Attrition in the Swiss Household Panel: Are Vulnerable Groups more Affected than Others?

Martina Rothenbühler and Marieke Voorpostel

1 Introduction

One of the major difficulties of longitudinal surveys is panel attrition, the loss of panel members from one wave to the next. Attrition can lead to selective samples and make the interpretation of estimates problematic. The central concern in the analysis of attrition is therefore selection bias, that is, a distortion of the estimation results due to non-random patterns of attrition. Attrition may be especially problematic for vulnerable populations that are more difficult to survey, as they are more likely to drop out of panel surveys.

A common distinction is made between attrition that is completely at random, attrition that is selective on variables observed in the data and attrition that is selective on variables unobserved in the data. If the attrition is random, then the erosion of the sample does not lead to biased estimates. However, in most cases the attrition is non-random (Alderman et al. 2001). If attrition introduces a bias in the estimates of interest, selective attrition on observable variables is more amenable to statistical solutions than attrition on unobserved variables. Weighting strategies can help to reduce or even completely repair the effects of attrition on estimations. Using information from prior waves to model appropriate weights can contribute to reducing the amount of unexplained variation in the data due to attrition, but selective attrition on unobservable variables remains a problem.

M. Rothenbühler (✉)
Clinical Trials Unit, University of Bern, Bern, Switzerland

Swiss Centre of Expertise in the Social Sciences FORS, Lausanne, Switzerland
e-mail: martina.rothenbuehler@ctu.unibe.ch

M. Voorpostel
Swiss Centre of Expertise in the Social Sciences FORS, Lausanne, Switzerland

We investigate the problem of attrition using the Swiss Household Panel (SHP). The SHP is a longitudinal survey with annual repetition and its main objective is the analysis of socio-economic change within households, in particular the dynamics of living conditions of the population in Switzerland.

In this contribution, we present two kinds of attrition analysis: first we show in Sect. 4 how variables can be affected by attrition by comparing means and frequencies of the answers given in the first wave by all respondents (all longitudinal sample members) with the longitudinal sample members still participating at a later wave. We also assess to what extent the use of weights correct for any bias. Second, we show which sociodemographic characteristics are most related to attrition patterns. This second part focuses on the relation between indicators of latent vulnerability and panel attrition and is addressed in Sect. 5. Using a discrete-time competing risk model, we analyse the impact of these variables on the dropout probability. Before turning to the analytical part, we describe in Sect. 2 the relation between being at risk of vulnerability and several demographic characteristics by using different theoretical approaches and point out how in turn these characteristics are related to attrition by presenting results from other studies. In Sect. 3 we give a short overview of the data used for our analyses. The last part of this contribution, Sect. 6 concludes the chapter by discussing the potential and limits of both analysing and countering selective attrition.

2   Attrition in Relation to Vulnerability

Nonresponse and panel attrition may be especially a problem for vulnerable populations. In this study we define people as vulnerable or at risk of being vulnerable if they possess traits that position them in low levels within the socioeconomic stratification. Several characteristics related to participation in surveys are associated with vulnerability, such as having a lower level of education, foreign nationality, being unemployed, in poor health, or being divorced (Loosveldt and Carton 2001; Kleiner et al. 2012; Stoop 2005; Watson and Wooden 2009). These characteristics indicate latent vulnerability; they tend to go together with a deficit in resources, creating a risky environment.

There are several reasons why vulnerable groups are in general more likely to drop out of surveys.

First, some vulnerable groups may be more difficult to locate or to contact, especially if they are more likely to move. For example, negative life events such as divorce usually involve the move of at least one of the partners, increasing the risk of losing track of the respondent. Unemployed and inactive people, older people, and people with children are more likely to be found at home, and these groups tend to include more women (Stoop 2005; Watson and Wooden 2009).

After locating and contacting sample members, they still have to be willing to cooperate. In making their decision sample members take the costs and benefits associated with participating into account (Dillman et al. 2002). Vulnerable groups

may expect higher costs and lower benefits for several reasons. First, expected benefits may be lower or the costs higher if past experiences in a previous wave have been unpleasant, thereby decreasing the likelihood of participation (Loosveldt and Carton 2001). Respondents from vulnerable groups, such as those who experience a combination of factors such as unemployment, poverty and health problems, may experience talking about their negative situations as unpleasant, making the interview situation uncomfortable.

Another reason is that certain vulnerable groups may lack the skills needed to successfully complete the survey and make it a pleasant experience. Loosveldt and Carton (2001) provide evidence that participation in the second wave of a panel study is related to the respondent's ability to perform the task; an ability that is found to be lower for less educated respondents (Loosveldt 1997). Also language proficiency is an important skill necessary to successfully participate in an interview, which poses problems when interviewing certain minority groups (Kleiner et al. 2012).

Finally, benefits of participation are expected to be lower for groups who are less socially integrated, which is related to vulnerability as well. Studies have shown social integration and isolation to be correlated with the likelihood of responding to surveys (Stoop 2005; Watson and Wooden 2009). For the SHP, this correlation has been established in earlier studies as well (Lipps 2007; Voorpostel 2010). People who score high on social integration are more likely to participate in surveys for several reasons. One reason is that they tend to be guided by the norms of the dominant culture, in which participation in a survey may be seen as a "civic duty" (Dillman et al. 2002; Johnson et al. 2002). Second, less social integration is related to more cynicism about established institutions, an attitude which expected to influence response rates to surveys as well Stoop (2005). Third, individuals who score higher on social integration are more likely to find the topics that are usually covered in surveys relevant or important. One potentially useful indicator of social involvement is political interest. Research has shown that politically interested people are more cooperative irrespective of the survey topic (Stoop 2005). Possibly a wide variety of topics covered in most surveys are of interest to members' level of interest, however, is also judged on experiences made in previous waves. Generally speaking, more interest in the topic will lead to a higher probability of responding Groves et al. (2000, 2004).

Many of the usual background characteristics in nonresponse analyses can be linked to these processes behind non contact and noncooperation. For instance, employed people and people with a higher socio-economic status are often harder to contact, as they are less often found at home, but they are more willing to cooperate than people from a lower socio-economic stratum or unemployed people, as employed people and people with a higher socio-economic status might have better skills, experience lower opportunity costs and be more interested in the survey. Also, holding a paid job can be perceived as a way of participating in society, whereas unemployed people face more social isolation (Gallie and Paugam 2004). Older people, on the other hand, are easier to contact, because they are more likely to be at home, but they tend to be more reluctant to cooperate. Some groups are harder

to contact and more reluctant to cooperate, such as men, singles, ethnic minorities, younger persons, and big city dwellers (Stoop 2005).

