Overall, the χ^2 tests suggest taht the two samples are not statistically different from one another on three out of four socio-demographic variables, allowing a further comparison of the results of the two calendars.

5.3 Between-Survey Differences on Completeness

Given the similarities between the surveys, but the different adminsitration mode, we expected to find more events reported in the interviewer-administered *Family tiMes* calendar. We hypothesise that the five indicators of completeness (i.e., number of moves, number of intimate partners, number of children, number of changes in job[3], number of experienced episodes of unemployment) would would be larger for *Family tiMes* than the SHP data.

Figure 9 shows at the glance the comparison between the two surveys. The graphs show the distribution of each indicator of completeness. In contrast with the litterature that posits that interviewer-respondent interaction help to have more complete calendars, our results show that the distribution of the events collected with the self-administered SPLC mirrors the one obtained with the in-depth interview of *Family tiMes*. The slightly higher variability in the SHP distribution (the red lines) is most likely due to the smaller sample size of the SHP. The SPLC recordered the same number of events, if not slightly more for certain domain.

However, particular attention should be given to the professional trajectory. The graph shows that whereas the SPLC seems to collect the same number of job changes, the rate of respondents with no job changes is larger than in the Fanily tiMes. In other words, a slightly highe number of respondents reported a single job for the all trajectory in the SPLC.

[3]Given the sample of adults, all respondents but one had had at list one job. The variable was considered then as the number of job changes, scoring zero for individuals that never changed employment

Variable		Family T.	%	SHP	%	χ^2	p
Sex	Men	397	0.49	49	0.55		
	Women	406	0.51	40	0.45	0.80	0.37
Education	Primary (acheived or not)	106	0.13	9	0.10		
	Secondary	542	0.68	55	0.62		
	Tertiary	148	0.19	25	0.28	4.80	0.09
Income per year	< 50000	139	0.19	9	0.12		
	50000-89000	218	0.30	29	0.38		
	89000-143000	192	0.27	24	0.31		
	>143000	169	0.24	15	0.19	4.39	0.22
Age	36	50	0.06	8	0.09		
	37	68	0.08	3	0.03		
	38	60	0.07	6	0.07		
	39	69	0.09	5	0.06		
	40	68	0.08	7	0.08		
	41	67	0.08	10	0.11		
	56	81	0.10	8	0.09		
	57	68	0.08	16	0.18		
	58	61	0.08	12	0.13		
	59	72	0.09	4	0.04		
	60	64	0.08	4	0.04		
	61	75	0.09	6	0.07	20.21	0.04

Table 11: Crosstabulation between Mode and Gender, Level of Education, Income, and Age

To further investigate the differences between the two surveys, a series of hurdle models was performed with the five indicators as separate dependent variables (see section 4.2). Both the zero (Table 12) and the count models (Table 13) confirm what observed in Figure 9, showing no between-survey differences but for the number of job changes in both zero and count models. That is SHP respondents were more likely to not report changes of job at all (*Odds Ratio* = 0.14). However, among respondents who reported job change, the SHP's had reported on average a higher number. The difference between the two surveys remained significant and the size of the effect unchanged even after controlling for gender, cohort, level of education and income (Table 14 and Table 15).

Despite these differences, the SPLC seems having recorded as many events as *Family tiMes*. Results seems to suggest that the self-administered mode performs similarly to the interviewer-administered calendar on our indicators of completeness. The result is quite surprising and in contrast with the literature. It would merit a deeper understanding and it could open the path to new application of the LHC method, reducing sensibly costs and time of data collection.

Table 12: Zero Hurdle Model for the Survey Comaprison (Passage from Zero to Non-Zero Event)

	Dependent variable:				
	Moves	Partners	Children	Jobs	Unemployment
	(1)	(2)	(3)	(4)	(5)
Intercept	5.00^{***}	4.00^{***}	1.20^{***}	5.20^{***}	-2.00^{***}
	(0.45)	(0.28)	(0.09)	(0.50)	(0.12)
SHP (ref. Family tiMes)	-0.62	16.00	0.12	-2.00^{*}	-0.10
	(1.10)	(2,021.00)	(0.29)	(0.77)	(0.39)
Observations	789	789	789	789	789

Note: *p<0.05; **p<0.01; ***p<0.001

Table 13: Count Hurdle Model for the Survey Comparison (Events from One to Infinite)

	Dependent variable:				
	Moves	Partners	Children	Jobs	Unemployment
	(1)	(2)	(3)	(4)	(5)
Intercept	1.60^{***}	0.43^{***}	0.87^{***}	1.60^{***}	1.00^{***}
	(0.03)	(0.06)	(0.03)	(0.02)	(0.28)
SHP (ref. Family tiMes)	0.01	-0.26	-0.16	0.26^{***}	-0.59
	(0.08)	(0.14)	(0.10)	(0.07)	(0.54)
Observations	789	789	789	789	789
Log likelihood	$-1,915.00$	$-1,187.00$	$-1,409.00$	$-1,852.00$	-488.00

