Deutsches Institut für
Wirtschaftsforschung

DIW BERLIN

SOEPpapers
on Multidisciplinary Panel Data Research

164

Oliver Lipps

Attrition of Households and Individuals in Panel Surveys

Berlin, March 2009

SOEPpapers on Multidisciplinary Panel Data Research
at DIW Berlin

This series presents research findings based either directly on data from the German Socio-Economic Panel Study (SOEP) or using SOEP data as part of an internationally comparable data set (e.g. CNEF, ECHP, LIS, LWS, CHER/PACO). SOEP is a truly multidisciplinary household panel study covering a wide range of social and behavioral sciences: economics, sociology, psychology, survey methodology, econometrics and applied statistics, educational science, political science, public health, behavioral genetics, demography, geography, and sport science.

The decision to publish a submission in SOEPpapers is made by a board of editors chosen by the DIW Berlin to represent the wide range of disciplines covered by SOEP. There is no external referee process and papers are either accepted or rejected without revision. Papers appear in this series as works in progress and may also appear elsewhere. They often represent preliminary studies and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be requested from the author directly.

Any opinions expressed in this series are those of the author(s) and not those of DIW Berlin. Research disseminated by DIW Berlin may include views on public policy issues, but the institute itself takes no institutional policy positions.

The SOEPpapers are available at
http://www.diw.de/soeppapers

Attrition of Households and Individuals in Panel Surveys

OLIVER LIPPS[*]

Attrition is mostly caused by not contacted or refusing sample members. On one hand it is well-known that reasons to attrite due to non-contact are different from those that are due to refusal. On the other hand does non-contact most probably affect household attrition, while refusal can be effective on both households and individuals. In this article, attrition on both the household and (conditional on household participation) the individual level is analysed in three panel surveys from the Cross National Equivalent File (CNEF): the German Socio-Economic Panel (GSOEP), the British Household Panel Study (BHPS), and the Swiss Household Panel (SHP). To follow households over time we use a common rule in all three surveys. First, we find different attrition magnitudes and patterns both across the surveys and also on the household and the individual level. Second, there is more evidence for reinforced rather than compensated household level selection effects if the individual level is also taken into account.

Keywords: CNEF, household attrition, attrition bias, reference person, household head, methods.

---

[*] Swiss Foundation for Research in Social Sciences (FORS), University of Lausanne, Vidy, CH – 1015 Lausanne, oliver.lipps@fors.unil.ch.

1. Introduction

Attrition is an important quality criterion in large household panel surveys (Zabel 1998, Watson and Wooden 2004, Behr et al. 2005, Gramlich 2008, Lipps 2007, Spiess and Kroh 2008, Uhrig 2008). These studies examine household panel surveys, in most of which all individuals in a household should answer the survey. Also common to these studies is that attrition analyses are limited to the individual level, although drop-out may well happen on the "level" of the household[1]: Usually drop-outs occur if households cannot be tracked or contacted, or - once contact is established - the first questionnaire (household enumeration grid) is refused. As noncontact or refusal of the survey request in a new wave is often the result of a common family decision[2], the reasons for "household attrition" might be different from that for individual attrition. The latter is conditional on a completed household listing and can be assumed to be the decision of the single individual concerned.

Although it is common practice to use the individual as the only longitudinal unit in household panel surveys, we wonder if a formal definition of "household attrition" exists. This is especially interesting in the case of a household split between waves. A simple GOOGLE-Scholar search for the term "household attrition" results in 94 hits, a standard GOOGLE search in even 404 hits (accessed 17FEB2009). This term is however used in a rather lax way in most articles. Even the site of the American Association for Public Opinion Research (AAPOR) does not include any definition, not even a description. Only one source provides a definition of "household attrition". EUROSTAT[3] defines household response rates in the European Community Household Panel (ECHP) as "the ratio of the number of interviewed households to the target number for interview. … The latter is the number of households forwarded from the previous wave, minus those no longer existing, plus the newly formed ones." (cited in Peracchi 2002: 74). Accordingly, "the household attrition rate between wave s and wave t > s is defined as one minus the ratio between the number of households interviewed in both waves and the number of households interviewed in wave s." (Peracchi 2002: 78). A problem with this definition is that the attrition rate might be higher than 1 – if a large number of both (or possibly more) parts of split households continue to take part, while "usual" households mainly continue the survey. While this is admittedly rather unrealistic it is

---

[1] Also substantive analyses often use household aggregated measure, especially in economic studies (e.g. Frick, Grabka and Sierminska 2007).
[2] For the SHP, from interviewer notes, this appears to be the case in about 80% of all dropping-out households.
[3] On the EUROSTAT website no definition can be found.

not entirely impossible. Peracchi (2002) for example, reports that between 2 and 6 % of all households split between waves 1 and 2 in the ECHP (p. 83).