In Sect. 5 we examine the importance of several factors related to attrition and latent vulnerability, most importantly age, education, income, working status, civil status, nationality, health and political interest. In addition we also include gender, the presence of children in the household and whether the respondent is a home owner or a tenant. It should be noted that our indicators of latent vulnerability are not absolute proof of vulnerability, but only provide an indication of being potentially at risk for vulnerability.

3 Data

The analyses in this chapter are based on the Swiss Household Panel (SHP). Currently, the SHP consists of three samples. The first sample SHP_I started in 1999, when 5074 households were interviewed for the first time. In 2004 a refresher sample was added with 2538 additional households (SHP_II). The third sample (SHP_III) started in 2013. The interviews are done both at the household and the individual level using the computer assisted telephone interviewing (CATI) method. Every household member aged at least 14 is eligible to answer to the individual questionnaire.

In 2012, that is the 14th wave of SHP_I and the 9th wave of SHP_II, the combined panel SHP_I and SHP_II contained 4390 households. The SHP_I counted 2923 households who answered to the household questionnaire. This corresponds to 58 % of the number of household interviews conducted in the first wave in 1999. The second panel, SHP_II, had 1467 valid household questionnaires in 2012. This is also equivalent to 58 % of the number of household interviews conducted in 2004. In terms of personal interviews, there were 7446 individuals who answered the questionnaire, 5032 of them belonging to the SHP_I and 2414 to the SHP_II.

At the individual level we differentiate between the overall number of individual interviews and the fully longitudinal ones. The former include original sample members (OSM), that is individuals that were already present in the first wave, as well as their cohabitants, individuals who joined the household after the first wave and were thus not part of the sample as it was drawn. The fully longitudinal individual interviews refer to OSM who answered at each consecutive wave.

As can be seen in Table 1 the overall number of individual interviews in the SHP_I diminished until 2005. A number of changes[1] in the rules of follow-up in

---

[1]Since 2006, participants receive an unconditional incentive. If all household members participate, the household receives another, this time conditional, incentive. The incentives contribute to motivate the people to maintain cooperation throughout the duration of the panel. Together with the unconditional incentive, participants also receive some information about the survey in general and some results from the previous wave.

Table 1 Number of valid individual and household interviews in SHP_I and SHP_II for the years 1999–2012

Year	SHP_I			SHP_II		
	Individual interviews	Household interviews	Fully longitudinal individual interviews	Individual interviews	Household interviews	Fully longitudinal individual interviews
1999	7799	5074	7799			
2000	7073	4425	6335			
2001	6601	4139	5429			
2002	5700	3582	4480			
2003	5220	3227	3888			
2004	4413	2837	3076	3654	2538	3654
2005	3888	2457	2622	2649	1799	2395
2006	4091	2537	2399	2568	1684	1930
2007	4630	2817	2209	2350	1494	1601
2008	4494	2718	2060	2410	1546	1400
2009	4800	2930	1952	2309	1476	1289
2010	5057	2985	1876	2489	1557	1220
2011	5103	2977	1811	2481	1520	1155
2012	5032	2923	1739	2414	1467	1102

2006 and 2008 had a positive impact on the participation rate. Since 2006, the number of individual interviews has increased more or less steadily until 2011 and decreased slightly in 2012. On the other hand, the number of fully individual longitudinal interviews decreases steadily after the first wave. In 2012, 22 % of the OSM of the SHP_I had answered the individual questionnaire every year since 1999, compared with 30 % of the SHP_II.

The first part of the analysis, the comparison of means and frequencies (see Sect. 4) as well as the multivariate analysis of the characteristics of individuals in Sect. 5 is performed using only the data of the SHP_I. Furthermore, we only include the observations of OSM. Despite having an influence on the sample size, this restriction makes sure that we do not confound the effects of attrition with evolution, that is time and period effects, by allocating the same starting point to all the subjects included in the analysis.

4 Differences in Means and Frequencies due to Attrition

To analyse the effect of attrition we examine whether means and frequencies of a series of variables change if the sample composition changes due to dropout of panel members. One can assume that if attrition would be at random, means and distributions of variables would not change following attrition.

For each wave, we can define a new sample that is composed of individuals who participated both at the first wave and at the wave in question. Because of dropout this sample is smaller than the original one in the first wave. We then compare the answers given in the first wave using each of these samples on one side and the control sample, which is the sample of the longitudinal sample members of the first wave, on the other side. This means that we analyse the data of the first wave, but use different samples: the sample composed by all longitudinal sample members of the first wave on one hand and a subsample of longitudinal respondents in any later wave on the other hand. Differences in the statistics between these two samples indicate that dropout from the SHP is not random.

In order to identify the variables affected by attrition, we examine all variables that were included in the latest wave and in any of the previous waves. We consider attrition from the SHP_I.[2] We compare the means and frequencies calculated with the value of the first year of the variable ££ in $(99, \ldots, 12)$ on the sub-populations of longitudinal respondents still present in the latest wave as follows:

$$sL^{R££} = sL^{££}$$

$$sL^{R\$\$} = sL^{££} \cap sL^{\$\$}$$

$$\ldots$$

$$sL^{R12} = sL^{££} \cap sL^{12} \tag{1}$$

where $sL^{££}$ represent the longitudinal respondents (OSM) in 1999 and $sL^{\$\$}$ are the longitudinal respondents in year 20$\$\$. Basically, we test to see if samples that still respond in a later year are representative of the same individuals that responded in the first year. The tests run through the SHP data of the wave 14 (year 2012).

The effect of attrition is analysed both on weighted and unweighted means and frequencies. The idea behind this approach is that the weight should correct for attrition e.g. that there should be no difference in means and frequencies when using different subsamples. The weights of the SHP represent a mixture of design weights and adjustment for nonresponse. The latter consists of several sociodemographic variables such as sex, age, civil status and nationality. As initial nonresponse to the first wave is also considered in the weighting scheme, the choice of the variables used for the adjustment to nonresponse is limited to information available in the sample frame.

As shown in Table 2, there are in total 1108 variables that appear in at least one wave of the SHP_I. Out of these, there are 306 variables that cannot be tested, either because they are proxy variables,[3] variables with the same response in all waves,

---

[2]The analysis is done for the first panel SHP_I that started in 1999. The analysis for the combined panel SHP_I and SHP_II was done in the same way and produced comparable results (not presented).

[3]Proxy-variables involve reports on other household members.