Note: *p<0.05; **p<0.01; ***p<0.001

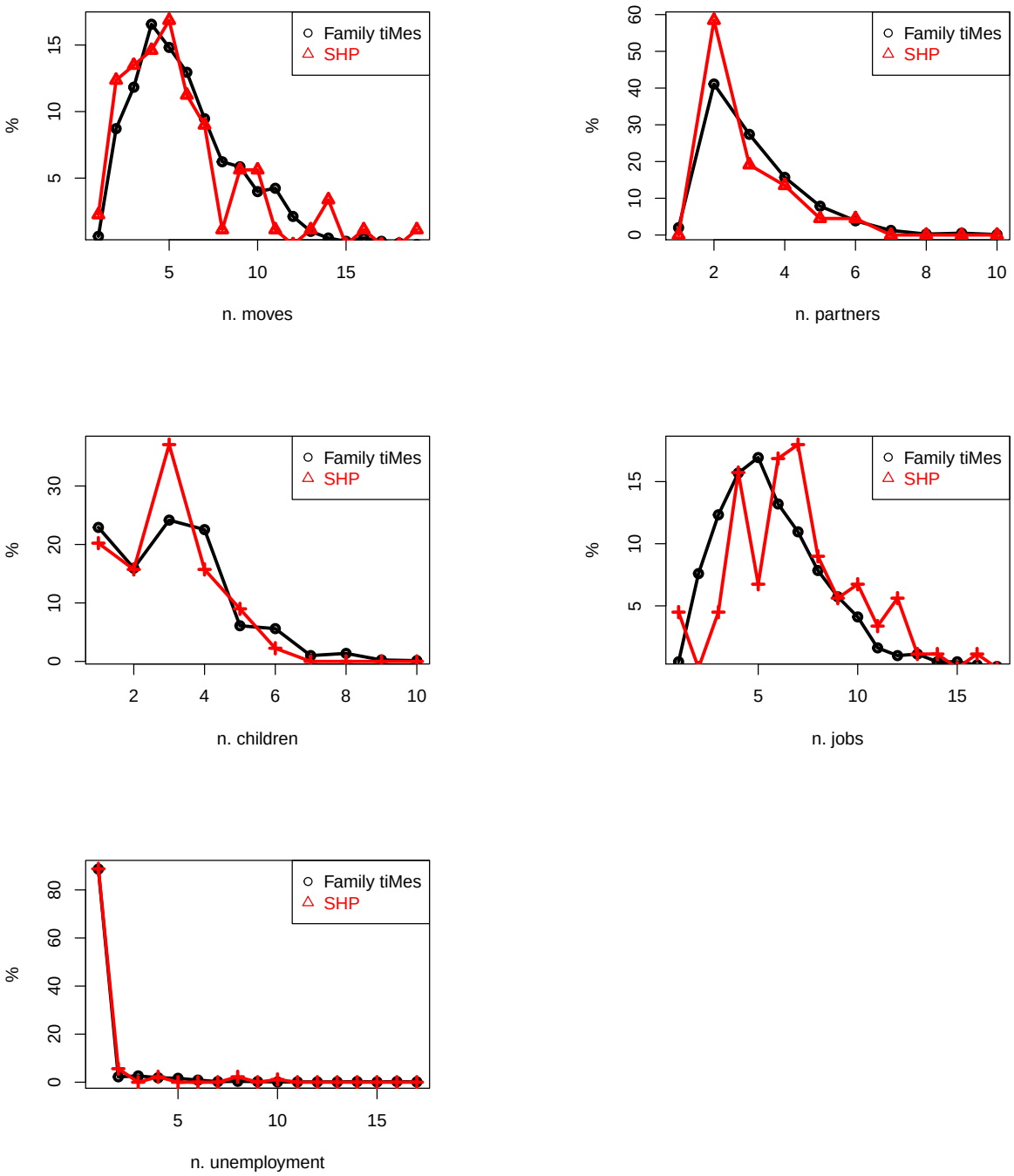

Figure 9: Survey Comparison

Table 14: Zero Hurdle Model for Job Change

	Dependent variable:
	Job Changes
Intercept	22.00
	(3,688.00)
SHP (ref. Family tiMes)	−2.00*
	(0.86)
Women	−1.70
	(1.10)
Cohort 1956-61	−0.01
	(0.80)
Secondary educ. (ref. Primary)	2.10*
	(1.00)
Tertiary educ..	1.10
	(1.20)
Income 50000-89000 (ref. less 50000)	−17.00
	(3,688.00)
Income 89000-143000	−18.00
	(3,688.00)
Income +143000	−0.89
	(4,897.00)
Observations	789
Log likelihood	−1,813.00
Note:	*p<0.05; **p<0.01; ***p<0.001

6 Analysis of the Interviewer's Check of the Calendar

This section investigates the interviewing protocol and the check list used by interviewers to spot errors and incongruences in the calendars. As explained earlier, the SPLC was self-administered in all modes. In F2F the interviewer was present during the whole time and used an interview protocol aimed to guide the respondent throughout the questionnaire. This protocol included a series of checks of the respondent's answers and registered refusals. In the mail mode, the interviewer went to collect the calendar performing a check of the answers in presence of the respondent. Thus, analysing the interviewer's checks allow us to understand the state of the data before they were handled and corrected by the interviewer and the respondents together. To assess the data quality, the analyses focused on two indicators from the interviewer's protocol and check-list: the number of errors and the number of refusals.

Table 15: Count Hurdle Model for Job Change

	Dependent variable:
	Job Changes
Intercept	1.20^{***}
	(0.09)
SHP (ref. Family tiMes)	0.25^{***}
	(0.06)
Women	0.17^{***}
	(0.04)
Cohort 1956-61	0.12^{**}
	(0.04)
Secondary educ. (ref. Primary)	0.23^{**}
	(0.07)
Tertiary educ..	0.46^{***}
	(0.08)
Income 50000-89000 (ref. less 50000)	-0.10
	(0.06)
Income 89000-143000	-0.14^{*}
	(0.06)
Income +143000	-0.12
	(0.07)
Observations	789
Log likelihood	$-1,813.00$
Note:	*p<0.05; **p<0.01; ***p<0.001

6.1 Response errors

Concerning response errors, seven indicators were extracted from both F2F and mail modes. They measured missing values and consistency of the responses in different columns of the calendar.

Two indicators were related to the residence trajectory. The first one (*check_year_residence*) measured errors made by respondents in terms of missing (i.e., blank) answers. Given that people always live somewhere, the residential trajectory should not have missing values, thus blanks in this column were to be considered as errors. In the F2F mode, this indicator was specified by the item "Check that there are no unfilled period or blanks in the residence column", while in the mail mode, the item linked to *check_year_residence* was: "There are information for each year since the birth in the residence column."

The second indicator (*check_place_residence*) controlled whether all municipalities and administrative districts (canton) were correctly written on the calendar. In the F2F mode, interviewers were given the instruction to "check that every location is indicated by munic-

ipality, the canton's initials and the country in the residence column". In the mail mode, interviewers checked whether "all the locations are indicated by the municipality + the canton (or by the country if abroad)".

Concerning cohabitation, two indicators were used. Similarly to *check_year_residence*, the indicator *check_year_cohabitation* measured the errors made in terms of missing values. In the F2F mode, interviewers were asked to "check that there aren't any unfilled period or blanks in the cohabitation column", while in the mail mode, they had to check whether "there is an information for every year since birth in the cohabitation column". The second indicator for cohabitation, *check_residence_cohab*, measured the consistency between the residence and cohabitation columns. In the F2F mode, interviewers had to "check information coherence between the residence and the cohabitation columns", and in the mail mode whether "information from residence and cohabitation columns are consistent". *Check_year_cohabitation* assessed errors within the cohabitation column, while check_residence_cohab controlled for the consistency of answers between two life domains.

For the intimate partnership domain, only an indicator of between column coherence was used to check the respondent's errors. *Check_relation_cohabitation* measured the consistency between the intimate partnership and cohabitation domains. In the F2F mode interviewers checked "the information coherence between intimate partnership life domain and cohabitation episodes", and in the mail "the informations of the cohabitation and intimate partnership columns are consistent."

Check_family_cohabitation controlled the consistency between family and cohabitation domains. For the F2F mode, *check_family_cohabitation* was measured by the interviewer's instruction "Check the consistency of information between the family and the cohabitation column", and "the informations between the family and the cohabitation columns are consistent" in the mail mode. 4.55% of respondents provided incoherent answers to these two columns.

Concerning activity and education, the *check_activity_education* indicator provided information about respondent's mistakes in terms of missing values. Alike cohabitation and residence, in this column there should not be blank years. Interviewers had to "check that there are no unfilled period or blanks in the activity and education column" in the F2F protocol, and they were asked whether "all the activities have been mentioned in the activity and education column."

Table 16 shows the frequencies, as well as the percentage of errors detected by each indicator. At first sight, the error's rates were very small. The highest number of errors (5.87% of errors) was found in the activity and education domain, which confirms that this column might have been more problematic or difficult to fill than the others.

Check indicator	Errors	Correct	Perc. of Errors
check_year_residence	17	477	3.44
check_place_residence	23	467	4.69
check_year_cohabitation	22	471	4.46
check_residence_cohab	18	468	3.70
check_relation_cohabitation	26	456	5.39
check_family_cohabitation	22	462	4.55
check_activity_education	29	465	5.87

Table 16: Frequencies of Errors

6.2 Refusals in F2F interviews

Unlike the previous paragraph, this part focuses on refusals during the F2F interview. For this reason, the indicators used in these analyses were present only in the F2F questionnaire and not in the mail mode. However, the analysis of refusal may allow us to better understand delicate aspects of the calendar throughout all the life domains.

The indicator *refusal_age* was relative to refusal of telling one's own age. The interviewers were asked to indicate whether the information concerning age was given or not.

Three indicators were used in the residence trajectory: refusal about residence at birth *refusal_residence_birth* concerned the refusal to reply to the question "In the column residence, mention please, in the space related to your year of birth, the municipality as well as the initials of the canton (or the country if born abroad) where you were physically living at the time of your birth"; *refusal_move* tapped refusals to the question "For each time you moved, please mention the municipality, the canton's initials (or country if abroad) even if you moved several times in the same municipality (or the same country)"; *refusal_nationality* concerns the refusal to providing the nationality.

Concerning cohabitation, *refusal_cohabitation* was related to the respondent's refusal to reply a general question about cohabitation: "In the cohabitation column, mention the people with whom you have lived since your birth and also if there were times when you lived alone".

Two indicators were used measuring for refusals in the field of couple matters. The first one, *refusal_couple* was related to the question: "In the column intimate partnership, mention the initial of the first name or a nickname for each partner you had until now in front of the corresponding year. For each person draw a point near the year relative to the beginning of the relationship and another at the end of it, draw a line to relate both points." The second indicator *refusal_marital_status* assessed refusals to marital status changes.

Concerning family history, *Refusal_birth_death* was related to the refusal to give information about the births and/or deaths of parents and children, and *refusal_parents_couple* was related to the respondents' refusals to give any information about their parents' separations, divorce, marriages, and death.