In the analyses to follow we identify one part of the split household as the one to follow for attrition analyses. In the ECHP, for example, the part of the split household that retains the 'parent' household ID is the one which remains at the same address as the parent household. Otherwise it is the household where the reference person of the parent household now lives. If such person is no longer in the survey population, then the present household of the person listed with the smallest line number in the parent household membership roster retains its ID (Locatelli et al. 2001).

The article is organised as follows: first we briefly present the three panel surveys used together with their household following rules before we propose a harmonised household following rule which allows for attrition analyses in comparative panel surveys. To analyse possibly different reasons to attrite, we compare the reason for attrition on both the household and the individual level in two surveys. Next we analyse attrition on the level of both the household and the individual, first based on descriptive statistics and then applying multivariate survival models. We summarise at last.

2. Data

To analyse household attrition based on common household following rules, we use data from the German Socio-Economic Panel[4] (GSOEP; Wagner, Frick, and Schupp 2007), the British Household Panel Study (BHPS; Buck 2006), and the Swiss Household Panel (SHP; Budowski et al. 2001). All surveys use the same rules as to following original sample members[5]. For the analyses, we include all adults who answered the individual questionnaire in the first wave and consider them until first drop-out, i.e., ignoring possible re-participation. For comparison purposes, we use the same number of initial panel waves in all three surveys. Because the SHP, which started in 1999, is the most recent survey with data available until the 9th (2007) wave, we use BHPS data from 1991 through 1999, and GSOEP data from 1984-1992. During these waves, the BHPS conducted face-to-face interviews using paper and pencil questionnaires (PAPI; see Uhrig 2008), as well as the GSOEP (Wagner, Frick, and Schupp 2007), while the SHP used CATI from the very beginning (Latouche and Naud 2001).

---

[4] We use the 95% scientific use sample of the GSOEP.
[5] See Frick et al. (2007), and specifically Gramlich (2008) and Kroh et al. (2008) for the GSOEP, Uhrig (2008) for the BHPS, and Latouche and Naud (2001) for the SHP.

Household identification across waves works as follows: while the BHPS circumvents the problem by assigning new household IDs across waves even without a changed household composition (Buck et al. 2006), such households retain their ID in the other surveys and can easily be tracked. In case of a split, one part of a split household usually keeps the household ID across waves. In the GSOEP, the part that does not move keeps the household ID. In the event that all parts leave an address, the part, in which the previous reference person (see below) lives, keeps the ID (Rendtel 1995). In the SHP, in case of a split, generally the part in which the previous reference person lives, keeps the ID. If the previous reference person does no longer live in one of the split households, the household which keeps the address also keeps the household ID. If the address is left by all split-off households, the largest split-off household keeps the former household ID. If even the number of individuals is the same, the split-off household with the oldest reference person keeps the ID (SHP 2003).

These examples show that the assignment of household IDs over waves follows different concepts in household panel surveys: while some drop former household IDs completely (BHPS), others base the ID mainly on the address (GSOEP, ECHP), or use the former reference person as an anchor (SHP). Also telephone numbers as another alternative in telephone surveys are imaginable.

3. The Reference Person as the Household's Anchor

In the surveys considered, one person has been designated as the household reference person in the first wave, and should preferably keep this status during the subsequent panel waves. Usually the appointment of the reference person follows a clear definition. A formal definition of the household reference person can be found on the BHPS website of the Institute for Social and Economic Research (ISER): "The household reference person … is the person legally or financially responsible for the accommodation or the elder of two people equally responsible."[6] The GSOEP uses a more subjective concept: "the household head [is] defined as the person who knows best about general conditions in the household …" (Knoppik 2002: 3). In the SHP the choice of the household reference person in the first panel wave was more at the discretion of the household (Budowski et al. 2001: 101f.) although s/he should be "the same person from one wave to the next" (Budowski et al. 2001: 111). While first wave household reference persons tend to be male in the face-to-face surveys BHPS (62.3%) and especially the GSOEP (79.5%), in the telephone SHP, only 36.3% are male. The

---

[6] http://www.iser.essex.ac.uk/ulsc/bhps/faqs/households.php

4

higher share of female reference persons in the SHP also reflects better accessibility of women in telephone surveys and the fact that women, more than men, tend to be responsible for answering the telephone (Groves and Couper 1998).

