Studies on Data Quality in Complex Household Surveys

Dr. Oliver Lipps

Habilitationsschrift

Universität Basel

Philosophisch-Historische Fakultät

Fachbereich Soziologie

Lausanne, September 2010

Table of Contents

Background: Introduction

"Scientific surveys represent an indispensable instrument of social research … [and] have a significant social influence if their results contribute to public opinion formation, to decision making in politics and jurisdiction as well as to administrative acts, if they lead into scientific diagnoses, causal analyses and prognoses or if they lead into economic decisions.". Scientific surveys are often considered as all-in-one multipurpose devices. Reality, however, shows that it is increasingly difficult to satisfy rigorous survey quality requirements (Biemer and Lyberg 2003). Meeting the research needs of various social science disciplines comes at the price of long and burdensome questionnaires. Guaranteeing high statistical precisions of estimates and keeping design effects small means high data collection efforts, a large sample with high response rates and small measurement biases. This all requires a fair amount of time, money and know-how.

Various organizations have begun to focus on methodological issues towards improving data quality. It was only recently that the German Council for Social and Economic Data (Rat für Sozial- und Wirtschaftsdaten - RatSWD)[1] acknowledged that the problem of quality of data from scientific surveys calls for more effort: "… the RatSWD aims to shift its main focus from increasing data access to improving data production and data quality" (Solga and Wagner 2007: 4). In January 2008, the Priority Programme on Survey Methodology (PPSM), funded by the German Research Foundation (Deutsche Forschungsgemeinschaft – DFG), got underway. On the PPSM website, the scientific objectives of this large program are explained: "… it is of essential importance, that no wrong, distorted, vague or misleading conclusions are drawn from the results of scientific surveys… In practice, however, the required survey quality cannot be attested unquestionably, either because the quality is not always achieved or because there is simply no attempt to judge it. In view of all the potential sources of error surveys may suffer from, this is a remarkable situation."[2] Also in January 2008, the Swiss Foundation for Research in Social Sciences (FORS) was founded. Methodological research is an essential part of the FORS mandate, with a focus on the improvement of data quality in Swiss scientific surveys.[3] In addition, various research networks and projects on specific

[1] "The Council's main purpose is to advise in the development of the German data infrastructure for empirical research in the social and economic sciences. The Council is working … to sustainably improve data quality, but increasingly also in the development of long-term data surveys" (RatSWD 2008).

[2] From the German Priority Programme on Survey Methodology (PPSM) website: http://www.survey-methodology.de/en/schwerpunkt_en.html

[3] http://www2.unil.ch/fors/spip.php?rubrique23&lang=en

survey data quality problems, especially the generally increasing proportion of nonresponding sample members in random sample surveys, have come into being. In Great-Britain, for example, a new project, "Understanding Non-Response and Reducing Non-Response Bias", was recently funded by the UK Economic and Social Research Council (ESRC) - Survey Design and Measurement Initiative (ESRC 2008). To take an example of scientific exchange, the "International Workshop on Household Survey Nonresponse", at which the world's leading scientists in the field meet once a year, is an institution since 1990.[4]

In the present studies, I investigate selected aspects of data quality in some current large "complex" household surveys. Three dimensions of complexity will be considered: first, repeated cross-sectional[5] surveys, in which different sample members are asked the same questionnaire several times. Second, panel surveys, in which sample members are observed several times over a longer time span. In addition to initial unit[6] nonresponse, there is the risk that sample units drop out of the survey after having participated for some time.[7] The third level of complexity involves panel studies conducted in several countries. Because of institutional and survey specific differences, harmonizing such studies contain a great deal of complexity. In the surveys considered, all household members are interview eligible. Obtaining interviews from all household members is an additional challenge and an additional focus of the studies. Due to length and complexity of the questionnaire, and the need to achieve high response rates, all surveys examined use interviewers. To analyze interviewer effects or to take into account that time points are clustered in respondents in panel surveys, multilevel analyses are needed. This is because the residuals over and across interviewers (or respondents) are not independent from each other. Multilevel models are the modeling focus of the studies. Meeting high data quality standards is a challenge particularly in surveys that contain such complex elements.

This work consists of two parts: first, a background paper gives an overview of the field, including a summary of the current state of research on survey quality. I start with a short history of social science surveys, focusing on their increasing importance and complexity and

[4] http://www.nonresponse.org/

[5] A *cross-sectional* survey is one that takes place at a single point in time. A *longitudinal* or *panel* survey is one that takes place over time – there are at least two (and often more) waves of measurement [of the same objects] in a longitudinal design (Trochim 2006).

[6] [*unit* or Total] "*nonresponse* occurs when all or almost all data for a sampling unit are missing" (Statistics Canada 2003: 32). Formal definitions and various forms of nonresponse can be found in AAPOR (2009).

[7] A content-related problem specific to panel surveys, possible influence of the survey participation to behavior ("panel conditioning") is not considered here (see e.g., Warren and Halpern-Manners 2008).

the need to develop adequate survey research methods. Next, I describe survey quality areas in which the studies are embedded. Following this, my own contributions to some specific problems in three areas of complex surveys are summarized, using own articles published in a variety of journals on survey methodology. The full papers are attached to the end of the background paper.

The paper summaries are divided into three sections: In Section I survey quality is examined in a cross-national context. Three contemporary surveys are used: the Survey of Health, Ageing, and Retirement in Europe (SHARE), the European Social Survey (ESS) and the Cross National Equivalent File (CNEF). In section II I analyze nonresponse and attrition in the Swiss Household Panel (SHP) survey and discuss results from an incentive experiment, targeted to decrease attrition. Section III is devoted to examining the question of respondent and interviewer effects in centralized telephone surveys. First, I propose and study new measures of interviewer performance for obtaining sample members' cooperation. Second, I analyze possible efficiency gains by reassigning interviewers to specific contacts. Third, I analyze response quality issues in both a cross-sectional and longitudinal way.

Table 1 shows selected factors that influence the so called total survey error (Groves 1989). Issues studied in more detail are marked in bold. Quality of questionnaire design issues, like scale effects (Alwin and Krosnick 1985) or qualitative methods (Bergman and Coxon 2005) are not considered in the studies.

Start --> End

<table>
<tr><th>Specification / Questionnaire</th><th>Sample</th><th>Data Collection/ Measurement</th><th>Processing</th><th>Data Analysis</th></tr>
<tr><td>Validity & Reliability of Concepts</td><td>Sampling Frame</td><td>Mode, Instrument</td><td>Coding</td><td rowspan="3">Adequate Methods: e.g., Multilevel</td></tr>
<tr><td rowspan="2">Translation, Cross-national Comparability</td><td>Sampling</td><td>Interviewer</td><td>Editing</td></tr>
<tr><td>Observation Units</td><td>Item/Unit Nonresponse, Attrition</td><td>Weighting, Imputation</td></tr>
<tr><td>…</td><td>…</td><td>…</td><td>…</td><td>…</td></tr>
</table>

Table 1: Total Survey Error Factors and Issues analyzed in the Studies (bold).

Background: Data Quality in complex Surveys

1. Advances in Survey Methodology

"Survey research is an inclusive term used to describe a rather wide range of methods of social research. … [and] may be defined as the study of a population sample by means of interviews or questionnaires; the transformation of the categorized responses into numerical form; and the characterization of various aspects of the population" (Sills 1992: 93). At the time when modern social surveys emerged in the early 20th century, mainly social problems like causes of poverty or sickness were investigated in social surveys (Riley 1911). U.S. sociologists complained early that in the then conducted surveys important aspects were not considered. Some argued that the contemporary social surveys "fall short of the requirements of scientific sociology. The scientific survey … should be significant for the development of a science of sociology. … We need, therefore, in any scientific social survey … also a study of … common belief, common opinion, and the like." (Ellwood in Riley 1911: 834 f.).

Since then, numerous methodological improvements have been achieved. Examples are scales, developed by Paul Lazarsfeld, "the founder of modern empirical sociology" (Jeřábek 2001: 229). Scales that measure attitudes could be cross-tabulated with socio-demographic characteristics. Rensis Likert developed the now widely used Likert scales[8]. Concerning samples, George Gallup largely improved representativeness by replacing large straw polls with comparatively small quota[9] samples (Gallup 1944). This design led to both a smaller bias[10] and at the same time lowered costs because it allowed for a much smaller sample size. A classical example is the reelection of Franklin D. Roosevelt in 1936, which was correctly forecasted by Gallup. The then often used large and expensive straw polls however failed.

Today, survey data contributes to the understanding of the state and development of a broad range of social science disciplines and interdisciplinary areas (Woolfrey 2007). For example "in the area of politics, survey research has drastically revised understanding of the causes and consequences of electoral attitudes and behaviors, both over the life course and across generations" (House et al. 2004: 3). The availability of high quality data from social science surveys are taken for granted in industrialized countries (Fowler 2002). Also

[8] "*Likert Scaling* is a unidimensional scaling method" (Trochim 2006), leading to e.g. a 1-to-5 Disagree-Agree response scale.
[9] "*Quota* random sampling involves dividing the population into homogeneous subgroups and then taking a simple random sample in each subgroup" (Trochim 2006).
[10] "*Bias* denotes a fixed (over replications) departure from some underlying true value for the statistic" (Groves 1987: 157).

governmental agencies now rely on social statistics generated by surveys conducted by public and private institutions as a basis for policy formulation, both within and across nations.

With increasing costs, social surveys become more and more institutionalized and borrow production methods common in large infrastructure projects in the natural sciences (Maynard and Schaeffer 2000). This is also acknowledged by funders of repeated or longitudinal[11] social surveys, which are funded for an increasingly longer period, similar to institutionalized infrastructures like established large natural science projects. For example, the "European Strategy Forum on Research Infrastructures … [put together a] road map to work out collaborations on big projects … [that range] in cost from the € 9 million European Social Survey[12] to the € 1.2 billion Facility for Antiproton and Ion Research" (Clery 2006: 399).

2. Sources of Errors in Surveys

Today, survey methodology is in fact interdisciplinary, with researchers having conflicting ideas about data quality: "Survey research is not itself an academic discipline, with a common … set of principles. … [Rather, it] … has evolved through the … uncoordinated contributions of researchers trained as statisticians, psychologists, political scientists, and sociologists. [Consequently this] mélange of workers … produces disagreement about the importance of various components of quality. … There are the measurers who try to build empirical estimates of survey error and the reducers who try to eliminate survey error. …They rarely confront each other" (Groves 1987: 156). One common concept of quality however is bias and variance, with replication being a key term: members of the different disciplines consider different features of a survey fixed and variable over replications. For example, "to those interested in sampling error alone, replication means a different implementation of the survey using a different sample drawn in the same manner. … Those focusing on interviewer variance, the errors of interest are the variation in results … obtained if a different set of interviewers had done the work" (Groves 1987: 157). Today survey methodologists however realize that "current best practice … requires examination of all of the … design features. If there is a major compromise or weakness in any aspect of the survey design, major investments in other portions of the survey are not sensible" (Fowler 2002: 8). This shows the

[11] Repeated cross sectional and longitudinal surveys together allow for the identification and separation of age, period, and cohort effects (Palmore 1978).
[12] The European Social Survey (ESS) is "an academically-driven social survey designed to chart and explain the interaction between Europe's changing institutions and the attitudes, beliefs and behavioral patterns of its diverse populations" (http://www.europeansocialsurvey.org/). The ESS is conducted every other year since 2002.

importance of identifying all possible error sources and their relationships: The accuracy of statistics can be viewed in terms of *total survey error* (e.g., Groves 1989), which is the total effect of various specific sources of error associated with the survey process. The sources can be viewed as operations performed in a certain sequence, such as specification of a research problem and the questionnaire design, defining a target population and a sampling scheme, developing the fieldwork protocol, and choosing a set of data processing operations such as data capture, editing, coding, weighting, and disclosure avoidance (see Table 1). Each operation adds to an estimate's mean squared error (Japec 2005: vi).

In general, errors can be distinguished according to whether or not the sample unit was observed. There are three major errors of nonobservation (Groves 1987: 159f):

(1) coverage error: refers to the discrepancy between sample survey results and the results of a full enumeration of the population under study which arises because some members of the population are not covered by the sampling frame[13]

(2) sampling error: discrepancies between population characteristics and those estimated from a sample survey which arise because some members of the population were excluded from the survey through selection of a subset

(3) nonresponse error: discrepancies between the population characteristics and those estimated from a sample survey which arise because some members of the sample were not measured in the survey

In addition, there exist three types of observation error due to different respondent stimuli:

(1) by the interviewer

(2) by the questionnaire

(3) by the mode of interview

Finally, post data-collection errors arise from coding, editing and imputation[14] (Groves et al. 1992, Statistics Canada 2003). In the present collection, I examine errors with respect to both nonobserved units and items[15], arising from nonresponse and sampling design. The observed errors investigated are due to the interviewer, the respondent, the survey mode, and from the imputation of missing items.

[13] "A *sampling frame* provides the means of identifying and contacting the units of the survey population … [e.g.] telephone numbers" (Statistics Canada 2003: 22).

[14] "Imputation is a process used to determine and assign replacement values to resolve problems of missing, invalid, or inconsistent data" (Statistics Canada 2003: 209).

[15] Item nonresponse occurs "if a respondent may not know or may not wish to answer a particular question during their interview" (Starick and Watson 2005: 3). Mostly, sensitive or difficult questions, like income questions are concerned (e.g. Pickery and Loosveldt 2001).

3. Complex Surveys: the longitudinal and the Cross-country Perspective and the Household Context

Before discussing data quality problems specific to “complex” surveys, I first define “complexity” as it is understood here. Survey replication over time plays an important role. *Cross-sectional* surveys are designed to observe a sample at one single point in time. *Repeated cross-sectional* surveys are conducted repeatedly, using the same design but different sample members. Such surveys allow for the investigation of population trends. One step further towards complexity is to observe the *same* sample members over several occasions (waves) in so called *longitudinal* or *panel* surveys (Statistics Canada 2003). Panel surveys are “designed to investigate short period dynamics and to relate these to longer period life course developments and to social … change” (Buck 2006: 1).

Today, “it is more apparent than ever that longitudinal analysis is crucial - not only to test life course models, but also to establish the causes of social phenomena and evaluate public policy programs” (Wagner et al. 2007: 1). Before the advent of micro data on private households, treatment of the life course was limited to theory and models (Wagner et al. 2006, Giele and Elder 1998). The oldest still existing longitudinal population representative social science household survey, the U.S. Panel Study on Income Dynamics (PSID), was commenced in 1968. It is interesting how the PSID research team went through a learning process on the nature of longitudinal data during the early stage of the project. In a pilot study the household living in the same dwelling than the initially interviewed household was re-interviewed after one year: “no attempt was made to interview the same *family* …” (Morgan and Smith 1969: 2f.). Since then, a number of social science driven general purpose household panel surveys were developed in various countries: in Europe, these include for example the German Socio-Economic Panel (GSOEP; Frick et al. 2008), which started in 1984, the British Household Panel Survey (BHPS; Lambert 2006), which started in 1991, and the Swiss Household Panel (SHP; Budowski et al. 2001), which started in 1999. The aim of these surveys is to provide high quality, timely and useful data on the dynamics of change on an individual level (Burkhauser 2006). Topics covered include working conditions, family life, leisure activities, well-being, health, attitudes and opinions, educational and occupational trajectories, gender roles, and social and economic mobility.

An additional dimension is *cross-national comparable* data. Comparability is mostly motivated by the scientific research community (Solga and Wagner 2007: 4), currently especially on the European level (Elias 2008). The benefit from comparative cross-national *panel* data is even greater because they allow for “causal inferences to be drawn based on the natural experiments sometimes created through inherent differences between institutions and countries" (Wagner et al. 2007: 1). Börsch-Supan et al. (2003) regard the Survey of Health, Ageing and Retirement (SHARE), a new, input-harmonized[16] European panel survey, as a laboratory, in which the participating countries benefit from a better common understanding of social phenomena: "the diversity in institutional histories, policies, and cultural norms, represents a unique living laboratory in which the various determinants of the current economic, health and socio-psychological conditions can be understood much easier than in the more homogeneous environment of a single country." (p. 203). Berthoud and Iacovou (2002) emphasize the importance of cross-country comparable survey data in order to identify 'best practices': "Imagine a world composed of nation-states functioning independently of one another, where scholars and policymakers were only concerned with events within the borders of their own country. Even in this sort of world, researchers should be interested in whether their findings are generalizable to human society as a whole, or to a group of similar countries …, or whether they are unique to a single country. Even in this sort of a world, we may want to learn from research in another country – but this can only be interpreted properly if the social situation in the other country is understood." (p. 5)

Increasing survey complexity, however, means greater vulnerability for data quality. "It is easier to pass a camel through the eye of a needle, than for a government statistical agency to successfully create and market an internationally comparable long-term, social-science-based longitudinal data set." (Burkhauser 2006: 1, also Lynn 2003). The failure of the European Community Household Panel (ECHP), whose goal was to create comparable panel data for all European Community countries using a common survey instrument, is a well-known example. The ECHP, abandoned in 2001, "was plagued by problems from the outset. In part, these problems arose because the ECHP was developed by Eurostat and implemented by each country's statistical agency with little or no consultation with the research community. … end users played a minor role in the creation and implementation of the survey instrument. Most troubling, the ECHP project failed to utilize the long experience of researchers who were running mature panel surveys in European Community countries" (Burkhauser 2006: 4f.)[17].

---

[16] *Input harmonization* is based on the idea of one survey centrally designed conducted in all member states under a centralized support and general coordination (Clemenceau and Museux 2008). It means that the same questionnaire, mode of data collection, field procedures, data editing and coding rules, and documentation is used (Wolf 2008). The sampling frames however may differ between countries.

[17] "EU-SILC (Statistics on Income and Living Conditions), the follow-up survey of ECHP … [has] a reduced panel component of just four waves focusing on short-term measurement of income and poverty dynamics. EU-SILC will not, however, allow the kind of in-depth life course analyses necessary for testing theoretical concepts and hypotheses in the social sciences" (Wagner et al. 2007: 2).

A more successful ex-post[18] harmonized cross-national panel study is the Cross National Equivalent File (CNEF; Frick et al. 2007). Currently, the CNEF contains nationally representative samples of the residential populations from six existing national panel surveys, with only a subset of the variables from the original panel surveys included. The panel surveys, which commenced between 1968 (PSID) and 2001 (HILDA), are the following:

1. the Panel Study of Income Dynamics (PSID) from the United States
2. the German Socio-Economic Panel (GSOEP) study
3. the British Household Panel Survey (BHPS)
4. the Survey of Labour and Income Dynamics (SLID) from Canada
5. the survey of Household, Income and Labour Dynamics in Australia (HILDA)
6. the Swiss Household Panel (SHP) survey, as the most recent CNEF member

Ex-post harmonizing survey data is a major challenge, not only because of the use of different measurement scales or substantive institutional differences, but also different data collection protocols, with possibly different effects on data quality and comparability. For example, while the rules for following the original household members (OSM)[19] are similar in most CNEF surveys, different rules apply about how other household members are followed (Spiess et al. 2008). This has consequences not only on the possibilities, for example, to study family formation or dissolution, but also on sample composition in the longer term. Further, different weighting designs must be taken into account. In addition, not only are different survey modes used by the different studies, possibly creating mode effects (Holbrock, Green and Krosnick 2003), but the mode of interview sometimes even changes *within* countries. For example, while the older CNEF surveys generally started with a paper and pencil questionnaire, they increasingly switched to computerized survey modes in the course of time. The Australian survey has increased the use of telephone interviews for households that move away from clusters of households in initial sample areas simply because of the high costs associated with following them to conduct face-to-face interviews. Also, different survey climates[20] in the participating countries may have different effects on data quality (Voogt and Saris 2005).

A further dimension of complexity in *household* surveys is that not only a single target person, but all members from a certain age on are interview eligible. This allows for the

[18] This means, given a number of target variables, original variables from different surveys are converted to the (common) target variables (Clemenceau and Museux 2008, Wolf 2008), possibly with a loss of information.
[19] OSM are in principle those observed in the first wave of data collection, plus children not (yet) age eligible.
[20] The "public willingness to participate in surveys" (Harkness 1999: 128).

analysis of individual attitudes or behaviors in both an intra- and an inter-generational context. Considering intergenerational transmissions of behavior and social structures allows for "disentangling the impacts of 'nature' and 'nurture'" (Wagner, Frick and Schupp 2007: 140f, also Lynn et al. 2006). As for the between-generation context, for example Wernli (2007) investigates the influence of parents on the political orientation of their children. Within generations, for example Homish and Leonard (2005) study the influence of the spouse on smoking behavior, and Wernli (2006) examines the voting behavior of spouses in partner households.

4. Item- and Unit-Nonresponse and Attrition

One of the most severe problems in contemporary household surveys is decreasing response rates, i.e., failure of eligible sample members to participate in the survey (Groves 2006). In panel studies, this includes not only (initial) unit nonresponse, but also attrition, which means temporary or definitive drop-out of eligible sample members who once participated in the survey. Attrition is a threat to the long term survival of a panel survey.

Both unit and item nonresponse, and attrition pose serious problems. Firstly, and obviously, the sample size is reduced, which increases the standard error[21] of estimated values. In panel surveys, an increasingly smaller sample size may make the analysis of dynamics of smaller subsamples impossible. Missing data also results in a loss of efficiency if cases that contain missing items are deleted in analyses ("listwise" deletion) or if variables containing missing items are left out while keeping the case ("pairwise" deletion). Secondly, respondents and nonrespondents may have different characteristics in the variable(s) of interest. That is, the data are not "missing completely at random" or - if the missing data are at random once controlled for *observed* variables - "missing at random".[22] This may apply both with respect to item (Little 1988) and unit (Groves 2006, Billiet et al. 2007) nonresponse. Weighting is mostly not a solution to reduce bias because respondent and nonrespondent means for the available weighting variables – usually the "sex-age" types – are not predictive of the difference between respondent and nonrespondent means for substantive variables (Peytcheva and Groves 2009).

---

[21] "The standard error is the spread of the averages around the average of averages in a *sampling distribution*" (Trochim 2006).

[22] See for the 'missing at random' concept Little and Rubin (1987). Only 'not missing at random' data cause efficiency problems.

Last but not least, nonresponse and bias lead to image problems of both the survey and the organization that conducted the survey (De Leeuw et al. 2005). That high response rates are politically important is made explicit by defining "target response rates" in project proposals and in fieldwork guidelines. For example, for both the European Social Survey (ESS; e.g., Philippens and Billiet 2004,) and the initial wave of the SHARE (Börsch-Supan and Jürges 2005), a target response rate of 70% was fixed. In the end, the target was achieved by only a few countries in both surveys (opt. cit., Billiet and Pleysier 2007, Symons et al. 2008).

Unit Nonresponse

Low, and in most countries declining (De Heer 1999, Tourangeau 2004, Fuchs 2006) response rates and high attrition are major challenges for survey research (Couper 2005, Hansen 2007a, Jäckle and Lynn 2007). Unit nonresponse occurs because either the eligible sample unit cannot be reached (non-contact), or because the sample unit refuses to participate in the survey, once contacted (non-cooperation).[23] Not contacting sample members is often due to insufficient fieldwork efforts in cross-sectional surveys (Lynn and Clarke 2002, Symons et al. 2008) or unsuccessful tracking[24] efforts in panel surveys for sample households that move or otherwise change contact information, (Couper and Ofstedal 2006). The reasons for non-contact and non-cooperation are different (Groves, Cialdini and Couper 1992, Nicholetti and Buck 1992, Tourangeau 2004, Spiess and Kroh 2008, Loosveldt and Storms 2008, Uhrig 2008).

Apart from information possibly included in the sampling frame, usually no information is available on not responding sample members. Therefore, it is difficult to calculate nonresponse bias for survey variables. In countries with central individual registers, if matching with individual sample members is possible and allowed, register information can be used to calculate bias from nonresponse (Chintex 2003, Røed 2006). In recent scientific surveys (e.g., SHARE, ESS, SHP), paradata[25] are being collected (Philippens and Billiet 2003, Couper 2005, de Luca and Lipps 2005, Blom et al. 2008, Blom and Carton 2008, Kreuter 2009). Interviewers enter information on all contact *attempts* (that is, also noncontacts) into contact forms (Carton 2007) during the recruitment of sample members to participate in the survey. This information includes not only the outcome[26] and the date and

[23] Breaking-off an interview is a special case of non-cooperation, but is relatively rare in surveys with a live interviewer (Tourangeau 2004).
[24] "locating," "tracking," and "tracing" are variously used in the literature" (Couper and Ofstedal 2006: 3).
[25] "*para*"data, "call"data, or "contact"data (Groves 2006).
[26] E.g., no contact, call-back or appointment, interview, refusal.

time of the contact attempt, but may also contain what has been said on the telephone or at the doorstep (Couper 1997, Groves and Couper 2002). In face-to-face surveys, often visible information of the sample member's house or environment is recorded (Das et al. 2005, Carton 2007). Information thus retrieved from contact forms or from the survey sample management system[27] can be used to analyze nonresponse (e.g., De Keulenaer 2005). To get an idea of the characteristics of the nonrespondents, often a "response-continuum" is hypothesized, which assumes that the more difficult it is to convince sample members to cooperate,[28] the more similar they are to sample members who belong to the "hard refusals" (Stoop 2005). The latter are people who are by no means ready to participate. Possibilities to obtain information on difficult cases are follow-up surveys: not responding units are re-approached in a special nonresponse survey after the regular fieldwork, mostly using a very short questionnaire and a stricter fieldwork protocol (e.g., Stoop 2004, 2005).

Recently, the response rate as the one and only quality criterion in cross-sectional surveys has been put into question: There is evidence that while low response rates are positively correlated with high nonresponse bias, low response rate does *not necessarily* mean a high nonresponse bias and vice versa (Groves 2006, Groves and Peytcheva 2006, 2009, Singer 2006, Groves et al. 2006).[29] Suppose, for example, one likes to measure the duration of the daily physical activity of a population, whose mean is, say 30 min. The 50% easy-to-get respondents report a mean of 30 minutes. More fieldwork effort would bring to the sample 20% points more respondents who are physically very active and therefore hard to reach, and result in a mean duration of physical activity of 40 min. The 30% (final) refusers are physically rather inactive; their inclusion would decrease the mean to the rate already achieved without the extra effort. This hypothetical example shows that investing low fieldwork effort could result in an unbiased mean, while an extra effort would bias this value. Concerning concrete measures to decrease bias, using *different* survey protocols rather than just increasing the *level* of fieldwork effort prove to be more successful (Peytchev et al. 2009).