For the activity and education domain, three indicators were used: *refusal_education_activity* concerned respondents' refusals to answer a very general question about his activity and education: "Could you please mention now, all the steps of your education and of your activity. Mention also the period where you had no activity at all"; *refusal_activity_rate* tracks respondents' refusals about activity rates; *refusal_employee_independant* measured refusals to the question whether the respondent was employee or independent.

Concerning health, refusals to the question: "In the health column, mention for each problem of health (surgery, diseases, psychological problems like depression or burn-out)" were summarized in the *refusal_health_problems* indicator, and *refusal_health_problem_length* provided indications about the length of period with health problems.

Frequencies and percentage of refusals for each indicator are reported in Table 17. Similarly to the errors, also the refusal rates were extremely small. The largest rate was recorded for the *refusal_health_problem_length* (3.48%), but it is still relatively small and not exceed the 5% cut-off used to identify items with problematic missings data (Schafer, 1999).

Refusal indicator	Answered	Refusals	Perc. of Refusals
refusal_age	295	4	1.34
refusal_residence_birth	294	4	1.34
refusal_move	291	7	2.35
refusal_nationality	285	6	2.06
refusal_cohabitation	296	9	2.95
refusal_couple	295	4	1.34
refusal_marital_status	292	6	2.01
refusal_birth_death	294	6	2.00
refusal_parents_couple	297	5	1.66
refusal_education_activity	297	8	2.62
refusal_activity_rate	297	5	1.66
refusal_employee_independant	294	7	2.33
refusal_health_problems	283	5	1.74
refusal_health_problem_length	277	10	3.48

Table 17: Frequencies of Refusals

6.3 Latent Class Analysis

As mentioned earlier, the aim of this section was to analyse which conditions could induce the respondents to make errors or refuse to answer to some questions. According to Auriat (1993), errors and misdating provided at early stages of the LHC interview increase the probability of making mistakes in later parts of the calendar. Thus, we wanted to explore whether refusing to answer or making an error in a column or life domain would have an impact on subsequent answers. Unfortunately the very low number of mistakes and refusals made a proper sequence analysis not possible or unreliable. For this reason we used latent class analysis (LCA), with the aim of estimating groups of respondents with similar patterns of errors and refusals. LCA can be used as a clustering technique, in which group membership is estimated as the probability of each respondent to belong to each classes (Magidson & Vermunt, 2002). In LCA this probability is estimated on the basis of patterns of responses to a given number of observed indicators. In other words, by mean of a data mining approach we were interested to see whether respondents could have been classified according to different patterns of errors and refusal. Two separate exploratory latent class models were performed with the `R` package `poLCA` (Linzer & Lewis, 2011). The first analysis looked at the error indicators and the second one at the refusals.

The adequate number of classes was estimated using the Bayesian Information Criterion (BIC), the chi-square test of log-likelihood difference, and the model entropy (Collins & Lanza, 2010). The BIC offsets the fit of the model with the number of estimated parameters, assuming that a model is penalized by the number of estimated parameters. The model with the lowest values of BIC is the one with the best model fit. The log-likelihood difference test compares the improvement in fit between neighbouring class models (i.e., comparing k -1 and k class models) and provides a p value that indicates whether including one more class in the analysis produced a significant (i.e., larger than zero) improvement in the model fit. Finally, the entropy indicator gives information on the probability of respondents of being classified into more than one cluster. Values of entropy near 1 indicate high certainty of classification while small values of entropy indicate that the members of one class have a high probability

to be member also of other classes.

After having assessed the number of classes, standard procedures assign each individual to the most probable class using the latent class posterior distribution obtained during the LCA estimation. In a second step, the most likely class is treated as an observed variable to estimate differences on a number of independent variables (e.g. Clogg, 1995; Morselli & Passini, 2012). Not having a strong hypothesis on the independent variables that may predict the errors, the extracted classes were contrasted against socio-demographic indicators, namely age, gender, education and incomes.

6.3.1 Latent Class Analysis with Errors Indicators

The first LCA was performed on the error indicators in order to estimate groups of respondents with similar patterns of errors. The best fit was obtained wit the two-classes model. Table 18 shows the proportion of errors for each indicator. Class 1 grouped respondents who made nearly no errors, while class 2 grouped respondents who had given some wrong answers. The entropy of the model was 0.77, 445 respondents were classified in the first class, and 22 in the second one .

	Class 1	Class 2
check_place_residence	0.01	0.77
check_year_cohabitation	0.01	0.80
check_residence_cohab	0.00	0.69
check_relation_cohabitation	0.02	0.77
check_family_cohabitation	0.03	0.75
check_activity_education	0.01	0.70

Table 18: Proportion of Errors in Each Class

After having chosen the number of classes with the best fit of the data, the most likely class membership was cross-tabulated with some socio-demographic variables. None of those variable were strongly associated with the class membership: age, $\chi^2 = 1.57$, $p = 0.67$, Cramer's $V = 0.06$; gender, $\chi^2 = 0.56$, $p = 0.45$, Cramer's $V = 0.04$; education, $\chi^2 = 2.13$, $p = 0.55$, Cramer's $V = 0.07$; incomes, $\chi^2 = 0.66$, $p = 0.88$, Cramer's $V = 0.04$.

This result suggests that age, gender, education and incomes were not strongly associated to one class or the other, thus to the probability of making errors when filling the life calendar. However, the very small number of errors found by the interviewers reduce the statistical power of these analyses.

6.3.2 Latent Class Analysis with Refusals Indicators

The second latent class analysis was performed with the refusal indicators. In this analysis, we were interested to test whether respondents could be grouped on the criteria of their patterns of refusal.

Similarly to the previous model, also for refusals the best fitting model was the one with two classes that grouped respondents with higher rates of refusals (Class 1) to the others (Class 2, see Table 19). The entropy of the two-classes model was 0.98, with 11 respondents in the first class, and 244 in the second one. Thus, alike the previous model, the number of respondents with high rates of refusals was very small.

	Class 1	Class 2
refusal_residence_birth	0.26	0.00
refusal_move	0.44	0.00
refusal_nationality	0.18	0.01
refusal_cohabitation	0.53	0.01
refusal_couple	0.26	0.00
refusal_marital_status	0.28	0.00
refusal_birth_death	0.27	0.01
refusal_parents_couple	0.18	0.01
refusal_education_activity	0.62	0.00
refusal_activity_rate	0.35	0.00
refusal_employee_independant	0.36	0.01
refusal_health_problems	0.35	0.00
refusal_health_problem_length	0.19	0.03

Table 19: Proportion of Refusals in Each Class

Again, no significant relationship was found between refusals and specific socio-demographic characteristics: age, $\chi^2 = 1.57$, $p = 0.67$, Cramer's $V = 0.06$; gender, $\chi^2 = 0.56$, $p = 0.45$; Cramer's $V = 0.04$; level of education, $\chi^2 = 2.13$, $p = 0.55$, Cramer's $V = 0.07$; incomes, $\chi^2 = 0.66$, $p = 0.88$, Cramer's $V = 0.04$.

6.3.3 A Different Perspective: Complexity of Life Trajectories

To further investigate possible factors influencing refusals and errors, the two latent class models were put in relation to the complexity of respondents' life trajectories as reported in the calendar. The five indicators of completeness used in sections 4 and 5 were used here as indicators of complexity of the information reported in the calendar: number of moves, number of different life partners, number of children, number of different jobs, and finally number of years the respondents were unemployed. The principal idea was that the larger the number of events to be reported, the larger the probability of making mistakes. Alternatively we could suppose that the completeness of the calendar would not depend on the complexity of life trajectories but on the respondent's willingness to fill the calendar. We could thus expect that the less the events reported in the calendar, the more the respondent had executed the task with superficiality and errors were more likely to increase. Similar reasoning could be applied to refusals. The more the trajectory was complex, the probability of feeling like not answering the questionnaire. Alternatively, the less interested in the survey, the less the events they reported and the more the number of refusals.

A LCA was performed on the five indicators of completeness. The best fit was given by the three-classes model, which groups respondents on the basis of different styles of calendar completeness: respondents reporting simple life trajectories ($n = 56$), respondents with complex trajectories ($n = 63$), and the larger group of respondents somehow in between the other two classes ($n = 363$). Table 20 reports the mean by class for each indicator.