We define the household longitudinally as that part of the household in which the previous reference person lives. Other eventually split household parts are no longer followed. Though still living in the household the previous reference person may not cooperate. In this case, the household still exists. If the previous reference person cannot be identified the household is considered to attrite; if s/he drops out of the sample, the same holds for the household. Some household socio-demographic characteristics like age are represented by the household reference person in the current or – if the household attrites – the previous wave. This is in line with previous research, in which, for example, the age of the reference person often characterises the 'age' of the household. For example, Sefton and van de Ven (2004) simulate the household saving behaviour and labour decisions depending on household age. Wave to wave changes of reference persons are rather rare: in the GSOEP, in less than 2% of all wave to wave transitions, individuals change their status from a non-reference person to a reference person or vice versa. This figure amounts to 7% in the BHPS and 13% in the SHP. Apart from a better control possibility by the interviewer through face-to-face, there is high interviewer continuity in the GSOEP households over years and thus familiarity between the interviewer and the reference person.

4. Magnitude of Attrition of Households and Individuals

In this section, we depict the magnitude of attrition of households and – given the household can be contacted and cooperates (i.e., the grid is completed) – attrition rates of individuals who participated in all previous waves. To obtain an interview from all interview eligible household members is an explicit target in all surveys, but is achieved to a different extent.
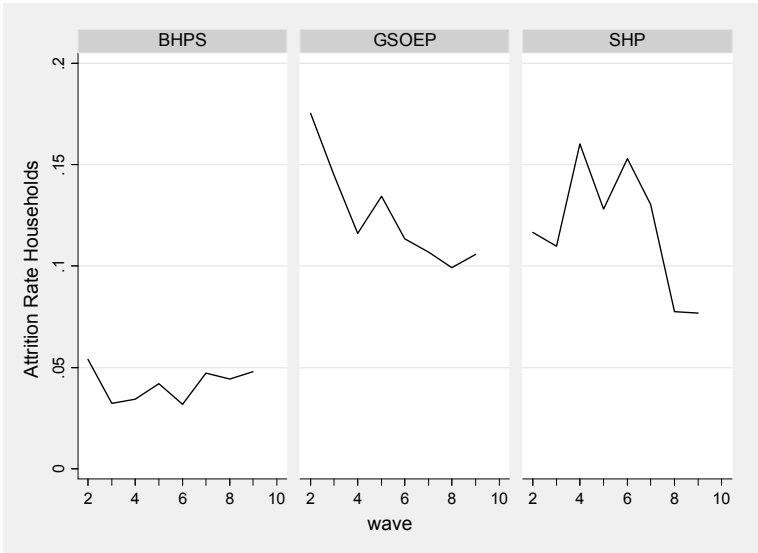
Figure 1: Attrition Rates of first Wave Households over Wave-to-Wave Transitions until Wave 9, by Survey.

Household level attrition patterns are different across the three surveys. They reflect tracking efforts and cooperation performances across surveys and waves.

Attrition rates on the individual level conditional on household grid completion is generally higher in the SHP:
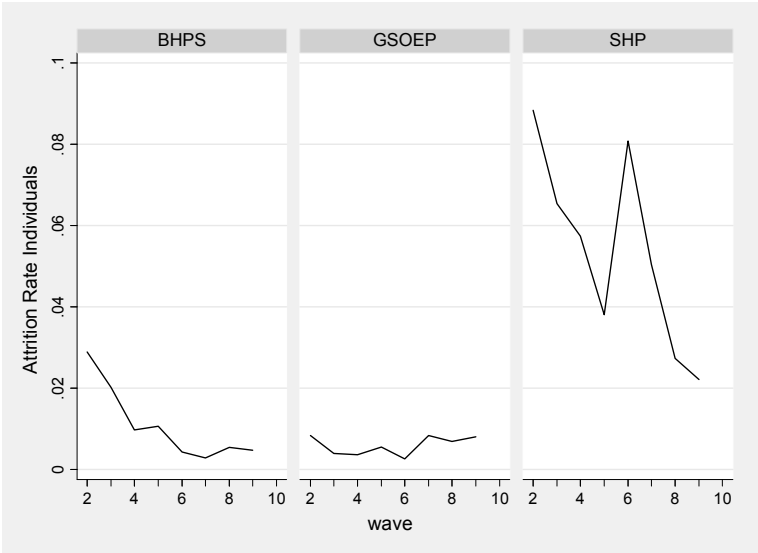

Figure 2: Attrition of first Wave Respondents over Wave-to-Wave Transitions until Wave 9, by Survey.