[27] Data necessary to manage the coordination of the fieldwork, e.g. monitoring fieldwork progress, or coordinating call-backs or appointments with sample members (Hansen 2007b).
[28] Stages of difficulties to convince someone can be: ready at once – ready after discussion – reluctant, but converted by another interviewer – refusal, but take part at a subsequent nonresponse study – definitive ("hard") refusal. Also the number of contacts can be used.
[29] To measure the relationship between response rates and response biases so called R-indexes may be used (Schouten and Cobben 2007).

There are means to combat low response and especially non-cooperation. In the nineties, Groves and Couper (1998) developed a theory of survey participation, according to which the decision to participate in a survey is influenced by a number of interacting factors: the social environment, social-psychological attributes, survey design, and the interviewer. Cialdini (1984) found compliance principles which are important to explain survey participation. A factor that might help to convince someone to participate is 'authority': requests coming from a properly constituted authority are more likely complied with. The 'Consistency' principle states that a sample member is more likely to cooperate if the position, to which one commits to, is consistent with the survey request. 'Scarcity' means that surveys are more attractive if requests are seen as scarce. 'Social Validation' works if one believes similar others cooperate as well. Finally, the principle of 'Liking' states that one more likely complies if the request comes from liked others. In addition to trying to reduce the respondent burden, most effects analyzed in the literature are based on these principles.

A vast literature exists on measures undertaken to enhance response rates. First, there are (positive) effects from informing sample members of the importance of the survey by brochures and newsletters, announcing the survey request by advance letters (Singer, van Hoewyk and Maher 1998, De Leeuw et al. 2007), using different modes of data collection (Voogt and Saris 2005), offering incentives (Singer 2002, Ryu et al. 2005), timing calls (Purdon et al. 1999), or training interviewers to tailor their behavior during the introductory interactions with sample members (Morton-Williams 1993, Groves and Couper 1996, 1998, 2002, Groves et al. 2004). Although targeting modes may lead to increased response rates, using different modes in panel surveys is rare (but see Burton, Laurie and Uhrig 2008). Regarding incentives, there are in general more alternatives for survey designers. Incentives work primarily by reducing refusals and have smaller effects on non-contact rates (Singer and Kulka 2000). Money is usually more effective than gifts, and prepaid incentives are more effective than contingent incentives (Ryu et al. 2005). Singer (2002), however, concludes from a number of experiments aiming to test the effects of incentives that care must be taken when trying to generalize reported findings: "many of the findings are based on one or a few experiments, and may not be replicable over time and across survey contexts" (p. 25). Also in panel surveys the effects of incentives on survey cooperation have been investigated by numerous authors (Lengacher et al. 1995, van Hoewyk and Maher 1998, Jäckle and Lynn 2007, Laurie and Lynn 2008). Most importantly, introducing an incentive for the first time after some panel waves may have little effect on response behavior, since attrition has left a sample which is essentially fairly cooperative. On the other hand, there is evidence that increases in monetary incentives may have positive effects on response rates, especially for those who did not respond in the wave before the incentive increase (Laurie and Lynn 2008). This suggests that an incentive increase may be an effective strategy especially for reluctant respondents. Finally, there seems to be little or no carryover effect, i.e., response rates are the same in a panel wave irrespective of an incentive offered in the preceding wave.

Item Nonresponse

Analyzing data with missing values may lead to biased estimates due to selection effects. In cross-national surveys, item nonresponse biases might even be effective towards opposite directions in different countries, thus making comparability difficult or impossible (Rendtel et al. 2004). Generally, there is evidence that item-nonresponse tends to lead to underestimates of inequality measures in cross-sectional surveys, and to an underestimates of changes in longitudinal surveys (Chintex 2003). For example, with respect to income, "results on *inequality* suggest that using observed values only … produces downward biased estimates. Likewise, analyses of earnings *mobility* based only on cases with observed information significantly understate income variability over time." (Frick and Grabka 2007: 23). Imputing missing income values is a possible solution but – particularly in cross-national surveys – may be an additional source of error when data are compared. Frick and Grabka (2007), showing that the use of different techniques may increase inequality differences, state that this "confirms the importance of further harmonizing the methods used to handle missing (income) data in (panel) surveys" (24).

5. Interviewers: another Source of Error

Generally, large population representative panel surveys, and in particular the complex surveys we consider here, (still) use interviewers. Interviewers have a strong effect on data quality (van Meter 2005). For example, interviewers can influence respondents in a certain direction ("interviewer effects"). Also the correctness of the respondent's answer depends on the interviewer (Pickery and Loosveldt 1998, 2001, 2004). Indications of bad answer quality are item nonresponse (Billiet and Loosveldt 1988, Pannekoek 1988), or variations of substantive answers (Brick et al. 1995, Huddy et al. 1997, Andersen et al. 2002), especially to difficult or sensitive questions. In addition to possibly biased estimated means, such systematic variations between interviewers increase survey design effects, leading to an increased survey variance. The interviewer design effects increase linearly with the product of the interviewer clustering effects and the size of the clusters, i.e. the mean number of interviews per interviewer.[30] Heeb and Gmel (2001), for example, using a Swiss telephone survey about alcohol consumption, find an interviewer portion of the total variance of .023.[31] Given a mean number of interviews by interviewer of 70.4, this leads to a design effect of $deff_{\text{interviewer}} = 1 + .023\ (70.4 - 1) = 2.6$. Therefore the variance increase from interviewers amounts to 160 %. The message is that even small interviewer effects can produce substantial increases in variances when combined with large interviewer workloads, as is usual in telephone surveys.

As for unit nonresponse, previous studies show that interviewers may have an effect on both obtaining contact (Purdon et al. 1999, Blohm et al. 2007) and especially obtaining cooperation of sample members (Singer et al. 1983, Hox et al. 1991, Campanelli and O'Muircheartaigh 1999, O'Muircheartaigh and Campanelli 1999, Snijkers et al. 1999, Pickery et al. 2001, Pickery and Loosveldt 2002, Hansen 2007a, Blohm et al. 2007). Only a few interviewer characteristics have been found to explain this variation. Some positive effects on obtaining contact are reported from training and experience (Snijkers et al. 1999, Groves and McGonagle 2001, Pickery and Loosveldt 2002, Durand 2005, Hansen 2007a) and – in face-to-face panel surveys – interviewer continuity (Laurie et al. 1999, Spiess and Kroh 2008). Negative effects may come from a high interviewer burden (Japec 2005), however the effects are not always consistent and the extent of explained variance rather small. As for interviewer effects in general, it appears that most of the interviewer variation is independent of *observable* characteristics. Rather, the quality of the interaction of interviewers with sample members seems to be relevant (Groves and Couper 1996, 2002, Groves et al. 2004), for example the ability to tailor their behavior (Morton-Williams 1993) or to maintain the interaction with reluctant sample members (Groves and Couper 1998).

6. Modeling clustered Data

Measuring effects in data with a clustered structure is not straightforward. In clustered (for example, survey-) data, lower level measurements (e.g., from respondents) are clustered in higher level units (e.g., interviewers) and thus are not independent. Multilevel models (Snijders and Bosker 1999, Steenbergen and Jones 2002, Rasbash et al. 2004) recognize the

[30] *Design effects* from Clustering (see below) C (e.g. interviews in interviewers) lead to an increase of the variance by a factor of $deff_C = 1 + .var_C\ (numb_C - 1)$ with var_C the portion of the total variance due to C and $numb_S$ the mean number of members in the clusters (see Groves et al. 2004).

[31] In face-to-face surveys these Effects are usually higher (e.g., Hox et al. 1991, Japec 2005).

existence of data clustering by "allowing for residual components at *each* level"[32]. Ignoring clustered effects by applying ordinary least squares (OLS) techniques (like for example in Analysis of Variance (ANOVA)) to model these data would result in biased coefficients and standard errors. The reason is that clustering violates standard OLS assumptions on independence of - using the example of survey data - survey answers over and across interviewers (Goldstein 2003). I briefly explain the main difference between the standard ANOVA and the easiest (2-)multilevel model, in which only the intercept is modeled as random.[33] Suppose y is the survey variable of interest, j the interviewer, and i the respondent. Standard ANOVA assumes that y can be modeled by:

$$y_{ij} = \mu + \mu_j + \varepsilon_{ij}\,, \qquad \varepsilon_{ij} \sim N(0,\sigma_\varepsilon^2)$$

with μ the "grand mean" and μ_j interviewer j's individual deviation. If the number of interviewers is large, to estimate all μ_j would mean to lose too many degrees of freedom.[34] If in addition the set of interviewers can be considered as a smaller but representative sample of the total interviewer population, an alternative is to use a 2-level model

$$y_{ij} = \mu + [\mu_j + \varepsilon_{ij}]\,, \qquad \mu_j \sim N(0,\sigma_u^2) \quad \text{and} \quad \varepsilon_{ij} \sim N(0,\sigma_\varepsilon^2)$$

That is, the interviewer's deviation from the grand mean is taken into account by adding a normally distributed random variable. Because there is more than one residual to be estimated, OLS cannot be used, but more complicated iteration techniques have to be applied (Kenny et al. 2002).

Sometimes, the data, though clustered, are not hierarchical, but rather "cross-classified". This occurs for example in the case of a centralized telephone survey when recruiting sample members to participate. Usually one sample member may be called by different interviewers. Calls are clustered in both sample members and interviewers, but neither are sample members clustered in interviewers, nor vice versa. In addition to the first (call) level, there is *one* cross-classified second (sample members and interviewers) level. To model such cross-classified data requires appropriate modeling techniques (Browne 2005, Fielding and Goldstein 2006).

---

[32] From the website of the Centre for Multilevel Modeling at the University of Bristol, available at http://www.cmm.bristol.ac.uk/learning-training/multilevel-models/what-why.shtml

[33] In multilevel models, not only the intercept, but each regression variable can be modeled as fixed (i.e., assumed to be measured without measurement error) or random (assumed to be measured with measurement error) on the higher levels. The most important example for panel data are multilevel models for growth (Singer and Willett 2003), where time is modeled as random on the individual level, allowing slopes to vary randomly.

[34] In so called "fixed effects models" - that are very popular in econometrics – a dummy variable is modeled for each higher level unit to control for fixed "case effects". This is possible if the higher units are a small and possibly not representative subsample (from the University of Princeton "Data and Statistical Service" Website).

Summary Section I: Cross-National Surveys

Section I of this collection comprises four[35] papers on "Data Quality in Cross-National Surveys". I consider two issues in the area of data collection and one in the area of data preparation. In the first paper, interviewer data collection efficiency is considered, using paradata from the first wave of an input harmonized face-to-face cross-country panel survey (SHARE) observing the population aged 50 and over. The aim is to learn more about "best-practices" for both making contact and obtaining cooperation. The second paper deals with measuring and analyzing the effect of interviewer experience in an output harmonized repeated cross-sectional survey (ESS). Aim is to get insight in the effect of interviewer experience in the different national survey contexts. The third paper deal with item nonresponse and unit nonresponse in cross-national surveys, respectively. A best-practice approach is followed to make use of longitudinal and cross-sectional imputation procedures to impute missing income data. I describe the chosen procedures and their implementation in the SHP. In the fourth paper, I examine how different data collection procedures in different member surveys of the CNEF lead to both a different magnitude and bias of attrition. This may make comparisons difficult or necessitate complex a posteriori weighting adjustments.

7. Cross National Contact Strategies

Research Question

In the *first paper* of this section (Lipps and Benson 2005), we analyze the efficiency of interviewers with respect to both establishing contact and obtaining cooperation early during fieldwork. To make contact and to obtain cooperation as early as possible is important with regard to both reducing costs and achieving high response rates. We are especially interested in comparing the effects of countries and interviewers, but also in the identification of efficient calling strategies, to learn more about best-practice behavior. In particular we examine the efficiency of the mode (telephone or face-to-face) and of different time windows. While we expected differences in optimal contact strategies for different segments of the population – for example, for the retired compared to the fully employed – we did not necessarily expect large cross-national differences.

Data and Methods

[35] I count the three articles from Frick et al. (2007, 2008) and Lipps (2010) as one paper on the Cross National Equivalent File (CNEF). Most of my own contribution to the former two articles is described in the latter.

We use paradata from the first wave of the SHARE that employs the same design and instruments in each country to contact and interview individuals aged 50 or over. The countries with valid paradata are Austria, Denmark, Germany, Greece, Italy, the Netherlands, Spain, Sweden, and Switzerland. To assess the efficiency of making early contact at the initial call and to eventual cooperation at the initial contact, we examine first calls and first contacts on households only. The outcome variables measure success or failure of obtaining contact, or final efficiency of initial contact, respectively. We use standard hierarchical binary multilevel models with a logit link and random intercepts on the level of interviewers and countries.

Results

Both as for obtaining contact and efficiency of initial contact, by far most of the variance is due to the interviewers, rather than to countries. This holds true even though country specific differences regarding survey related issues, for example the sampling frame used and final response rates achieved, are considerable. The high interviewer variance persists even after controlling for when (day of the week and daytime) and how (face-to-face or telephone) the contact (attempt) is made, as well as the interviewer reported conditions of the environment and the house type of the sample household. Consistent with previous experiences from the U.S. and the U.K., first contact attempts on evenings, especially on Sundays, seem to be more efficient. As for sample members who finally participate, working day afternoons by face-to-face, working day evenings irrespective of the mode used, and Saturday evenings by telephone are favorable time windows. Sunday and Saturday morning first contacts by face-to-face however appear to generate annoyance. The significance of the time a first contact is made cannot be confirmed by other studies and might be due to the SHARE sample consisting of people aged 50 and over.

Contribution

My contribution to this paper was twofold:

1.) Preparation of the paradata that was retrieved from the SHARE - sample management system (Das et al. 2005, De Luca and Lipps 2005). The final data contains all relevant information together with the final disposition codes[36] of the sample members.

2.) Review of the literature, design and analysis of the multilevel models, and discussion of the model results.

[36] Basically no contact, refusal, no sample case, and cooperation (AAPOR 2009).

8. Effects of Interviewer Experience on Components of Nonresponse in the European Social Survey

Research Question

In the *second paper* of section I (Lipps and Pollien 2010) we analyze effects of interviewer experience on three nonresponse components in face-to-face surveys: ultimate noncontact of households, ultimate noncontact of target persons given household contact and ultimate noncooperation of target persons given target person contact. Experience is one of the very few variables that turns out to be significant to explain interviewer effects in most research articles. In addition, it is widely accepted that country specific differences play a role. Following the literature, we first hypothesize that experienced interviewers are more successful in both establishing contact and obtaining cooperation. Second, we expect a decrease of interviewer performance heterogeneity with experience in that there is a tendency toward conformity with colleagues and agency standards over time.

Data and Methods

To be able to take into account cross-national differences we use data from the European Social Survey (ESS). We define two measures of experience: interviewers who conduct refusal conversion calls, and previous experience working with the ESS.[37] The latter measure restricts the possible countries to those that used the same survey agency over rounds, which in turn used the same interviewer IDs for the interviewers: Switzerland, Germany, Denmark, and Finland. In addition, we have to exclude round 1 (2002), which leaves us with data from rounds 2 (2004) and 3 (2006). To analyze noncontact of households, we use the interviewers' first calls on households, for noncontact with target persons, we use the interviewers' first contacts with households, and for obtaining cooperation of target persons, we use the interviewers' first contacts with target persons. To model interviewer experience effects, we use 2-level logit models, with households/target persons the first, and interviewers the second level. We analyze performance by comparing the interviewers based on previous ESS experience (yes/no) as well as on refusal converter (yes/no) status. Interviewer homogeneity of the different experience groups is examined by comparing the interviewer specific portion of the total variance ("intraclass correlation coefficient"). We build models for each nonresponse component separately by each country, and also pool all four countries.

Results

[37] While the first measure is objective, derived from the contact data, the second is subjective and at the discretion of the survey agency.

Generally, we find positive effects from interviewer experience and refusal converters whenever there is potential to improve a nonresponse component. For example, in Switzerland where a household sampling frame is used, the household must be screened to identify the target person first. This needs additional contacts with the danger of higher noncooperation. Here, interviewers with previous ESS experience perform better. Overall, although refusal converters perform better than 'ordinary' interviewers, those with previous ESS experience perform surprisingly well. This is surprising since refusal converters should be the best interviewers available to the survey agency. It might however be the case that refusal converters are especially trained for the refusal conversion phase, and might not perform better during the normal fieldwork phase that is considered here. We conclude with the recommendation to survey agencies to try to decrease the (still high) turn-over of interviewers, and possibly train novice and experienced interviewers in a different way.

Contribution

The majority of the contribution was mine with the exception of communication with the survey firms about the following of interviewers in the ESS over time, and the literature review on interviewer homogeneity.

9. The Cross-National Equivalent File (CNEF) and its Member Country Household Panel Studies

The CNEF and the SHP

The *third paper* of section I (Frick, Jenkins, Lillard, Lipps and Wooden 2007, 2008, Lipps 2010) is descriptive and presents key issues of the CNEF, in particular the income imputation in the SHP. To extend the small Swiss social science research community analyzing the SHP and to improve international scientific exchange, SHP joined the CNEF in 2006.[38] New CNEF members must be able to construct a sufficient part of the CNEF variables and provide complete, i.e., imputed income data. Unlike the other CNEF surveys, the SHP is not primarily designed to serve the socio-*economic* research community. Therefore both tasks are a major challenge for the SHP.

CNEF Income Components in the SHP

On the household level, income variables to be delivered are the total pre- and post tax incomes, aggregated over all individuals in the household. Single income components to be

[38] Information on the SHP-CNEF variables can be found in the 1999-2007 codebook (Lipps and Kuhn 2009).

delivered separately are income from labor, from assets, from private and public transfers, and from social security pensions. Taxes to be calculated are the total taxes, federal taxes, and social security taxes. On the individual level, labor earnings must be delivered. To construct CNEF-compatible variables, first suitable aggregates have to be constructed from the information available in the SHP (Kuhn 2008). Income components of *unit non-responding* individuals are only imputed if it is likely that income from the source considered was earned.

Income Imputation

To impute missing income data, I take advantage of the longitudinal character of the data. Similar to methods used by Frick and Grabka (2007) to impute missing SOEP income data, and Starick and Watson (2007) to impute missing HILDA income data, I mainly use a modified "row-and-column" imputation technique (Little and Su 1989). This method takes longitudinal as well as cross-sectional information into account. Imputed values are the product of the "row effect", the "column effect", and a residual effect. The column effects are calculated as the standardized mean income in each wave and represent the yearly inflation. The row effects are computed for each sample member and represent the mean individual income. Sorting by the row effect, borrowing the income from the nearest reporting neighbor, and multiplying by the column effect yields the predicted individual mean. This value is then multiplied with the stochastic component of the imputation from the complete case with the closest row effect, to give the imputed value. I distinguish the recipients and donors by education classes.[39] For income components that are stable over waves, I impute with the "carryover" method, i.e., I take the reported value from another wave of the individual in question, starting with the closest wave. It must be noted that both longitudinal approaches (Little & Su, Carryover) fail if no income information is available for a given case in either wave. In these cases, the missing income value must first be "initialized" based on cross-sectional information. I use a stochastic imputation method which is based on a set of linear regressions. If a value cannot be imputed in one of the waves due to missing predictors the regression is repeated by stepwise reducing the predictors, until no predictor is missing and the regression yields an imputed value.[40]

Contribution

[39] Frick and Grabka (2004) propose age groups, but I find that education is a stronger discriminator for income.
[40] In addition, I run regressions from both the earliest to the latest wave (left-right chain) and from the latest to the earliest wave (right-left chain). The final imputed value is taken from the left-right chain, if the regression has a higher explanation power than the regression from the right-left chain, and vice-versa.

My main contribution to this research was to propose the CNEF integration of the SHP, to initialize contact with the CNEF group in Cornell, and to design, program and conduct all missing income imputation in the SHP as described in Lipps (2010). I also constructed all socio-demographic CNEF variables and documented the SHP-CNEF codebook (Lipps and Kuhn 2009). For the Frick et al. (2007, 2008) papers, I contributed the SHP part.

10. Attrition of Households and Individuals in Panel Surveys

Research Questions

The *fourth paper* of section I (Lipps 2009a) investigates attrition in selected CNEF member data, where all household members are interviewed. First, I extend the concept of attrition, which is usually only applicable for individuals, to households, by using the previous' wave household reference person as the household "anchor". Second I analyze attrition on both the household and the individual level, with the latter conditional on household participation. Specifically I analyze whether different survey procedures cause unequal attrition of the individuals in multi-person households. This would be a serious threat to conducting cross-national intra- and intergenerational comparisons.

Data and Results

Using data from the second through the ninth wave of the (face-to-face) BHPS, the (face-to-face) GSOEP, and the (telephone) SHP, I find that the attrition patterns are different across surveys both on the household and the individual level. While on the household level, the BHPS exhibits very small attrition, on the individual level, the SHP has the highest attrition rates. The latter can be expected because 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 appointments) are necessary to work a household, which increases the possibilities for individuals to drop out.

Using duration models, I find generally smaller attrition selection effects 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. Selectivity on the household level is never compensated for on the subsequent individual level: I 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 distinguishing household and individual level attrition separately, especially for comparison analyses across different surveys. Differences that are related to different survey procedures (for example, different survey modes) need to be taken into account.

Recommendations

In order to reduce bias in future waves, 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 analyze household and individual attrition effects on other than socio-demographic variables. It would be interesting to consider socio-economic, attitudinal or satisfaction ("well-being") variables, or variables measuring social activities which all have causal effects on nonresponse.

Summary Section II: Attrition and Incentives in the SHP

The second section of the studies consists of two papers that focus on unit nonresponse, attrition, and incentive effects to fight attrition. The SHP is not an exception to the typically low response rates in Switzerland (for example, Zimmermann and Joye 2003, Billiet and Pleysier 2007). The first paper of this section analyzes attrition in the SHP, which is comparatively high on both the household and the individual level. I find that individual characteristics of individuals who drop out tend to be similar in many respects across waves. This is indicative of an attrition bias which is cumulative over waves. In the second paper, I report on an incentive experiment conducted in the 2006 wave of the SHP that aimed to study attrition on both the household and the individual level. On both levels I find surprisingly small effects from the incentives tested on response rates, and the expected small sample selection effects. There are however some positive cost saving effects with regard to fieldwork efforts.

11. Attrition in the Swiss Household Panel

Specifics of the SHP for the Attrition Analysis

In the *first paper* of section II (Lipps 2007a) I analyze both initial wave (1999) non-response and subsequent attrition in the SHP, analyzing first wave (1999) respondents until 2005. From the sampling frame used for the SHP, only residential geography is known (ZuWallack et al. 2004), leaving few variables to analyze initial wave nonresponse. The SHP survey asks household information of the household reference person before each individual aged 14 and above is asked to complete the individual questionnaire. Thus, in addition to survey data from previous waves, information from the household stage can be used to analyze characteristics of individual drop-outs in the same wave.

Initial Nonresponse

The basis for the initial wave nonresponse analysis is the gross sample, that is, randomly drawn telephone numbers minus not eligible cases (AAPOR 2009). I find an initial household response rate of 49% (i.e. households with at least one responding household member), a noncontact rate of about 6%, and a 45% refusal rate. The individual response rate, conditional on a completed grid questionnaire, amounts to 83%. Compared with the population census, foreign households, males, and younger people tend to be underrepresented.

Attrition

Generally, I find that the characteristics of attritors are similar over waves, i.e., selection effects cumulate rather than balance out over time. The high attrition rate in the SHP is especially pronounced in the younger age group: For example, while the ratio of the proportion of all 15-24 years old in the first wave to the proportion of those among them who stay in the survey until (at least) the fifth wave amounts to 1.45 in the SHP, this value amounts to 1.05 in the BHPS. In this age group, noncontact is a major reason for attrition. This indicates that there are problems to track mobile young people, presumably moving out of the parental home. In the longer run, especially attrition of young people is a problem for long-term survival of the panel. This is because it is particularly this group that forms new households with the new household members, mostly the partner, and eventually own children, who in turn become new members of the panel, help to maintain representativeness and the sample size.

Apart from younger people, attrition bias is not particularly high in the SHP. Generally, attrition is higher in person groups who are difficult to contact (have the intention to move) or tend to be socially excluded (for example, foreigners, people who have little interest in society and politics, people dissatisfied with their living conditions). Also poor answer quality in previous waves, for example an overuse of extreme answer categories on subjective questions, is more often followed by a panel drop-out. These results are in line with findings from other panel surveys, and confirm the theories of 'social isolation' and – related to it – 'topic interest' (Groves et al. 2004).

Recommendation

Based on the results of the attrition analysis, I recommended introducing (monetary equivalent) incentives in addition to the annual newsletter of the SHP. Results from an incentive experiment are described in the next paper of this section.

12. Incentive Effects on Household and Individual Behavior in Telephone Panel Surveys

Continued high attrition on both the household and the individual level in the SHP led to several new measures in the 2006 wave (MIS-Trend 2007), amongst others an incentive experiment, of which some specific results are discussed in this paper.

The Incentive Experiment Design

In the *second paper* of this section (Lipps 2010a), I analyze cooperation effects from a randomized incentive experiment. Incentives should preferably be effective on both response stages: when the reference person is asked to answer the household grid (1st response stage), and when eligible individuals are asked to answer the individual questionnaire (2nd response stage). The SHP tested three mid-sized incentives on households that cooperated in the previous wave: for each random quarter of the sample households, the following incentives were announced in the advance letter:

- 12 stamps (à 1 Sfr.) announced and unconditionally sent with the advance letter
- a lottery participation with prizes of 5,000 Sfr., 3,000 Sfr., and 2,000 Sfr, conditional on individual participation
- a donation of 10 Sfr. to a charity institution to be chosen from a list, also conditional on individual participation

In addition, a no incentive ('control') group is used

Hypothesized Incentive Effects

1.) a.) Higher household level cooperation in the stamps group, and b.) an insignificant composition effect

2.) a.) Conditional on household (grid questionnaire) participation, higher individual participation rates in the conditional incentive groups, and b.) different composition effects depending on the incentive used

3.) Reduced fieldwork efforts from incentives on both levels

Results and a Consequence

As for answering the household grid, there are only small effects depending on household size and maturity (duration of stay), with stamps performing slightly better. Relative to the control group, there are neither sample composition nor reduced fieldwork effort effects in any of the incentive groups. Given the household grid is filled, there are also slightly higher individual participation rates in the stamps group, and some weak sample composition effects. Finally as for cost saving effects, relative to the control group, all incentive groups appear to significantly shorten the fieldwork duration.