The LCA model of complexity was then cross-tabulated with both errors and refusal LCA models to test whether the complexity of of reported life trajectory was linked to higher or lower probabilities to make errors or refuse to answer. Neither the χ^2-test nor the Cramer's V supported the hypothesis of an association between complexity and errors ($\chi^2 = 1.97$, $p =$

	Complex	Simple	Medium
N. Moves	8.83	2.11	4.67
N. Unemployment	1.40	1.02	1.28
N. Jobs	8.25	2.86	5.64
N. Children	2.40	1.00	3.13
N. Relationships	4.38	1.36	2.27

Table 20: Average Number of Events for Each Class

0.37, $V = 0.06$). Although weak, an association was found between complexity and refusals ($\chi^2 = 6.76$, $p = 0.03$, $V = 0.16$), with respondents with high rate of refusals more likely to be in the complex class ($n = 4$) than the in the medium ($n = 6$) or simple ($n = 1$) life history classes. However, the number of effectives is so small that even the Cramer's V may not be reliable, and further data would be required to support this hypothesis.

Overall, the analyses of interviewers' checks and interviewing protocol highlighted very small rates of errors and refusals, confirming that the SPLC procedure was quite a robust way of collecting data. No strong evidence was found in support of a domino effect of errors and refusals that would suggest that respondents refusing or making errors at certain point were more likely to further refuse or mistake the rest of the questionnaire. The LCA contrasted respondents with errors or refusals against those with clean interviews. However, the errors and refusals classes were not associated to any of the investigated socio-demographic characteristics.

Similar conclusion can be drawn on the complexity hypothesis. Refusals could not be completely at random, but associated to the complexity of the trajectories and may be linked to the respondent's tiredness. Further analysis and different data are however needed to fully validate this hypothesis

7 Coherence with the Grid Data

Although in this study was not possible to assess the validity of the SPLC with external precise data nor to include a real test-retest design in the survey, as explained in former sections the SHP pilot study was structured in different parts, some of which asked a few questions in common with the SPLC. In particular, both the grid questionnaire and the SPLC tapped age, civil status, level of education, and employment. The difference between the two tools is that the SPLC asked the whole personal history, while the grid accounted only fro the current situation. In addition, with the exception of the F2F contact mode in which the interviewer administered, if possible, all parts of the questionnaire in a single day, the grid was administered at the moment of contact and the SPLC several days later. In the CATI+F2F mode the SPLC interview was concluded on average 26 days later (Median = 15.5). Thus the grid and the calendar can be used to analyse the coherence between answers as in test-retest designs. Being the tools different, it is not a real test of reliability of SPLC, nevertheless it can provide information on potential problems of the SPLC in collecting basic demographic data.

In the SHP pilot study, the grid questionnaire was administered to one single reference person in the household, while the life calendar could have been filled by more than one

person per household. Thus, to use the grid information as in test-retest procedures, a subsample of respondents that filled both the grid and the life calendar was selected (n=289)

7.1 Age

According to the data gathered by the calendar questionnaire respondents were born between 1918 and 1998. The mean date of birth was 1960. However in the grid questionnaire people resulted to be born between 1918 and 2010. The mean date of birth was 1961. 7.96% of the responses between the grid and the calendar questionnaire were not congruent. 23 cases were concerned by this incongruence (see Table 21). Interestingly the main problem of incoherence was found in the F2F mode, and in particular the when contact was done on F2F. However, some of the mismatches are attributable to a wrong transcription of the dates, probably in the grid questionnaire which was either CATI or CAPI interviewer-administered.

Calendar	Grid	Difference	Mode
1966	1996	-30	F2F+F2F
1973	1985	-12	F2F+F2F
1967	1977	-10	F2F+F2F
1962	1972	-10	F2F+F2F
1935	1943	-8	F2F+F2F
1973	1980	-7	F2F+F2F
1943	1948	-5	F2F+F2F
1966	1970	-4	F2F+F2F
1956	1960	-4	F2F+F2F
1949	1952	-3	F2F+F2F
1966	1968	-2	CATI+F2F
1979	1981	-2	F2F+F2F
1984	1985	-1	F2F+F2F
1986	1987	-1	F2F+F2F
1980	1981	-1	F2F+F2F
1963	1962	1	CATI+Mail
1948	1947	1	F2F+F2F
1967	1965	2	CATI+F2F
1939	1936	3	CATI+Mail
1964	1961	3	CATI+F2F
1991	1986	5	F2F+F2F
1966	1954	12	F2F+F2F
1998	1967	31	CATI+Mail

Table 21: Cases of mismatched birth year

7.2 Civil Status

33.22% of the responses about civil status were incoherent between the grid and the calendar data. Some of these differences could be attributed to the fact that requested data were sligthly different between the grid and the calendar. In the grid questionnaire respondents were asked for their actual civil status whereas the SPLC asked for status changes across the

life history. Indeed, the 76.04% of this difference was explained by the number of missing in the calendar data set: nearly all the missing in the calendar were registered as unmarried in the grid. Initial respondent's status in the calendar was indeed supposed to be unmarried, even if it was not openly declared. Table 23 shows a certain coherence (61.64%) between missing in the calendar and unmarried respondents in the grid, although 36.99% did not report any civil status change in the calendar even if we know from the grid that they were married.

The relatively high number of incongruence does not mean that the calendar is less reliable than the grid, but that the information collected by the two tools did not match. In certain cases the calendar seemed to be more correct than the grid. For instance, one respondent declared to be unmarried in the grid, but in the calendar he/she resulted as divorced. Thus, the civil status declared in the grid might no be completely reliable because yet based on respondent self-perception and self-presentation. Similar considerations also apply to incongruence between separation and divorce. Separation is not always considered as a civil status by respondent and might be confounded with divorce. From the respondents' point of view, separation might indeed be considered as the act of separating from someone, while from administrative point of view in Switzerland it denotes a specific civil status which precede divorce. However the effectives were so small that the interpretation could be misleading. Statistically the overall coherence between the two tools seemed stable (Cramer's $V = 0.82$)

Surprisingly, the incongruence between the grid and the calendar was larger in the F2F+F2F mode, as shown in Table 22. That rises some questions about the interviewers' effect. It is indeed possible that the F2F+F2F interview setting was more likely to be stressed by the urgency to terminate all for sections of interview in a single appointment. This might have led respondents and interviewers to generate a higher number of errors.

	CATI+Mail	CATI+F2F	F2F+F2F
Congruent	61	66	66
Incongruent	19	23	54

Table 22: Total Number of Incongrences in Civil Status by Mode

7.3 Education

Concerning education, the information gathered by the grid questionnaire and the information gathered by the calendar questionnaire were not identical. The level of education at the moment of the interview was the information which differed the most between the calendar and the grid questionnaires. These differences can be attributed to the complexity of the Swiss educational system, which varies from canton to canton. Thus some school names that apply in one canton may indicate a slightly different thing in another. If this is controllable in a prospective interview, by adjusting the school list per each canton, it become more problematic in the life calendar because it is not clear whether the respondent indicated the school name used in the canton where he/she attended it or the canton were he/she lived at the moment of the interview. Consequently when we compared the grid and the calendar, different names were coded even if they might actually indicate the same information.