The discrepancy between BHPS and GSOEP on one hand, and the SHP on the other is probably mostly due to the mode of interview. In face-to-face panel surveys, usually the

interviewer makes sure that all interview eligible household members are at home at the time of the visit and has a comparatively high control over individual response. In telephone surveys, generally more contacts (and visits) are necessary to work a household, which increases the possibilities for individuals to drop out. As mentioned above, in the GSOEP this higher interviewer control is increased by comparatively high interviewer continuity in the GSOEP households over waves.

5. Reasons for dropping-out

Because the reasons driving noncontacts differ from those for refusals (e.g. Uhrig 2008), it is necessary to check reasons for no contact and for refusal separately. We perform this based on last contact or last contact attempt results, in each wave, available in the GSOEP and the SHP. In the GSOEP, pooling wave 2 to wave 9, we find the following reasons for household and individual level dropping-out. Again, individual drop-out is analysed conditional on household grid completion.

N [household- and person- years] Reason for dropping-out [%]	Households (N = 1,854)	Individuals (N = 253)
Unsuccessful Approach/ not reached	21.9	24.1
Refusal	69.6	75.1
Out of Sample	8.4	.8
All (temporary) Drop-out	100.0	100.0

Table 1: Reasons for first Wave participating Households and Individuals (given Household Participation) to (temporarily) dropping-out of the GSOEP (2nd through 9th Wave).

By pooling nonresponse reason in the SHP, we obtain the results in Table 2

N [household- and person- years]	Households	Individuals
Reason for dropping-out [%]	(N = 3,786)	(N = 1,612)
Unsuccessful Approach/ not reached	24.4	25.8
Refusal	74.4	73.0
Out of Sample	1.2	1.2
All (temporary) Drop-out	100.0	100.0

Table 2: Reasons for first Wave participating Households and Individuals (given Household Participation) to (temporarily) dropping-out of the SHP (2nd through 9th Wave).

Apart from the much higher share of individual dropping-out *occurrence* in the SHP, one difference is the higher proportion of non-sample cases in GSOEP households. It is likely that a higher portion of sample members who moved out of the country were identified as such in the GSOEP (see Infratest 2002), compared to the SHP. The similarity of drop-out reasons of both households and individuals and across both surveys is surprisingly high.

6. Duration Modelling on Household and Individual Level

We first present the amount of attrition bias with respect to five socio-demographic variables available in the CNEF. They will be included later in a multivariate regression analysis. In Table 3 we depict means for (initial) age, male (0/1), partner (0/1), household size, and working status (0/1), distinguished by all first wave adults and those among them who stay until the ninth wave. If the differences are significant this is indicated by asterisks.

Individuals in 1st wave: all vs. 9th wave stayers	BHPS all N=8,636	BHPS stay N=6,197	SHP all N=7,390	SHP stay N=2,347	GSOEP all N=11,203	GSOEP stay N=3,921
Age	43.62	43.98	44.68	46.01**	42.72	41.97*
Male	.472	.448**	.436	.405**	.492	.488
Lives with Partner	.751	.773**	.689	.743**	.821	.839*
Household Size	2.898	2.890	2.823	2.872	3.193	3.250*
Working	.605	.636**	.684	.691	.593	.622**

Table 3: Mean Age, Sex, Partner, Household Size and Working Status of 1st Wave Persons vs. 9th Wave Stayer Sample. **=significantly different from Sample with all Persons at 1%-level, *=at 5%-level.

We find significant bias for age in the GSOEP and especially in the SHP, for sex in the BHPS and the SHP, for the variable "living together with a partner" in all surveys (GSOEP only 5%), for the household size in the GSOEP (5%) and for the working status in the BHPS and the GSOEP. Apart from age, bias point in the same direction in all surveys once significant. Given the comparatively small attrition in the BHPS it is surprising that the attrition bias is highly significant for three of the five variables considered. Overall, attrition bias in the SHP is not as high as could be expected from the high attrition magnitude.