Table 2 Effect of attrition of means and frequencies for the SHP_I

Difference without weight	Difference with weight	Occurrences out of the 1108 variables (percent out of the 802 variables tested in parentheses)
–	–	306
No	No	644 (80.3 %)
No	Yes	9 (1.1 %)
Yes	No	85 (10.6 %)
Yes	Yes	64 (8 %)

variables with too few observations,[4] variables of which the modality is too high[5] or because it does not make sense to compare the variable.[6]

In the 14th wave of the SHP_I there are 644 variables out of the 802 tested variables that are not affected by attrition. This means that in 80.3 % of the variables there is no difference in means or frequencies when the sample of the first wave is compared to any subsequent longitudinal sample. The variables considered do not appear to be biased after attrition. The mean or the frequencies of 85 of the tested variables are different, but the difference disappears once the weights are used. These variables are therefore touched by attrition but the weighting corrects the phenomenon. In 64 cases we observe a difference without the weight and the difference persists even with weighting. The variable is therefore affected by attrition without the possibility of correction by weighting. With nine variables the estimates are biased only if the weights are applied. In these cases the weights clearly fail to correct for selective attrition.

The variables having been identified as being biased by attrition (in particular variables related to leisure and politics) need to be studied with care by the researchers who use them in their analyses. These results do not mean that these variables are unusable. However, they show that the phenomenon of attrition can certainly not be ignored. Most of the variables that are affected by attrition concern information on leisure activities, political and social participation, professional integration and health, which is in line with previous studies on attrition. One variable affected by attrition is, for example, political interest. Because politically uninterested individuals tend to drop out, the means of the political interest in the original sample SHP_I in 1999 and in a subsample consisting of individuals still present in 2004 or onwards differ even when using the weights that should correct for attrition. Other variables that have been affected by attrition are satisfaction with health, associational membership and working status.

---

[4]For categorical variables, at least one category needs to have 30 observations. Numeric variables are not tested if the total number of observations is less than 30.

[5]This is for categorical variables with more than 100 distinct responses, such as the isco job classification.

[6]This is the case for identification variables, dates or weights.

5 Participation Patterns and Sociodemographic Characteristics of Nonrespondents in the SHP

We now turn to the different causes for dropout and examine the characteristics of the individuals who participate in the SHP and those who drop out. We first give a short overview of the methods used to do this analysis (Sect. 5.1). We then describe the dropout rates according to the different reasons for nonresponse (Sect. 5.2). In Sect. 5.3 we describe the characteristics of the individuals for each participation pattern. The fifth part of this section is dedicated to the joint analysis of various characteristics in relation to dropping out of the SHP. We include the following characteristics: gender, age, education, working status, civil status, children in the household, Swiss nationality and legal status, income,[7] home ownership, political interest and satisfaction with health status. We analyse the effect of these variables on several causes of nonresponse.

5.1 Methodological Note

Participation in the SHP can be considered as exposure to the risk of dropping out. This risk arises each year when the interviewers call to realise an interview. Furthermore, there are different kinds of risks: people can drop out because they refuse to participate, because of missing contact information or because illness or frailty prevents them from answering to the questions. We differentiate seven causes for dropping out:

1. Not eligible
2. Left the household
3. Problems related to health and/or age
4. Family-related problems
5. Refusal
6. Non contact
7. Other reason

   We grouped the individuals who are not eligible anymore because of death, emigration or institutionalisation into one group. The second group considers individuals who left the household either temporarily or permanently. There is no other information about the reason for nonresponse available for these panel members. A third group comprises the individuals who cannot participate due to health or age problems, whereas the fourth group includes individuals who state family-related problems such as the death of a family member or taking care of other household members as reasons for not participating. The fifth group contains

---

[7]To measure the effect of income we use the natural logarithm of the personal gross income.

the individuals who refuse to participate without any specific reason, either at the individual or the household level. Individuals who could not be contacted anymore due to missing contact information are included in the sixth group. The last group considers various causes for nonresponse with too few observations to be analysed separately, such as language problems or technical difficulties.

We provide a description of the variables of interest both separately for each participation pattern and by reporting the overall distribution. Continuous variables are presented as means and standard deviations (sd) if they are distributed normally, otherwise as medians and interquartile ranges (IQR). The corresponding statistical association with the participation pattern is evaluated using analysis of variance or the Kruskal-Wallis-test, respectively. Categorical variables are described by counts and percentages and are compared using the χ^2 test, Fisher's exact test or multinomial logistic regression where appropriate. All hypotheses are two sided and a p-value less than 0.05 is deemed statistically significant. We performed all the analyses using the statistical packages STATA version 13.1 and R version 3.1.0.

In our case, participation is equal to survival whereas the different reasons for nonresponse are treated as competing failures. In order to investigate the dropout patterns according to these different causes of failure, we use Kaplan-Meier estimates. The Kaplan-Meier procedure enables us to examine the distribution of the length of participation by estimating conditional probabilities for each cause of failure at each point in time (at each wave in our case) and by using these probabilities to estimate the corresponding survival rates (Kaplan and Meier 1958). Applying a survival model enables us to take into account that the probability of dropping out in any subsequent wave depends on the probability of having participated in all the previous waves. The probability of survival until the end of a specific interval (wave) corresponds to the product of the probabilities of not dropping out at one of the previous waves: the discrete time hazard function is the probability of an event occurring during interval t, conditional on the fact that the event did not occur before t (Mills 2011). We can therefore calculate a probability of not responding for each wave, given that there was no nonresponse up to this given wave.

In presence of competing events, the different dropout rates can also be understood as a cumulative incidence function. Cumulative incidence corresponds to the expected proportion of panel members experiencing a specific event in the presence of other risks (Beyersmann and Schumacher 2008; Gooley et al. 1999). Let us suppose that we observe an event of interest such as family-related problems which lead to nonresponse. This event has competing risks, events whose occurrence preclude or alter its own probability to happen.

As we interview the individuals once a year, there is only once in a year the possibility to participate or to nonrespond. We therefore have a limited number of occurrences and the duration data is intrinsically discrete. Because of the nature of this underlying transition process, we use a discrete time model and control for the intra-individual clustering of the data. As we are not only interested in general nonresponse, but also in the reasons leading to dropout, we will apply a competing risk discrete time model. Competing risk models make it possible to distinguish

between different kinds of events and are thus applied in situations with more than one cause of failure (Allison 1982; Prentice et al. 1978; Putter et al. 2007). This is especially useful if there is reason to believe that the effect of the explanatory variables differs among the various types of competing failures.