It is likely that the size of the incentives is too small to induce effects in a mature, rather cooperative household panel. In the literature there is evidence that incentives are effective to increase response in particular among households with low response propensities (Rodgers 2002). The SHP decided to send a 50 Swiss Francs voucher with the advance letter to previously refusing households from 2007 on. Results suggest that completion rates among these households seem to be higher, when compared with similar households in former waves.

Summary Section III: Interviewer Effects in central CATI Panels

While the papers in the first two sections of the studies deal with survey quality aspects based on information from sample members, the three papers of the third section include effects from interviewers who try to obtain cooperation, in the special case of a centralized telephone panel survey. Unlike in face-to-face or decentralized telephone surveys, the advantage of centralized telephone surveys for analyzing interviewer effects is that usually all contacts of interviewers with sample members are at random. This guarantees an experimental design that can be tested for design efficiency and possible improvements.

The first paper presents new measures to assess interviewer performance for obtaining cooperation. The new measures are superior to existing ones, because they take into account all contacts of interviewers on a sample case, and not only the first contact. The second paper examines whether cooperation rates can be increased by reassigning better interviewers to particularly critical contacts with sample cases. I show that first contacts are subject to interviewer effects, especially when reference persons are contacted. The last paper of these studies investigates respondent and interviewer effects on substantive question response behavior, both in a cross-sectional and longitudinal way. I find that both interviewer and time effects are different depending on the answer quality measure considered.

All three papers involve clustered data with three levels: respondents and interviewers are at the higher level, and contacts (first two papers) or waves (last paper) at the lower level. The two higher levels are cross-classified (Figure 1 and Figure 2).

13. A Note on Interviewer Performance Measures in centralized CATI Surveys

The *first paper* of this section (Lipps 2008) presents and discusses three new methods to measure interviewer performance to try to obtain cooperation from sample members in a centralized telephone survey. I describe two of the methods here. Besides substantive aspects, the aim of the paper is to demonstrate the potential of paradata for investigating quality issues.

Problem Description

Unlike in face-to-face surveys, in centralized telephone surveys, contact attempts to convince sample members to participate are not assigned to one interviewer only. Instead, each call is assigned to interviewers at random. Because several interviewers may work one sample case, success or failure to convince a sample member to cooperate cannot be simply attributed to one interviewer. The assignment of calls and contacts to interviewers in face-to-face surveys vs. centralized telephone surveys is schematized in the figures that follow:

First in face-to-face surveys:

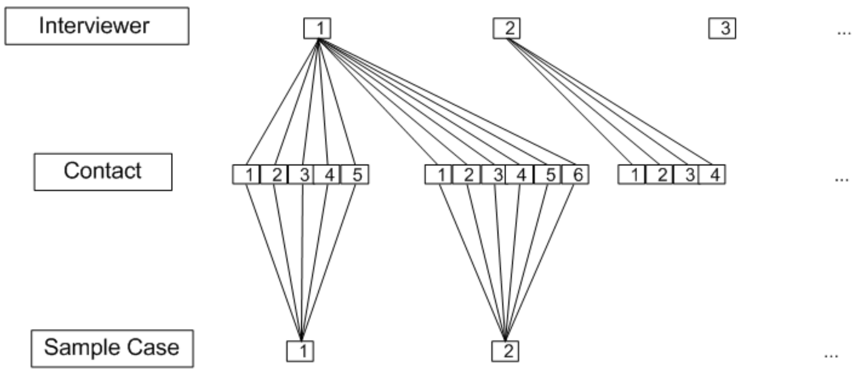


Figure 1: Interviewer-Sample Case Assignments via Contacts in face-to-face Surveys.

And, in centralized telephone surveys:

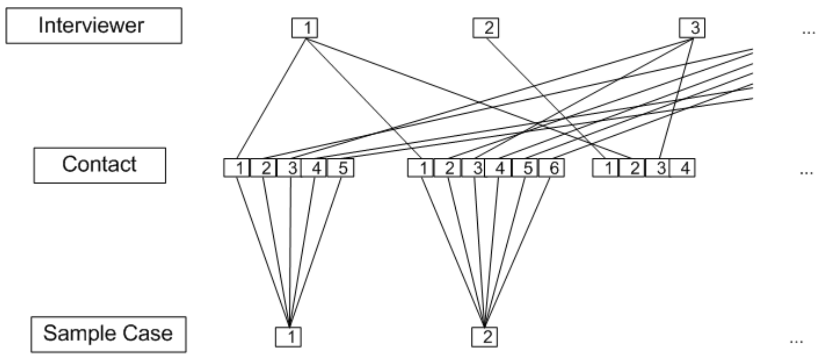


Figure 2: Interviewer-Sample Case Assignments via Contacts in Surveys with Random Assignment.

To avoid contaminating the measure with the performance of a previous interviewer, to measure interviewer performance in central telephone surveys usually only first contacts with sample cases are used. However, interviewers working later contacts on a sample case also contribute to whether or not the case finally cooperates. In addition, only using first contacts reduces the sample of interviewers examined. In this paper all contacts are taken into account.

Measures proposed and Data Analysis

For the first measure, a contact is defined as successful if the sample member *eventually* cooperates, irrespective of the specific contact outcome. In the second measure, a contact is considered successful if the outcome of this contact is *not a refusal*. In the analyses, to account for different effects an interviewer can have, contacts are weighted according to the interviewer proportion of the total variance (interviewer intra-class-correlation). The number of the contact is a good discriminator for the interviewer effect. I use paradata from the 2004 and 2005 waves of the SHP and the Swiss pilot of the Statistics on Income and Living Conditions (CH-SILC).

Findings

The benefit of these measures is that interviewer performance can be assessed in a more comprehensive and equitable way by including *all* contacts. While the first measure appears to measure interviewer effects more realistically, the second measure is available much earlier during fieldwork. The latter therefore allows to quickly react on interviewer failures while "to [create] the opportunity to alter the design during the course of [process] data collection" (Groves and Heeringa 2006: 439). This is in the sense of a 'responsive design' (opt. cit.).

14. Cooperation in centralised CATI Household Panel Surveys- A Contact-based Multilevel Analysis to Examine Interviewer, Respondent, and Fieldwork Process Effects

The second paper of the third section (Lipps 2009) deals with optimization possibilities to assign special interviewers to single contacts in order to increase cooperation. I take interviewer and respondent characteristics, and the fieldwork process into account.

Central Idea

The interpenetrated design in centralized telephone surveys allows analyzing if a re-assignment of interviewers to specific contacts is able to reduce noncooperation. The idea is to use better interviewers for contacts that imply high interviewer effects.

Data and Assessment of Single Contact Results

I use paradata from the Swiss Household Panel (SHP) and the second wave of the Swiss part of the Survey on Income and Living Conditions (CH-SILC) pilot study from 2005, both conducted by the same fieldwork agency. I consider those response stages (cases) in the surveys which are most "critical" with respect to attrition: first when the household reference person is asked to complete the household grid questionnaire, and second when eligible individuals "other" than the reference person are asked to complete their individual questionnaire. In addition, I distinguish first and later contacts.

Unlike in the previous paper, to measure performance of each contact, a refusal gives a "0" and a completed interview a "1". For the intermediate contact outcomes, like vague (without) or fixed appointments (with a fixed date and time), the mean of the *final* cooperation probabilities is taken, by case type (reference person / other persons), first / later contacts, and contact phase (regular / refusal conversion).

Results

I find that the interviewer effects are highest for first contacts, especially when reference persons are contacted. For later contacts, the interviewer share of the total variation is negligible. As expected, contact performance during the refusal conversion phase is worse; however better for later contacts with reference persons. Contrary to existing research, even if the result of the previous contact is controlled for, socio-demography and previous response propensity are still important in later contacts especially for reference persons. Later contact performance with reference persons becomes increasingly worse with the number of contacts, while later contact performance with other persons improves. The result of the previous contact is much more decisive for cooperation of reference persons than of other persons.

Fieldwork Recommendations

More effort (i.e., both with respect to interviewer persuasion skills and speed) should be put in contacts with reference persons, and particularly first contacts with them. On the other hand, as for other persons, the principle of maintaining interaction appears to be more important. It would also be interesting to use other contact performance measures, such as those of the previous paper.

15.Interviewer and Respondent Survey Quality Effects in a CATI Panel

Research Question

In the *third paper* of section III (Lipps 2007) I examine interviewer and respondent fixed and especially random effects on (substantive) response quality, both in a cross-sectional and – a novelty – in a longitudinal way. Specifically I investigate three response indicators, susceptible to interviewer effects:

1.) Satisficing, meaning not expending the cognitive effort necessary to give an *accurate* response (Krosnick 1991, Pickery and Loosveldt 1998, 2001). This is quantified by the proportion of extreme value answers on subjective questions,
2.) Giving socially desirable answers, quantified by the proportion of (strong) agreement with positive statements on selected political questions, and
3.) Not reporting income.

In the cross-sectional analysis I try to identify fixed interviewer effects that possibly explain variance on the interviewer level. The focus of the longitudinal analysis is on possible interviewer learning effects and whether a specific respondent answering behavior is a respondent trait that is stable over time or rather depends on situational factors. The second objective of the longitudinal analysis is to assess the portion of variance on each of the levels involved (interviewer, respondent, and wave).

Cross-sectional Analysis

For the cross-sectional models I use data from the SHP 2004 wave, in which an interviewer survey on their background, experience, and attitudes was conducted. Using a 2-level hierarchical multilevel modeling approach, I find very small interviewer random effects for social desirability, a medium effect for satisficing, and a high effect for income nonresponse. Interviewer characteristics do not play a role except for experience that plays a positive role on reporting income. There is no respondent attrition effect. Similarly, the interviewer-respondent matching variables sex, age, and education have no effect, once the respondent main effects are controlled. This is consistent with the theory that it is rather the quality of the interaction with sample members that is relevant for response quality. I find within-wave 'late case' effects (Kennickell 2000).

Longitudinal Analysis

Based on the SHP 2000-2005 original sample, I find both fixed personal traits and variation on the level of the respondent, depending on the indicator analyzed: while satisficing and especially not reporting income appears to be a fixed personal trait, giving socially desirable answers varies over time. Also in the longitudinal models, I do not find an effect of respondent attrition on data quality reported.

Background: Closing Remarks

I started this background paper with a brief historical overview of survey methodology, which is characterized by a steady increase in the complexity of both survey design and methods of analysis. More complex surveys mean greater challenges to data quality. This has been acknowledged by the scientific community, which is more frequently being supported by national and international research foundations and funding agencies, either through the establishment of research institutions or direct funding of survey quality related projects.

The studies presented here deal with data quality issues in complex interviewer mediated surveys, most of which are longitudinal, and most of which are conducted across different countries. I mainly analyze item and unit nonresponse and attrition problems from both a cross-national and a longitudinal perspective. In addition, interviewer effects on obtaining contact and cooperation of respondents, and on survey answer quality are examined.

To find innovative ways to improve survey data quality is necessary, given decreasing budgets for surveys, oversurveying as an increasing threat to data quality of social surveys, and data protection laws increasingly prohibiting the use of more informative sampling frames and interesting context data. All this necessitates more sophisticated survey designs, especially taking greater advantage of new and creative modeling possibilities, mixing survey modes, using responsive designs, available paradata, and the experiences from surveys conducted in other countries.

Referring back to the schema in Table 1, one might now have a better idea of some of the main data quality problems in complex social surveys and possible ways to analyze them. This implies at the same time that we are far from knowing how to design and conduct an error-free survey. To the contrary, new problems and challenges are generating new avenues of research. The present collection of papers does not aim to completely fill any of the gaps in understanding on the methodological research front. Rather, it shows how single focused contributions are needed to increase understanding in a stepwise fashion. Indeed, I would argue that under current circumstances new advances in survey methodology must occur broadly in such small steps. Further, I am well aware that in doing so I am not only compounding ever more research in the field, but am also opening new research areas to be explored.

Background: References

16. Own Contributions

Frick, J., S. Jenkins, D. Lillard, O. Lipps and M. Wooden (2007) The Cross-National Equivalent File (CNEF) and its Member Country Household Panel Studies. *Journal of Applied Social Science Studies* (Schmollers Jahrbuch) **127** (4): 627-654, accessed on March 16 2010 at: http://www.diw-berlin.de/documents/dokumentenarchiv/17/77260/schmoller_Frick_etal_2007.pdf

Frick, J., S. Jenkins, D. Lillard, O. Lipps and M. Wooden (2008) Die internationale Einbettung des Sozio-oekonomischen Panels (SOEP) im Rahmen des Cross-National Equivalent File (CNEF). In: Frick, J., O. Groh-Samberg, J. Schupp und K. Spieß (Hrsg.) 25 Wellen Sozio-oekonomisches Panel, *Vierteljahrshefte zur Wirtschaftsforschung*, DIW Berlin , 77 (3), Berlin: Duncker & Humblot: 110-129

Lipps, O. and G. Benson (2005) Cross National Contact Strategies. *Proceedings of the American Statistical Association, Section on Survey Research Methods*. Alexandria, VA: American Statistical Association: 3905–3914, accessed on October 3 2008 at http://www.amstat.org/sections/SRMS/proceedings/y2005/Files/JSM2005-000549.pdf presented at the 60th annual AAPOR Conference, Miami Beach, USA, May 2005

Lipps, O. (2007) Interviewer and Respondent Survey Quality Effects in a CATI Panel. *Bulletin de Méthodologie Sociologique* **95**: 5-25, presented at the second conference of the European Survey Research Association (ESRA), Prague, Czech Republic, June 2007: http://esra2007.vse.cz/home/

Lipps, O. (2007a) Attrition in the Swiss Household Panel. *methoden – daten – analysen* **1 (**1): 45-68, presented at the 3rd International Conference of Panel Data users in Switzerland, Neuchâtel, March 2006, accessed on March 16 2010 at: http://www.swisspanel.ch/net/conference.php?lang=en&pid=57

Lipps, O. (2008) A Note on Interviewer Performance Measures in centralised CATI Surveys. *Survey Research Methods* **2** (2): 61-73, presented at the 19th International Workshop on Household Survey Nonresponse, Ljubljana, September 2008: http://www.nonresponse.org/

Lipps, O. (2009) Cooperation in centralised CATI Household Panel Surveys- A Contact-based Multilevel Analysis to Examine Interviewer, Respondent, and Fieldwork Process Effects. *Journal of Official Statistics* **25** (3): 323-338, presented at the 20th International Workshop on Household Survey Nonresponse, Lausanne, September 2009: http://www.nonresponse.org/

Lipps, O. (2009a) Attrition of Households and Individuals in Panel Surveys. *SOEPpapers* **164***,* accessed on March 16 2010 at: http://www.diw.de/documents/publikationen/73/diw_01.c.96125.de/diw_sp0164.pdf presented at the International Conference on Survey Methods in Multinational, Multiregional, and Multicultural Contexts (3MC), Berlin, July 2008: http://www.3mc2008.de/

Lipps, O. and U. Kuhn (2009) Codebook for CNEF variables in the SHP (1999 - 2007). *Discussion paper 5-09,* Swiss Household Panel, Lausanne, accessed on January, 20 2010 at: http://aresoas.unil.ch/workingpapers/WP5_09.pdf

Lipps, O. (2010) Income Imputation in the Swiss Household Panel 1999-2007. *FORS Working Paper Series* 2010-1. FORS, Lausanne, accessed on March 16 2010 at: http://www2.unil.ch/fors/IMG/pdf/FORS_WPS_2010-01_Lipps-3.pdf

Lipps, O. (2010a) Effects of Different Incentives on Attrition and Fieldwork Effort in Telephone Household Panel Surveys. *Survey Research Methods* **4** (2): 81-90, presented at the 18th International Workshop on Household Survey Nonresponse, Southampton, September 2007

Lipps, O. and A. Pollien (2010) Effects of Interviewer Experience on Components of Nonresponse in the European Social Survey. To appear in *Field Methods,* presented to the FORS Scientific Board, and at the the Seventh International Workshop on Comparative Survey Design and Implementation (CSDI), Lausanne, March 2010

17. Own related work

Lipps, O. (2010b) Does Interviewer-Respondent Socio-demographic Matching increase Cooperation in centralized CATI Household Panels? *Survey Practice* (August 2010)

Lipps, O. (2010c) A Note on improving Contact Times in Panel Surveys. To appear in *Field Methods*

Lipps, O. and K. Kissau (2010) Nonresponse in an Individual Register Sample Telephone Survey in Lucerne/Switzerland. To appear in Häder, M., S. Häder, and M. Kühne (eds.): Telephone Surveys in Europe

Lipps, O. and G. Lutz (2010) How Answers on Political Attitudes are shaped by Interviewers: Evidence from a Panel Survey. Special Issue (Lipps, O., R. Tillmann, U. Kuhn, D. Lillard, and M. Bergman (eds.): "Social Science Research with Panel Data in Switzerland") of the *Swiss Journal of Sociology* 2/2010: 345-358

18. Reviewed Literature

AAPOR (The American Association for Public Opinion Research) (2009) *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 6th edition*. Lenexa, Kansas: AAPOR, Accessed on October, 22 2009 at: http://www.aapor.org/AM/Template.cfm?Section=Standard_Definitions1&Template=/CM/ContentDisplay.cfm&ContentID=1814

Alwin, D. and J. Krosnick (1985) The Measurement of Values in Surveys: A Comparison of Ratings and Rankings. *Public Opinion Quarterly* **49** (4): 535-552

Andersen, A.-M. and J. Olsen (2002). Do Interviewers' Health Beliefs and Habits Modify Responses to Sensitive Questions? A Study using Data Collected from Pregnant Women by Means of Computer-assisted Telephone Interviews. *American Journal of Epidemiology* **155** (1): 95-100

Becker, R., R. Imhof and G. Mehlkop (2007) Die Wirkung monetärer Anreize auf den Rücklauf bei einer postalischen Befragung und die Antworten auf Fragen zur Delinquenz (Effects of prepaid monetary incentives on the return of mail survey and selfreporting about delinquency). *Methoden, Daten, Analysen* **1** (2): 131-15

Bergman, M. and A. Coxon (2005) The Quality in Qualitative Methods. *Forum: Qualitative Social Research* **6** (2), Art. 34

Berthoud, R. and M. Iacovou (2002) Diverse Europe - Mapping Patterns of Social Change across the EU. Institute for Social and Economic Research, University of Essex

Biemer, P. and L. Lyberg (2003) *Introduction to Survey Quality*. Wiley, New York

Billiet, J. and S. Pleysier (2007) Response Based Quality Assessment in the ESS - Round 2. An update for 26 countries. Center of Sociological Research. University of Leuven

Billiet, J. and G. Loosveldt (1988) Improvement of the Quality of Responses to Factual Survey Questions by Interviewer Training. *Public Opinion Quarterly* **52** (2): 190-211

Billiet, J., M. Philippens, R. Fitzgerald and I. Stoop (2007) Estimation of Nonresponse Bias in the European Social Survey: Using Information from Reluctant Respondents. *Journal of Official Statistics* **23** (2): 135-162

Blohm M., J. Hox and A. Koch (2007) The Influence of Interviewers' Contact Behavior on the Contact and Cooperation Rate in face-to-face Household Surveys. *International Journal of Public Opinion Research* **19** (1): 97-111

Blom, A. and A. Carton (2008) Introduction Special Interest Session Paradata. 19th International Workshop on Household Survey Nonresponse, Ljubljana (Slovenia), September 15-17, 2008. Unpublished Manuscript

Blom, A., P. Lynn and A. Jäckle (2008) Understanding Cross-National Differences in Unit Non-Response: The Role of Contact Data. *ISER Working Paper* 1/2008. Colchester: University of Essex

Börsch-Supan, A., A. Brugiavini, H. Jürges, J. Mackenbach, J. Siegrist and G. Weber (eds.) (2005) *Health, Ageing and Retirement in Europe – First results from the Survey of Health, Ageing and Retirement in Europe*. Mannheim: Mannheim Research Institute for the Economics of Aging

Börsch-Supan, A. and H. Jürges (eds.) (2005) *The Survey of Health, Ageing and Retirement in Europe - Methodology*. Mannheim: Mannheim Research Institute for the Economics of Aging

Börsch-Supan, A., H. Jürges and O. Lipps (2003) SHARE – Measuring the ageing Process in Europe. *ZUMA Nachrichten* **53**: 96-113

Brick, J., R. McGuinness, S. Lapham, M. Cahalan, D. Owends and L. Gray (1995) *Interviewer Variance in two Telephone Surveys*. SESTAT Methodological Research Compendium, WESTAT, Accessed on October, 3 2008. available at: http://srsstats.sbe.nsf.gov/

Buck, N. (2006) General Household Panel Studies: The British Household Panel Survey. Unpublished Presentation for the Conference on Longitudinal Surveys in International Perspective, Montreal, 25-27 January 2006

Budowski, M., R. Tillmann, E. Zimmermann, B. Wernli, A. Scherpenzeel and A. Gabadinho (2001) The Swiss Household Panel 1999-2003: Data for Research on micro-social Change. *ZUMA Nachrichten* **49**: 100-125

Burkhauser, R. (2006) Creating Internationally Comparable All-Age or Older-Age Cohort Long-Term, Social-Science-Based Longitudinal Data Sets in Canada. Unpublished Presentation for the Conference on Longitudinal Surveys in International Perspective, Montreal, 25-27 January 2006

Burton, J., H. Laurie and N. Uhrig (eds.) (2008) Understanding Society. Some preliminary Results from the Wave 1 Innovation Panel. Institute for Social and Economic Research, University of Essex

Campanelli, P. and C. O'Muircheartaigh (1999) Interviewers, Interviewer Continuity, and Panel Survey Nonresponse. *Quality & Quantity* **33**: 59-76

Carton, A. (2007) Using paradata to cope with survey nonresponse. Unpublished paper prepared for the 18th Workshop on Nonresponse, Sept 3-5, 2007, Southampton

Chintex (2003) CHINTEX (The Change from Input Harmonisation to Ex-post Harmonisation in National Samples of the European Community Household Panel – Implications on Data Quality) - Synopsis. Accessed on October, 3 2008 at: http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Wissenschaftsforum/Chintex/ResearchResults/Downloads/Synopsis,property=file.pdf

Cialdini, R. (1984) *Influence: The new Psychology of modern Persuasion*. Quill, New York

Clemenceau, A. and J.-M. Museaux (2008) Harmonisation of household surveys from an Official Statistics Perspective: experience gained from HBS, ECHP and EU-SILC. Unpublished Presentation for the 3MC Conference Berlin, Germany, June 25-28, 2008

Clery, D. (2006) Panel draws up shopping List. *Science* 20OCT2006 **314** (5798): 399

Couper, M. (1997) Survey Introductions and Data Quality. *Public Opinion Quarterly*, **61** (2): 317-338

Couper, M. (2005) Technology Trends in Survey Data Collection. *Social Science Computer Review* **23** (4): 486 - 501

Couper, M. and M. B. Ofstedal (2006) Keeping in Contact with Mobile Sample Members. Unpublished Presentation for the Methodology of Longitudinal Surveys (MOLS) conference, University of Essex, Colchester, UK, July 12-14, 2006

Das, M., C. Vis and B. Weerman (2005) Instruments: LMU, CAPI, Drop.off, and CMS. In: A. Börsch-Supan et al. (eds.), *Health, Ageing and Retirement in Europe: First Results of the Survey of Health, Ageing and Retirement in Europe,* MEA: Mannheim: 351f.

De Heer, W. (1999) International Response Trends: Results of an International Studies. *Journal of Official Statistics*, **15** (2): 129-142

De Keulenaer (2005) Using Process Data to Predict Attrition from a Panel Survey: a Case Study. *Proceedings of the Survey Research Methods Section, American Statistical Association.* Accessed on December 3, 2008 at: http://www.amstat.org/sections/SRMS/proceedings/y2005/Files/JSM2005-000426.pdf

De Leeuw, E., M. Callegaro, J. Hox, E. Korendijk and G. Lensvelt-Mulders (2007) The Influence of Advance Letters on Response in Telephone Surveys. *Public Opinion Quarterly* **71** (3): 413-443

De Leeuw, E., M. Couper and W. De Heer (2005) Nonresponse and Bias. Unpublished Presentation prepared for the "NOSMO Methodologendag Nijmegen November, 25, 2005. Accessed on October, 3 2008 at: http://nosmo.nl/leeuw.pdf

De Luca, G. and O. Lipps (2005) Fieldwork and Survey Management in SHARE. In: Börsch-Supan and Jürges (eds.): *The Survey of Health, Ageing and Retirement in Europe - Methodology*. Mannheim: Mannheim Research Institute for the Economics of Aging,

Durand, C. (2005) Measuring Interviewer Performance in Telephone Surveys. *Quality & Quantity* **39**: 763–778

Elias, P. (2008) Providing Data on the European Level. *RatSWD Working Paper Series* **46**

ESRC (2008) Research Projects: Understanding Non-Response and Reducing Non-Response Bias. ESRC Research Centre on Micro-social Change Accessed on November, 25 2008 at: http://www.iser.essex.ac.uk/misoc/projects/project-details.php?projectID=233

Fielding, A. and Goldstein, H. (2006) Cross-classified and Multiple Membership Structures in Multilevel Models: An Introduction and Review. *Research Report* **RR791**, department for education and skills, University of Birmingham

Fowler, F. (2002) *Survey Research Methods (p 3/ed)*. Sage

Frick, J. and M. Grabka (2004) Missing Income Data in Panel Surveys: Incidence, Imputation and its Impact on the Income distribution. *DIW Discussion Papers* **376**

Frick, J. and M. Grabka (2007) Item Non-Response and Imputation of Annual Labor Income in Panel Surveys from a Cross-National Perspective. *IZA Discussion Paper* **3043**

Frick, J., O. Groh-Samberg, J. Schupp and K. Spieß (eds.) (2008) 25 Wellen Sozio-oekonomisches Panel, *Vierteljahrshefte zur Wirtschaftsforschung*, DIW Berlin , 77 (3), Berlin: Duncker & Humblot: 110-129

Fuchs, M. (2006) Mobile Web Survey: Möglichkeiten der Verknüpfung von Online-Befragung und Handy-Befragung. *Zuma Nachrichten Spezial* **13**: 105-126

Gallup, G. (1944) *A Guide to Public Opinion Polls*. Princeton University Press, Princeton.