Next, we check whether the bias is rather due to household or individual attrition. We use discrete survival models to analyse household and individual participation. Checking the logits of the household attrition rates (here distinguished by the "lives with a partner" variable), we model wave by a dummy variable because of non-monotone attrition logits.
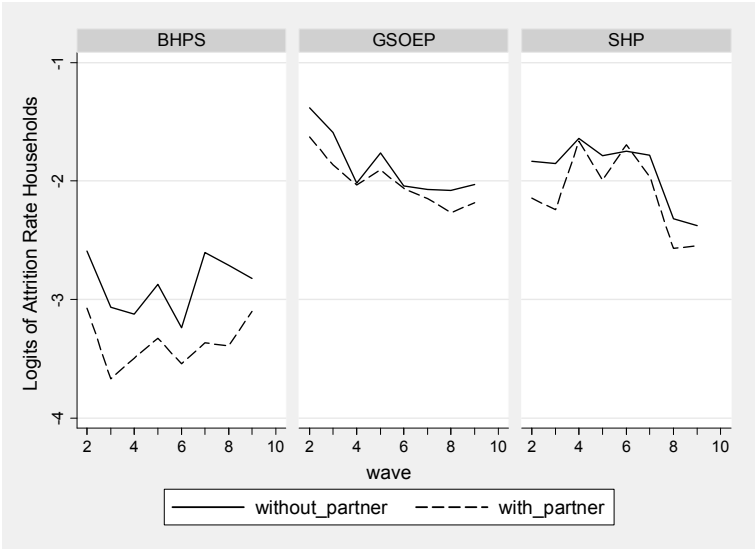

Figure 3: Logits Attrition Rates of first wave Households over wave-to-wave transitions until wave 9, by survey.

Logit Household Attrition	[Households]	BHPS	SHP	GSOEP
N (household waves at risk)		36,534	26,058	30,393
Age (in first wave)		-.098**	-.070**	-.027**
Age squared (in first wave)		.0009**	.0006**	.0004**
Male		.374**	.137**	.170*
Lives with Partner		-.464**	.015	-.176**
Household Size		.018	-.088**	-.007
Working		-.457**	-.098*	-.020
McFadden Pseudo R^2		.028	.018	.011

Table 4: Discrete Duration Logit Model (wave effect controlled): Attrition of Households. **=significant at 1%-level, *=significant at 5%-level, -=not included.

All surveys have a u-shaped age-attrition pattern on the household level, which reflects higher nonresponse in young (mostly movers; resulting in a non-contact) and older (mostly refusers; resulting in a non-cooperation) households. In the SHP and especially the BHPS, male headed households show higher attrition. Larger households tend to attrite to a smaller extent in the SHP; those with a working reference person in the BHPS. Although household in the BHPS have by far the smallest attrition rates (Figure 1), the explanation power is highest in the BHPS. This also holds if the wave dummies are dropped. Therefore attrition selection is comparatively strong on the BHPS household level, compared to both other panel surveys.

Also for the logits of the individual attrition rates (conditional on household participation) we find a mostly non-monotone effect of wave.
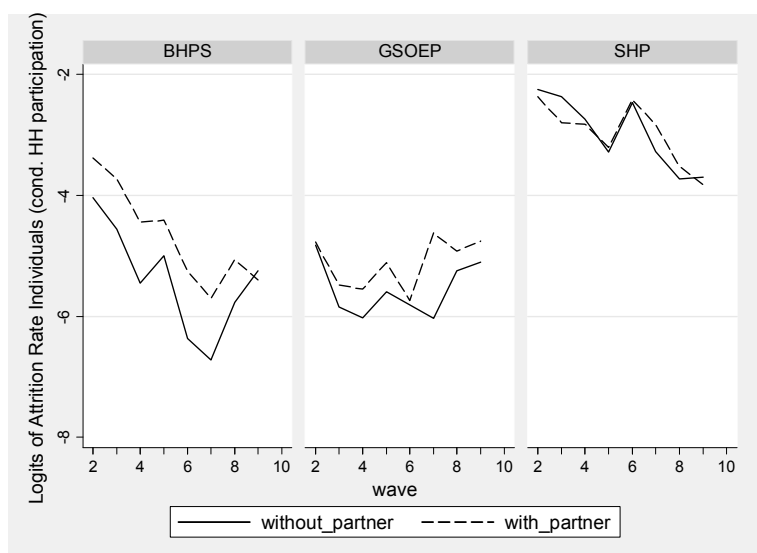
Figure 4: Logits Attrition Rates of first wave Individuals over wave-to-wave transitions until wave 9, by survey.