It is worth of note that, despite these problematic aspects, most of the time the information for the first episode of education was coherent. However, 11.68% (n =59) of the

Calendar	Grid						Total
	unmarried	married	registered partnersh	separated	divorced	widowed	
missing	45	27	0	0	0	1	73
	61.6%	37.0%	0.0%	0.0%	0.0%	1.4%	
	83.3%	15.7%	0.0%	0.0%	0.0%	4.5%	
married	2	144	0	1	1	2	150
	1.3%	96.0%	0.0%	0.7%	0.7%	1.3%	
	3.7%	83.7%	0.0%	12.5%	3.1%	9.1%	
registered partnersh.	0	0	1	0	0	0	1
	0.0%	0.0%	100.0%	0.0%	0.0%	0.0%	
	0.0%	0.0%	100.0%	0.0%	0.0%	0.0%	
separated	2	0	0	4	2	1	9
	22.2%	0.0%	0.0%	44.4%	22.2%	11.1%	
	3.7%	0.0%	0.0%	50.0%	6.2%	4.5%	
divorced	1	1	0	1	27	0	30
	3.3%	3.3%	0.0%	3.3%	90.0%	0.0%	
	1.9%	0.6%	0.0%	12.5%	84.4%	0.0%	
widowed	0	0	0	0	0	18	18
	0.0%	0.0%	0.0%	0.0%	0.0%	100.0%	
	0.0%	0.0%	0.0%	0.0%	0.0%	81.8%	
other	4	0	0	2	2	0	8
	50.0%	0.0%	0.0%	25.0%	25.0%	0.0%	
	7.4%	0.0%	0.0%	25.0%	6.2%	0.0%	
Total	54	172	1	8	32	22	289

Table 23: Civil Status Congruence Between The SPLC and the Grid Questionnaire

participants at the calendar declared an incoherent start for their education trajectory, but yet the biggest part of these errors was attributable to data coding mistakes. In particular, "unachieved compulsory school" and "achieved only compulsory school" were confounded. If these errors were parcelled out the number of errors is reduced to n =5, that is 0.99% of the total answers.

8 Between-Samples Error Check: the Vulnerable Population

The task of filling the calendar is quite compelling and may be particularly difficult for the most vulnerable strata of the society. To test the potential difficulties that vulnerable social classes could find in this task, a subsample ($n = 92$) of respondents with either a low level of education (i.e., primary education or post-obligatory training) or low incomes (i.e., household incomes less than 35000 CHF per year) were compared to the rest of the sample on the same indicators of coherence than the previous section. If the task of filling the SPLC was too cognitively demanding, we would expect to find a higher rate of errors among vulnerable respondents than in the rest of the population.

Compared to the general sample, the vulnerable sub-sample is composed by a higher percentage of women, and a higher percentage of older respondents. The two groups did not differ in the distribution of the modes of interview (Table 24).

Variable		General Population	%	Vulnerable Sample	%	χ^2	p
Sex	Men	191	0.53	49	0.37		
	Women	169	0.47	82	0.63	9.48	0.02
Cohort	>1988	14	0.04	37	0.28		
	1979-88	51	0.14	8	0.06		
	1969-78	52	0.14	8	0.06		
	1959-68	96	0.27	15	0.11		
	1949-58	79	0.22	17	0.13		
	<1949	68	0.19	46	0.35	90.64	0.00
Mode	CATI+Mail	120	0.33	45	0.34		
	CATI+F2F	100	0.28	38	0.29		
	F2F+F2F	140	0.39	48	0.37	0.21	1.00

Table 24: Crosstabulation between Samples and Gender, Cohort, and Mode

8.1 Age

Concerning the date of birth, 10.87% of the vulnerable sample had incongruent birth dates between the grid and the calendar, against the 6.6% of the general sample. However the statistical tests were not significant: Cramer's V =0.07; χ^2 =1.56, p =0.21, suggesting an equal distribution of errors between the two samples.

8.2 Civil Status

Concerning errors in the civil status, the vulnerable respondents were not more likely of reporting wrong first episodes of civil status in the calendar (namely, a first episode of separation, marriage or widowhood) than the non-vulnerable respondents (see Table 25). Similarly, in respect to the difference between the grid and the calendar 73.91% of the vulnerable sub-sample gave a coherent answer to both questionnaires, while for the general sample the rate of coherence was 63.78%. Against our expectations, the vulnerable sample was more precise than the general one. The χ^2-test on the cross-tabulation between number of incongruences and type of sample tended to significance (χ^2 =2.91, p =0.09), although Cramer's V was very small (V =0.1). Thus, we could assume none or marginal difference in the number of incongruences between the two samples, with the vulnerable sub-sample performing slightly better.

An explanation of this result could be that low status respondents might have given simplified answers to both the grid and the calendar, minimizing the complexity of their life trajectories, or that this result depended on a selection bias at earlier stages of the study. However, not enough evidences are available to support these hypotheses. What we are left with is that the vulnerable sub-sample tended to give more correct and coherent answers than the general sample.

8.3 Education and Employment

Only one of the vulnerable respondents had a wrong entry to the first education event. More interestingly, 10.69% of the subsample did not answer to this column of the calendar, while non response in the general sample was 8.38%. It might be that this column was more problematic for low social class respondents. At this stage we could not compare neither of the differences between the grid and the calendar in respect of the highest education level

First Status Change	General Population	Vulnerable Sample	
missing	103	45	148
	28.6%	34.4%	
married	241	83	324
	66.9%	63.4%	
registered part.	2	0	2
	0.6%	0.0%	
separated	2	0	2
	0.6%	0.0%	
divorced	3	0	3
	0.8%	0.0%	
widowed	1	0	1
	0.3%	0.0%	
other	8	3	11
	2.2%	2.3%	
Total	360	131	491

Table 25: Differences in the First Civil Status Episode

and actual employment status because they need recoding to comparable categories (see comments on section 7.3)

Overall the difference between the two sub-samples in the number of incongruence was marginal, suggesting that for certain aspects the SPLC was more coherent with the grid for the low status respondents than for the general population. Although these results may not be robust, given the very low number of errors, they are particularly interesting and counter-intuitive. Results suggest that the SPLC can be used for surveying vulnerable populations, and that, once understood the logic of the life calendar, the tool is not too cognitively demanding for low social status respondents.

9 Analysis of Subjective Evaluation of Life-Periods

This section explores the subjective measures included in the SHP pilot attached to the SPLC. At the end of the calendar, all respondents were asked to answer to the Subjective Well-Being Scale (SWBS, Diener, Emmons, Larsen, & Griffin, 1985), a five-items battery of questions scoring from 1 to 7. Figure 10 displays the distribution of the SWBS which is tendentially skewed towards the right (high level of well being). This is not surprising, given that this kind of distribution is quite common with well-being scales (e.g., Cummins, 2003; Springer & Hauser, 2006). Item non response for this scale was reasonably low (7.27%, $n = 36$).

Afterwards, half of the sample was selected to give an evaluation on the three different domains of the SPLC: partnership, family, and activity. Respondents were asked to divide those three life domains into periods of different length and to express an evaluation of each period on a 5-point differential scale from very negative to very positive.

This method should allow to capture the counterbalance between evaluations on one life-sphere (e.g., family) and what happened or was perceived in the others (e.g., health, job, etc..). Moreover, by asking to divide the whole trajectory into periods that might not be

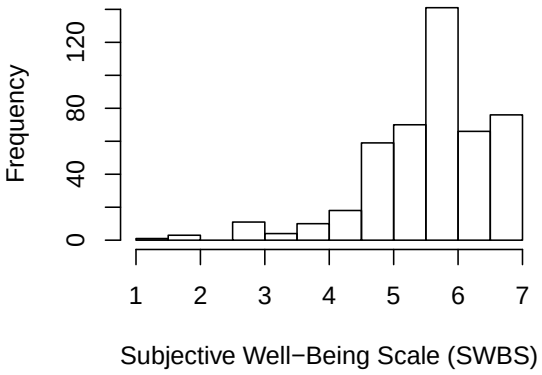

Figure 10: Distribution of the Subjective Well-Being Scale

linked to the specific events asked in the calendar (e.g., moves, jobs, births), this method may be particularly suitable for tapping extra information that would otherwise be lost. For instance, as seen in the section 4 young respondents may have calendars with very little factual data. They are likely to live with their parents and not having experienced moves, not having civil-status changes, not having children, and not yet being inserted into the labour market. The division of life domains into periods allows detecting if and when things got better or worse, even in absence of specific events.

Thus, we expected that the number of events and the number of periods did not fully match. In other words we expected that respondents did not provide evaluations of single events, but that they told a different story of their life by dividing their trajectories into periods. We also expected that the evaluation of events would have similar measurement characteristic of the SWBS. In particular, we expected the evaluation of life domains to have distributions similar to the SWBS.