A major aspect in favour of applying survival analysis to model attrition is that discrete time models make it possible to include time-varying covariates in the analysis. Applying a standard method could introduce bias and lead to a loss of information (Allison 1982). If we would use fixed values or a standard method, it would be difficult to decide which value to consider. Should we consider the last available value? This would mean that we would compare values of different years. Should we rather consider the value of the variable when the spell started, that is at the first wave of participation? This would mean that we do not take into account the evolution, like changing education, age or marital status, and thus ignore much information. By applying a survival model, we can overcome this problem, as we use for each transition the information of the previous wave, that is the most current information available independent of the participation status.

It should be noted that in all our analyses we only consider participation until the first dropout and ignore therefore the influence of the changes in the follow-up rules mentioned in Sect. 3. Those respondents who drop out temporarily and come back at a later wave are disregarded in this analysis after the first time they fail to respond. Because of these restrictions the analysis tends to overestimate attrition, as it cannot distinguish between permanent attrition or temporary nonresponse.

5.2 Dropout Patterns According to Causes for Nonresponse

If we consider the overall survival curve, we can see in Fig. 1 that the participation rate diminishes over the analysis period. The survival curve portrays that the dropout or first nonresponse rate is highest after the first wave of participation. The survival curve becomes flatter as the duration of the panel increases. This is illustrated by the fact that the vertical distance at each time point, that is the change in cumulative probability, becomes lower as the analysis time increases.

If we turn to the different reasons for dropping out, we can observe in Fig. 2 that the survival rates differ slightly according to the reasons for nonresponse. A vertical gap between two curves indicates that one group has a greater proportion of panel members surviving. The lower curves illustrate that the dropout rate among those participants is higher.

We can see in Fig. 2 that the dropout rate due to ineligibility was highest for this group at the start of the analysis period, but became the lowest compared to the other failure causes as time passes. The survival rate of the participants who state that the reason for their nonresponse are health problems is, compared to the other causes, mostly higher. We can also observe that non contact becomes a less important reason for dropping out: while its curve was steeper during the first five transitions, it became flatter afterwards. This change might be explained by analysis
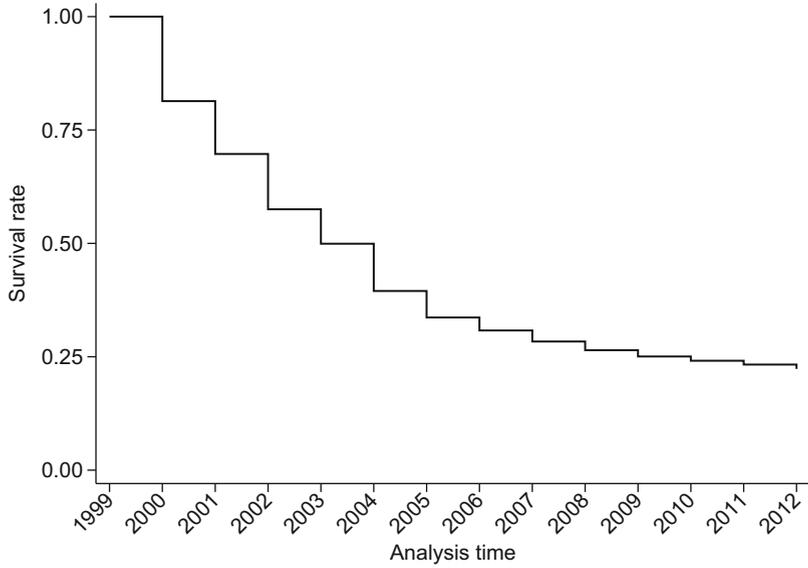
Fig. 1 Overall Kaplan-Meier survival estimates
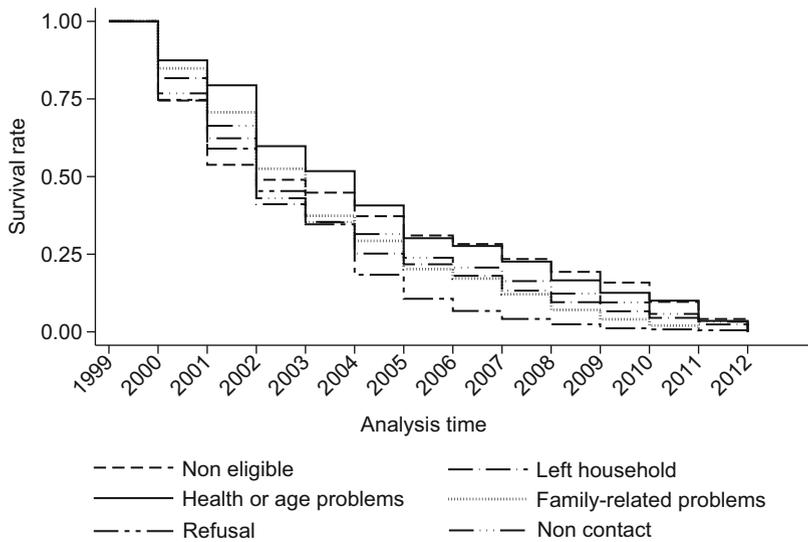

Fig. 2 Kaplan-Meier survival estimates according to failure cause

time: hard to reach individuals tend to drop out early. Once the panel members have participated during a few consecutive waves, their nonresponse is not because of a missing contact but rather because of other reasons like health problems, refusal or family-related difficulties. At the same time we can also suppose that it became easier to trace individuals due to the internet. Another explanation for the lower dropout rates due to missing contact information is that the interviewers put more effort in finding panel members after the change of the follow-up rules in 2006 and 2008 mentioned in Sect. 3. Figure 2 also suggests that refusal conversion and the use of unconditional incentives, two other measures enhanced in 2006 to counter attrition, had a positive impact on the participation rate, as the failure rate due to this cause becomes less important after 2006.