Giele, J. and G. Elder (1998) *Methods of Life Course Research. Qualitative and Quantitative Approaches*. Sage, Thousand Oaks-London

Goldstein, H. (2003). *Multilevel statistical models* (3rd edition). Arnold, London

Groves, R. (1987) Research on Survey Data Quality. *Public Opinion Quarterly* **51** (2), Supplement 50th Anniversary Issue: 156-172

Groves, R. (1989) *Survey Errors and Survey Costs*. Wiley, New York

Groves, R. (2006) Nonresponse Rates and Nonresponse Bias in Household Surveys. *Public Opinion Quarterly* **70** (5), Special Issue: 646-675

Groves, R., R. Cialdini and M. Couper (1992) Understanding the Decision to participate in a Survey. *Public Opinion Quarterly* **56** (4): 475-495

Groves, R. and M. Couper (1996) Contact-Level Influences on Cooperation in Face-to-Face Surveys. *Journal of Official Statistics*, **12** (1): 63-83

Groves, R. and M. Couper (1998) *Nonresponse in Household Interview Surveys.* Wiley, New York

Groves, R. and M. Couper (2002) Introductory Interactions in Telephone Surveys and Nonresponse. In: Maynard, D., H. Houtkoop-Steenstra, N. Schaeffer and J. van der Zouwen (eds.) *Standardization and Tacit Knowledge*. Wiley, New York

Groves, R., F. Fowler, M. Couper, J. Lepkowski, E. Singer and R. Tourangeau (2004). *Survey Methodology.* Wiley Series in Survey Methodology.

Groves, R. and S. Heeringa (2006) Responsive Design for Household Surveys: Tools for actively controlling Survey Errors and Costs. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **169** (3): 439-457

Groves, R., M. Couper, S. Presser, E. Singer, R. Tourangeau, G. Acosta and L. Nelson (2006) Experiments in producing Nonresponse Rates. *Public Opinion Quarterly* **70** (5), Special Issue: 720-736

Groves, R. and S. Heeringa (2006) Responsive design for household surveys: tools for actively controlling survey errors and costs. *Journal of the Royal Statistical Society A* **169** (3): 439-457

Groves, R. and K. McGonagle (2001) A Theory-guided Interviewer Training Protocol regarding Survey Participation. *Journal of Official Statistics* **17** (2): 249-265

Groves, R. and E. Peytcheva (2006, April). The impact of non-response rates on non-response error: A meta- analysis. Unpublished Paper presented at the European Conference on Quality in Survey Statistics, Cardiff

Hansen, K. (2007a) The Effects of Incentives, Interview Length, and Interviewer Characteristics on Response Rates in a CATI-Study. *International Journal of Public Opinion Research* **19** (1): 112-121

Hansen, S. (2007b) CATI Sample Management Systems. In: J. Lepkowski et al. (eds.) *Advances in Telephone Survey Methodology*, John Wiley & Sons: 340-358

Harkness, J. (1999) In pursuit of quality: issues for cross-national survey Research. *International Journal of Social Research Methodology* **2** (2): 125-140

Heeb, J. and G. Gmel (2001). Interviewers' and respondents' effects on self-reported alcohol consumption in a Swiss health survey. *Journal of Studies on Alcohol*, July: 434-442

Holbrock, A., M. Green and J. Krosnick (2003) Telephone versus face-to-face Interviewing of national Probability Samples with long Questionnaires. *Public Opinion Quarterly* **67**: 79-125

Homish, G. and K. Leonard (2005) Spousal Influence on Smoking Behaviors in a US Community Sample of Newly Married Couples. *Social Science & Medicine* **61** (12): 2557-2567

House, J, E. Singer, R. Kahn, H. Schuman and F. Juster (2004) *A Telescope on Society: Survey Research and Social Science at the University of Michigan & Beyond.* Ann Arbor, Michigan

Hox, J., E. De Leeuw and I. Kreft (1991) The Effect of Interviewer and Respondent Characteristics on the Quality of Survey Data: A Multilevel Model. In: Biemer, Groves, Lyberg, Mathiowetz and Sudman (eds.): *Measurement Errors in Surveys*. New York: Wiley

Huddy, L. J. Billig, J. Bracciodieta, L. Hoeffler, P. Moynihan and P. Pugliani (1997) The Effect of Interviewer Gender on the Survey Response. *Political Behavior* **19** (3): 197-220

Japec, L. (2005) *Quality issues in interviewer surveys: some contributions.* PhD Thesis. Department of Statistics, Stockholm University

Jäckle, A. and P. Lynn (2007) Respondent Incentives in a Multi-Mode Panel Survey: Cumulative Effects on Nonresponse and Bias. *ISER Working Paper* **1**/2007. Colchester: University of Essex

Jeřábek, H. (2001) Paul Lazarsfeld – The Founder of Modern Empirical Sociology: A Research Biography. *International Journal of Public Opinion Research* **13** (3): 229-244

Kennickell, A. (2000) What Do the "Late" Cases tell us? Evidence from the 1998 Survey of Consumer Finances. Unpublished Presentation for the 1999 International Conference on Survey Nonresponse, Portland, Oregon, October 28-31,1999

Kenny, D., N. Bolger and D. Kashy (2002) Traditional methods for estimating multilevel models. In: D. Moskowitz and S. Hershberger: Modeling intraindividual variability with repeated measured Data. Mahwah: Lawrence Erlbaum Associates

Kreuter, F. (2009) Survey Methodology: International Developments. *RatSWD Working Paper Series* **59**

Krosnick, J. (1991) Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys. *Applied Cognitive Psychology* **5**: 213-236

Kuhn, U. (2008) Collection, construction and plausibility checks of income data in the Swiss Household Panel. *Swiss Household Panel Working Paper* **1_08**. Neuchâtel, Accessed on October, 3 2008. available at http://aresoas.unil.ch/workingpapers/WP1_08.pdf

Lambert, P. (2006) The British Household Panel Survey: introduction to a longitudinal data resource. Working Paper 2 of 'Longitudinal Data Analysis for Social Science Researchers', ESRC Researcher Development Initiative training programme. University of Stirling, Accessed on October, 3 2008. available at: http://www.longitudinal.stir.ac.uk/wp/lda_2006_2.pdf

Laurie, H. and P. Lynn (2008) The use of respondent incentives on longitudinal surveys. Forthcoming in Lynn, P. (ed.) *Methodology of Longitudinal Surveys*, Colchester: John Wiley & Sons

Laurie, H., R. Smith, and L. Scott (1999) Strategies for Reducing Nonresponse in a Longitudinal Panel Survey, *Journal of Official Statistics* **15**: 269-282

Little, R. (1988) A Test of Missing Completely at Random for Multivariate Data with Missing Values. *Journal of the American Statistical Association* **83** (404): 1198-1202

Little, R. and D. Rubin (1987) *Statistical analysis with missing data*. Wiley, New York

Little, R and Su, H.-L. (1989) Item Non-Response in Panel Surveys. In: Kasprzyk, D., Duncan, G., Kalton, G. and Singh, M. P. (eds.): Panel Surveys. Wiley, New York: 400-425.

Loosveldt, G. and V. Storms (2008) Measuring Public Opinions about Surveys. *International Journal of Public Opinion Research* **20** (1): 74-89

Lynn. P. (2003) Developing quality standards for cross-national survey research: five approaches. *International Journal of Social Research Methodology* **6** (4): 323-336

Lynn, P., N. Buck, J. Burton, H. Laurie, N. Uhrig (2006) *Quality Profile: British Household Panel Survey*. Institute for Social and Economic Research, Colchester: University of Essex

Lynn, P. and P. Clarke (2002) Separating Refusal Bias and non-contact Bias: Evidence from UK National Surveys. *The Statistician* **51** (3): 319-333

Lynn, P., P. Clarke, J. Martin and P. Sturgis (2002) The Effects of extended interviewer Efforts on Nonresponse Bias. In: Groves , R., D. Dillman, J. Eltinge and R. Little (eds.): *Survey Nonresponse*,, Wiley, New York: 135-148

Maynard, D. and N. Schaeffer (2000) Toward a Sociology of Social Scientific Knowledge : Survey Research and Ethnomethodology's asymmetric Alternates. *Social Studies of Science* **30** (3): 323-370

MIS-Trend (2007) Panel Suisse de Ménages 2006-2007. Rapport technique réalisé pour Le Panel Suisse de Ménages – Neuchâtel. Unpublished Field Report, Lausanne.

Morgan, J. and J. Smith (1969) A Panel Study of Income Dynamics - Study Design, Procedures, and Forms. 1968 interviewing year (wave I). A Report on Research in Process under Contract to the Office of Economic Opportunity. Survey Research Center, University of Michigan

Morton-Williams, J. (1993) *Interviewer Approaches*. Dartmouth Publ. Comp. Ltd., Aldershot

Nicoletti, C. and N. Buck (2004) Explaining Interviewee Contact and Co-Operation in the British and German Household Panels. *ISER Working Paper* **6**/2004. Colchester: University of Essex

O'Muircheartaigh, C. and P. Campanelli (1999) A Multilevel Exploration of the Role of Interviewers in Survey Non-Response. *Journal of the Royal Statistical Society A* **162** (3): 437-446

Palmore, E. (1978) When can Age, Period, and Cohort be separated? *Social Forces* **57** (1): 282-295

Pannekoek, J. (1988) Interviewer Variance in a Telephone Survey. *Journal of Official Statistics* **4** (4): 375-384

Peytchev, A., R. Baxter and L. Carley-Baxter (2009) Not all Survey Effort is equal. Reduction of Nonresponse Bias and Nonresponse Error. *Public Opinion Quarterly* **73** (4): 785-806

Peytcheva, E. and R. Groves (2009) Using Variation in Response Rates of Demographic Subgroups as Evidence of Nonresponse Bias in Survey Estimates. *Journal of Official Statistics* **25** (2): 193–201

Philippens, M. and J. Billiet (2004) Monitoring and Evaluating Non-response Issues and Fieldwork Efforts in the European Social Survey. Results of the analysis of contact form data. Paper presented at the European conference on quality and methodology in official statistics, Mainz

Pickery, J. and G. Loosveldt (1998) The impact of Respondent and Interviewer Characteristics on the Number of "No Opinion" answers. *Quality & Quantity* **32**: 31-45

Pickery, J. and G. Loosveldt (2001) An Exploration of Question Characteristics that mediate Interviewer Effects on Item Nonresponse. *Journal of Official Statistics* **17** (3): 337-350

Pickery, J. and G. Loosveldt (2002) A Multilevel Multinomial Analysis of Interviewer Effects on various Components of unit Nonresponse. *Quality & Quantity* **36**: 427-437

Pickery, J. and G. Loosveldt (2004) A Simultaneous Analysis of Interviewer Effects on Various Data Quality Indicators with Identification of Exceptional Interviewers. *Journal of Official Statistics* **20** (1): 77-89

Pickery, J., G. Loosveldt and A. Carton (2001) The Effects of Interviewer and Respondent Characteristics on Response Behavior in Panel Surveys. *Sociological Methods & Research* **29** (4): 509-523

Purdon, S., P. Campanelli and P. Sturgis (1999) Interviewer's calling Strategies on Face-to-face Interview Surveys. *Journal of Official Statistics*, **15** (2): 199-216

Rasbash, J., F. Steele, W. Browne and B. Prosser (2004) A User's Guide to MLwiN. Version 2.0. Centre for Multilevel Modelling, University of Bristol, Accessed on October, 3 2008 at: http://www.cmm.bristol.ac.uk/MLwiN/download/manuals.shtml

RatSWD (2008) Website of the German Council for Social and Economic Data, accessed on November, 16 2008 at: http://www.ratswd.de/eng/index.html

Rendtel, U., A. Behr, E. Bellgardt, T. Neukirch, M. Pyy-Martikainen, J. Sisto, R. Lehtonen, T. Harms, E. Basic and I. Marek (2004) Report on Panel Effects. Chintex Working Paper 22 (The Change from Input Harmonisation to Ex-post Harmonisation in National Samples of the European Community Household Panel – Implications on Data Quality)

Riley, T. (1911) Sociology and Social Surveys. *The American Journal of Sociology*, **16** (6): 818-836

Rodgers, W. (2002) Size of Incentive Effects in a Longitudinal Study. In: American Association for Public Research: *Strengthening Our Community - Section on Survey Research Methods,* (pp. 2930-2935)*,* Alexandria, VA

Røed K. (2006) Longitudinal Administrative Registers in Economic Research – A Norwegian Experience. Unpublished Presentation for the Conference on Longitudinal Surveys in International Perspective, Montreal, 25-27 January 2006

Ryu, E., M. Couper and R. Marans (2005) Survey Incentives: Cash vs. In-Kind, Face-to-Face vs. Mail, Response Rate vs. Nonresponse Error. *International Journal of Public Opinion Research* **18** (1): 89-106

Schouten, B. and F. Cobben (2007) R-indexes for the Comparison of different Fieldwork Strategies and Data Collection Modes. *Statistics Netherlands Discussion Paper* **07002**

Sills, D. (1992) Transformation in Social Research. *Science* **255**: 93-94

Singer, E. (2002) The Use of Incentives to Reduce Nonresponse in Household Surveys. *Survey Methodology Program* **051**, University of Michigan, Institute for Social Research Survey Research Center

Singer, E. (2006) Introduction: Nonresponse Bias in Household Surveys. *Public Opinion Quarterly* **70** (5), Special Issue: 637-645

Singer, E., M. Frankel and M. Glassman (1983) The Effect of Interviewer Characteristics and Expectations on Response. *Public Opinion Quarterly* **47** (1): 68-83

Singer, E. and R. Kulka (2000) Paying respondents for survey participation. *Survey Methodology Program Working Paper* No.**092**, Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbour

Singer J. and J. Willett (2003) *Applied longitudinal Analysis – Modeling Change and Event Occurrence.* Oxford University Press

Snijders, T. and R. Bosker (1999) *Multilevel analysis*. Newbury Park, California: Sage

Snijkers, G., J. Hox and E. De Leeuw (1999) Interviewer's tactics for Fighting Survey Nonresponse. *Journal of Official Statistics*, **15** (2): 185-198

Solga, H. and G. Wagner (2007) A Modern Statistical Infrastructure for Excellent Research and Policy Advice – Report on the German Council for Social and Economic Data during its First Period in Office (2004 - 2006). *Working Papers Series des Rates für Sozial- und Wirtschaftsdaten (RatSWD)* **2**/2007

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

Spiess, M., M. Kroh, R. Pischner and G. Wagner (2008) On the Treatment of Non-Original Sample Members in the German Household Panel Study (SOEP) — Tracing, Weighting, and Frequencies. *SOEPpapers* **98**, DIW Berlin

Starick, R. and N. Watson (2007) Evaluation of Alternative Income Imputation Methods for the HILDA Survey. *Hilda Project Discussion Paper Series* **1/07**

Stata Library. Multiple Imputation using ICE. Accessed on October, 3 2008 at http://www.ats.ucla.edu/stat/stata/library/ice.htm

Statistics Canada (2003) Survey methods and practices. Statistics Canada, Ottawa

Steenbergen, M. and B. Jones (2002) Modelling Multilevel Data Structures. *American Journal of Political Science* **46** (1): 218-237

Stoop, I. (2004) Surveying Nonrespondents. *Field Methods* **16** (1): 23-54

Stoop, I. (2005) *The Hunt for the last Respondent: Nonresponse in Sample Surveys*. PhD Thesis, Utrecht University

Stoop, I., S. Devacht, J. Billiet, G. Loosveldt, and M. Philippens (2003) The development of a uniform Contact Description Form in the ESS. Unpublished Presentation for the 14th International Household Survey Nonresponse Workshop, Leuven, Belgium

Symons, K., H. Matsuo, K. Beullens and J. Billiet (2008) Response based Quality Assessment in the ESS – Round 3. An update for 19 countries. Version - March 16 – 2008. Center of Sociological Research (CeSO), K.U.Leuven

Tourangeau, R. (2004) Survey Research and Societal Change. *Annual Review of Psychology* **55**: 775-801

Trochim, W. (2006) Research Methods Knowledge Base. 2nd Edition. Internet WWW Accessed on October, 3 2008 at: http://www.socialresearchmethods.net/kb/ (version current as of October 20, 2006)

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

van Meter, K. (2005) Studying Survey Interviewers: A Call for Research and an Overview. *Bulletin de méthodologie sociologique* **88**: 51-71

Voogt, R. and W. Saris (2005) Mixed Mode Designs: Finding the Balance between Nonresponse Bias and Mode Effects. *Journal of Official Statistics* **21** (3): 367-387

Wagner, G., J. Frick and J. Schupp (2006) Enhancing the Power of Household PanelStudies - The Case of The German Socio-Economic Panel Study (SOEP). *Data Documentation* **13**, DIW Berlin

Wagner, G. J. Frick and J. 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

Warren, J. and A. Halpern-Manners (2008) Panel Conditioning Effects in Longitudinal Social Science Surveys: Implications for Measuring Social and Economic Well-Being. Paper for the annual meetings of the Population Association of America, Detroit, April 2009, Version September 2008

Wernli, B. (2006) Homo- et hétérogamie dans les attitudes et le comportement politique en Suisse. *Swiss Political Science Review* **12** (1): 33-72

Wernli, B. (2007) La transmission intergénérationnelle de l'orientation idéologique en Suisse dans les familles à plusieurs générations. *Swiss Political Science Review* **13** (2): 237-259

Wolf, C. (2008) Harmonizing Survey Data. Unpublished Presentation for the FORS Workshop, November 6, 2008, Lausanne

Woolfrey, L. (2007) A Survey Data Archive Network in Africa. *IASSIST Quarterly* (Spring 2007): 14-20

ZuWallack, R., L Hall and R. Iachan (2004) Analysis of Nonresponse in Cross-Sectional RDD Surveys. Proceedings of the American Statistical Association, Section on Survey Research Methods. Alexandria, VA: American Statistical Association: 4703–4710

Published Papers

Cross National Contact Strategies

Oliver Lipps, Grant Benson
Swiss Household Panel, Neuchâtel, oliver.lipps@unine.ch
University of Michigan Institute for Social Research, gdbenson@umich.edu

Abstract

Using paradata gathered from the 11-nation Survey of Health, Ageing and Retirement in Europe (SHARE), this paper examines the impact of the first contact attempt, and the first contact properties, respectively, on contact and response efficiency using logistic multilevel models. We find that despite the different sample frames and interviewer compensation structure between countries, there are no considerable country effects with respect to make contact, once interviewer effects are controlled. Moreover, results point to an increased efficiency associated with evenings especially on Sundays, at least on the very first contact attempt. For attempts that result in initial contact, Saturday afternoons are most likely to eventually lead to completed interviews, followed by initial contact on weekdays during the daytime. We hypothesize that this may be due to the SHARE sample being composed of people aged 50 and over.

Keywords: Contact strategies, cooperation rates, cross-national survey, interviewer effects.

1. Introduction

With the ever-increasing cost of conducting surveys in the face of declining contact and response rates, many survey agencies resort to implementing contact strategies. These include directing interviewers on when and how to attempt contact and on how to respond to statements and questions made by the respondent. While we would expect differences in optimal contact strategies for different segments of the population – for example, for the retired compared to the fully employed – it is not clear that we would expect cross-national differences. That is, can we take lessons on contact strategies from the U.S. and U.K. and apply them with similar effects to continental European countries, for example? Or are there significant cultural and demographic differences that will impact the relative efficiency of contact strategies?

This paper analyzes data from the Survey of Health, Ageing, and Retirement in Europe (SHARE), conducted in 11 countries during the second half of 2004. Respondents were selected from among the non-institutionalized, resident persons aged 50 and over, as well as their spouses. Data was collected at every interviewer attempt, including time and day of the attempt, mode of the attempt, and outcome of the attempt (contact, no contact, resistance, completed interview, etc.). Nine of the countries used a sample management system designed specifically for the project, while two countries (Switzerland and Belgium) used their own sample management systems to track effort and progress.

During centralized trainings for all participating survey agencies, the importance of working in optimal "call windows" was stressed. This was based largely on existing research from the American and British experiences. Several survey agency representatives remarked that the European context was simply different from the American experience. In particular, you could not call on a household after 6 PM without antagonizing household members, and never on Sundays. Taking these concerns seriously, this paper is first concerned with unraveling the effects of *when* attempts were made on successfully obtaining contact with a target respondent. That is, are evening and weekend attempts more likely to yield contact than weekday attempts? In a second step, we examine the impact of time, day, and mode of first contact on ultimately interviewing the respondent. For example, if the survey agency anecdotal evidence is correct, we would expect an annoyance factor to result from Sunday attempts, particularly if it resulted in the very first interaction with a household member. Subsequently, we would expect initial contacts on Sundays to result in a net decrease in final interviews with those sample lines relative to cases where the initial contact was obtained on a different day.

2. Background

Much of the literature for both in-person surveys (e.g., Groves and Couper 1998) and telephone surveys (e.g., Greenberg and Stokes 1990) points to net efficiencies associated with obtaining initial contact by optimizing interviewer contact effort. The data generally show that attempts made on working day evenings and weekends are most likely to yield successful contact with a household member. Presumably, this is the case because traditional work patterns keep respondents out of the home weekday daytime. However, it is not clear that these patterns would continue to hold either for a sample drawn from an aging population[1] or for countries other than the United States and Britain. That is, with an aging population, we might expect that we would be more likely to reach a respondent during the day. Moreover, if the contact attempt is in-person (face-to-face or F2F), then evening attempts may be *less likely* to obtain contact, potentially because of a perception of respondent fear. Additionally, in the United States and Britain, there is a culture for the general population of working in blocs of eight (or more) continuous hours (typically 9 AM to 5 PM). This is not necessarily the case for other countries, where the work day may be interrupted by a two-hour lunch break allowing employees return to their place of residence. In this case, making contact during the day may increase the chance of reaching a household member at home.

In addition to country effects and the impact of when attempts are made, we are particularly interested in potential interviewer effects on establishing contact. That is, do some interviewers find better ways to obtain contact with respondents, even when controlling for time of day and day of the week that the attempt is made? If this is the case, then it suggests that we must do more thinking about recruiting and training interviewers.

3. Data

The analyses draw upon the early release of data ("Release 1") for the Survey of Health, Ageing and Retirement in Europe (SHARE). SHARE was conducted in 9 European countries (Austria, Denmark, Germany, Greece, Italy, Netherlands, Spain, Sweden, Switzerland) between April and October 2004, with another 2 (Belgium and France) starting data collection late (Börsch-Supan 2005). In general, survey agencies were required to collect interviews from around 1,500 households at an unweighted household response rate of greater than 50 percent. The cross-national aspect of the study presented numerous challenges. The most

[1] See for example Kirgis et al. (2004). Moreover, the principle underlying a contact window effort is predicated on knowing nothing about the household composition. Thus, Groves and Couper (1998, pp. 89-94) show that there is a reduced effort associated with contacting household members if there are more adults in the household, if there are children in the household, or if there are adults over 70 in the household.

significant of these were to establish baseline contact rules and flexible sample designs that both accommodated standard practices within each country and retained probability selection methods. Each survey agency was required to record every contact attempt and make minimally five attempts per household before coding out the case as non-finalized.[2]

Table 1: Household Response Rates and Sampling Frames

Country	Sampling Frame	Household Response Rate
Austria	Telephone (TEL) Directory, Household (HH)	57.3%
Denmark	National (Nat'l) Registry, HH	61.1%
Germany	Regional Registry, Individual (Ind)	60.2%
Greece	TEL Directory, HH	60.2%
Italy	Regional Registry, Ind	54.1%
Netherlands	Regional Registry, Ind	61.6%
Spain	Nat'l Registry, Ind	50.2%
Sweden	Nat'l Registry, Ind	42.1%
Switzerland	TEL Directory, HH	37.6%
Average		53.8%

Source: Adapted from De Luca and Peracchi (2005).

Due to variability in available data, participating countries relied on several different sampling frames.[3] For most countries (Denmark, Germany, Italy, the Netherlands, Spain, and Sweden), population registers were available. Sweden selected a reference person, Denmark a reference household randomly from national population registers. The other countries relied on multi-stage designs due to cost concerns and registers being administered at the local level. Three countries (Austria, Greece, and Switzerland) relied upon telephone directories to conduct screenings identifying households with at least one respondent aged 50 or over (Klevmarken 2005).

[2] For more details on the fieldwork procedures and SHARE's sample management system, see Lipps and De Luca (2005).
[3] Häder and Gabler (2003) cite Leslie Kish as saying of comparative research that „Sample designs may be chosen flexibly and there is no need for similarity of sample designs. Flexibility of choice is particularly advisable for multinational comparisons, because the sampling resources differ greatly between countries. All this flexibility assumes probability selection methods: known probabilities of selection for all population elements."

There were 71,114 in-person household contact attempts ("calls") recorded for 14,040 first household interviews at an average of 5.07 attempts per household interview. An additional 72,945 attempts were recorded for 17,342 non interviews. Almost 29 percent of the households refused to participate, leaving 17.3 percent of households not participating for other reasons, including non-contact, language barrier, and other reasons.

Table 2: Individual Response Rates and Contact Attempts

Country	Individual Response Rate (within household)	Average In-Person Attempts per Household
Austria	87.4%	4.9
Denmark	93.0%	9.9
Germany	86.5%	5.4
Greece	91.8%	6.0
Italy	79.7%	4.7
Netherlands	87.9%	4.3
Spain	73.8%	4.8
Sweden	83.8%	9.6
Switzerland	86.9%	6.1
Average	85.6	6.46

Source: Adapted from De Luca and Peracchi (2005) and own calculations.

