**Attrition rates in longitudinal household surveys: does sampling scheme matter?**

Michela Coppola[[1]](#footnote-1), Benedikt Alt

The use of longitudinal studies measuring the same sample of individuals at different point in time has become increasingly popular in the last years, not least because of the analytical advantages they offer over cross-sectional surveys. The quality of the analyses performed with such data, however, is possibly threatened by attrition and in particular by the possible selective drop-out of participants from the panel.

While several studies have already thoroughly examined the relationship between respondent characteristics (such as age, education or attitudes) and attrition behavior, much less is known about how attrition is related to sampling scheme. More specifically, there is scarce evidence on possible differences in attrition rates between access panels and traditional random samples. The use of access panels to conduct household surveys is gaining momentum in the last years, partially thanks to the rapid increase of web surveys. Understanding if attrition behavior differs between access panels and traditional random samples is therefore relevant.

Using data from the German survey “Saving and Old-age provision” (SAVE) I examine the effect of sampling scheme on attrition behavior. SAVE is a longitudinal study focused on households’ saving decisions and asset choices. Besides collecting information on relevant socio-psychological determinants of saving and old-age provision behavior, the SAVE questionnaire asks respondents detailed quantitative questions on all aspects of their household balance sheet (e.g. financial and real assets, income, liabilities). It started in 2001 with a set of experiments on different interview modes conducted on a quota sample, which was then dismissed in 2003 after being re-interviewed a second time to collect data on the willingness to participate in a long-term panel study on financial matters.

The main scientific SAVE Random Sample (RS) started in 2003 (Figure 1) and after the refreshment in 2005, it has been conducted on a yearly base. The 2003 random sample of SAVE was drawn by a multiple stratified multistage random route procedure (Heien and Kortmann, 2003). Since this turned out to be costlier than expected, the refreshment to the random sample in 2005 used a large sample drawn from the community-based German population registers (“Einwohnermeldeamtsstichprobe”) in a multistage procedure. At the same time, SAVE includes also a third sample, the so-called TPI Access Panel (AP). It consists of a standing panel of household surveyed at regular intervals, operated by the company TNS Infratest TPI (Test Panel Institute, Wetzlar).

Figure 1: Evolution of the different SAVE-subsamples over time.

The SAVE survey, consisting of a random sample and of an access panel, offers a unique opportunity to compare attrition rates within the two subsamples, while keeping constant other aspects of the survey which might affect attrition (e.g. length of the questionnaire or type of questions). The two subsamples, however, slightly differ in their interview mode. For the RS it was applied a mixed mode strategy. The first wave (2003 for the first sample and 2005 for the refresher) was conducted as computer assisted personal interview (CAPI), with the sensitive financial questions being asked in a paper-and-pencil (P&P) drop-off questionnaire that was picked up by the interviewer afterwards. From the second wave on, participants in the RS receive first a self-administered P&P questionnaire via mail and, if the questionnaire is not mailed back after a certain amount of time and a series of reminder, respondents are contacted by an interviewer. Eventually a CAPI is conducted (again with an additional drop-off questionnaire for the sensitive items). The self-administered P&P design, on the contrary, have been used for the respondents in the AP from the very first wave, based on the argument that they are already used to participating in a survey. These respondents receive a series of reminder as well in case they do not mail back the questionnaire after a certain amount of time. However, they are not contacted by an interviewer and no CAPI is conducted. Different incentives have been also used for the households in the two subsamples over the years (Table 1).

Table 1: Incentives in the RS and in the AP over time

|  |  |  |
| --- | --- | --- |
|  | Random Sample | Access Panel |
| 2005 | 15€ cash (after completed interview) | Present 3€/5€ (before interview) |
| 2006 | Present 5€ (before interview); | Present 3€/5€ (before interview) |
| 20€ bank transfer after completed interview; |  |
| 2007 | 20€ cash (before interview) |  5€ cash (before interview) |
| 17 out of 258 (**6.6**%) paid again |  |
| 2008 | 20€ cash (before interview) |  5€ cash (before interview) |
| 68 out of 198 (**34**%) paid again | for attriters after 8 weeks: lottery (10X150€) |
|  | 5€ cash, consolation prize |
| 2009 | 20€ cash (before interview) | 5€ cash (before interview) |
|  | for attriters after 8 weeks: lottery (10X150€) |
|  | 5€ cash , consolation prize |
| 2010 | 20€ cash (before interview) | 10€ cash (before interview) |
| for attriters after 8 weeks: ARD lottery (5€) + pen | for attriters after 8 weeks: ARD lottery (5€) + pen |

Aim of this work is to compare the attrition rates in the RS and in the AP and to see if the determinants of attrition in the two subsamples differ. To this scope we just look at the RS refresher sample in 2005 and at the AP refresher sample in 2006. Figure 2 plots the Kaplan-Meier survival estimates for the two subgroups (analysis time is in this case number of waves survived). A test of the equality of the two survival curves rejects the null hypothesis of equality: RS and AP are thus different in their survival in the study. In particular, individuals in the AP appear to have a lower survival probability than respondents in the RS.