The cumulative incidence plot (Fig. 3) illustrates the dropout rates according to the different events. As in Fig. 2, we can see that refusal at individual or household level was the most cited reason for nonresponse, followed by non contact. It is also visible that both curves become flatter as time of analysis increases, which is partly due to a general lower dropout rate, but also due to competing risks that become more important such as health or age problems.
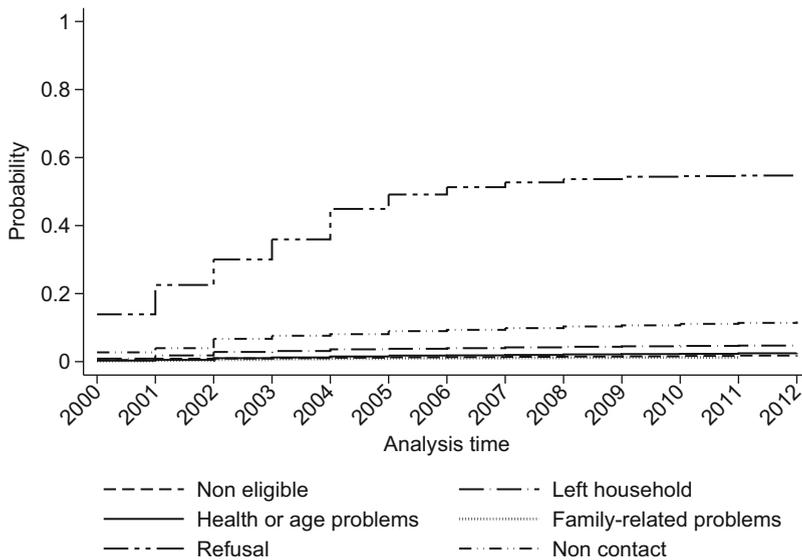

Fig. 3 Cumulative incidence of the different reasons for attrition

5.3 Description of the Characteristics of Nonrespondents Within the SHP

Before we turn to the competing risk model, we first provide a description of the variables of interest in relation to the different participation patterns. In total, we use the data of 7788 individuals. More than half of the panel members (55 %) refused at least once to participate during the observed period. The mean duration of participation among the nonrespondents varies from 2.7 ± 1.9 years for the individuals who dropped out because of different reasons than the ones specified here to 5.4 ± 3.5 years for the panel members who stopped participating because of health- or age-related problems.

Table 3 contains several variables related to vulnerability: level of education, income, working status, legal status and satisfaction with health. In addition to these variables we also included some general demographic variables (gender, age, civil status, children in the household, home ownership) and political interest as it is known to be a determinant of attrition and nonresponse (Groves et al. 2000, 2004). Moreover, political interest can be seen as a measure of social involvement, which in turn can be related to vulnerability. The association between the participation pattern and the variables in Table 3 is statistically significant in all cases, which is a first indication that causes of nonresponse differ according to both demographic characteristics and those related to vulnerability.

The overall participation rate is higher among women than men. However, nearly 70 % of the participants who state having family-related difficulties that prevent them from participating are women. The proportion of women is also higher among the participants having health problems or who do not participate anymore due to their age. The median age of the respondents is 57 (IQR $= 48 - 68$), whereas it is 25 (IQR $= 21 - 39$) among the individuals who left the household temporarily or permanently. This distribution is not surprising, as young adults tend to have a high geographic mobility due to educational reasons or in order to establish their own household. We can also see that the median age is highest among the group with health- or age-related problems (75, IQR $= 65 - 82$), followed by the ineligible sample members (62, IQR $= 41 - 77$). This last group includes deceased persons.

The groups also differ according to the education of their members. The lowest proportion of individuals having completed the obligatory school without any further education can be found among the individuals who have never dropped out until 2012 (11.7 %). At the same time the participants that are still involved have the highest proportion (37 %) of individuals with a tertiary education. The lowest proportion (16 %) of group members with a tertiary education can be found among the individuals who dropped out because of health and age problems. This group also has the largest share of individuals with a low level of education (35 %).

The median income, as measured by the personal gross income, is highest among the individuals who never dropped out (64,550; IQR $= 32,600 - 100,800$) and lowest among the individuals having health- or age-related problems that prevent them from participating (30,000; IQR $= 19,900 - 59,200$). The median income

Table 3 Description of the characteristics according to each participation pattern

Variable	Participation	Ineligible	Left household	Health- or age-related problems	Family-related problems	Refusal	Non contact	Other
Pattern: n (%)	1739 (22.3)	145 (1.9)	377 (4.8)	199 (2.6)	99 (1.3)	4283 (55.0)	918 (11.8)	28 (0.4)
Duration (years): mean (sd)	14.0 (0.0)	4.9 (4.0)	4.2 (3.2)	5.4 (3.5)	4.4 (2.9)	3.6 (2.4)	4.4 (3.5)	2.7 (1.9)
Men: n (%)	694 (39.9)	80 (55.2)	193 (51.2)	77 (38.7)	30 (30.3)	1887 (44.1)	433 (47.2)	15 (53.6)
Age (years): median (IQR)	57 (48–68)	62 (41–77)	25 (21–39)	75 (65–82)	46 (35–57)	45 (33–57)	38 (30–50)	26 (20–53)
Education: n (%)								
Primary level	204 (11.7)	36 (24.8)	83 (22.0)	69 (34.7)	16 (16.2)	1033 (24.1)	162 (17.7)	7 (25.00)
Secondary level	884 (50.8)	70 (48.3)	210 (55.7)	98 (49.3)	64 (64.7)	2331 (54.4)	507 (55.2)	16 (57.1)
Tertiary level	651 (37.4)	39 (26.9)	84 (22.3)	32 (16.1)	19 (19.2)	919 (21.5)	249 (27.1)	5 (17.9)
Income (thousands): median (IQR)	65 (33–101)	42 (20–70)	39 (14–63)	30 (20–59)	35 (18–64)	48 (19–79)	56 (31–80)	47 (24–75)
Home ownership: n (%)	1078 (62.0)	80 (55.2)	196 (52.3)	94 (47.2)	47 (47.5)	2234 (52.3)	251 (27.4)	10 (35.7)
Working status: n (%)								
Active occupied	1089 (62.6)	62 (42.8)	286 (75.9)	33 (16.6)	68 (68.7)	2892 (67.5)	698 (76.0)	23 (82.1)
Unemployed	12 (0.7)	2 (1.4)	10 (2.7)	2 (1.0)	2 (2.0)	67 (1.6)	29 (3.2)	0 (0.0)
Not in labour force	638 (36.7)	81 (55.9)	81 (21.5)	164 (82.4)	29 (29.3)	1324 (30.9)	191 (20.8)	5 (17.9)
Legal status/nationality: n (%)								
Swiss	1481 (85.2)	119 (82.1)	292 (78.3)	178 (89.5)	86 (86.9)	3387 (79.8)	661 (73.6)	19 (70.4)
Naturalised	186 (10.7)	8 (5.5)	41 (11.0)	14 (7.0)	8 (8.1)	363 (8.6)	100 (11.1)	4 (14.8)
Permit C	72 (4.1)	16 (11.0)	32 (8.6)	6 (3.0)	3 (3.0)	413 (9.7)	107 (11.9)	4 (14.8)
Permit B or other	0 (0.0)	2 (1.4)	8 (2.1)	1 (0.5)	2 (2.0)	79 (1.9)	30 (3.3)	0 (0.0)
Civil status: n (%)								
Single	206 (11.9)	32 (22.1)	253 (67.1)	21 (10.6)	16 (16.2)	1105 (25.8)	364 (39.7)	17 (60.7)
Married	1178 (67.7)	83 (57.2)	103 (27.3)	96 (48.2)	66 (66.7)	2629 (61.4)	348 (37.9)	9 (32.1)
Not married anymore	355 (20.4)	30 (20.7)	21 (5.6)	82 (41.2)	17 (17.2)	549 (12.8)	206 (22.4)	2 (7.1)
With children	689 (39.6)	35 (24.1)	268 (71.3)	24 (12.1)	53 (54.1)	2390 (55.8)	398 (43.5)	17 (60.7)
Political interest: mean (sd)	6.3 (2.5)	5.4 (2.9)	5.3 (2.8)	5.5 (2.8)	4.7 (2.9)	5.1 (2.9)	5.1 (2.9)	4.6 (3.6)
Satisfaction with health: mean (sd)	7.7 (1.6)	6.4 (2.7)	8.1 (1.6)	7.0 (2.2)	8.1 (1.6)	8.0 (1.8)	7.8 (2.0)	8.4 (2.0)