4. Issues

In general, we are concerned with accounting for the impact of country and interviewer effects on the likelihood of obtaining contact with a household member. We would expect country effects to impact likelihood of contact for two reasons. First, two countries (Greece and Switzerland) had unknown eligibility on the initial attempt. This would lead us to hypothesize that evening call windows would be more efficient in these countries than in the other countries. Second, Europeans sometimes refer to cultural differences between the "north" and the "south" – in this case, Greece, Italy, and Spain belonging to the south. The southern countries are more likely to have businesses that permit for extended mid-day breaks, while the northern countries are more likely to work continuous 8-hour workdays interrupted only by one-hour lunch breaks. Third, interviewers were compensated using different schemes between countries. Switzerland alone paid interviewers on an hourly basis, while most of the other countries compensated interviewers per completed interview or household listing, and Denmark employed a mix of hourly wages and piecemeal compensation. There were also differences in interviewer recruitment, with some countries using independent contractors as interviewers and other countries using a professional, in-house staff. Each of these factors could lead to differences in interviewer behavior between countries, due to differential incentive schemes.

With respect to interviewers, some interviewers clearly develop ways of customizing their responses to potential respondents that enable them to achieve higher response rates. But there is no reason to expect interviewer effects in terms of obtaining initial contact, at least once one controls for time of day and day of the week that the attempt is made. If there is an interviewer effect, then it suggests either that some interviewers are doing something different – waiting for longer times at the door, ringing the bell more insistently, or trying multiple entrances – or that the characteristics of particular interviewers (being of a certain age or sex, or wearing clothing appropriate to a particular neighborhood) are more likely to have respondents open the door. If, contrary to our expectations, there are significant interviewer effects with respect to make contact, then we must do a better job of uncovering why that is the case.

4.1 Issue 1: Contact Efficiencies

A growing body of literature points to the importance of when contact attempts are made on obtaining initial contact (e.g., Groves and Couper 1998; Weeks, Kulka, and Pierson 1987; Purdon et. al. 1999). In general, the research points to increased efficiencies associated with working weekday evenings and weekends relative to weekday daytime. Additionally, O'Muircheartaigh and Campanelli (1999) explore the impact of interviewers on non-response in the British Household Panel Survey, using a multilevel approach. By employing multilevel analysis, they disentangle the clustering effects of region from interviewer effects on non-response. They find that there was, after controlling for several household variables and indicators of cooperation and contactability, almost no region effects on refusal and non-contact rates, but still some (although not statistically significant), evidence for an impact of interviewer effects.

For most of the SHARE sample, households had known eligibility, with at least one household member over the age of 50. Subsequently, we would expect that this population might have an increased chance of being contacted during the daytime. However, it is unclear whether there might also be country effects, particularly due to differences in laws or common practice in terms of age of retirement, home maker rates, and other cultural norms. We would expect that interviewer effects would be reduced if we control for time and day of the week that calls were made.

4.2 Issue 2: Contact Effect on Respondent Cooperation

We continue the analysis of contact efficiencies by attempting to determine whether the mode (face-to-face versus telephone) and timing of the contact has any impact on cooperation rate. Thus, if Saturday afternoons are optimal times to reach a respondent at home, does reaching a respondent during the weekend have an impact on their ultimate likelihood to cooperate? The multilevel analysis accounts for the effects of the timing of the first successful contact on obtaining an interview with the household at any point in time.

5. Results

5.1 Contact Efficiencies

Using the the *MLwiN* software (See Rasbash et. al. 2004) to conduct the multilevel logistic regression, we create a series of models iteratively building upon each other. We progress by conducting a step-wise (not necessarily hierarchical) inclusion of additional exploratory variables, in each case, increasing specificity depending on the results. The variables used in the analysis are summarized in Table3.

For the initial model, all contact attempts are included. Random effects of the intercept for country, interviewer and household levels are introduced for all models. These are modeled on either the initial contact attempt or the initial attempt resulting in contact with a household member. Note that - when considering only the initial attempt – or the initial contact – with a household, there would be no option of differentiating between respondents within a household.

The dependent variable in the first approach is "contact", modeled binomially and using a logit link function. Attempts resulting in reaching a person in the household are coded "1". These would include refusals, completed interviews, as well as appointments. All attempts resulting in no contact, such as no one being home or no one answering the door or the telephone, are coded "0". The base model equation is given in Equation 1.

The random components $v_{j,\ ctry}$, $u_{j,\ ctry,\ iwer}$ are expected to follow a normal distribution $N(0,\Omega_{var})$, the random component $e_{j,\ ctry\ iwer,\ hh}$ is expected to follow an extreme value distribution, $var \in \{v,u,e\}$. If several independent variables are modeled as random on the same aggregation level (i.e. Ω is a matrix), the non-diagonal elements of Ω are the covariances of the respective coefficients.

Table3: Variable Descriptions

Dependent Variables

<u>Contact:</u> Binary variable that equals 1 if the contact attempt investigated resulted in some interaction with a household member.

<u>CompletedHH:</u> Binary variable that equals 1 if there was ever a completed interview with at least one household member, for the contact investigated.

Independent Variables

<u>Work:</u> Binary variable that equals 1 if the day of the week of the contact attempt was Monday through Friday.

<u>Sat:</u> Binary variable that equals 1 if the day of the week of the contact attempt was Saturday

<u>Sun:</u> Binary variable that equals 1 if the day of the week of the contact attempt was Sunday.

<u>Morning:</u> Binary variable that equals 1 if the time of day of the contact attempt was between 8 AM and 11:59 AM.

<u>Afternoon:</u> Binary variable that equals 1 if the time of day of the contact attempt was between 12 noon and 5:59 PM.

<u>Evening:</u> Binary variable that equals 1 if the time of day of the contact attempt was between 6 PM and 9:59 PM.

<u>Night:</u> Binary variable that equals 1 if the time of day of the contact attempt was between 10

PM and 7:59 AM.

F2F: Binary variable that equals 1 if the mode of the contact attempt was in-person (face-to-face) rather than by telephone.

Condition: Ordinal variable scoring interviewer evaluation of the physical appearance of the selected housing unit. If 0 then the physical appearance was evaluated as being in "Good" condition. If 1 then it was evaluated as being "Average" and if 2 then it was evaluated as "Poor".

Environment: Ordinal variable scoring interviewer evaluation of the physical appearance of the neighborhood in which the selected housing unit is located. If 0 then the physical appearance was evaluated as being in "Good" condition. If 1 then it was evaluated as being "Average" and if 2 then it was evaluated as being "Poor".

Impediment: Binary variable indicating whether the interviewer observed any impediments to access to the housing unit. If 1, then the interviewer observed barriers to access, such as a locked entrance to an apartment complex.

Equation 1: Base Model

$$
\begin{aligned}
&\text{logit}(call1\_success)_{ctry,iwer,hh} = \\
&\sum coefficient_i * (fixed)\,\text{variable}_i + \\
&\sum coefficient_{j,ctry,iwer,hh} * (random)\,\text{variable}_j \\
&\quad with \\
&coefficient_{j,ctry,iwer,hh} = c_j + v_{j,ctry} + u_{j,ctry,iwer} \\
&+ e_{j,ctry,iwer,hh}
\end{aligned}
$$

We test for potential under- or overdispersion (Rabash et. al. 2004) by relaxing the assumption of binomial variation, but found only slight deviations from a binomial variation. For parsimony, we opted to keep the lowest level variation binomial. This choice also facilitated the comparison of coefficients across models.

Table 4: Null model using all contact attempts (N=83,150 Release 1: 106,469) Model 0

Dependent variable:	Fixed Effects	Random Effects (between[4])

[4] Here and in the following tables, we omit the random effects within the lowest level (in this case: call level), which - due to binomial variation assumed - always equals 1.

logit Contact		*Country*	*Interviewer*	*Household*
Constant (Intercept)	1.203	0.174	0.369	1.268
(std err.)	(0.150)	(0.090)	(0.027)	(0.025)

The first (null model) includes only the intercept, modeled both as a fixed and a random effect on each of the other hierarchical levels, being country, the interviewer, and the household level. Such a base model is primarily designed to calculate the amount of (disjoint) variation of the dependent variable on the different levels. Thus, we expect significant random variation on the dependent variable "contact" on all three (nested) levels.

Model 0 (Table 4) shows that the likelihood of obtaining contact is quite high. Overall, interviewers have a 77 percent likelihood of obtaining contact on any given attempt.[5] There is small significant random variation on the country level. However, the variance on the interviewer and household levels are highly significant, well above the $p<0.05$ level. The interviewer level variation more than doubles that on the county level, while the former is in turn less than a third of that on the household level. Thus, most of the variation on contact is due to differences of the households.

Due to the compounding effects of contact with multiple respondents, we drop analysis at the single respondent level and consider only the first contact attempt for the remaining analyses. Subsequently, the lowest level of analysis is the household[6]. We would expect that this would significantly decrease the likelihood of contact, as the interviewers would have no previous knowledge about the household.

Table 5: Contact Achieved -- Results of Multilevel Logit Models, First Contact Attempt (N=22,447), Models 1 and 2

Dependent variable: logit Contact	**Model 1**		**Model 2**	
	β	*std. err.*	*β*	*std. err.*
Fixed Effects				
const	.554	.129	.562	.129
Sat			-.144	.056
Sun			.054	.071
	Random Effects: Country			
Constant	.123	.067	.122	.066

5 Calculated as the antilogit of 1.203.
6 The random effect on the household level is in the following standardized at 1.

	Random Effects: Interviewer			
Constant	.603	.044	.603	.044

Comparing the two-level variance components in Model 1 with the all-call model above (see Table 4), on the country level, we find only a slightly smaller variance (and significance). This suggests that the binary variable contact success varies between countries in the same way, irrespective whether all calls are considered or only first calls. As suspected, the chances of a successful contact drop considerably if only first contact attempts are considered.

The interviewer level variation is much lower when all calls are taken into account. This indicates that for first calls only, the interviewer's behaviour has a greater impact on the call result. In particular, we would suspect that an interviewer's decision to make initial effort on particular days of the week or during particular times of day would have a significant impact on the likelihood to obtain contact.

In Model 2, we begin to attempt to decrease variations through a non-hierarchical stepwise inclusion of further explanatory variables. Here we included the day of the week, in a first step modeled as fixed effects. Model 2 generates a significantly negative Saturday effect on obtaining contact with a household member. There are (statistically insignificant) negative Sunday effects as well. Both of the weekend effects are measured against traditional working days (Monday through Friday). The coefficients thus indicate a somewhat smaller contact success rate on weekends relative to weekdays. The country and interviewer specific variation basically remain unchanged by the inclusion of the day of the week.

In Model 3 (Table 6), we extend the model and take the *random* effects of Saturday and Sunday on the interviewer level into account. The hypothesis behind the model is that interviewers gaining contact on the first attempt during the weekend may be different from other interviewers. First, the fixed effects changed somewhat. More interestingly, the variation of Sunday and Saturday (vs. weekdays) successes on the interviewer level are highly significant (.558 and .618); with a variance much higher than the fixed coefficient (-.051 and .076). This indicates that around half of the interviewers being more efficient on the weekends than during the weekdays at an interviewer-specific level.

In addition, we find a weakly significant positive covariance between the interviewer's successes on Saturday and Sunday on the one hand, and negative covariances of these weekend days with the interviewer's overall performance (intercept) (-.119 and -.192). Thus, interviewers who are successful on Saturdays are more likely to be successful on Sundays as well, and vice-versa, and interviewers, whose first calls are more successful overall, tend to be less successful on weekends, and vice-versa, respectively. We subsequently added time of day to the analysis, and abandoned the randomness of the weekend days on the interviewer level. The literature generally indicates that weekday evenings are best for obtaining initial contact, followed by weekend afternoons. We follow Kulka and Weeks (1988) for definitions of time of day, extending the afternoon time slot from 5 PM to 6 PM, following Groves and Couper (1998). The results are controlled against a "night" time slot (10:00 PM – 7:59 AM, all days).[7]

Table 6: Contact Achieved -- Results of Multilevel Logit Models, First Contact Attempt, Models 3 and 4

Dependent variable: logit Contact	**Model 3** β	*std. err.*	**Model 4** β	*std. err.*
Fixed Effects				
const	.560	.130	.833	.166
Sat	-.051	.076		
Sun	.108	.090		
Work Morning			-.411	.105
Work Afternoon			-.302	.101
Work Evening			-.116	.103
Sat Morning			-.609	.143
Sat Afternoon			-.395	.121
Sat Evening			-.236	.155
Sun Morning			-.615	.172
Sun Afternoon			-.233	.137
Sun Evening			.139	.167
Random Effects: Country level				
Constant	.125	.067	.136	.073
Random Effects: Interviewer level				
Constant	.617	.047	.597	.044

[7] Given that this characterizes first attempts, we consider this to be a curious choice for first attempts in any case. We believe that there are three acceptable interpretations for attempts registered during the night time slot. First, it can simply be very poor judgment on the part of the interviewer. He or she might be attempting a first look at the neighborhood, but this would not be particularly optimal due to the lack of daylight in most cases. Second, it is possible that the interviewer simply miscoded the time of day. We have no reason to believe that this miscoding would systematically have come from one of the other time slots. Third, there is the possibility that interviewers recorded the time that they made the note, rather than the time that they attempted the contact. However, there is little evidence of any interviewer systematically doing so.

Sat	.558	.130
Sun	.618	.173
	Covariances: Interviewer level	
Constant, Sat.	-.119	.068
Constant, Sun.	-.192	.084
Saturday, Sun.	.272	.138

In Model 4, only Sunday evenings have a higher probability of obtaining contact on the first attempt than the control group (late nights), which is not surprising. Weekend mornings perform worst, followed by Saturday afternoons.

We are surprised by the lack of country effects. That is, there is no decrease in the random constant coefficient on the country level when we include the day and time of day variables. The given differences in interviewer compensation and sample frames suggest some more explanation of the country level variation. In particular, with three countries using the telephone for initial household listing, and other countries selectively allowing interviewers to attempt initial contact by telephone, we hypothesize that country level explanation could be increased by including call mode (face-to-face or telephone) as a dummy variable in Model 5 (Table 7).

Table 7: Contact Achieved -- Results of Multilevel Logit Models, First Contact Attempt, Models 5 and 6

Dependent variable: logit Contact	**Model 5**		**Model 6**	
	β	*std. err.*	*β*	*std. err.*
Fixed Effects				
const	.325	.287	.429	.294
Work Morning	-.277	.052	-.276	.052
Work Afternoon	-.180	.041	-.181	.041
Sat Morning	-.446	.118	-.444	.118
Sat Afternoon	-.248	.085	-.237	.085
Sun Morning	-.509	.156	-.510	.156
Sun Afternoon	-.172	.108	-.168	.108
Sun Evening	.254	.144	.260	.144
F2F	1.120	.410	1.131	.410
Impediment			-.196	.046
House average			-.112	.036
House poor			-.079	.077
	Random Effects: Country level			
Constant	.597	.325	.625	.339
F2F dummy	1.255	.672	1.254	.672
	Random Effects: Interviewer level			

Constant	.673	.068	.676	.068
F2F dummy	1.076	.167	1.075	.167
		Covariances: Country level		
Constant, F2F	-.605	.404	-.624	.414
		Covariances: Interviewer level		
Constant, F2F	-.500	.096	-.504	.096

With the introduction of a fixed and random effect mode dummy for the initial contact attempt being in-person, the daytime fixed effects are either reduced or made insignificant. Even the intercept becomes insignificant by controlling for mode. However, making the attempt in person (F2F) produces a positive impact on the likelihood of obtaining contact on the first call attempt.

The random effect of F2F on the country level is slightly significant (1.255), while the effect on the interviewer level is highly significant (1.076). Interestingly, the covariances with the intercept on both levels are negative even if statistically significant on the interviewer level. This points to an interesting counter-intuitive observation. In particular, it suggests that interviewers develop specialized skills such that if they have good skills contacting respondents by telephone, then those skills do not transfer to in-person attempts, and vice-versa.

We test this hypothesis by means of a simple logit model, with first contact attempts resulting in contact as the dependent variable, proportion of face-to-face first calls and its square as independent variables (not shown). We find significant positive effects on first contacts from the proportion of in-person attempts, and significant negative effects from the square of the proportion of in-person attempts, and a nonsignificant intercept. This means that we have an inverse U-shaped relation between in-person proportion of calls and success of the first call, with the maximum contact rates for those interviewers, who realise a F2F proportion somewhere between 0 and 100%. This finding supports a hypothesis that well performing interviewers are able to apply the most appropriate mode of initial contact, depending on the special situation required.

We are curious whether interviewer evaluations of the “state” or condition of the selected household or neighborhood impacted their willingness to make contact on a household. In fact, we asked the interviewers to rate both the neighborhood and the housing unit in terms of being "good," "average" or "bad". The neighborhood evaluation does not show any significance. However, in Model 6 we see the highest impacts of evaluations of the housing unit being in an "average" state. Barriers to entry have an expected negative impact on likelihood to obtain contact with the household on the first attempt. The other coefficients only change to a very minor extend, once these housing observations are included in the model.

Finally, we are interested in whether we can determine anything about the relationship between the initial contact attempt and completing an interview with at least one household member. In short, are there interviewer actions that are particularly "off-putting" to respondents? Thus, we introduce the variable completedHH in Model 7 (Table 8), which equals 1 for those first calls on households that eventually deliver at least one individual interview. This variable is modelled with a fixed and a random effect on the interviewer level. We also keep the F2F variable, with a fixed effect, and a random effect on the interviewer level in the model.

Again, the time of day effects further decrease, as does the neighbourhood effects. Not surprisingly, the fixed effect of the variable completedHH is positive and highly significant on all levels. Households that eventually have a completed interview are more likely to have had contact on the first attempt, as opposed to not having contact on that attempt. Reversing the logic of the equation, an initial contact attempt resulting in contact is more likely to result in a completed interview than a call that does not obtain contact on the first attempt. Moreover, there is a highly significant random effect between interviewers (0.325), which is nevertheless smaller than the fixed effect (0.501). The fixed effect suggests that – everything else equal - even some of the relatively unsuccessful interviewers do better if the initial call yields contact (again, reversing the logic of the equation).

Table 8: Contact Achieved -- Results of Multilevel Logit Models, First Contact Attempt, Model 7

Dependent variable: logit Contact	**Model 7** β	*std. err.*
Fixed Effects		
const	-.104	.310
Work Morning	-.217	.051
Work Afternoon	-.137	.040

Sat Afternoon	-.161	.085
Sun Morning	-.466	.158
Sun Evening	.343	.147
F2F dummy	1.157	.411
Impediment	-.166	.047
House average	.-083	.037
House poor	-.011	.078
CompletedHH	.501	.136
	Random Effects: Country level	
Const.	.696	.377
F2F	1.257	.674
Compl.HH	.133	.074
	Random Effects: Interviewer	
Constant	.654	.084
F2F dummy	1.026	.166
Compl.HH	.325	.057
	Covariances: Country level	
Const, F2F	-.621	.428
Constant, F2F	-.039	.159
Const,Compl.HH	-.110	.125
	Covariances: Interviewer level	
Constant, F2F	-.471	.101
Constant, Compl.HH	-.048	.056
F2F, Compl.HH	-.065	.067

5.2 Issue 2: Contact Effect on Respondent Cooperation

Of course, the ultimate objective of any contact attempt is to obtain a completed interview with at least one eligible respondent, not just to obtain contact with a household member. We are particularly interested in whether there are aspects of the initial attempt yielding contact that would have an influence on the final outcome of respondent participation.

The dependent variable is thus set to the binary variable *completedHH*, equalling 1 if the household approached ultimately has at least one respondent agreeing to be interviewed, else 0. The database is reconfigured to include only attempts that result in the initial contact. Subsequently, the number of observations decreases from total initial contact attempts on all sample lines.

Again, we start with the base model, listed below as Model 8 (

Table 9), but proceed with the complexity models more quickly than in the first seven models. The fixed effect of the intercept is significant, as is the random effect on the interviewer levels, with the country level effect still only barely significant at the $p<0.05$ level.

Table 9: Null model on completed interview with at least one household member using first call attempt yielding contact, (N=20,486), Model 8.

Dependent variable: CompletedHH	**Fixed Effects**	**Random Effects** *Country*	*Interviewer*
Constant	-.838	.064	.604
(standard error)	(.097)	(.038)	(.046)

Again, following the iterative steps listed above, we include interaction variables for day (weekday, Saturday, Sunday) and mode of the attempt (either in-person [F2F] or not), dropping statistically insignificant terms and re-estimating the model until all terms have coefficients greater than their standard errors. The excluded combinations automatically serve as base category (aggregated)

Model 9 (Table 10) is suggestive of the importance of completing an initial phone contact on Saturday evenings, a contact on working day evenings with either mode, or on working day afternoons by F2F, or on Sunday afternoons by phone in order to have this household interviewed. A bad choice is weekend morning F2F, which is not very surprising.

Finally we are interested in determining whether the time of day and day of the week of the first contact is different for different interviewers or different countries. Forty-four and a half (44.5) percent of successful first contacts were performed on workday afternoons. Subsequently, we only consider this combination and investigate the random effects on both levels, after controlling the condition of the environment, the state of the house, and whether there are potential impediments:

Table 10: Time of day by type of day and F2F included (with fixed effects), first successful contact, (N=20,486), Model 9.

Dependent variable:	**Fixed Effects**	**Random Effects**	
Logit Completed HH		*Country*	*Interviewer*
Constant	.318	0.307	0.199
(std. err.)	(0.199)	(0.156)	(0.020)
Sun. Morn.*F2F	-.424		
(std. err.)	(0.181)		

Sun. Aftern*. phone	.259
(std.err.)	(0.154)
Sun.Even.*F2F	-.231
(std. err.)	(0.172)
WorkMorn*F2F	-.078
(std. err.)	(0.057)
WorkAftern*F2F	.137
(std. err.)	(0.046)
WorkEve.*phone	.176
(std. err.)	(0.057)
WorkEven.*F2F	.124
(std. err.)	(0.054)
Sat.Morn.*F2F	-.324
(std. err.)	(0.141)
Sat.Even.*phone	.560
(std. err.)	(0.213)
Sat.Even.*F2F	0.195
(std. err.)	(0.154))

Table 11: Working day afternoon fixed and random effects and F2F included (with fixed effects), first successful contact (N=20,486), Model 10.

Dependent variable:	**Fixed Effects**	**Random Effects**	
Logit CompletedHH		*Country*	*Inter-viewer*
Constant	.334	0.327	0.223
(std. err.)	(0.205)	(0.166)	(0.027)
WorkAftern.*phone	-.094	0	.186
(std. err.)	(0.056)		(.077)
WorkAftern * F2F	.113	.007	.099
(std. err.)	(0.051)	(.009)	(.039)
Conditions good	.112		
(std. err.)	(0.051)		
Conditions bad	-.073		
(std. err.)	(0.079)		
House good	.218		
(std. err.)	(0.036)		
House bad	-.249		
(std. err.)	(0.074)		
Impediments	-.228		
(std. err.)	(0.043)		

Contacting a household in-person on a working day afternoon increases the probability to realize an interview with this household, the more the environments and especially the house is subjectively evaluated to be in good condition and without impediments. More importantly, there are virtually no random effects of the choice of this time and mode on the country level, whereas there are significant random effects on the interviewer level. That is – controlling for physical appearance of the housing – time and mode does not have a variance across countries as regards to the dependent variable considered. All variance (more for phone calls) are therefore on the interviewer level.

6. Discussion and Conclusions

Surprisingly, we do not find any noteworthy significant country effects throughout our analyses, once the interviewer effects are controlled. That is, it simply did not appear to be the case that either the circumstances or the difficulties to obtain contact in some countries than in others are different. However, the interviewer effects within each country are quite significant on obtaining contact with a household, even as we introduce additional explanatory variables including day of the week and time of the day that contact was attempted.

In addition, the results point to an increased efficiency associated with evenings especially on Sundays, at least on the very first contact attempt. This is not surprising to us, as it confirms similar studies in the United States and Britain.

Sunday mornings, however, appear to be a call window with decreased initial contact efficiency. For attempts that result in initial contact, Saturday afternoons are most likely to eventually lead to completed interviews, followed by initial contact on weekdays during the daytime. However, Sunday and Saturday morning and partially Sunday evening attempts by F2F appear to generate an annoyance factor with household members, decreasing the willingness of respondents to ultimately participate.

Throughout the analysis, interviewer effects persist. That is, there are interviewers who more likely find ways to obtain initial contact than other interviewers, even when controlling for when and how the attempt is made. Clearly, more research needs to be conducted to determine what accounts for these differences.

Acknowledgements

We gratefully acknowledge funding for the SHARE project. The SHARE data collection has been primarily funded by the European Commission through the 5th framework programme (project QLK6-CT-2001-00360 in the thematic programme Quality of Life). Additional funding came from the US National Institute on Aging (U01 AG09740-13S2, P01 AG005842, P01 AG08291, P30 AG12815, Y1-AG-4553-01 and OGHA 04-064). Data collection in Austria (through the Austrian Science Foundation, FWF), Belgium (through the Belgian Science Policy Administration) and Switzerland (through BBW/OFES/UFES) was nationally funded. The SHARE data set is introduced in Börsch-Supan et al. (2005); methodological details are contained in Börsch-Supan and Jürges (2005). We also wish to acknowledge support of the Swiss Household Panel at Neuchâtel, and the University of Michigan Institute for Social Research, Survey Research Operations in Ann Arbor, Michigan.

References

Börsch-Supan, Axel (2005) "History of the Development Process" in *Health, Ageing and Retirement in Europe: First Results from the Survey of Health, Ageing and Retirement in Europe*. Börsch-Supan et al. (eds.) Mannheim: Mannheim Research Institute for the Economics of Aging.

Börsch-Supan, Axel and Hendrik Jürges (eds.) (2005) *Health, Ageing and Retirement in Europe-methodology,* Mannheim: MEA, 2005, forthcoming.