Logit Individual Attrition	[Individuals]	BHPS	SHP	GSOEP
	N (individual waves at risk)	55,472	31,071	46,106
	Household Reference Person	-.673**	-1.162**	-1.582 **
	Age (in first wave)	.014	-.060**	-.046*
	Age squared (in first wave)	-.0001	.0006**	.0004
	Male	1.144**	-.025	1.281**
	Lives with Partner	-.001	-.178**	-.173
	Household Size	.283**	-.009	.095*
	McFadden Pseudo R^2	.095	.066	.072

Table 5: Discrete Duration Logit Model[7] (wave effect controlled): Individual Attrition of Individuals conditional on Household Participation. **=significant at 1%-level, *=significant at 5%-level, -=not included.

The model for the BHPS has again the highest explanation power. In all surveys, first wave reference persons have a higher participation rate. This is in line with findings from De Keulenaer (2005) who analyses attrition in the Panel Study of Belgian Households (PSBH).[8] As for age, the bias from household attrition is aggravated by individual attrition in the SHP. In the GSOEP and especially in the BHPS, the age bias is not increased due to individual

---

[7] The variable "working" cannot be used because of co-linearity with the attrition variable in too many cases.
[8] Unfortunately she did not distinguish between the reference person and his/her partner.

attrition (but also not compensated for). Higher attrition by male headed households can be observed for individuals as well in the BHPS and the GSOEP. Attrition of individuals living together with a partner (whose households attrite to a smaller extent in the BHPS and the SHP) do not attrite to a different degree. This finding is to the contrary in the SHP, where attrition of partners is smaller only on the individual level. Household size is negatively correlated with attrition in the BHPS and the GSOEP, where this variable had no effect on the household level.

7. Summary

In the article we use a common following rule for split households across waves in three household panel surveys which allows for a comparative analysis of panel attrition on the level of the household. Specifically, we use the household reference person as the household longitudinal component. The high probability to keep the reference persons status over waves makes this concept plausible. If a household splits we follow just the part in which the former reference person lives, or define the household as not responding if the former reference person cannot be tracked.

To analyse attrition on both household and individual level, we use data from the second through the ninth wave of the BHPS, the GSOEP, and the SHP. We find that the attrition patterns are different across both levels and surveys. Using duration models, we find smaller attrition selectivity of socio-demographic variables on the household level compared to the individual level, conditional on household participation. Surprisingly, while the BHPS has the smallest overall attrition magnitude, selectivity is highest on both levels. We only find evidence for a reinforcement of selectivity from the household level, if the individual level is considered in addition. For example, sex bias is reinforced in the BHPS and the GSOEP, and age bias in the SHP. Other variables are significant on at most one level only.

The findings show that it is worth to distinguish household and individual level attrition separately. In order to reduce bias, knowledge of possible compensating or reinforcing effects of selectivity may give hints how to improve communication, incentive schemes, and fieldwork on both levels in household panel surveys. More research is needed to analyse household and individual attrition effects on other variables. It would be interesting to consider socio-economic or attitudinal variables, or variables measuring social activities which all have causal effects on nonresponse.

References

Behr, Andreas, Egon Bellgardt and Ulrich Rendtel (2005) Extent and Determinants of Panel Attrition in the European Community Household Panel. *European Sociological Review* **21** (5): 489-512

Buck, Nick (2006) General Household Panel Studies: The British Household Panel Survey. Paper prepared for the StatCan-Conference "Longitudinal Social Surveys in an International Perspective", McGill University, Montreal, 25-27 Jan 2006

Buck, Nick, Jonathan Burton, Heather Laurie, Peter Lynn, and Noah Uhrig (2006) Quality Profile: British Household Panel Survey Version 2.0: Waves 1 to 13: 1991-2003. Institute for Social and Economic Research, University of Essex