of the individuals who could not be contacted anymore was also relatively high (56,090; IQR = 31,380 − 80,340). Nearly three quarters of the individuals who could not be contacted anymore (73 %) lived in a rented home, whereas over 60 % of the individuals who never stopped participating live in a house or in a flat they own. This is an indicator that home ownership has a positive impact on the possibility to contact individuals, as they seem to be less mobile geographically, and thus, on the participation rate.

Most of the individuals who refused to participate without indicating any specific reason (68 %), who refused to participate because of family-related problems (69 %) or who could not be contacted anymore (76 %) were active in the labour market. Four out of five individuals who refused to answer because of health- or age-related problems were not in the labour force (anymore) whereas only one fifth of the individuals who could not be contacted anymore or who left the household temporarily or permanently were either unemployed or actively occupied.

The largest share (89 %) of Swiss without a second nationality can be found among the individuals having health- or age-related problems preventing them from participating at the survey, where as the lowest proportion is among the panel members who could not be contacted anymore. This group also contains the biggest proportion of individuals who became naturalised (11 %) and those who have a residence permit C (12 %).[8]

The civil status is another characteristic that differs among the participation patterns. Sixty-seven percent of the individuals who left the household that was selected for the SHP_I are single, whereas only 27 % among them are married. The lowest proportion of singles can be found among the individuals who do not participate anymore due to old age- or health-related problems (11 %). They also have the biggest proportion of individuals who are not married anymore (41 %), that is who are either divorced, separated or widowed. The group containing the individuals who never stopped participating has the largest share of married people (68 %) and nearly 40 % of them have at least one child living in the same household. Most of the individuals who left the household temporarily or permanently (71 %) lived in a household with at least one child before they moved out. It is very probable that the individuals of this group of nonrespondents are one of these children. Over half of the individuals who do not participate anymore because of family-related difficulties or who refused to participate without any specific reason live together with one or more child.

The findings suggest that the effect of the number of children in a household is not as straightforward as one might think. On one side, having children might be associated with being at home and therefore being accessible to answer to the questionnaire. On the other side, having children can prevent participation because of a tighter schedule or worries related to the children.

---

[8]Residence permit C is a long-term permit, whereas residence permit B is a short-term permit. This second category also includes other arrangements (for example asylum seekers or people with no permit).

On a scale between 0 and 10, the sample members who still participate have the highest political interest (6.3 ± 2.5) whereas those having family-related problems have the lowest (4.7 ± 2.9). Also on a scale going from 0 to 10, satisfaction with health was highest among individuals who had left the household (8.1 ± 1.6), while it was lowest for panel members who were no longer eligible (6.4 ± 2.7) and individuals who stated age- or health-related problems (7 ± 2.2).

5.4   Analysis of the Characteristics of Nonrespondents Within the SHP

As introduced in Sect. 2 and demonstrated in the previous parts, attrition is in general not random in longitudinal surveys. People with certain characteristics tend to drop out more often. Beside describing the participation rate of individuals according to unique variables, we can analyse the attrition using different characteristics at the same time. We do this using the same variables as in Table 3. Applying the discrete time competing risk model described in Sect. 5.1, we can see in Table 4[9] that panel attrition diminished with duration even when controlling for the other factors. Compared to individuals who drop out after the first wave, the probability of dropping out becomes lower with each transition (from one wave to the next).

If we turn to the individual characteristics, we can see that in general men drop out more often than women. The relative risk of dropping out because of ineligibility relative to non-stop participation increases by 2.50 (95 % CI: 1.45–4.31) for men compared to women. The relative risk of men having health-related problems preventing participation (1.61; 95 % CI: 1.04–2.50) or refusing to participate (1.13; 95 % CI: 1.03–1.24) is also higher compared to women.

The effect of age in all groups is not linear. We can see that age has the shape of an inverted "U", that is that the participation rate is higher as the age of an individual increases, but diminishes after a specific age. Correspondingly, getting older first has a negative impact on the probability of being among the nonrespondents, but later, when the maximum is reached, has a positive impact on the probability of dropping out. For the group of individuals who had left the household, this change in effect happened shortly after turning 55.[10] Until the age of 55, each additional year lowers

---

[9]Only results that are significant at least at the 5 %-level and where the variable has strictly positive counts are displayed in Table 4.