Campbell, P, Patrick Sturgis and Susan Purdon (1997) Can you hear me knocking: An investigation into the Impact of Interviewers on Survey Response Rates. The Survey Methods Centre at the Social and Community Planning Research, London. URL: http://joni.soc.surrey.ac.uk/~scs1ps/papers/ can%20you%20hear%20me%20knocking.pdf

De Luca, Giuseppe and Franco Peracchi (2004) *Survey participation in the first wave of SHARE*, unpublished manuscript, Mannheim 2004.

Greenberg, Betsy S. and S. Lynne Stokes (1990) "Developing an Optimal Call Scheduling Strategy for a Telephone Survey" *Journal of Official Statistics*, Vol. 6 (4), pp. 421-435.

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

Groves, Robert M. John Van Hoewyk, Grant Benson, Paul Schulz, M. Patricia Maher, Lynette Hoelter, William Mosher, Joyce Abma, and Anjani Chandra (2003) "Using

Process Data from Computer Assisted Face to Face Surveys to Make Survey Management Decisions." Paper presented at the 2003 AAPOR Conference, Nashville, TN.

Häder, Sabine and Siegfried Gabler (2003) "Sampling and Estimation." Chapter 8 in Harkness, Van de Vijver, and Mohler (eds.) *Cross-Cultural Survey Methods*. Wiley Series in Survey Methodology (Hoboken: Wiley & Sons).

Kalsbeek, William D., Steven L. Bottman, James T. Massey, and Pao-Wen Liu (1994) "Cost-Efficiency and the Number of Allowable Call Attempts in the National Health Interview Survey" *Journal of Official Statistics*, Vol. 10 (2), pp. 133-152.

Kirgis, Nicole et al. (2004) "Health and Retirement Study 2004 Study Manual." Unpublished interviewer training manual.

Klevmarken, Anders (April 2005) "Sample Design" in *Health, Ageing and Retirement in Europe: First Results from the Survey of Health, Ageing and Retirement in Europe*. Börsch-Supan et al. (eds.) Mannheim: Mannheim Research Institute for the Economics of Aging.

Kulka, Richard A. and Michael F. Weeks (1988) "Toward the Development of Optimal Calling Protocols for Telephone Surveys." *Journal of Official Statistics*, Vol. 4 (4), pp. 319-332.

Lipps, Oliver and Giuseppe De Luca (2005) "Fieldwork and Sample Management" in *Health, Ageing and Retirement in Europe: First Results from the Survey of Health, Ageing and Retirement in Europe*. Börsch-Supan et al. (eds.) Mannheim: Mannheim Research Institute for the Economics of Aging.

O'Muircheartaigh, Colm and Pamela Campanelli (1998) "The Relative Impact of Interviewer Effects and Sample Design Effects on Survey Precision" *Journal of the Royal Statistical Societies*, Vol. 161 (1), pp. 63-77.

O'Muircheartaigh, Colm and Pamela Campanelli (1999) "A Multilevel Exploration of the Role of Interviewers in Survey Non-Response" *Journal of the Royal Statistical Society*, Vol. 162 (3), pp. 437-446.

Purdon, Susan et. al. (1999) “Interviewers’ Calling Strategies on Face-to-Face Interview Surveys” *Journal of Official Statistics*, Vol. 15 (2), pp. 199-216.

Rasbash, J., F. Steele, W. Browne, and B. Prosser (2004). *A user's guide to MLwiN version 2.0*. London, Institute of Education.

SRC (2004) Survey of Health, Ageing and Retirement in Europe. Interviewer Project Manual. Main Test 2004. Unpublished manuscript, Survey Research Center, University of Michigan. URL: http://www.share-project.org/index.php?page=Documentation&menue=4&sub=

Weeks, Michael F., Richard A. Kulka, and Stephanie A. Pierson (1987) “Optimal Call Scheduling for a Telephone Survey” *Public Opinion Quarterly*, Vol. 51, pp. 540-549.

Effects of Interviewer Experience on Components of Nonresponse in the European Social Survey

Oliver Lipps*
Alexandre Pollien*

We analyze interviewer related nonresponse differences in face-to-face surveys distinguishing three types of interviewers: those who have previous experience with the same high standard cross-sectional survey ("experienced"), those who were chosen by the survey agency to complete *refusal conversions* ("seniors"), and *usual* interviewers. The nonresponse components are obtaining household contact, target person contact, and target person cooperation. In addition we examine if interviewer homogeneity with respect to these components is different across the three interviewer groups. Data come from the European Social Survey (ESS) contact forms from four countries which participated during the three rounds 2002/04/06 and used the same survey agency that in turn used to some extent the same interviewers. To analyze interviewer effects, we use discrete two-level models. We find some evidence of better performance by both senior and experienced interviewers and indications of greater homogeneity for nonresponse components, especially for those that contain room for improvement. Surprisingly, the senior interviewers do not outperform those experienced. We conclude that survey agencies should make more efforts to decrease the comparatively high interviewer turnover.

Keywords: ESS, interviewer effects, interviewer experience, interviewer turnover, nonresponse, cooperation, call data

- Swiss Foundation for Research in Social Sciences (FORS)
  c/o University of Lausanne, Vidy, CH - 1015 Lausanne
  Corresponding author e-mail address: oliver.lipps@fors.unil.ch

Acknowledgment: Support by the Swiss National Science Foundation is gratefully acknowledged. We thank three anonymous reviewers and the editor for comments on a previous version of this paper.

Introduction

To obtain sample unit response, survey interviewers first need to obtain contact with the household and – once contact is established – cooperation of the sample member (Lynn and Clarke 2002). Previous studies show that interviewers differ in both obtaining contact (Purdon et al. 1999) and especially obtaining cooperation (Hox et al. 1991, Groves and Couper 1996, Campanelli and O'Muircheartaigh 1999, O'Muircheartaigh and Campanelli 1999, Snijkers et al. 1999, Pickery et al. 2001, Pickery and Loosveldt 2002, Cantor et al. 2004, Lipps and Benson 2005). However, only a few interviewer characteristics have been found to explain this. Experience is among the few easy to observe interviewer properties that seem to have positive effects on obtaining high response rates (Chromy et al. 2003).

Usually interviewer experience is measured by the number of years in the job (e.g., Durrant et al. 2008). In this research, we study two concepts of "expertise" an interviewer may have:

- Being used to conduct refusal conversion calls (Hill and Willis 2001, Billiet et al. 2007). This difficult task is generally performed by "senior" interviewers who showed good performance during other contact phases.
- Past working experience from a high standard repeated cross-sectional survey. Cumulative experience in such a survey may result in better contact and in particular cooperation skills.

According to the ESS fieldwork recommendations, "refusal conversion attempts should not be made by the interviewer who previously received the refusal but should be reissued to another (senior) interviewer." Survey agencies are more likely to trust senior interviewers to be able to convert refusals. (Loosveldt et al. 2004: 2). The second measure takes working experience from the same survey into consideration without considering performance.

Interviewers working for the same survey agency will usually be assigned to the same geographical area, and have to work in a similar way.

There are several reasons why we examine interviewers in a cross-national survey. Survey conditions are different across countries (Harkness 1999). For example, while the target person can be approached directly in countries with an individual sampling frame, in countries with a household sampling frame, it is necessary to screen the household in order to select the target person first. This necessitates the cooperation of the household gatekeeper, if any. The characteristics of the interviewers (sex, age, professional status, etc.) are different across countries (ESS 2002-2003, 2004, 2006)[1]. Amongst others, this also depends on the survey agency (public or private, predominant survey mode, e.g., face-to-face or centralized telephone, working climate, etc.). In addition, it is likely that there are negative selection effects in some countries: “Interviewers are asked to carry out an activity that is considered to be scientific in nature - a job once reserved for specialists. However …the job [is not] considered a career” (Durand et al. 2004, also Raz and Blank 2007). Consequently, interviewers may quit their interviewer activity once higher paying and more reputable jobs are available, although the management is usually interested in keeping good interviewers. This presumably applies for the better interviewers who have better chances on the job market.

This paper is organized as follows: we first review the literature on interviewer effects on obtaining contact and cooperation of sample members, and interviewer homogeneity. Using data from the ESS, we analyze interviewer effects on obtaining contact with households and target persons, and cooperation of target persons, as well as interviewer homogeneity. We distinguish two different concepts of interviewer “expertise”. Finally, we summarize and discuss implications from our findings.

[1] The Documentation reports include information about the number of interviewers, their status (free lance, employed), type of training, payment, incentives used (if any), sample sizes, fieldwork dates, etc

Obtaining Contact

Interviewers may anticipate sample member's time-use patterns and adapt their timing behavior to obtain contact (Purdon et al. 1999, Lipps and Benson 2005, Billiet and Pleysier 2007). In addition, unlisted telephone numbers or addresses, a "beware of dog" sign, the presence of interphone or other barriers to access, or simply not answering the door may indicate householders' unwillingness to be contacted with a survey request. Interviewers may be able to link such conditions to characteristics of the people who live there (Garfinkel 1967) and adapt their contact behavior accordingly (Tourangeau 2004). It is reasonable to assume that interviewer experience may reduce the noncontact component of nonresponse in face-to-face surveys as interviewers can take advantage of experiences from previous waves to contact hard-to-contact sample members (Campanelli and O'Muircheartaigh 1999, Snijkers et al. 1999, Groves and McGonagle 2001, Pickery and Loosveldt 2002).

Obtaining Cooperation

Once successfully contacted, lack of time or interest, or concerns about privacy and not being able to adequately respond to survey questions are often mentioned reasons for refusal. The sample members' likelihood of cooperating depends on factors related to 1.) interest in the topic of the survey, 2.) a general sense of civic responsibility, 3.) the desire to express opinions or attitudes, 4.) the effect which an incentive might have, and 4.) a general propensity to answer to such requests (Groves, Cialdini and Couper 1992). How these elements add up and result in sample members' decision to cooperate or not has been the subject of numerous studies in social psychology (Groves, Singer and Corning, 2000). Interviewers differ with respect to their ability to tailor their behavior to sample members (Morton-Williams 1993) and to maintain interaction with reluctant individuals (Groves and Couper 1996, 1998). Both are found to be essential in obtaining cooperation (Groves et al. 2004). According to Cantor et al. (2004), "Interviewers tailor their statements by addressing the respondents' specific concerns; [they] maintain interaction by engaging respondents in conversation in order to learn more about those concerns" (p. 4935). Interviewers skilled in tailoring survey requests to respondents are also able to make survey issues salient to the individual (Groves and Couper 1998).

Carton (2008) reports strong interviewer effects on respondents' doorstep concerns mentioned in an annual face-to-face survey on changes in attitudes, opinions, and behaviors of the Flemish people. The only interviewer variable able to reduce variance on the interviewer level is their experience. Other research suggests that interviewer response rates correlate positively with years in the job (e.g., Couper and Groves, 1992). Experienced interviewers' success probably comes from their "larger number of combinations of behaviors proven to be effective for one or more types of householders" (Groves et al. 1992: 478–479). As Groves and colleagues (1992) suggest, expert interviewers have access to a large repertoire of cues, phrases, or descriptors to choose from during the survey request. The choice of whichever statement used to open the conversation is the result of observing the housing unit, the neighborhood, and immediate reactions during the first contact with the person who answers the door (Blom et al. 2008).

Interviewer Homogeneity

Economists provide evidence that with increased experience the productivity of employees stagnates or even decreases (Medoff and Abraham, 1980). Assumptions are that more experienced employees do not accept further training and become increasingly less motivated. Kiraz and Yildirim (2007) compare competences of experienced and novice teachers. They report that while experience is important, novices demonstrate better supervisory skills. The authors suggest that novices may be more motivated and more open minded to their environment. In the survey context, Olsen and Peytchev (2007) notice a changed interview perception with increased interviewer experience: the length of interviews decreases and the interviewers perceive the respondents to be less interested. To explain, the authors posit that experienced interviewers are less likely to engage in conversations unrelated to the interview.

We conclude that while novice interviewers are generally more enthusiastic and effusive in their work, they later try to improve the cost/benefit ratio over time because interviewers are commonly paid by completed case. At the same time, increasing experience improves work efficiency, for example through better tailoring (Groves and Couper 1998). Also, interviewers tend to get more insight into survey practice over time and – looking for similarities with their colleagues – continue to behave towards a “cultural” conformity (Festinger 1954). All this may lead to an equilibrium between costs and benefits, with a tendency toward conformity with colleagues and survey agency standards. In addition, it is usually the best or the worst interviewers who quit the job. The former find a more interesting job elsewhere, as explained above. The latter quit because they are fired or frustrated.

From these findings, we hypothesize that results from “expertised” interviewers is more homogeneous.

Effects of Interviewer Experience: Hypotheses

We test the following hypotheses. For more experienced interviewers, we expect

A. a higher likelihood of conducting first calls on households that are *ever* contacted

B. a higher likelihood of *ever obtaining contact with the target person, once contact with the household has been established*

C. a higher likelihood of *ever obtaining cooperation of the target person, once contact with the target person is established*

D. smaller interviewer effects.

Data

The European Social Survey (ESS), conducted every other year since 2002, was among the first to define rigorous and strictly standardized fieldwork protocols for participating countries (Philippens and Billiet 2003). For example, in all ESS rounds conducted so far, a target response rate of 70% was set out, with a maximum noncontact rate of 3% (Billiet and Philippens 2004, Billiet and Pleysier 2007, Symons et al. 2008, ESS 2007). Four countries (Switzerland, Germany, Denmark, and Finland) used the same survey agencies with some of the same interviewers, and – most importantly – provided the same IDs for interviewers between rounds. To use the repeated cross-sectional ESS rather than a panel survey where often the same interviewers are used avoids confusing effects from panel conditioning of *respondents*.

In Table 1, we depict the number of interviewers used with their respective experience, by round (percentages in Table 2). Because we compare performances based on previous ESS experience, only rounds two and three can be used.

< about here Table 1>

< about here Table 2>

In Finland mostly the same interviewers are used after round 1, whereas in Germany interviewer turnover is rather high. In Germany and Switzerland, the two interviewer "expertise" variables are highly positively correlated[2] with rho=.56 (std.err.=.07) and rho=.73 (std.err.=.09) respectively. In Denmark the correlation is weaker (rho=.30, std.err.=.12) and in Finland not significant on the 5% level (rho=.29, std.err.=.19).

[2] We use the command "polychoric" from STATA that is appropriate for categorical variables.

In the analysis, we exclude re-issued contacts, and pool rounds 2 and 3. At first, we present some descriptive statistics, treating all cases independently. Table 3 lists the mean percentages of ever contacted households, ever contacted target persons given that the household is contacted, and ever cooperating target persons given that this person has already been contacted, by country and "expertise" of the interviewer doing the call or contact. We use first calls on households, first contacts with households, and first contacts with target persons, respectively.

< about here Table 3 >

From this simple statistics, senior and experienced interviewers seem to perform better than those without "expertise", especially if the response rate is far from 1.

To adequately model interviewer contact and cooperation performance requires multilevel models, because respondents approached or contacted by the same interviewer are not independent. We will use two-level random intercept logit models that are briefly introduced in the next section.

Multilevel Modeling

Measuring effects in data with a clustered structure is not straightforward. In interviewer mediated survey data, measurements from respondents (lower level) are not independent from the respective interviewers (higher level). Multilevel models (e.g., Snijders and Bosker 1999) recognize the existence of data clustering by allowing for residual components at *each* level. Standard ANOVA models model only the intercept as random. Suppose y is the survey variable of interest, j the interviewer, and i the respondent. ANOVA assumes that y can be modeled as:

$$y_{ij} = m + \mu_j + \varepsilon_{ij}\,, \qquad \varepsilon_{ij} \sim N(0,\sigma_\varepsilon^2)$$

with m the "grand mean" and μ_j interviewer j's individual deviation. If the number of interviewers is large, to estimate all μ_j would mean to lose too many degrees of freedom. If in addition the set of interviewers can be considered as a smaller but representative sample of the total interviewer population, multilevel modeling approaches are more appropriate. The multi-(here: 2) level variant is to model

$$y_{ij} = m + [\mu_j + \varepsilon_{ij}]\ , \qquad \mu_j \sim N(0,\sigma_\mu^2) \quad \text{and} \quad \varepsilon_{ij} \sim N(0,\sigma_\varepsilon^2)$$

That is, the interviewer's deviation from the grand mean is taken into account by adding a normally distributed random variable. This approach allows for a nonzero covariance between μ and ε, with the "intraclass correlation coefficient" (ICC) the interviewer-specific proportion of the total variance $(\sigma_\mu^2/(\sigma_\mu^2+\sigma_\varepsilon^2))$.

In our analyses, the ICC is calculated separately for each "expertise" group, with the difference of the ICC of the "expertised" interviewers and the ICC of the novice interviewers tested for significance. To test interviewer "expertise" performance, we combine the two ("expertise"/complementary) groups and add a dummy variable $exper_j$ to the model that is tested for significance. (z-) Tests are based on the assumption of a normally distributed estimate of the estimated parameter that is tested.

Modeling Results

The figures in Table 4, Table 6, and Table 8 depict the odds ratios of the experienced/senior interviewer dummies ($exper_j$) in the models, distinguishing interviewers in one of the two "expertise" groups from the respective complementary group. These dummies quantify the odds of "expertise" compared to the complementary group, as for obtaining contact with the household (Table 4), conditional contact with the target person (Table 6), and conditional cooperation of the target person (Table 8). In tables Table 5, Table 7, and Table 9, the differences of the ICC between the senior/experienced interviewers and the respective complementary interviewers are shown for the respective nonresponse measures.

Ever making Contact with Household

The first performance measure analyses whether interviewers who make the first call on a sampled household are better at ever making contact with the household.

< about here Table 4 >

From Table 4, we find that experienced/senior interviewers have higher household contact rates overall, but not necessarily in all countries. Previous ESS experience seems to have slightly more positive effects on successfully calling households than seniority.

< about here Table 5>

In Table 5 we can see that only German interviewers with previous ESS experience obtain more similar contact rates, i.e., are more homogeneous.

Ever making Contact with Target Person once Contact with Household

This indicator measures performance with respect to overcoming the gatekeeper problem, i.e., whether obtaining contact with the target person is not possible once the household has already been contacted. Specifically, we consider the first household contacts as regards to whether the target person can *ever* be contacted.

< about here Table 6 >

With respect to ever making contact with the target person once the household has been contacted, Table 6 shows that interviewers with ESS experience perform better in Switzerland, and overall.

< about here Table 7 >

From Table 7 we find that the proportion of the interviewer-specific variance of making contact with the target person is smaller only among Danish senior interviewers.

Ever obtaining the Cooperation of Target Person once Contact with Target Person has been made

< about here Table 8 >

As far as the most important reason for nonresponse, noncooperation of target persons, is concerned, Table 8 shows that only German senior interviewers perform better. Also the pooled value for both senior and experienced interviewers is significantly positive.

< about here Table 9 >

According to Table 9, higher interviewer cooperation performance homogeneity can be reported in both "expertise" groups in Germany, and overall.

Conclusion

In this research we examine whether using interviewers with "expertise" pays off in producing higher response rates and whether they produce more homogeneous response rates. Using data from the European Social Survey (ESS), we define and compare two measures of "expertise":

- "Seniority", this means, whether the interviewer conducts refusal conversion attempts in the ESS (re-issued cases). This indicator of experience is very much at the discretion of the survey agency
- "ESS experience", i.e., whether the interviewer worked for the ESS in (at least) one of the previous rounds

The latter concept reduces the countries to be analyzed to those using the same survey agency and at least some of the same interviewers: Switzerland, Germany, Denmark, and

Finland. We analyze interviewer performance distinguished by the three main reasons for nonresponse: noncontact with the household, noncontact with the target person conditional on household contact, and noncooperation of the target person conditional on contact. In addition we analyze interviewer homogeneity with respect to each of these response components. We use data from the second and third round of the ESS, and exclude all cases that are re-issued to another interviewer.

We find that previous experience with the same survey instruments and protocols help to ever obtaining contact in Germany and Denmark. It may be that in Finland, general contact rates are so high that experience does not further pay-off. In Switzerland, respondents are not area-clustered. Contrary to Switzerland, where travel efforts may prevent this, it may be that in Germany and Denmark experienced interviewers take the time to visit the sample member's place first to check how respondents can be better accessed. A significantly more homogenous behavior can be concluded in Germany only.

Obtaining access to the target person is more difficult in Switzerland that uses a household sampling frame. It is therefore not surprising that only in this country experienced interviewers find better ways to accessing the potential respondent. This does however not hold for Swiss senior interviewers. With the exception of slightly more homogeneous Danish senior interviewers, homogeneity is not higher among "expertised" interviewers.

Senior interviewers perform better when trying to obtain cooperation of the target person in Germany only. In this country, the target person refusal rate is the highest among the countries examined (Table 3). It may be that senior interviewers perform the better the worse the overall situation. Also homogeneity is higher among interviewers of both "expertise" groups in Germany.

Pooling all countries increases the power to find positive effects for almost all nonresponse measures in both "expertise" groups. Exceptions are homogeneity of obtaining household and target person contact. It may be that usual contact rates are already too high to cause homogeneity differences across "expertise" groups.

Overall, interviewers with previous ESS experience perform slightly better than senior interviewers. Since senior interviewers are supposed to be the best interviewers available, it is surprising that they do not outperform survey experienced interviewers: it can be expected that senior interviewers are selected for refusal conversion by the survey agency just because they performed so well with "normal" (not refusal converted) sample cases. That senior interviewers are not better that survey experienced interviewers shows that experience is an equally good performance indicator – at least for the normal fieldwork phase. It must be noted that we did not investigate performance of refusal conversion calls, or treatments of otherwise difficult sample members, for which senior interviewers might be specially trained for. Our conclusions apply for the initial sample case only that may, in fact, later be re-issued to another interviewer.

Finding additional promising measures of interviewer "expertise" remains an issue for further research. A close collaboration of survey methodologists and research institutes could help to identify specific qualifications. Nevertheless, the good performance results of experienced interviewers show that survey agencies should do their best to decrease the turn-over of their interviewers, especially those who have more experience with similar surveys. Additionally, these results suggest that training could be different for new and experienced interviewers.

References

Billiet, J., M. Philippens, R. Fitzgerald and I. Stoop (2007) Estimation of Nonresponse Bias in the European Social Survey: Using Information from Reluctant Respondents. *Journal of Official Statistics* **23** (2): 135-162

Billiet, J. and M. Philippens (2004) Data Quality Assessment in Round 1 : Between Wishes and Reality. Center of Sociological Research (CeSO) K.U. Leuven, Belgium

Billiet, J. and S. Pleysier (2007) Response Based Quality Assessment in the ESS - Round 2. An update for 26 countries. Center of Sociological Research (CeSO) K.U. Leuven, Belgium

Blom, A., P. Lynn and A. Jäckle (2008) Understanding Cross-National Differences in Unit Non-Response: The Role of Contact Data. *ISER Working Paper* 1/2008. Colchester: University of Essex

Campanelli, P. and C. O'Muircheartaigh (1999) Interviewers, Interviewer Continuity, and Panel Survey Nonresponse. *Quality & Quantity* **33**: 59-76

Cantor, D., B. Allen, S. Schneider, T. Hagerty-Heller, and A. Yuan (2004) Testing an Automated Refusal Avoidance Training Methodology. In: 2004 proceedings of the American Statistical Association, Section on Survey Research Methods. Alexandria, VA: American Statistical Association, 2004:4935-42

Carton, A. (2008) Are our paradata always of good quality? Presentation for the 19th International Workshop on Household Survey Nonresponse 2008, Ljubljana, Slovenia

Chromy, C., D. Odom, J. Eyerman, and M.E. McNeeley (2003) The Effect of Interviewer Experience on the Interview Process in the National Survey on Drug Use and Health". In: 2003 proceedings of the American Statistical Association, Section on Survey Research Methods. Alexandria, VA: American Statistical Association, 2003:954-61

Couper, M. and R. Groves (1992). The role of the interviewer in survey participation. *Survey Methodology* **18**: 263–277

Durand, C., M.-E. Gagnon, C. Doucet and E. Lacourse (2004) An Experiment in Cognitive Training of Telephone Survey Interviewers. Presentation for the RC33 Sixth International Conference on Social Science Methodology, Amsterdam, The Netherlands

Durrant, G., L. Staetsky and F. Steele (2008) The Effects of Interviewer Characteristics and Attitudes on Refusal in Face-to-Face Surveys. Presentation for the 19th International Workshop on Household Survey Nonresponse 2008, Ljubljana, Slovenia

ESS 2002-2003 Documentation Report, Edition 6.0, European Social Survey, http://ess.nsd.uib.no/index.jsp

ESS 2004 Documentation Report, Edition 3.1, European Social Survey, http://ess.nsd.uib.no/index.jsp

ESS 2006 Documentation Report, Edition 3.1, European Social Survey, http://ess.nsd.uib.no/index.jsp

ESS 2007 European Social Survey, Round 4 Specification for participating countries. http://www.europeansocialsurvey.org/index.php?option=com_content&view=article&id=74&Itemid=350

Festinger, L. (1954) A theory of social comparison process. *Human Relations* **7** (2): 117-140

Garfinkel, H. (1967) Studies in ethnomethodology. Englewood Cliffs, NJ: Prentice-Hall

Groves, R., Cialdini, R., and M. Couper (1992). Understanding the decision to participate in a survey. *Public Opinion Quarterly* **56** (4): 475–495

Groves, R. and M. Couper (1996) Contact-Level Influences on Cooperation in Face-to-Face Surveys. *Journal of Official Statistics*, **12** (1): 63-83

Groves, R. and M. Couper (1998) *Nonresponse in Household Interview Surveys.* Wiley, New York

Groves, R., F. Fowler, M. Couper, J. Lepkowski, E. Singer and R. Tourangeau (2004). *Survey Methodology.* Wiley Series in Survey Methodology

Groves, R. and K. McGonagle (2001) A Theory-guided Interviewer Training Protocol regarding Survey Participation. *Journal of Official Statistics* **17** (2): 249-265

Groves, R., S. Presser and S. Dipko (2004) The Role of Topic Interest in Survey Participation Decisions. *Public Opinion Quarterly* **68** (1): 2–31

Harkness, J. (1999) In Pursuit of Quality: Issues for Cross-National Survey Research. *Int. J. Social Research Methodology* **2** (2): 125-140