9.1 Intimate Relationship Domain

Concerning the partner relationship, the SPLC asked to mention all the relationship significant to the respondent, independently from their duration or type. The 9.7% of the sample did not mention any romantic partner. Looking at the interview protocol for the F2F modes, only in 4 out of 316 interviews (1.27%) respondents openly refused to give an answer. The remaining missing values were empty entries, not refusals.

The number of mentioned partners ranged from zero to eight. The majority of the respondents (n =304, 60.2%) mentioned one partner, 89 persons (17.62%) mentioned two relationships. 34 and 17 persons respectively indicated 3 or 4 partners and 12 respondents (2.38%) had 5 or more partners. Overall there were no incoherent cases, in which a partner was indicated for a very unlikely age ($<$ 12 years old).

Respondents divided their partnership trajectory into different periods, ranging from 1 to 20. The most common division of this life domain was in two periods (Figure 11) (M =3.4, Med =3, SD =2.34). Interviewers registered no refusals to this task and the total number

of period was 935.

As expected, there was not a strong association between the number of partner and the number of period (Cramer's V=0.37). Indeed only 30 respondents reported the same number of periods than partners.

Considering the evaluation itself its worth of note that the mean evaluation (i.e., the mean across the whole life) follows a distribution similar to the SWBS (Fig. 11, right). However the number of refusal to the evaluation of the first period was very high: 132 respondents out of 275 (48%) refused to give an evaluation. Nevertheless, the rate of refusals dropped for the evaluation of the second period: 9 respondents out of 226, 3.98%. The different refusals rate between the first and second period suggest that refusals were not likely be caused by the sensitivity of the question, otherwise we would have had refusals equally distributed for all marked periods. The high rate of refusal to the first period indicates that some respondents probably thought it was senseless to evaluate the period previous to their first partner, even if they were supposed to. In principle the interviewer should have corrected this misunderstanding, but instructions may not have been sufficiently clear. Respondents who really refused to answer did not give evaluations for any of the marked periods (n =22, 95.64%).

9.2 Family Domain

Similarly to the partnership, also in the family domain no refusal was registered to the task of dividing the family trajectory into periods. The number of periods ranged from 1 to 20. The most common division of this life domain was in one single periods (25.09%), followed by the division in three (19.64%) and two (16%, see Figure 12). Mean number of periods was M =3.38, Median =3, and standard deviation SD =2.37. The total number of period was 930.

The average evaluation of the whole trajectory follows a SWBS like distribution, skewed to the right (Fig. 12, right). Also for the family, the number of refusal to the evaluation task was quite high. However, similarly to what seen for the partnership domain, 105 respondents out of 275 (38.18%) refused to give an evaluation to the first period, but not necessarily the subsequent ones. The second period was not evaluated indeed by only 8 respondents out of 206 (3.88%). Most of (but not all) the participants who had refused to give the evaluation on the first period of the partnership also refused on the first period of the family domain. Respondents that refused to provide an evaluation to all reported periods were 24 (4.75%).

9.3 Activity Domain

The same task was replicated on the education and occupation domain. The number of periods ranged from 1 to 19. The division of this section is on average more articulated than the previous ones (Figure 13). Mean number of periods was M =4.28, Median =4, and standard deviation SD =2.79. The total number of period was 1177.

The distribution of the average evaluation is more normal than the other two domains. Namely, the extreme left of the scale is not inflated as for the others (Fig. 13, right). However the evaluation has a slightly smaller variance ($\sigma^2_{ACTIVITY}$ = 18.49, $\sigma^2_{PARTNER}$ = 48.35, and σ^2_{FAMILY} = 52.54). The minimum point of the scale was given by respondents was -1 instead of -2. This could indicate an higher normativity of this life doamin to provude positive evaluations (Cummins, 2003).

Still, 40.73% of respondents refused to provide an evaluation of the first period (112 respondents out of 275). Similarly to what seen in the previous section, the majority the refusals (24.1%) are respondents that refused also the evaluations of the first periods in the other columns. These refusals concerned only the first period, refusals to evalute the every marked period happened in only 1 case.

To go further in the exploration of the refusal to the evaluation task, the "real" refusals to at least one of those tasks was contrasted against the willingness of participating to a second wave of the study. The subjective evaluation was indeed the last task in the pilot interview. Before concluding the interview, respondents were asked whether they would feel like participating to another wave. If the evaluation task stimulated a negative reaction in the respondents, we could expect an higher rate of unwillingness to be re-contacted among those respondents. Table 26 shows that the relationship between willingness to participate and having refused at least one of the evaluation task is virtually null. Although the number of real refusal was very small, there was virtually no difference in the willingness of participating again between who completed the task and who did not (Cramer's V=0.04).

Overall the analyses of the answers to the evaluation task suggest that this tool may indeed capture a different side of the life trajectories, producing different and complementary data. The high rates of refusal to the evaluation of the first period only suggest that the question per se was not too sensitive, otherwise we would have found refusals for all periods and not only the fist one. This result rather suggest that interviewer's instructions might not have been sufficiently clear and this might have been a source of error.

Willing to Participate Again	Evaluation Task		Total
	Completed Evaluation	Refused Evaluation	
Willing	282	9	291
	61%	56%	
Maybe	80	4	84
	17%	25%	
Unwilling	103	3	106
	22%	19%	
Total	465	16	481