[10]This example illustrates the inter-operation of the influence of age for the group who left the household:

$$y = \beta_0 + \beta_1 * x + \beta_2 * x^2$$

$$\frac{\partial y}{\partial x} = \beta_1 + \beta_2 * 2 * x = 0$$

$$x = -\beta_1/(2 * \beta_2) = 0.2669914/(2 * 0.0023951) = 55.287874$$

Table 4 Competing risk discrete time model for various causes of dropout from the SHP

Variable	Non eligible	Left household	Health- or age- related problems	Family- related problems	Refusal	Non contact	Other
Duration of participation	0.92*		0.91**	0.90**	0.83**	0.97*	0.70**
	(0.03)		(0.02)	(0.03)	(0.01)	(0.01)	(0.07)
Men (ref: women)	2.50*		1.61***		1.13***		
	(0.69)		(0.36)		(0.06)		
Age (centralised)	0.89*	0.77**			0.97*	0.92**	0.81***
	(0.04)	(0.02)			(0.01)	(0.02)	(0.07)
Age squared (centralised)	1.00**	1.00**	1.00**		1.00*	1.00***	1.00***
	(0.00)	(0.00)	(0.00)		(0.00)	(0.00)	(0.00)
Education (ref: primary school)							
Secondary level		1.53***					
		(0.32)					
Tertiary level			0.51***		0.76**		
			(0.17)		(0.06)		
Income (ln)		1.30*		0.72*		1.16***	
		(0.11)		(0.09)		(0.07)	
Home owner-ship (ref: tenant)					1.11***	0.53**	
					(0.05)	(0.06)	
Working status (ref: active occupied)							
Unemployed					1.54***		
					(0.29)		
Not in labour force			1.82***				
			(0.55)				
Legal status/nationality (ref: Swiss (only))							
Swiss and other					1.24*	1.36***	
					(0.10)	(0.20)	
Permit C					1.31**	1.54*	
					(0.10)	(0.23)	
Permit B or other					1.75*	2.36*	
					(0.32)	(0.68)	
Civil status (ref: single)							
Married		0.43**		3.08***	0.85***	0.63**	
		(0.09)		(1.40)	(0.06)	(0.08)	
Not married anymore		0.21**			0.78*	1.55*	
		(0.08)			(0.07)	(0.22)	
Children (ref: none)		1.87*			1.23**		
		(0.38)			(0.07)		
Political interest			0.91*	0.90***	0.95**	0.96***	
			(0.03)	(0.05)	(0.01)	(0.02)	
Satisfaction with health	0.66**		0.84**				0.92**
	(0.04)		(0.04)				(0.02)

$n = 7027$; * $p < 0.01$; ** $p < 0.001$; *** $p < 0.05$

the relative risk of dropping out because of leaving the household temporarily or permanently. Afterwards, getting older has a positive effect on the relative risk of dropping out because of having left the household. For the individuals who drop out because of ineligibility, the maximum of the negative age effect is reached at 38. After 38, each additional year increases the relative risk to becoming ineligible compared to the non-stop participants. For the group who refused to participate, the maximum is reached at 44, for the individuals who dropped out because of other reasons at 50 and for those who could not be contacted anymore at age 78.

The participation rate is also influenced by education. Depending on the cause of nonresponse, the influence is positive or negative. On one side, a higher educational level has a positive impact on the probability of leaving the household (1.53; 95 % CI: 1.02–2.30). On the other side, compared to individuals whose highest education is primary school, tertiary education is associated with a lower relative risk to drop out due to age- or health-related problems (0.51; 95 % CI: 0.27–0.99) or refusal (0.76; 95 % CI: 0.66–0.89).

If we consider the effect of a higher income, we can also observe different patterns according to the failure cause. The relative risk for individuals who dropped out because of family-related problems relative to the constant participants decreases by 0.70 (95 % CI: 0.57–0.91) for each unit change of income.[11] In this case, a higher income has a negative effect on the probability of dropping out. On the other hand, a higher income increases the probability of leaving the household temporarily or permanently (1.30; 95 % CI: 1.11–1.53) or of having missing contact information (1.16; 95 % CI: 1.03–1.32).

One aspect that is linked to income is home ownership. Besides acting as an indicator for income and wealth, home ownership also influences geographical mobility, in the sense that these individuals move less and can thus be contacted more easily. This is visible in our data: the relative risk for individuals who cannot be contacted anymore relative to the constant participants is lowered by home ownership (0.53; 95 % CI: 0.43–0.65). However, compared to tenants, individuals who own a house or a flat tend to refuse more often without specifying the reason (1.11; 95 % CI: 1.01–1.22).

Working status is another aspect that influences participation in the SHP positively, in the sense that unemployment affects the probability of refusing to participate without specific reason or for not responding for another reason. Compared to individuals who are employed, individuals who are unemployed have a lower participation rate (1.50; 95 % CI: 1.07–2.24). Compared to the individuals who are active on the labour market, not being in the labour force increases the relative risk of being in the group having health- or age-related problems relative to the participants (1.82; 95 % CI: 1.01–3.31).

---

[11]Because income is not distributed normally, we use the natural logarithm of it in this analysis.

The legal status or nationality is another characteristic of nonrespondents. Individuals who received the Swiss nationality or have for some other reason two nationalities tend to drop out more often due to refusal (1.24; 95 % CI: 1.06–1.44) or non contact (1.35; 95 % CI: 1.01–1.82) than those having only the Swiss nationality or who have it since birth. In general, a long- or short-term permit leads to higher dropout rates. Furthermore, the relative risk to drop out is higher for those having a short-term permit than for individuals with a long-term permit.

Another characteristic related to dropout is marital status. Compared to individuals that have never been married, the dropout rate of married individuals or of those that are no longer married (widowed, divorced or separated) is in general lower, except for the individuals who drop out because of family-related problems. Here, being married increases the relative risk for this group relative to the respondents by the factor 3.10 (95 % CI: 1.27–7.48) compared to the singles. Being no longer married also increases the relative risk of being among the individuals who cannot be contacted anymore (1.55; 95 % CI: 1.18–2.05). Living in a household with children influences the dropout rate positively when the reason for not responding is moving out of the household (1.87; 95 % CI: 1.25–2.79) or refusing without any specific reason (1.23; 95 % CI: 1.10–1.38).

Political interest, which is associated with social integration, is another aspect that influences loyalty to the survey. On a scale going from zero (not interested at all) to ten (very interested), each additional point leads to a lower dropout rate, except when nonresponse is due to ineligibility or because of moving out, where it has no statistically significant influence. Politically interested individuals seem to be more motivated to participate in the survey. The same applies to health satisfaction. On a scale going from zero (very unsatisfied) to ten (very satisfied), each additional point diminishes the probability of dropping out of the panel due to ineligibility (0.66; 95 % CI: 0.59–0.74), age- or health-related problems (0.84; 95 % CI: 0.77–0.91) and missing contact information (0.92; 95 % CI: 0.87–0.96). Individuals who are satisfied with their health status tend to participate more often in the survey.