To further analyze this issue we estimate a discrete time hazard rate model. A complementary log-log model is estimated first on the whole sample, and then on the two sub-samples separately. Table 2 reports the results. Even in a multivariate setting it turns out that respondents in the AP have a higher probability to drop-out of the panel. Furthermore, while income and gender of the respondent are significant determinants of attrition in the AP, these variables play no role in the RS. On the contrary, a switch toward a full-time job significantly increases the probability to drop-out of the panel only in the RS.

**Discussion points:**

* Is there any relevant variable missing to model attrition?
* Can the differences in the interview mode and in the incentives invalidate the comparison of the two subsamples?
* When introducing as regressor the number of questions in the questionnaire, the sign of the coefficient is negative (the more questions, the lower the probability to attrite): how to make sense of it?

Figure 2: Kaplan-Meier survival estimates RS and AP.



Table 2: Discrete Time Hazard Rate Model. *Cloglog* specification.

|  |
| --- |
| Hazard Ratios |
|  | All  |  | AP | RS |  |
| RP is female | 0.82 | \*\*\* | 0.72 | \*\*\* | 0.95 |  |
| RP lives with partner | 1.11 |  | 1.09 |  | 1.12 |  |
| # children | 1.05 |  | 1.05 |  | 1.07 |  |
| age in W1 < 30 |  |  |  |  |  |  |
| Age in W1 30-40 | 0.78 | \*\*\* | 0.90 |  | 0.62 | \*\*\* |
| Age in W1 40-50 | 0.53 | \*\*\* | 0.49 | \*\*\* | 0.62 | \*\*\* |
| Age in W1 50-60 | 0.43 | \*\*\* | 0.39 | \*\*\* | 0.49 | \*\*\* |
| Age in W1 60+ | 0.49 | \*\*\* | 0.44 | \*\*\* | 0.58 | \*\*\* |
| Hauptschulabschluss |  |  |  |  |  |  |
| Realschulabschluss | 0.77 | \*\*\* | 0.81 | \*\* | 0.74 | \*\*\* |
| (Fach-)Abitur | 0.74 | \*\*\* | 0.82 | \* | 0.71 | \*\*\* |
| No post-secondary educ |  |  |  |  |  |  |
| Vocational training | 0.80 | \*\*\* | 0.79 | \*\* | 0.83 | \* |
| University | 0.89 |  | 0.84 |  | 0.94 |  |
| East Germany | 0.79 | \*\*\* | 0.86 |  | 0.74 | \*\*\* |
| RP full-time employed | 1.08 |  | 0.95 |  | 1.26 | \*\* |
| Av. relevance sav. motive (W1) | 0.92 | \*\*\* | 0.94 | \*\*\* | 0.92 | \*\*\* |
| HH income Q1 (W1) |  |  |  |  |  |  |
| HH income Q2 (W1) | 0.81 | \*\* | 0.75 | \*\* | 0.90 |  |
| HH income Q3 (W1) | 0.75 | \*\*\* | 0.73 | \*\* | 0.78 |  |
| HH income Q4 (W1) | 0.76 | \*\*\* | 0.70 | \*\*\* | 0.85 |  |
| HH income Q5 (W1) | 0.89 |  | 0.91 |  | 0.80 |  |
| HH net worth Q1 (W1) |  |  |  |  |  |  |
| HH net worth Q2 (W1) | 0.98 |  | 1.16 |  | 0.91 |  |
| HH net worth Q3 (W1) | 0.94 |  | 1.15 |  | 0.81 | \* |
| HH net worth Q4 (W1) | 0.88 | \* | 0.99 |  | 0.81 | \* |
| Self-reported health: good |  |  |  |  |  |  |
| Self-reported health: average | 0.99 |  | 1.03 |  | 0.96 |  |
| Self-reported health: bad | 1.27 | \*\*\* | 1.12 |  | 1.56 | \*\* |
| time interval: 1 |  |  |  |  |  |  |
| time interval: 2 | 0.60 | \*\*\* | 0.78 | \*\*\* | 0.45 | \*\*\* |
| time interval: 3 | 0.60 | \*\*\* | 1.04 |  | 0.23 | \*\*\* |
| time interval: 4+ | 0.22 | \*\*\* | 0.23 | \*\*\* | 0.21 | \*\*\* |
| AP | 1.32 | \*\*\* |  |  |  |  |

Significance levels: \*\*\* p<0.01; \*\* p<0.05; \* p<0.10

Source: SAVE 2005-2010

1. Munich Center for the Economics of Aging at the Max-Planck-Institute for Social Law and Social Policy, Munich. Email: Coppola@mea.mpisoc.mpg.de [↑](#footnote-ref-1)