Table 26: Refuse to Evaluate and Willingnes to Participate Again

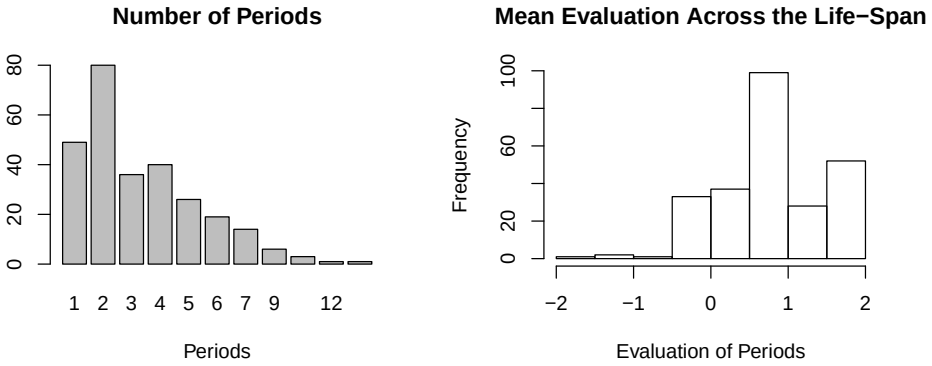

Figure 11: Periods of the Couple Relationship

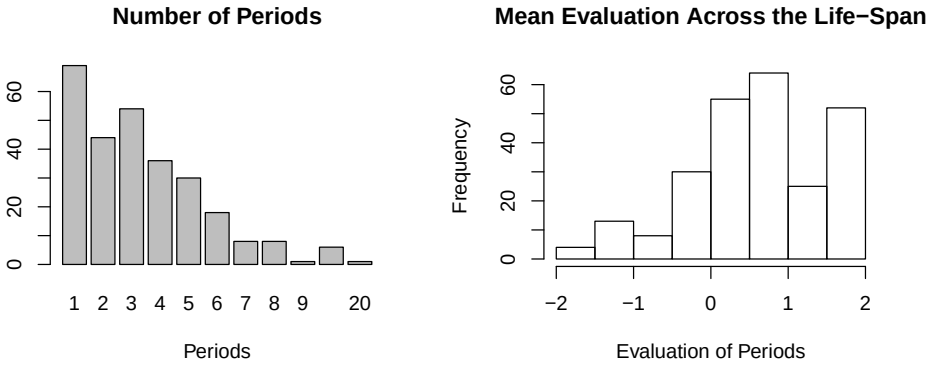

Figure 12: Periods of the Family Trajectory

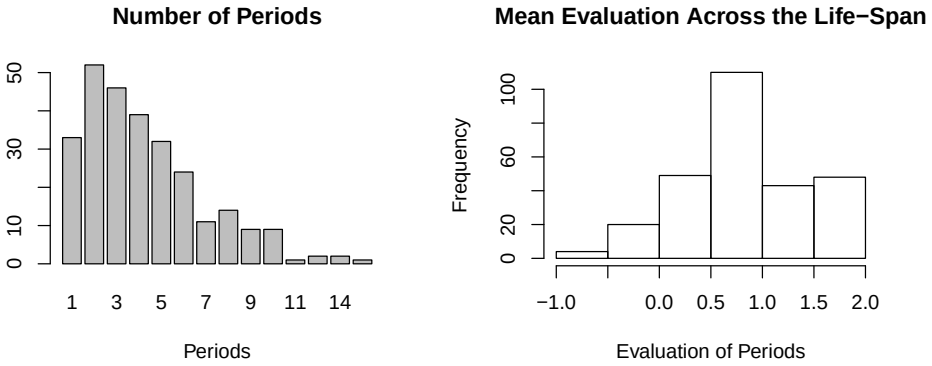

Figure 13: Periods of the Activity Trajectory

10 Conclusions

The analyses highlighted a certain robustness of the SPLC as well as some weak points. This study was meant to be a pilot test of the SPLC and it had some limitations in terms of sample size, which reduced the statistical power of some analyses. In some analyses we could only compare a small sample against a much larger one. However given that variance tends to be larger in small samples, we would be more likely to find between samples difference where it does not really exist. None or little difference was instead found in the several comparisons performed in this study. In addition, the results from different analyses were coherent, drawing an interesting picture of the SPLC performance.

The indicators of completeness showed a weak effect of the administration mode that is quite encouraging. Overall, respondents provided similar responses in the three contact and administration modes (CATI+Mail, CATI+F2F, F2F+F2F). This might be attributable to the fact that the tool was filled directly by the respondents also in presence of an interviewer. When the full interview (grid, calendar and subjective) was administered at the same time (F2F+F2F mode) the level of completeness tended to be slightly lower than when part of the interview was done by telephone and part in F2F.

In support of the self-administration mode, the comparison with the *Family tiMes* survey showed that the SPLC can collect the same amount of events of an interactive respondent-interviewer interview. The distributions in the indicators of completeness between the two survey were indeed similar, even if the type of administrations of the two calendars were substantially different. Although these conclusions can be drawn only in relation to the particular events accounted by our indicators of completeness, this result is particularly interesting and intriguing.

Contrary to the expectations, the life calendar method may be suitable for self-administration, performing equally well than in F2F modes. That may depend on a number of different reasons that go beyond the aims of this study. For instance, it may be that the concept of life history calendar is becoming more popular thanks to its implantation on recreational platforms (e.g., Facebook, Timelines). That may help respondents in better understanding the task and making calendar-based questionnaires cognitively less demanding. Further research should investigate whether it would be possible to obtain the same results also for other types of life events.

The feasibility of self-administration in LHCs was supported also by the other analyses of this study. The check of the calendars made by the interviewers showed a low level of errors made by respondents while filling the calendar. Similarly the number of incoherent information between the SPLC and the grid questionnaire did not highlight problems of the calendar over the conventional methods. On the contrary, it suggested that the SPLC performed reasonably well and that most of the incoherence might be attributable to coding problems rather than wrong answers.

Thus, contrary to expectations the F2F mode did not produced better quality data. On the contrary, the comparison with the data collected by the grid showed that the F2F+F2F mode can lead to higher rates of incongruence. That might depend on how interviewers managed the interview, or by the respondent's level of attention and tiredness for doing the interview at once. This result suggest that respondents probably need low pressures in answering the LHC, thus the paper-and-pencil self administration could enhance respondents' concentration. This task might be helped by the attention given to the graphical layout and clearness of the instructions in the SPLC.

The fact that not substantial difference were found in the level of coherence errors between the vulnerable sub-sample and the general sample support this point. LHCs may be indeed suitable for the surveying vulnerability (at least low educated and low incomes populations), but they need to be carefully designed and tested on the target population. The SPLC seemed to fulfil some of the needed characteristics, although a more in-depth study is needed.

Concerning the subjective evaluation the non correspondence between the single events and the general periods suggested that, in line with the expectations, the two tasks produced different and complementary information. In other words, asking to divide the three domains (partnership, family and occupation) into periods returned a richer information, especially when few events were marked (e.g., a single partner for the whole life).

Acknowledgement

This study is the result of research work developed within the framework of the Swiss National Centre of Competence in Research LIVES, which is financed by the Swiss National Science Foundation. The authors are grateful to the Swiss National Science Foundation for its financial support.

References

Auriat, N. (1993). My Wife Knows Best: A comparison of event dating accuracy between the wife, the husband, the couple, and the Belgium population register. *Public Opinion Quarterly*, *57*(2), 165–190.

Auriat, N. (1996). *Les défaillances de la mémoire humaine*. Paris: Presses Universitaires de France, INED.

Axinn, W. G., & Barber, J. S. (2001). Mass education and fertility transition. *American Sociological Review*, 481–505.

Axinn, W. G., Pearce, L. D., & Ghimire, D. (1999). Innovations in life history calendar applications. *Social Science Research*, *28*(3), 243–264.

Beck, U. (1992). *Risk society: Towards a new modernity*. London: Sage Publications.

Becker, S., & Sosa, D. (1992). An experiment using a month-by-month calendar in a family planning survey in Costa Rica. *Studies in Family Planning*, *23*, 386–391.

Belli, R. F. (1998). The structure of autobiographical memory and the Event History Calendar: Potential improvements in the quality of retrospective reports in surveys. *Memory*, *6*(4), 383–406.

Belli, R. F., Bilgen, I., & Al Baghal, T. (2013). Memory, communication, and data quality in calendar interviews. *Public Opinion Quarterly*, *77*(S1), 194–219.

Belli, R. F., Shay, W., & Stafford, F. (2001). Event history calendars and question-list surveys: A direct comparison of interviewing methods. *Public Opinion Quarterly*, *65*, 45–74.

Berney, L. R., & Blane, D. B. (1997). Collecting retrospective data: accuracy of recall after 50 years judged against historical records. *Social science & medicine*, *45*(10), 1519–1525.

Bidard, C. (2010). Bifurcations biographiques et ingrédients de l'action. In B. C. M., Bidard, & M. Grossetti (Eds.), *Bifurcations : Les sciences sociales face aux ruptures et à l'événement* (pp. 224–238). Paris: La Découverte.

Bilgen, I., & Belli, R. F. (2010). Comparison of verbal behaviors between calendar and standardized conventional questionnaires. *Journal of Official Statistics*, *26*(3), 481.

Blane, D. B. (1996). Collecting retrospective data: development of a reliable method and a pilot study of its use. *Social Science & Medicine*, *42*(5), 751–757.

Brändle, K. (2012, November). Well-being over the life course: Measuring subjective evaluation with life calendars.. (Poster presented at NCCR LIVES - 2nd Site Visit of the Review Panel, Geneva)

Brändle, K., Morselli, D., & Spini, D. (2013, June). Interviewer effects in life calendar data. flexible interviews vs standardized procedure: The assessment of subjective data.. (Poster presented at the SARMAC Conference, Rotterdam)

Caspi, A., Moffitt, T. E., Thornton, A., Freedman, D., Amell, J. W., Harrington, H., et al. (1996). The lhc: A research and clinical assessment method for collecting retrospective event-history data. *Journal of Methods in Psychiatric Research*, *6*, 101–114.

Champion, C. (2011). Switzerland: A latecomer catching up? In J. Clasen & D. Clegg (Eds.), *Regulating the risk of unemployment: National adaptations to post-industrial labour markets in europe* (p. 121). Oxford: Oxford University Press.

Clausen, J. A. (1995). *American lives.* Berkeley: University of California Press.

Clogg, C. C. (1995). Latent class models. In G. Arminger, C. C. Clogg, & M. E. Sobel (Eds.), *Handbook of statistical modeling for the social and behavioral sciences* (pp. 311–359). London: Springer.

Collins, L. M., & Lanza, S. T. (2010). *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences* (Vol. 718). Wiley. com.