6   Conclusion

We have shown that attrition in the Swiss Household Panel is not completely at random, but that individuals with certain characteristics tend to drop out more often, as is usually the case in panel surveys (Groves 2006; Watson and Wooden 2009). Many of these characteristics can be associated with the concept of "vulnerability". Individuals with a migration background, who are unemployed, who are socially less integrated or whose health status is poor drop out more frequently. Moreover, a lower income is also associated with higher dropout rates when the reason for not responding is family-related problems and a higher educational level tends to have a positive impact on participation, except if the reason for dropping out is leaving the household. This means that in general population surveys such as the SHP vulnerable groups tend to be underrepresented. Whereas a survey such

as the SHP allows for a comparison of vulnerable groups with other groups in society, researchers should be aware that such datasets underestimate the degree of vulnerability in the general population. Weights allow to adjust to a certain extent for unit nonresponse, but, as has been shown in Sect. 4, for a number of variables, the correction is not sufficient.

Furthermore, one has to be cautious with the results related to the bias in estimates resulting from dropout, as the variables are compared from their first year of appearance and not imperatively in the first year of interviewing the sample. So it is possible that there is left handed bias already introduced in the sample when the question is asked the first time, in the sense that the estimates are already biased due to selective attrition. That is to say that a selective attrition may have occurred before the variable was introduced. This is undetectable by this method. These calculations are done on the entire sample of longitudinal respondents. There are no comparisons on the sub-populations (by sex, age class, nationality, etc.). Such comparisons could reveal differences which are not observed at the aggregated level. The inverse is also possible.

A difficulty related to the analysis of attrition is that the underlying patterns are complex. Our results contribute to the existing studies on attrition by specifically analysing different causes for nonresponse. This has revealed that the relationship between variables and dropout is often complex. For example, married individuals are less likely to drop out than other groups, except when family-related issues prevent their continued participation. Similarly for home ownership, compared to tenants home owners are less likely to drop out because they cannot be contacted, but they are more likely to refuse participation. When introducing group-specific measures to counter attrition one should, therefore, bear in mind that a single intervention might have an impact on a specific reason for nonresponse, but might not affect all the causes.

Another difficulty related to this kind of attrition analysis is that the imminent factors leading to irregular participation or complete dropout might often be unobserved, because there is no data for the wave in question: because individuals drop out, we don't know their most current situation, but only the one from the previous year, when they last answered the questionnaire. If someone gets ill between two waves and cannot participate anymore because of this illness, it is not possible to consider this in the statistical model, because the last available observations of this individual would not reflect the illness. In our data, this individual might still state that the satisfaction with health is high, because the data refer to the situation in the previous year. Therefore it is likely that we tend to underestimate the influence of the explanatory variables. Although efforts are made to collect information about why a household or an individual does not want to participate anymore, the data is incomplete and cannot, therefore, be incorporated into the model. Also related to the variables that were used here is that when doing this kind of analysis, we suppose that attrition is based on variables that can be observed in the dataset. However, attrition is also likely to be partly due to variables that are not included in the questionnaire. If this is the case, models estimating the factors influencing attrition fall short. Moreover, the available weights, whose construction is based on

the variables available in the dataset, would run the risk of not fully correcting the bias introduced by attrition. Although this can be a problem, our analysis has shown that for over 90 % of the variables estimates are unbiased or the bias is corrected by applying the weights.

References

Alderman, H., Behrman, J., Watkins, S., Kohler, H.-P., & Maluccio, J. A. (2001). Attrition in longitudinal household survey data. *Demographic Research, 5*, 79–124.

Allison, P. D. (1982). Discrete-time methods for the analysis of event histories. *Sociological Methodology, 13*, 61–98.

Beyersmann, J., & Schumacher, M. (2008). Time-dependent covariates in the proportional subdistribution hazards model for competing risks. *Biostatistics, 9*(4), 765–776.

Dillman, D. A., Eltinge, J. L., Groves, R. M., & Little, R. J. (2002). Survey nonresponse in design, data collection, and analysis. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. Little (Eds.), *Survey nonresponse*. New York: Wiley.

Gallie, D., & Paugam, S. (2004). Unemployment, poverty and social isolation: An assessment of the current state of social exclusion theory. In D. Gallie (Ed.), *Resisting marginalization: Unemployment experience and social policy in the European Union*. Oxford, UK: Oxford University Press.

Gooley, T. A., Leisenring, W., Crowley, J., & Storer, B. E. (1999). Estimation of failure probabilities in the presence of competing risks. *Statistics in Medicine, 18*(6), 695–706.

Groves, R. M. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly, 70*(5), 646–675.

Groves, R. M., Singer, E., & Corning, A. (2000). Leverage-saliency theory of survey participation: Description and illustration. *Public Opinion Quarterly, 64*(3), 299–308.

Groves, R. M., Presser, S., & Dipko, S. (2004). The role of topic interest in survey participation decisions. *Public Opinion Quarterly, 68*(1), 2–31.

Johnson, T. P., O'Rourke, D., Burris, J., & Owens, L. (2002). Culture and survey nonresponse. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. Little (Eds.), *Survey nonresponse*. New York: Wiley.

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association, 53*(282), 457–481.

Kleiner, B., Lipps, O., & Ferrez, E. (2012). Satisficing and language proficiency. *FORS working paper series* (Paper 2012-1). Lausanne: FORS.

Lipps, O. (2007). Attrition in the Swiss Household Panel. *Methoden – Daten – Analysen, 1*(1), 45–68.

Loosveldt, G. (1997). Interaction characteristics of the difficult-to-interview respondent. *International Journal of Public Opinion Research, 9*(4), 386–394.

Loosveldt, G., & Carton, A. (2001). An empirical test of a limited model for panel refusals. *International Journal of Public Opinion Research, 13*(2), 173–185.

Mills, M. (2011). *Introducing survival and event history analysis*. London: SAGE Publications Ltd.

Prentice, R. L., Kalbfleisch, J. D., Peterson, A. V., Jr., Flournoy, N., Farewell, V. T., & Breslow, N. E. (1978). The analysis of failure times in the presence of competing risks. *Biometrics, 34*(4), 541–554.

Putter, H., Fiocco, M., & Geskus, R. B. (2007). Tutorial in biostatistics: Competing risks and multi-state models. *Statistics in Medicine, 26*(11), 2389–2430.

Stoop, I. A. L. (2005). *The hunt for the last respondent: Nonresponse in sample surveys*. The Hague: SCP.

Voorpostel, M. (2010). Attrition in the Swiss Household Panel by demographic characteristics and levels of social involvement. *Swiss Journal of Sociology, 36*(2), 359–377.

Watson, N., & Wooden, M. (2009). Identifying factors affecting longitudinal survey response. In P. Lynn (Ed.), *Methodology of longitudinal surveys* (pp. 157–182). London: Wiley.