Hill, D. and R. Willis (2001) Reducing Panel Attrition. A Search for effective Policy Instruments. *Journal of Human Resources* **36** (3): 416-438

Hox, J., E. De Leeuw and I. Kreft (1991) The Effect of Interviewer and Respondent Characteristics on the Quality of Survey Data: A Multilevel Model. In: Biemer, Groves, Lyberg, Mathiowetz and Sudman (eds.): *Measurement Errors in Surveys*. New York: Wiley

Keeter, S., C. Miller, A. Kohut, R. Groves and S. Presser (2000) Consequences of Reducing Nonresponse in a National Telephone Survey. *Public Opinion Quarterly* **64** (2): 125-48

Kiraz, E and S. Yildirim (2007) Enthusiasm vs. Experience in Mentoring: A Comparison of Turkish Novice and Experienced Teachers. *Asia Pacific Education Review* **8** (2): 250-261

Lipps, O. and G. Benson (2005) Cross National Contact Strategies. Proceedings of the American Statistical Association, Section on Survey Research Methods. Alexandria, VA: American Statistical Association: 3905–3914

Loosveldt, G., M. Philippens, I. Stoop and J. Billiet (2004) Refusal conversion procedures in the European Social Survey. Paper for the International Workshop on Household Nonresponse in Leuven, Belgium, 22-24 September, 2004

Lynn, P. and P. Clarke (2002) « Separating Refusal Bias and Non-Contact Bias: evidence from UK national surveys », *The Statistician* **51** (3): 319-333

Medoff, L. and Abraham, K. G. (1980) Experience, Performance, and Earnings. *The Quarterly Journal of Economics* **95** (4): 703-736

Morton-Williams, J. (1993) *Interviewer Approaches*. Dartmouth Publ. Comp. Ltd., Aldershot

Olsen, K. and A. Peytchev (2007) Effect of interviewer Experience on interview pace and interviewer attitudes. *Public Opinion Quarterly* **71**(2): 273-286

O'Muircheartaigh C. and P. Campanelli (1999) A Multilevel Exploration of the Role of Interviewers in Survey Non-Response. *Journal of the Royal Statistical Society A* **162** (3): 437-446

Philippens, M. and J. Billiet (2003) Monitoring and Evaluating Non-response Issues and Fieldwork Efforts in the European Social Survey: Results of the analysis of contact form data, University of Leuven

Pickery, J. and G. Loosveldt (2002) A Multilevel Multinomial Analysis of Interviewer Effects on various Components of unit Nonresponse. *Quality & Quantity* **36**: 427-437

Pickery, J., G. Loosveldt and A. Carton (2001) The Effects of Interviewer and Respondent Characteristics on Response Behavior in Panel Survey*s. Sociological Methods & Research* **29** (4): 509-523

Purdon, S., P. Campanelli and P. Sturgis (1999) Interviewer's calling Strategies on Face-to-face Interview Surveys. *Journal of Official Statistics* **15** (2): 199-216

Raz, A. E. and E. Blank (2007) Ambiguous professionalism: managing efficiency and service quality in an Israeli call centre, New Technology. *Work and Employment* **22** (1): 83 - 96

Snijders, T. and R. Bosker (1999) *Multilevel analysis*. Newbury Park, California: Sage

Snijkers, G., J. Hox and E. De Leeuw (1999) Interviewer's tactics for Fighting Survey Nonresponse. *Journal of Official Statistics* **15** (2): 185-198

Symons, K., H. Matsuo K. Beullens and J. Billiet (2008) Response Based Quality Assessment in the ESS – Round 3. An update for 19 countries. Center of Sociological Research (CeSO), K.U. Leuven, Belgium

Tourangeau, R. (2004) Survey Research and Societal Change. *Annual Review of Psychology* **55**: 775-801

Tables

Country		Switzerland		Germany		Denmark		Finland		All 4 c'tries	
ESS round		2	3	2	3	2	3	2	3	2	3
Refusal Conv. Exp. ("Seniors")	Juniors	19	18	72	105	65	75	18	26	174	224
	Seniors	36	24	95	69	23	17	108	108	260	218
Number of previous ESS-rounds	1. round	28	18	121	121	53	55	5	3	207	197
	2. round	27	11	44	30	35	19	121	20	227	80
	3. round	0	13	0	23	0	18	0	111	0	165

Table 1: Number of Interviewers working for the ESS, by Country and Round, and by Seniority (also works on re-issued Cases), and by Number of ESS Rounds.

Country		Switzerland		Germany		Denmark		Finland		All 4 c'tries	
ESS round		2	3	2	3	2	3	2	3	2	3
Refusal Conv. Exp. ("Seniors")	Juniors	35	43	43	60	74	82	14	19	40	51
	Seniors	65	57	57	40	26	18	86	81	60	49
Number of previous ESS-rounds	1. round	51	43	73	70	60	60	4	2	48	45
	2. round	49	26	27	17	40	21	96	15	52	18
	3. round	0	31	0	13	0	20	0	83	0	37

Table 2: Percentage of Interviewers by Seniority, and by Number of ESS Rounds, working for the ESS, by Country and Round.

	Switzerland	Germany	Denmark	Finland	All 4 c'tries
Ever obtaining Contact with Household					
Senior	.94 (.003)	.93 (.003)	.94 (.007)	.97 (.002)	.94 (.002)
Previous ESS experience	.94 (.003)	.94 (.004)	.95 (.004)	.97 (.002)	.95 (.002)
All Others	.90 (.007)	.92 (.004)	.92 (.006)	.98 (.002)	.92 (.003)
Ever obtaining Contact with Target Person conditional on Contact with Household					
Senior	.75 (.006)	.89 (.004)	.94 (.006)	.96 (.003)	.87 (.003)
Previous ESS experience	.76 (.007)	.89 (.005)	.95 (.005)	.96 (.003)	.89 (.003)
All Others	.66 (.012)	.88 (.006)	.94 (.005)	1	.85 (.004)
Ever obtaining Cooperation of Target Person conditional on Contact with Target Person					
Senior	.76 (.007)	.66 (.007)	.70 (.013)	.71 (.007)	.70 (.004)
Previous ESS experience	.79 (.007)	.64 (.008)	.67 (.010)	.71 (.006)	.71 (.004)
All Others	.74 (.015)	.57 (.009)	.66 (.011)	.72 (.064)	.63 (.006)

Table 3: Mean Performance Measures, Standard Errors in Brackets. Round 2 and 3 pooled.

Country	Switzerland	Germany	Denmark	Finland	All 4 c'tries
Senior	1.36	1.02	.97	1.09	1.22*
Previous ESS experience	1.11	1.41**	2.06**	.99	1.66**

Table 4: 2-Level Logit Models experienced/senior Dummy Odds Ratios of ever obtaining Contact with Households. * (**) = Dummy significant on 5% (1%) Level. Round 2 and 3 pooled.

Country	Switzerland	Germany	Denmark	Finland	All 4 c'tries
Senior	.026	.009	.116	-.025	.030
Previous ESS experience	.080	-.129*	-.056	.081	-.032

Table 5: 2-Level Logit Models: Delta of Interviewer Variance Proportion (ICC) of ever obtaining Contact with Households of Model with experienced/senior Interviewers vs. complementary Group. * (**) = delta significantly different on 5% (1%) Level. Round 2 and 3 pooled.

Country	Switzerland	Germany	Denmark	Finland	All 4 c'tries
Senior	1.16	1.18	.99	1.15	1.30**
Previous ESS experience	1.41**	1.01	1.03	1.66	1.59**

Table 6: 2-Level Logit Models experienced/senior Dummy Odds Ratios of ever obtaining Contact with Target Person. * (**) = Dummy significant on 5% (1%) Level. Round 2 and 3 pooled.

Country	Switzerland	Germany	Denmark	Finland	All 4 c'tries
Senior	-.023	-.022	-.071*	.026	.023
Previous ESS experience	.090	-.008	.036	-.137	.077

Table 7: 2-Level Logit Models: Delta of Interviewer Variance Proportion (ICC) of ever obtaining Contact with Target Person of Model with experienced/senior interviewers vs. complementary Group. * (**) = delta significantly different on 5% (1%) Level. Round 2 and 3 pooled.

Country	Switzerland	Germany	Denmark	Finland	All 4 c'tries
Senior	.99	1.60**	1.19	.91	1.36**
Previous ESS experience	1.14	1.18	.97	1.00	1.23**

Table 8: 2-Level Logit Models experienced/senior Dummy Odds Ratios of ever obtaining Cooperation of Target Person given Contact with Household. * (**) = Dummy significant on 5% (1%) Level. Round 2 and 3 pooled.

Country	Switzerland	Germany	Denmark	Finland	All 4 c'tries
Senior	.050	-.100*	.028	-.007	-.068*
Previous ESS experience	.033	-.092*	-.015	-.030	-.064*

Table 9: 2-Level Logit Models: Delta of Interviewer Variance Proportion (ICC) of ever obtaining Cooperation of Target Person given Contact with Household (ICC) of model with experienced/senior interviewers vs. complementary Group. * (**) = delta significantly different on 5% (1%) Level. Round 2 and 3 pooled.

The Cross-National Equivalent File and its Member Country Household Panel Studies[1]

By Joachim R. Frick, Stephen P. Jenkins,
Dean R. Lillard, Oliver Lipps and Mark Wooden

1 *Introduction*

Over the past thirty years industrialized nations have increasingly invested resources to develop and maintain general purpose social science surveys of households and individuals. This investment, in many cases driven by the scientific communities, has allowed academic and government researchers to document and track how socio-economic characteristics of a country's population are evolving, to measure how behavior changes when social policies are introduced or changed, and to build models that can be used to estimate how alternative social policies might change behavior. These data have not only sparked policy and behavioral studies within each country, but have also increased studies of policy and behavior across countries.

To use country-based survey data for cross-national research, researchers must determine the extent to which the information in the data sets is or can be made comparable. That task involves substantive methodological issues, most of which involve equilibrating already collected data. Because the country surveys have been established with national policy and research goals in mind, they have generally not been designed ex-ante to explicitly

[1] The Cross-National Equivalent File (CNEF) has been funded over the years by the US National Institute on Aging, the German Institute for Economic Research (DIW Berlin) and Cornell University. This project is a collaborative effort with researchers at the six CNEF partner institutions: Cornell University; SOEP at DIW Berlin; Statistics Canada; the Institute for Social and Economic Research (ISER) at the University of Essex; the Melbourne Institute of Applied Economic and Social Research at The University of Melbourne; and the University of Neuchâtel. Our thanks to Richard V. Burkhauser, Gaétan Garneau, Robert Schoeni and Gert G. Wagner for their comments on previous drafts of this paper.

generate data that are comparable across countries. The two exceptions are the European Community Household Panel (ECHP) and a cohort study – the Survey of Health, Ageing, and Retirement in Europe (SHARE). The ECHP was only partly successful and has been abandoned. SHARE has been more successful but, because it focuses on older respondents, cannot be used to study the broader population.[2]

Because most data have to be harmonized ex-post, cross-national researchers must invest considerable time and effort to define variables that measure equivalent concepts and behavior. This task is straightforward for basic concepts like age and gender. The task of creating equivalent measures is much more complicated for concepts that are defined in the context of country-specific institutions or have a cultural basis. Cross-nationally comparable measures of many concepts, such as economic well-being, education, employment and health, can only be derived with considerable effort ex-post because the data collected on them in each country-based survey flows from, and is shaped by, culture and country-specific institutions. That effort requires researchers to learn the institutions, laws, and cultural patterns of each country.

One of the first efforts to create cross-nationally comparable data was the Luxembourg Income Study (LIS). Begun in 1983, the LIS harmonizes nationally representative micro-level survey data for over 30 countries (see www.lisproject.org and Smeeding, Jesuit, and Alkemade 2002). Because the LIS bears the substantial costs of harmonizing data, it dramatically reduces the burden individual researchers must bear.

While the standardized LIS data are impressive, it cannot meet some goals of the cross-national research community. For example, the LIS allows researchers only indirect access to the underlying confidential microdata which in several cases is official data. Further,

[2] See Burkhauser and Lillard (2005) for a detailed discussion of the successes and failures of efforts to create both ex ante and ex post harmonized data sets for cross-national research, and Lillard and Burkhauser (2006) for an evaluation of SHARE's success in creating ex-ante harmonized data.

researchers cannot easily get access to the original data sources. This limitation means that most researchers must accept the LIS standardization rules. Finally, and perhaps most importantly, the LIS data are cross-sectional, and so do not serve researchers interested in longitudinal analyses.

Here we describe a project built on the LIS model that overcomes the above limitations. This project is the Cross-National Equivalent File (CNEF), a cooperative effort of individuals and institutions that collect panel survey data in (as of 2007) six different countries: the Panel Study of Income Dynamics (PSID) for the United States; the Socio-Economic Panel Study (SOEP) for Germany; the British Household Panel Survey (BHPS) for Great Britain; the Survey of Labour and Income Dynamics (SLID) for Canada; the Household, Income and Labour Dynamics in Australia (HILDA) Survey for Australia; and the Swiss Household Panel (SHP) for Switzerland.[3] The CNEF harmonizes data common to two or more of the country-based surveys, allows researchers access to both the harmonized and original data, provides all harmonization algorithms to interested researchers, and focuses on some of the most successful nationally representative ongoing longitudinal micro-data sets in the world.

The CNEF differs from other standardization projects not only because it includes data from ongoing panel studies, but also because the development and expansion of the equivalized variable set is largely driven by research questions. Equivalently defined variables are added when researchers develop cross-nationally comparable measures as part of a particular research project. Because those researchers are experts on the topic of their study, they not only inform themselves of specific country institutions but also bring their topic-

[3] The CNEF is administered at Cornell University in close collaboration with researchers at the Socio-Economic Panel Study at the German Institute for Economic Research (DIW Berlin) in Berlin, the Institute for Social and Economic Research (ISER) at the University of Essex, Statistics Canada in Ottawa, the Survey Research Center at the University of Michigan, the Melbourne Institute of Applied Economic and Social Research at the University of Melbourne, and the University of Neuchâtel.

specific expertise to bear. Consequently, the harmonized data included in the CNEF are an amalgam of the knowledge of many researchers answering a diverse set of questions. Just as importantly, the CNEF continuously evolves as researchers refine and add to the set of harmonized variables.

The CNEF is also distinguished by its inclusion of data on the same person over many years. These longitudinal data make it possible for cross-national researchers to use more powerful statistical methods to better control for otherwise unobserved person-specific heterogeneity in behavior. Furthermore, these panels allow researchers to exploit policy variation not only across countries but also over time; variation that yields a richer understanding of human behavior. Finally, the design of each country's survey allows researchers to follow families across multiple generations. Consequently, the CNEF is increasingly used to study, from a cross-national as well as a cross-disciplinary perspective, how socio-economic status is correlated and transmitted across multiple generations[4].

2 Evolution of the CNEF

Begun in 1991 with funding from the National Institute on Aging[5], the CNEF has expanded from a set of variables harmonized across just two countries – the US and Germany – to a set of variables harmonized across six countries. Data from the BHPS in Britain and SLID in Canada were added in 1999, with data from the HILDA Survey in Australia following in 2007. Data from the SHP in Switzerland will be added in late 2007.

The set of harmonized variables included in the CNEF has grown from a core set of income and demographic variables to a set that includes multiple measures of health, geographic residence, and other characteristics. The original core variables to be harmonized

[4] See e.g. Butz and Torrey (2006).

[5] Principal investigators were Richard V. Burkhauser and Gert G. Wagner, the director of the German SOEP. Special thanks go to Richard Hauser, then University of Frankfurt, for his important initial support of this cross-national endeavour.

were income and demographic characteristics of respondents to the PSID and SOEP, and reflects the objectives of the original project that motivated the creation of the CNEF – to compare and understand income-based inequality and income mobility in the US and Germany.[6]

Because this research topic was of interest more broadly, the CNEF naturally expanded to include both the BHPS and SLID. This extension was natural because much of the income focus of the PSID and SOEP was also present in the BHPS and SLID – surveys with designs that were informed by the experiences of the PSID and SOEP. As a consequence, many studies began to also compare economic well-being, wage, and income mobility in the US, Canada, Great Britain and Germany (see, for example: Jenkins and Schluter, 2003; Jenkins, Schluter, and Wagner, 2003; Burkhauser, Giles, Lillard, and Schwarze, 2005). Over time, additional variables have been harmonized. The most recent expansion of the CNEF variables took place in 2003 when harmonized versions of health variables available in any two of the then four country-based panel studies were created (see Lillard and Burkhauser 2005). Data from the HILDA Survey were added for the first time in 2007 and used to compare how employment and earnings of workers with and without disabilities varies across time and countries (see Burkhauser, Schmeiser, and Schroeder 2007). The most recent addition to the CNEF – the SHP – is in the process of taking place as this article goes to press. Data from the SHP will be included in the next release of CNEF, scheduled for late 2007 / early 2008.

At its next release, the CNEF will include data from 1980-2005 for the PSID, 1984-2006 for the SOEP, 1991-2005 for the BHPS, 1992-2003 for SLID, 1999-2005 for the SHP, and 2001-2005 for the HILDA Survey. Sample sizes of individual respondents (adults and

[6] Much of the early work comparing economic well-being and wage and income mobility in the United States and Germany in the 1980s and early 1990s used these harmonized data. See Burkhauser, Frick, and Schwarze (1997) and Burkhauser, Crews-Cutts, and Lillard (1999).

children) by year up to 2005 are listed in Appendix 2 – pooled across all six surveys the total number well exceeds 2.3 million person-year observations.

3 Design and content of the CNEF

The CNEF is designed to facilitate cross-national research by social scientists, regardless of their experience with panel data methods. To achieve this goal, which includes research as well as "capacity building", the CNEF collects data from the different surveys that can be used to create comparably defined variables in a most userfriendly manner. It puts these variables into data files – one for each year for each country – which researchers can analyze either as stand-alone data files or, as is commonly the case, with other data of interest. Frequently researchers merge country-specific policy information into the CNEF files. Often they extract other data from one or more of the original country data files and merge them into the CNEF file which in this case is used as some kind of *navigation* or *master* file.

The design of the CNEF facilitates the work of less experienced researchers because the variables in each data file have identical names, labels, and value formats. The variable names reflect the variable's content. The first letter of the variable name represents the variable's category – demographic (D), employment (E), household composition (H), income (I), weighting (W), sample identifiers (X), location (L), health (M), and macro-level indicators (Y) – and the last four digits of each variable name indicate the survey year from which the variable was drawn. This parallel structure allows researchers to use the same computer programs to analyze data from all panels – eventually by just one single run.

The CNEF is also designed so that more experienced researchers can quickly and easily modify algorithms used to create variables or add other data to supplement existing variables. A CNEF codebook identifies the algorithm used to construct each comparably defined variable. That algorithm names the variables from the original files that are used. It also allows researchers to modify the way any particular variable is constructed. To allow researchers to supplement existing data with data from the original "parent" surveys, the CNEF includes the unique person and yearly household identifiers from the original surveys. This aspect of the data thus allows researchers to check whether particular results are robust to small changes in how variables are defined and it allows them to develop their own measures when they believe the existing variable construction can be improved.

In addition to the algorithm used to construct variables, each variable is assigned a reliability code that represents the degree of cross-national comparability that the surveys permit. For example, a code of "1" indicates that the variables are judged to be completely comparable, whereas a code of "4" indicates that there is no comparable variable between the surveys. CNEF researchers set these reliability codes using their experience, judgment, direct comparisons of the survey instruments, and knowledge of institutional differences across the countries.

A distinguishing feature and major innovation of the CNEF is that it includes a set of constructed variables that are not directly available in any of the original surveys. These variables include measures of household income before and after taxes, estimated household tax burdens and household size adjusted median income for the population. Many of these variables cannot be computed without significant effort on the part of individual users because they require the estimation of taxes paid by each household. The construction of the tax burdens is one of the innovative contributions of the CNEF that make it possible to compare disposable income across countries. It is also an example of how the CNEF, like LIS, reduces the burden each cross-national researcher faces.

The effort required to compute after-tax income varies across the different country panel surveys. In SLID and SHP[7] taxes paid are collected as part of the survey. In the other data sets household tax burdens have to be estimated.

Tax simulation programs for the BHPS, SOEP, and HILDA Survey were written by researchers in each institute responsible for the survey data. Stephen Jenkins and coauthors at the University of Essex wrote and update the tax estimation routine for the BHPS (Levy et. al. 2006);[8] Johannes Schwarze of Bamberg University wrote, and Markus Grabka of the DIW Berlin updates, the tax routine for the SOEP (Schwarze 1995); and Bruce Headey of the Melbourne Institute at Melbourne University wrote the tax simulation program for the HILDA Survey (Headey 2003). In the case of the PSID, prior to1993 tax burdens were estimated by the PSID staff and included in the public data release. Since 1993, however, the PSID data have not included tax burden estimates. To estimate household tax burdens in the PSID, Dean Lillard at Cornell University uses the National Bureau of Economic Research tax simulation program, TAXSIM (see Feenberg and Coutts, 1993).[9] TAXSIM has thus been used to estimate PSID household taxes for all years in CNEF.

Even more effort is required to compute measures of post-tax income in the SOEP, since all income variables in the SOEP are reported as average monthly amounts received during the previous year. Thus, for cross-national comparability, income must be annualized by calculating the number of months in each year various types of income are received and multiplying this number by the reported respective average monthly amount. The tax simulation program produces estimated annual tax burdens for all households in the SOEP.

---

[7] With the exception of social security income which are estimated by the SHP researchers.
[8] BHPS "Net income" files: can be downloaded directly from the UK Data Archive at http://www.data-archive.ac.uk/findingData/snDescription.asp?sn=3909 with documentation at: http://www.data-archive.ac.uk/doc/3909/mrdoc/pdf/3909userguide.pdf).
[9] Butrica and Burkhauser (1997) discuss in detail the NBER and PSID tax calculation algorithms and compare PSID taxes estimated by TAXSIM with the PSID estimates from 1980 through 1992.

These annual tax values are combined with the annualized components of income to create a measure of household post-government income.

The construction of tax burdens and the collection of income from public and private transfers make it possible for the CNEF to produce and distribute unique measures of household income. For example, the CNEF produces a measure of total household income after taxes and transfers (and simply labeled post-government income). This measure is the sum of labor earnings, asset flows, private transfers, public transfers, and other income of all individuals in a given household minus income and payroll taxes (non-cash income advantages given by imputed rental value of owner-occupied housing are available as a separate variable). All household-level income variables[10] are assigned to each individual in the household.

Appendix 1 lists the variables currently included in the CNEF. For each variable we describe the variable, indicate which country data files have valid data, and list the variable name, and unit of analysis for which data are measured. Note that the CNEF codebooks also include some relevant macro-level information for each country, such as the consumer price index for each year. Because these data do not vary across sample members, they are only included in the codebooks. Appendix 2 lists the basic sample sizes included in each of the CNEF country files.

4 Household panel studies in CNEF

All six panel surveys in CNEF collect information on household composition, income, employment, housing, and demographic characteristics. However, differences exist not only in the type and manner of the questions asked across surveys but also within those surveys

[10] In general, the definition of CNEF income variables follows the recommendations of the "Canberra Group on Household Income Measurement" (Canberra Group 2001). Making use of the longitudinal nature of the underlying data missing income information arising from item-non-response is corrected for by means of imputation routines. See Frick and Grabka (2007) for a comparative analysis focusing on the need of harmonized imputation techniques in cross-national databases.

over time. Hence some variables that are comparable across surveys in some years will not be comparable in other years.

To provide some flavor of the overall comparability of data across the six country data sets, Table 1 compares their key features. All surveys except SLID follow members of the original sample households and all offspring of those sample members.[11] The surveys use different rules about which other household members are followed and they differ in who is interviewed. The BHPS, SOEP, HILDA Survey, and SHP interview all adults in each household. All six surveys collect information about adults who join an existing sample household. The BHPS, SOEP, HILDA Survey, and SHP collect that information directly because they interview all adult household members. The PSID only interviews one member of the household while SLID allows proxies to be interviewed. SLID also differs from the other surveys in that its sample consists of respondents to two six-year panels that overlap by three years.

Five surveys have also varied the method they use to collect data during the life of the panel. The older surveys initially interviewed respondents using face-to-face paper and pencil interviewing (PAPI) techniques before switching, mostly in the 1990s, to computer assisted methods. Perhaps the most important mode distinction concerns whether interviews are conducted in person (i.e., face-to-face) or by telephone. The BHPS, SOEP, and HILDA Survey are primarily conducted in person, increasingly with the assistance of a laptop computer (computer-assisted personal interviewing, or CAPI). Mixed-mode surveying takes place in SOEP and HILDA, with self-completion becoming more prevalent in SOEP while the HILDA Survey has been slowly increasing its use of telephone interviews because of the costs of following respondents over time as they move away from clusters of households in

[11] SLID follows only original household members but not their offspring for a maximum of six years. However, they are included as a joiner/cohabitant. They have positive cross-sectional weights but longitudinal weights are equal to 0. A new panel that represents half of the sample is started every three years.

the initial sample area. Almost 7 percent of all wave six HILDA interviews were conducted by telephone. The PSID converted from PAPI to telephone interviewing in 1973 and switched to computer-assisted telephone interviewing (CATI) in 1993. Both SLID and the SHP, which began in 1992 and 1999 respectively, have used a CATI system since their inception.

The period within a year over which each survey is in the field varies across surveys.[12] Data collection for the SOEP and PSID is concentrated in the first four months of the year. In contrast the BHPS concentrates data collection in the autumn of each year. The main fieldwork period for the HILDA Survey is September through December and it is September to February for the SHP. In part, these differences are motivated by the varying national definitions of the financial year.

The studies also vary in their experiences with respect to response and, to a slightly lesser degree, attrition. Across the six surveys, wave 1 response rates appear to average somewhere around 70 percent depending on how it is measured. Full household response rates (i.e., the proportion of sampled households where all eligible members responded) vary from about 50 percent in the SHP[13] and 59 percent in the HILDA Survey, up to 76 percent in the case of the PSID. In the BHPS and SOEP, interviews were completed with all household members at 69 and 65 percent of cases respectively.[14] Wave 1 response rates for both the BHPS and SOEP compare quite favorably with the PSID, especially given that in the PSID an interview is only required from one family member. Wave 1 response rates are lower in the more recently fielded samples, both across countries and, as can be seen in the case of refreshment samples in the SOEP, within countries.