Conway, M. A. (1992). A structural model of autobiographical memory. In M. A. Conway, D. Rubin, H. Spinnler, & W. A. Wagenaar (Eds.), *Theoretical perspectives on autobiographical memory.* Dordrecht: Kluwer Academic Publishers.

Conway, M. A. (1996). Autobiographical knowledge and autobiographical memories. In D. Rubin (Ed.), *Remembering our past: Studies in autobiographical memory.* Cambridge: Cambridge University Press.

Cummins, R. A. (2003). Normative life satisfaction: Measurement issues and a homeostatic model. *Social indicators research*, *64*(2), 225–256.

Dannefer, D. (2003). Cumulative advantage/disadvantage and the life course: Cross-fertilizing age and social science theory. *Journal of Gerontology*, *58*, 327–337.

Dannefer, D. (2009). Stability, homogeneity, agency: Cumulative dis/advantage and problems of theory. *Swiss Journal of Sociology*, *35*, 193–210.

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of personality assessment*, *49*(1), 71–75.

Engel, L., Keifer, M., & Zahm, S. (2001). Comparison of a traditional questionnaire with an icon/calendar-based questionnaire to assess occupational history.

American Journal of Industrial Medicine, *40*, 502–511.

Freedman, D., Thornton, A., Camburn, D., Alwin, D., & Young-DeMarcco, L. (1988). The LHC: A technique for collecting retrospective data. *Sociological Methodology*, *18*, 37–68.

Glasner, T. J. (2011). *Reconstructing event histories in standardized survey research: Cognitive mechanisms and aided recall techniques (Unpublished doctoral dissertation) Vrije Universiteit*. Amsterdam.

Goldman, N., Moreno, L., & Westoff, C. F. (1989). Collection of survey data on contraception: an evaluation of an experiment in Peru. *Studies in Family Planning*, *20*, 147–157.

Jackman, S. (2012). pscl: Classes and methods for R developed in the political science computational laboratory, stanford university [Computer software manual]. Stanford, California. Retrieved from http://pscl.stanford.edu/ (R package version 1.04.4)

Jeanneret, B., & Goebel, V. (2012). *Les disparités régionales en suisse*. Neuchatel: Swiss Federal Statistical Office.

Lelièvre, E., Roubaud, F., Tichit, C., & Vivier, G. (2006). Données factuelles et perceptions, la place du flou dans l'observation et l'analyse. In A. Philippe, L. Eve, & G. D (Eds.), *Etats flous et trajectoires complexes : observation, modélisation, interprétation* (pp. 29–59). Paris: INED/CEPED.

Levy, R., Ghisletta, P., Le Goff, J.-M., Spini, D., & Widmer, E. (2005). Incitations for interdisciplinarity in Life Course Research. In R. Levy, P. Ghisletta, J.-M. Le Goff, D. Spini, & E. Widmer (Eds.), *Towards an interdisciplinary perspective on the life course* (pp. 361–391). Amsterdam-Boston: Elsevier.

Linzer, D. A., & Lewis, J. B. (2011). poLCA: An R package for polytomous variable latent class analysis. *Journal of Statistical Software*, *42*(10), 1–29. Retrieved from http://www.jstatsoft.org/v42/i10/

Magidson, J., & Vermunt, J. (2002). Latent class models for clustering: A comparison with k-means. *Canadian Journal of Marketing Research*, *20*(1), 36–43.

Means, B., Nigam, A., Zarrow, M., Loftus, E. F., & Donaldson, M. S. (1989). Autobiographical memory for health-related events: Enhanced memory for recurring incidents. In *Vital health stat* (Vol. 6).

Morselli, D., Dasoki, N., Gabriel, R., Gauthier, J.-A., Henke, J., & Le Goff, J.-M. (in press). Using life-history calendars to survey vulnerability. In D. Joye, M. Oris, M. Ernst-Staehli, & C. Roberts (Eds.), *Surveying vulnerability*. Berlin and New York: Springer.

Morselli, D., & Passini, S. (2012). Rights, democracy and values: A comparison between the representations of obedience and disobedience to the authority on italian and finnish students. *International Journal of Intercultural Relations*, *36*.

Mortimer, J. T., & Shanahan, M. J. (2003). *Handbook of the life course*. New York: Springer.

Oesch, D., & Lipps, O. (2012). *Does unemployment hurt less if there is more of it around? A panel analysis of life satisfaction in Germany and Switzerland*. European Sociological Review.

Perren, S., Keller, R., Passardi, M., & Scholz, U. (2010). Well-being curves across

transitions: The development of a retrospective measure. *Swiss Journal of Psychology/Schweizerische Zeitschrift für Psychologie/Revue Suisse de Psychologie*, *69*(1), 15.

Ranci, C. (2010). Social vulnerability in Europe. In *C* (pp. 3–24). Basingstoke: Palgrave Macmillan.

Ranci, C., & Magliavacca, M. (2010). Social vulnerability: A multidimensional analysis. In *C* (pp. 219–249). Basingstoke: Palgrave Macmillan.

Reimer, M. (2001). *Die Zuverlässigkeit des autobiographischen Gedächtnisses und die Validität retrospektiv erhobener Lebensverlaufsdaten.* Berlin: Max-Planck-Insitut für Bildungsforschung.

Rosenberg, M., Layde, P., Ory, H., Strauss, L., Rooks, J., & Rubin, G. (1983). Agreement between women's histories of oral contraceptive use and physician records. *International journal of epidemiology*, *12*(1), 84.

Schafer, J. L. (1999). Multiple imputation: a primer. *Statistical methods in medical research*, *8*(1), 3–15.

Schutz, A. (1962). *Collected papers.* The Hague: Martinus Nijhoff.

Scott, J., & Alwyn, D. F. (1998). Retrospective vs prospective measurement of life histories in longitudinal research. In J. Z. Giele & G. H. Elder (Eds.), *Methods in life course research. qualitative and quantitative approaches* (pp. 98–127). Thousand Oaks: Sage.

Searles, J. S., Helzer, J. E., & Walter, D. E. (2000). Comparison of drinking patterns measured by daily reports and timeline follow back. *Psychology of Addictive Behaviors*, *14*(3), 277.

Spini, D., Hanappi, D., Bernardi, L., Oris, M., & Bickel, J.-F. (2013). Vulnerability across the life course: A theoretical framework and research directions. *LIVES Working Papers*, *2013/27*. doi: 10.12682/lives.2296-1658.2013.27

Springer, K. W., & Hauser, R. M. (2006). An assessment of the construct validity of ryff's scales of psychological well-being: Method, mode, and measurement effects. *Social science research*, *35*(4), 1080–1102.

Tourangeau, R. (2000). Remembering what happened: Memory errors and survey reports. In A. Stone, J. Turkkan, C. Bachrach, J. Jobe, H. Kurtzman, & V. Cain (Eds.), *The Science of Self-Report: Implications for research and practice* (pp. 29–48). Englewood Cliffs, N. J.: Lawrence Erlbaum.

Van der Vaart, W. (2004). The timeline as a device to enhance recall in standardized research interviews: A split ballot study. *Journal of Official Statistics*, *20*(2), 301–317.

Vivier, G. (2006). Comment collecter des biographies? de la fiche ageven aux grilles biographiques, principes de collecte et innovations recentes. *Population et travail–Dynamique et travail*, 119–131.

Widmer, E. D. (2010). *Family configurations: A structural approach to family diversity.* (L.

Widmer, E. D., & La Farga, L.-A. (2000). Family networks: a sociometric method to study relationships in families. *Field Methods*, *12*(2), 108–128.

Willis, G. B. (2005). *Cognitive interviewing: A tool for improving questionnaire design.* Thousand Oaks: Sage.

Yoshihama, M., Gillespie, B., Hammock, A. C., Belli, R. F., & Tolman, R. M.

(2005). Does the life history calendar method facilitate the recall of intimate partner violence? comparison of two methods of data collection. *Social Work Research*, *29*(3), 151–163.

Zeileis, A., Kleiber, C., & Jackman, S. (2008). Regression models for count data in R. *Journal of Statistical Software*, *27*(8). Retrieved from http://www.jstatsoft.org/v27/i08/