Budowski, Monica, Robin Tillmann, Erwin Zimmermann, Boris Wernli, Annette Scherpenzeel, and Alexis Gabadinho (2001) The Swiss Household Panel 1999-2003: Data For Research On Micro-Social Change. *ZUMA-Nachrichten* **49**: 100-125

Burkhauser, Richard, and Dean Lillard (2007) The expanded Cross-National Equivalent File: HILDA Joins Its International Peers. *The Australian Economic Review* **40**:2: 208–215

De Keulenaer, Femke (2005) Using Process Data to Predict Attrition from a Panel Survey: a Case Study. *Proceedings of the Survey Research Methods Section, American Statistical Association,*

Frick, Joachim, Markus Grabka, and Eva Sierminska (2007) Representative Wealth Data for Germany from the German SOEP: The Impact of Methodological Decisions around Imputation and the Choice of the Aggregation Unit. *DIW Discussion Papers* **672**

Frick, Joachim, Stephen Jenkins, Dean Lillard, Oliver Lipps, and Mark Wooden (2007) The Cross-National Equivalent File (CNEF) and its Member Country Household Panel Studies. *Schmollers Jahrbuch* **4**: 627-654

Gramlich, Tobias (2008) Analyse der Panelausfälle im Sozio-Oekonomischen Panel SOEP (Analysis of Panel Attrition in the German Socio Economic Panel Study GSOEP). *SOEPpapers* **129**

Groves, Robert and Mick Couper (1998) *Nonresponse in Household Interview Surveys.* New York: Wiley

Infratest (2002) Verbleibstudie bei Panelausfällen im SOEP. Infratest Sozialforschung München, accessed Dec 3, 2008, available at:
http://www.diw.de/documents/dokumentenarchiv/17/40135/verbleibstudie.357184.pdf

Knoppik, Christoph (2002) A Semi-Parametric Analysis of Downward Nominal Wage Rigidity in the GSOEP 1984-2000. *University of Regensburg Discussion Paper* **374**

Latouche, Michel and Jean-François Naud (2001) Weighting of the Swiss Household Panel. Wave 2 detailed Description. *Swiss Household Panel Working Paper* 8/2001

Lipps, Oliver (2007) Attrition in the Swiss Household Panel. *methoden – daten – analysen* **1** (1): 45-68

Locatelli, Marilena, Vito Moscato and Silvia Pasqua (2001) The European Community Household Panel (ECHP): elements for users with special focus on labour and household economics. *Child centre working paper* **24**/2001, University of Torino.

Peracchi, Franco (2002) The European Community Household Panel: a Review. *Empirical Economics* **27**: 63-90

Rendtel, Ulrich (1995) Lebenslagen im Wandel. Panelausfälle und Panelrepräsentativität. Campus

Sefton, James and Justin van de Ven (2004) Simulating Household Savings and Labour Supply: an Application of Dynamic Programming. *National Institute Economic Review* **188**: 56-72

SHP (2003) Numérotation automatique des Ménages. Unpublished Manuscript, Swiss Household Panel, Neuchâtel.

Spiess, Martin and Martin Kroh (2008) Documentation of Sample Sizes and Panel Attrition in the German Socio Economic Panel (SOEP) (1984 until 2006). *Data Documentation* **27**, DIW Berlin

Kroh Martin, Rainer Pischner, Martin Spiess and Gert G. Wagner (2008) On the Treatment of Non-Original Sample Members in the German Household Panel Study (SOEP). *methoden – daten – analysen* **2** (2) 179-198

Uhrig, Noah (2008) The Nature and Causes of Attrition in the British Household Panel Survey. *ISER Working Paper* **5**/2008

Wagner, Gert, Joachim Frick and Jürgen Schupp (2007) The German Socio-Economic Panel Study (SOEP) – Scope, Evolution and Enhancements. *Journal of Applied Social Science Studies* (Schmollers Jahrbuch) **127** (1): 139-169

Watson, Nicole, and Mark Wooden (2004) Sample Attrition in the HILDA Survey. *Australian Journal of Labour Economics* **7** (2): 293-308

Zabel, Jeffrey (1998) An Analysis of Attrition in the Panel Study of Income Dynamics and the Survey of Income an Program Participation with an Application to a Model of Labor Market Participation. *Journal of Human Resources* **33** *(2), Special Issue: Attrition in Longitudinal Surveys*, (Spring): 479-506