[12] While all six panels collected data annually when they started, the PSID moved to a biennial interview schedule in 1997.
[13] Note that in the SHP with the CATI technique, all households that could not be contacted are treated as not responding, irrespective of eventual nonsample cases.
[14] The initial response rates for the two original sub-samples in the SOEP were 61 per cent for "West Germans" and 68 percent for "foreigners".

Because, in most panel surveys, attrition typically stabilizes after a few waves at quite low rates (typically at around 4 per cent or better per year), attrition rates do not vary as much across the CNEF country samples. For example, response rates (for the unbalanced panel) for wave 5 in the SOEP, BHPS, and HILDA Survey range between 71 and 74 percent. Wave 5 response rates, however, are much lower in the SHP (56 percent) and much higher in both the SLID and PSID (around 80 percent). In part, these higher response rates reflect the collection of information from only one household member, in the case of the PSID, and permitting one household member to be a proxy respondent for other household members, in the case of the SLID (about 30 percent of cases are reported by proxy). Nevertheless due to demographic losses (death and emigration) as well as panel attrition there is a consistent deterioration in the size of the original sample over the life of all panel surveys. With respect to the development of cross-sectional sample size these negative developments are at least partly countered by births and new persons joining existing survey households.

The surveys differ with respect to sample enhancements and the introduction of top-up samples. Partly in response to questions of whether the PSID sample failed to adequately represent the immigrant population, the PSID added a Latino sample in 1990 (later dropped) and a general immigrant sample in 1997 that continues. The SOEP has a well established tradition of adding new representative samples (in 1998, 2000, 2006) and in over-sampling specific subgroups of interest and policy relevance such as immigrants (in 1984, 1995) and high-income households (in 2002). Similarly, the BHPS has both added and dropped new sub-samples targeted to represent low-income households and the UK population. Sample replenishment is largely irrelevant for SLID given it uses overlapping panels of relatively short duration, and is not yet relevant for the HILDA Survey given its young age. The SHP, however, is also relatively young, but because of high attrition, recruited a refreshment sample in 2004 that was representative of the non-institutionalized Swiss population.[15]

There are several arguments in favor of such sample additions, especially in long-running panels. In addition to simply enhancing sample size, refreshment samples can be used to empirically test for panel effects in the old samples (see e.g. Frick et al., 2006). Refreshment samples also help correct for the loss of cross-sectional representativeness that occurs because of recent immigration (since the "old" samples were drawn).

The addition of refreshment samples supplements the birth of new households in each panel as household members split off to form their own households. Both sources of new households and natural sample attrition mean that sample sizes for each country file (see Appendix Table 2) in the most recent wave of CNEF data differ considerably from the sample sizes that were present in each surveys' first wave. For example, the birth of new households and the addition of new (refreshment as well as top-up) subsamples in the SOEP resulted in about 12,500 household interviews in 2006, up from roughly 6,000 in wave 1 in 1984.

Finally, the studies also vary markedly with how they are governed and administered. The SLID is run by a national statistical agency and hence internalizes all data collection functions. Similarly, the PSID scientific leadership and data collection activities are managed and conducted by the same academic institution – the Institute for Social Research at the University of Michigan. The institutes that administer the SOEP, BHPS, the HILDA Survey, and SHP contract with private firms to collect the data for them. Once the data are collected, they are also coded and edited in different ways. The host organizations of the SLID, HILDA, SHP, PSID, and BHPS for the most part code and edit data at their respective institutions. By contrast, data editing and coding for the SOEP is largely left to the contracted fieldwork agency while imputation and weighting procedures are in-house activities.

---

[15] Sections 5.1-5.6 below provide more details.

5 Specificities of the national panels contributing to the CNEF

Above and beyond the survey characteristics mentioned above, the CNEF country panels are living surveys that are continually evolving in emphasis and range of the surveyed concepts. These changes are driven by the needs of policy makers and researchers in their own countries. This evolution will necessarily require CNEF to continually work to harmonize these evolving data for cross-national research. The next section provides a short overview of survey specific developments not yet considered in the CNEF.

5.1 The PSID *(psidonline.isr.umich.edu/)*

The PSID began in 1968 with a sample of 5,000 households, which, by design, comprised a disproportionate number of low-income individuals. All current PSID families contain at least one member who was either part of the original 5,000 families or born to a member of one of these families. As of 2005, the PSID has collected information on more than 67,000 individuals spanning as much as 37 years of their lives. The original sampling scheme disproportionately selected individuals from low-income families. A sub-sample of 1,872 low-income families was drawn from an earlier survey conducted by the US Census Bureau. The sample does represent the 1968 United States population if this low-income over-sample is excluded or – more efficiently – if researchers use sample weights. Two-thirds of the low-income oversample was dropped in 1997. The PSID added a Latino sample in 1990 but dropped it in 1995 because the sample did not represent all post-1968 immigrants. In 1997 the PSID added a sample of individuals who immigrated to the US after 1968 regardless of their country of origin and these individuals continue to be interviewed. Starting in 1997 the PSID administers its survey every other year.

The content of the PSID has historically focused on the dynamic aspects of economic and demographic behavior, but its content over the past two decades has broadened, including sociological, psychological, and health measures. The central focus of the PSID has been to maintain a clean and consistent time series of core content – income sources and amounts, employment, family composition changes, and demographic events. Other important topics covered by the PSID include housing and food expenditures, housework time, health, and consumption, wealth, pensions and savings. Wealth data for the PSID were collected in 1984, 1989, 1994, and every wave since 1999.

Like the other country surveys, the PSID has evolved in innovative ways. In addition to collecting the wealth information and other new data mentioned above, the PSID added a Child Development Supplement (CDS) first fielded in 1997. This study, which focuses on the human capital development of approximately 3,600 children aged 0-12 in PSID families, includes measures of their cognitive, emotional and physical functioning. These same children where surveyed again in 2002 and 2007. The PSID has also been a leader in tracking information about sample members who have died. PSID staff have worked together with the US Public Health Services, using the National Death Index to obtain information about the date and causes of death of PSID sample members. The long time-series and intergenerational nature of the PSID has also led to special files of the PSID that link household members across multiple generations. These family relationship files are available as public use files.

5.2 *The SOEP (www.diw.de/english/sop/index.html)*

The SOEP fielded its first survey in 1984 with a sample of almost 6,000 households and about 16,000 individuals in the then Federal Republic of Germany. In 1990, only half a year after the fall of the Berlin wall, the SOEP introduced a new sample of almost 2,200 East German households successfully coping with the unique event of the extension of its survey territory. In 2008, the SOEP will collect its 25th year of data. Over the period 1994 to 2001 (i.e., in SOEP waves 11 to 18) SOEP data was harmonized into the format of the European Community Household Panel (ECHP). In 2001 the SOEP began using age-triggered survey instruments when a special questionnaire for teenagers was developed and introduced. In

2003 the SOEP started to collect information from the parents on the lives of their children up to the age of 16 to complement the individual level data that will be collected annually from the youth once they reach age 17. For instance, mothers of newborn babies are now being asked for information on their children beginning at inception. These data are enhanced by follow-up questionnaires once these children reach age two to three (the time they start moving to pre-school institutions), enter school (around age six), move from primary to secondary school (around ages 10 to 12) and in the year before they become respondents on their own. At the same time, SOEP is testing in 2008 the introduction of death-triggered "exit interviews" to capture a final picture of the deceased as well as the economic and social effects of death on surviving household members.

A second strand of current SOEP initiatives focuses on collecting more and better instruments to proxy otherwise unobserved heterogeneity. Thus, in addition to the self reported health related measures (e.g., smoking, alcohol consumption and the introduction of the SF-12), in 2006 the SOEP began to collect measures of grip strength, personality traits, risk awareness, trust and trustworthiness, and cognitive abilities.[16] Discussion about further improvements is underway, for example the introduction of biomarkers (see Lillard and Wagner, 2006).

In 2002, and again in 2007, wealth data was collected at the individual level which – unlike many other studies surveying wealth at the aggregated household level – supports the analysis of intra-partnership wealth inequality. Multiple imputation techniques have been applied to correct for missing data arising from item- and partial-unit-non-response. Finally,

[16] See Wagner, Frick and Schupp (2007) for a fuller discussion of these changes that were developed in collaboration with researchers working in these areas to further ensure their competent and rigorous empirical testing. Comprehensive documentation of the SOEP data is available from www.diw.de/gsoep and in Haisken-DeNew and Frick (2005).

SOEP micro data has been complemented by a survey of the interviewer staff in 2007, thus greatly improving the potential of interviewer-respondent effects.

5.3 The BHPS (www.iser.essex.ac.uk/ulsc/bhps/doc/)

The BHPS began in its fieldwork in the autumn of 1991 and has been following and re-interviewing respondents ever since. The wave 1 sample consists of some 5,500 households and 10,300 individuals drawn from 250 areas of Great Britain. The BHPS was supplemented in wave four to include direct data collection from children in sample households aged 11–15 inclusive, and this survey design has been maintained in subsequent waves. These respondents form what is known as the British Youth Panel (BYP) – these data are not in the CNEF.

From wave 7 the BHPS began providing data for the United Kingdom European Community Household Panel (ECHP). As part of this effort, it incorporated a sub-sample of the original UKECHP, including all households still responding in Northern Ireland, and a ‘low income’ sample of the Great Britain panel. The low-income sample was selected on the basis of characteristics associated with low income in the ECHP. When funding stopped, the sample was discontinued (after wave 11). A major development at wave 9 was the recruitment of two additional samples to the BHPS in Scotland and Wales, containing over 2,000 extra households in each country. At wave 11, the survey was extended to Northern Ireland with the introduction of a sample of around 2,900 households (5,200 persons). Thus from 2001 onwards the survey has therefore been a truly UK-wide survey.[17]

The current tranche of funding for the BHPS, from the UK Economic and Social Research Council, covers fieldwork until Wave 18. Thereafter it is planned that the BHPS sample will be incorporated into a major new household panel survey – the United Kingdom

[17] These samples are included in CNEF. Special weights are also included that researchers must use to generate statistics that represent particular populations.

Household Longitudinal Study, also financed by the ESRC and run by ISER. For further information, see http://www.iser.essex.ac.uk/ukhls/.

The UKHLS is intended to collect data at regular intervals over time about the same 90,000 individuals, from a sample of 40,000 households, making it the largest household panel survey in the world. Initial funding (£15.5 million over five years) supports collection of the first two rounds of interview with each sample member. The study is planned to continue over several decades.

There will be a number of substantial innovations relative to the BHPS and, indeed, many other household panels. First there is the very large sample size, which greatly increases the capacity for research on small-sized groups in the population (e.g., lone parents), or for tailored questions directed at particular subgroups. There is to be an over-sample of ethnic minority groups, where existing UK data is inadequate. Second, it is intended to support collection of a wider range of biomarkers and health indicators than any previous social-science focused survey in Britain. Third, there are to be innovations in data collection, including linkage to external data from administrative data records (e.g., on taxes and benefits received; hospital records and vital statistics) and geo-coded data. There are likely to be additional modes of interviewing other than CAPI. Also being discussed for the future is collection of qualitative and visual data to supplement the quantitative data. In addition there is to be a special panel that will consist of 2,000 households. Known as the "Innovation Panel," it is designed to allow for experiments and continuous methodological development of new survey questions and interviewing techniques.

At the time of writing (September 2007), extensive consultation with potential UKHLS users is underway, with the first fieldwork with the new sample planned for 2008. Current plans are for the BHPS sample to be incorporated in UKHLS wave 2.

5.4 The SLID *(www.statcan.ca/start.html)*

The SLID began in 1993 with a sample of about 15,000 households, containing approximately 30,000 adults. It is run and administered by Statistics Canada. The SLID panel differs from the other surveys in that each panel lasts only six years. In part, the limited length of the panel was chosen to keep the sample population representative of the national population. In 1996, three years after the first panel was surveyed, a second six-year panel was started and the sample sizes were substantially increased as SLID took on the role of providing data for the purpose of cross-sectional estimation of population statistics. Since then a new six-year panel has been launched every three years. This three-year overlap was chosen to maintain continuity in the data. In 2003 more than 95,000 individuals living in more than 38,000 households were interviewed. As in the other surveys, all current SLID families contain at least one member who was part of or born to one of the original household samples that begin each six-year panel.

One of the distinguishing and attractive features of SLID, in addition to its very large sample sizes, is that it links administrative tax records to supplement income data that respondents provide. This feature of SLID means that it has very high quality data on post-government income for SLID respondents who have consented to have their tax information appended (currently about 80 percent of SLID respondents give their consent). While SLID focuses primarily on income and employment (and therefore lacks rich data on health), the quality of its income data is superb.

An exciting development for cross-national research is that, in fall 2008, Statistics Canada will pilot test a new longitudinal survey, the Canadian Household Panel Survey (CHPS). The design and content of the CHPS will be similar to that of the SOEP, the BHPS and HILDA Survey. It will collect information from all household members, follow these respondents for an indefinite period of time, and will collect information on a broader set of topics (including health) than the current edition of SLID. Like SLID, the CHPS will link to administrative records to collect income data. While this survey has not yet been launched or incorporated into the CNEF, the expected design and content of the CHPS will more closely align with the CNEF country surveys.

5.5 The HILDA Survey *(http://melbourneinstitute.com/hilda/)*

The HILDA Survey began in 2001 with a sample of almost 7,700 households. The wave 1 sample includes data on 19,914 individuals ages 15 and older from all but the remotest parts of Australia. Now in its 7th wave, the HILDA Survey has continued to evolve and mature.

The design and structure of the HILDA Survey parallels the design and structure of its older siblings, especially BHPS and SOEP. Nevertheless there are important differences. For example, most of the panels now collect data on household wealth but none of the other panels collected such data so early in the life of the panel (wave 2) or collect as much detail. The HILDA Survey also now collects (starting wave 5) much more detail about household expenditure than any of the other studies. This is achieved by means of a supplementary self-administered questionnaire, as is also done in the BHPS, but the amount of information collected via this instrument is far greater in the HILDA Survey.

The HILDA Survey is also governed differently than the BHPS, PSID and SOEP. Like SLID, the HILDA Survey is owned and controlled by its government. As such, the design and content of the HILDA Survey is dictated as much by policy needs as it is by research questions. While all CNEF member panels serve both policy and research needs to varying degrees, the more direct governance of the Australian government means that the HILDA Survey must respond to emerging policy issues. At times this dual focus creates tension between the need to collect data to answer short-term policy questions and the desire to collect data to meet longer-term research objectives, especially given the limited interview time available.

While the funding for HILDA, as with other panel studies, depends in part on the will of political leaders, the immediate future of the HILDA Survey seems secure. Not only has the Australian Government recently committed additional funds to ensure the continuation of the survey until at least wave 12, it increased the level of funding to allow additional respondents to be recruited. A new refreshment sample of about 2,000 households selected from across Australia is thus being planned for wave 9 or 10. This refreshment sample will help ensure the representativeness of the sample in the face of high rates of immigration to Australia.[18]

Attempts will also be made to expand on the limited amount of health-related data currently collected. The main vehicle for achieving this will be a questionnaire module dedicated to health and planned for wave 9.

Finally, and like other surveys, the HILDA Survey is also expecting to switch from pen-and-paper methods to CAPI in the near future. Indeed, a small split sample test was conducted in conjunction with the pilot test for wave 7.

5.6 The SHP (http://www.swisspanel.ch/)

Although the SHP is largely research driven, and funded by the Swiss Science Foundation, it complements data collected by the Swiss Federal Statistical Office. Its main purpose is to ensure a solid database for social reporting about stability and changes in living arrangements and well-being in Switzerland.

Like the HILDA Survey, the design and structure of the SHP Survey both parallels and differs from the design and structure of its older siblings. Perhaps most importantly, the SHP is designed primarily to cover data needs from sociologists and political scientists rather than economists (Zimmermann and Tillmann 2004). Thus income related variable requirements from the CNEF are only partly met in the first few SHP panel waves, but some

[18] Estimates reported by Watson (2006), for example, suggest that after 10 years about 7 percent of the Australian population will be excluded from the coverage of the original HILDA Survey sample.

questionnaire modifications, especially in the 2002 wave, enable satisfactory harmonization possibilities thereafter. Unlike the other panels, the SHP does not employ modularized questionnaires with topics changing between waves, and thus asks the same questions every year. On the other hand, more so than its siblings, the SHP data contain rich subjective measures (e.g. in the health section).

The SHP started in 1999 with a representative sample of more than 5,000 households, in which all individuals aged 14 years or over are to complete the individual questionnaire. A weakness of the SHP is the relatively high attrition which did not decline and stabilize after several waves. Non response seems to be a common problem for surveys in Switzerland. On one hand, this is possibly due to “over-surveying” by market research and administrative surveys in a small country. On the other hand, the highly developed federal system together with the strong tradition of direct democracy fosters a culture where any centralized institution, including surveys, is treated with skepticism and suspicion. As previously noted, the high attrition made a refreshment sample necessary in 2004, adding some randomly selected 2,500 new households. Incentives and other measures introduced since the 2006 wave have facilitated the reintegration of households and individuals who had refused to participate in earlier waves and have also appeared to have reduced the rate of attrition.

Starting in 2008 the SHP will be part of a newly created Centre for Research Infrastructures, tentatively named the *Forschungszentrum Sozialwissenschaften* (FORS). FORS will be housed at the University of Lausanne. Besides the SHP, FORS will also contain the former Swiss Data Archive (SIDOS) and other international surveys in which Switzerland takes part, such as the European Social Survey, the Eurobarometer, and the International Social Survey Program. The housing of FORS at the University of Lausanne is expected to facilitate easy access to the data it houses and generate fruitful exchanges with the national and international academic social science research communities.

6 Looking ahead

The CNEF allows experienced and novice users with an interest in cross-national socio-economic research to perform cross-sectional and longitudinal comparative analyses of Australia, Canada, Germany, Great Britain, Switzerland, and the United States. In contrast to other cross-sectional data files, the CNEF allows researchers substantial freedom to modify the data by providing detailed descriptions of how all variables were created. Since the creation of functionally equivalent variables across countries in the CNEF is research-driven, the data file is accompanied by numerous examples of how each variable is used in a research application. Because the CNEF is continually searching for best practice methods for harmonizing data, future comparative research may result in a revised version of the harmonization procedures currently applied to generate a given variable as well as the addition of new variables.

The CNEF only contains a small subset of the variables included in the PSID, SOEP, BHPS, SLID, HILDA Survey, and SHP data. The number, however, is growing each year as international researchers explore new areas and contribute carefully considered equivalently defined variables, a procedure which only recently began to focus on health.[19] At the same time, the improved interaction of data providers and data analysts currently contributing to the ex-post harmonization of existing survey data will eventually also improve future ex-ante harmonization of new survey features, which in turn will improve cross-country comparability of the micro data and thus will enhance the quality of research results.

[19] Future extensions may consider subjective measures such as "Satisfaction with Life in General" and additional non-cash income components to complement the currently available measure on "Imputed Rental Value of owner-occupied housing" (Variable I11105_xxxx).

7 How to get access to CNEF data

Data availability is influenced by national data privacy regulations. Because the original PSID data are publicly available, we are able to post PSID-CNEF files via the CNEF-website for public use. To access BHPS-CNEF, GSOEP-CNEF, HILDA-CNEF, or SHP-CNEF files you must first apply for and be approved to use these data by the respective country's data manager.[20] Once approved, e-mail or fax the approval documentation to the CNEF-office at Cornell University and you will be sent the CNEF CD. To access SLID-CNEF files you must first be a registered CNEF user. SLID-CNEF data are not included on the CNEF CD but all registered CNEF users can submit their programs to Statistics Canada. Staff at Statistics Canada will run these programs and return log and output files that meet confidentiality requirements.

The one-time registration fee to become a CNEF user is $125 (US), payable to Cornell University. For greater detail on how to access these data, visit the CNEF web page at http://www.human.cornell.edu/che/PAM/Research/Centers-Programs/German-Panel/cnef.cfm or send an e-mail message to <cnef@cornell.edu>.

[20] In principle, this is a formal but rather simple procedure, accomplished in a rather short period of time. A detailed description of the relevant steps is available from the CNEF website at Cornell University.

References

Burkhauser, R.V. / *Crews-Cutt* A. / *Lillard* D.R. (1999) How Older People in the United States and Germany Fared in the Growth Years of the 1980s: A Cross-Sectional versus a Longitudinal View. Journal of Gerontology, September, Vol. 54B, No. 5, S279-S290.

Burkhauser, R. V., *Frick*, J. R. and *Schwarze*, J. (1997): A Comparison of Alternative Measures of Economic Well-Being for Germany and the United States. Review of Income and Wealth, 43(2), 153-171.

Burkhauser, R.V. / *Giles*, P. / *Lillard*, D.R., / *Schwarze*, J. (2005) Until Death Do Us Part: An Analysis of the Economic Well-Being of Widows in Four Countries. Journal of Gerontology: Social Sciences. Vol. 60B No. 5, S238-S246.

Burkhauser, R.V. / *Lillard*, D.R. (2005): The Contribution and Potential of Data Harmonization for Cross-National Comparative Research, Journal of Comparative Policy Research 7, 313-330.

Burkhauser, R.V. / *Lillard*, D.R. (2007): The Expanded Cross-National Equivalent File: HILDA joins its International Peers, The Australian Economic Review 40, 208-215.

Burkhauser, R. V. / *Schmeiser*, M. D. / *Schroeder*, M. 2007. The Employment and Economic Well Being of Working-Age Men with Disabilities: Comparing Outcomes in Australia, Germany, and Great Britain with the United States. Paper presented at the HILDA Survey Research Conference 2007, July 19-20, 2007 University of Melbourne

Butrica, B.A. / *Burkhauser*, R. V. (1997): Estimating Tax Burdens in the PSID Using the TAXSIM Model, Aging Studies Program Project Paper No. 12, Center for Policy Research, The Maxwell School, Syracuse University, Syracuse NY.

Butz, W.B. / *Torrey*, B.B. (2006): Some frontiers in Social Sciences, Science 312 (30 June 2006), 1898-1900

Canberra Group (2001): Expert Group on Household Income Statistics: Final Report and Recommendations, Ottawa.

Feenberg, D. / *Coutts*, E. (1993): An Introduction to the TAXSIM Model, The Journal of Policy Analysis and Management 12, 189-194.

Frick, J. R. / *Goebel*, J. / *Schechtman*, E. / *Wagner*, G. G. / *Yitzhaki*, S. (2006): Using Analysis of Gini (ANoGi) for detecting whether two sub-samples represent the same universe: The German Socio-Economic Panel Study (SOEP) Experience, Sociological Methods & Research 34 (4), 427-468

Frick, J. R. / *Grabka*, M.M. (2007): Item Non-Response and Imputation of Annual Labor Income in Panel Surveys from a Cross-National Perspective. IZA Discussion Paper No. 3043, September 2007, Bonn: IZA. http://ftp.iza.org/dp3043.pdf

Haisken-DeNew, J. P. / *Frick*, J.R. (2005): Desktop Companion to the German Socio-Economic Panel Study (GSOEP), Version 8.0 – Update to Wave 21, DIW Berlin.

Headey, B.W. (2003): How Best to Impute Taxes and Measure Public Transfers? HILDA Discussion Paper Series no. 2/03, Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

Jenkins, S.P. / *Schluter*, C. (2003): Why are Child Poverty Rates Higher in Britain than in Germany? A Longitudinal Perspective, Journal of Human Resources 38, 441-465.

Jenkins, S.P. / *Schluter*, C. / *Wagner*, G. (2003): The Dynamics of Child Poverty: Britain and Germany Compared, Journal of Comparative Family Studies 34, 337-353.

Levy, H. / *Zantomio*, F. / *Sutherland*, H. / *Jenkins*, S.P. (2006): Documentation for Derived Current and Annual Net Household Income Variables, BHPS Waves 1-14. [Retrieved from: http://www.data-archive.ac.uk/doc/3909/mrdoc/pdf/3909userguide.pdf]

Lillard, D. / *Burkhauser*, R.V. (2006): Evaluation of the Cross-National Comparability of the Survey of Health, Ageing, and Retirement in Europe, the Health and Retirement Study, and the English Longitudinal Study of Ageing, Report prepared for the National Institutes of Aging, Washington DC.

Lillard, D / Wagner, G.G. (2006); The Value Added of Biomarkers in Household Panel Studies. Data Documentation No. 14, Berlin: DIW Berlin.

Schwarze, J. (1995): Simulating the Federal Income and Social Security Tax Payments of German Households Using Survey Data, Cross-National Studies in Aging Program Project Paper No. 19, Center for Policy Research, The Maxwell School, Syracuse University, NY.

Smeeding, T.M. / *Jesuit*, D. K. / *Alkemade*, P (2002): The LIS / LES Project Databank: Introduction and Overview. Schmollers Jahrbuch (Journal of Applied Social Science Studies) 122, 497-717.

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

Watson, N. (2006): Options for a Top-up Sample to the HILDA Survey, Paper presented at the ACSPRI Social Science Methodology Conference, University of Sydney, 10-13 December.

Zimmermann E. / *Tillmann*, R. (eds) (2004): Vivre en Suisse 1999-2000 _ Leben in der Schweiz 1999-2000. Une année dans la vie des ménages et familles en Suisse. Ein Jahr im Leben der Schweizer Familien und Haushalte (Vol. 3. Population, Famille et Société), Peter Lang Verlag AG, Berne.

Table 1: Key Features of the CNEF Member Panels

Feature	*PSID*	**SOEP**	*BHPS*	*SLID*	*HILDA Survey*	*SHP*
Host organization	Institute for Social Research, University of Michigan	SOEP at German Institute for Economic Research (DIW Berlin)	Institute for Social and Economic Research, University of Essex	Statistics Canada	Melbourne Institute of Applied Economic and Social Research, University of Melbourne	Swiss Household Panel, University of Neuchâtel
Funding source	National Science Foundation, National Institute of Health, plus range of other organizations[(a)]	1984 to 2002: German National Science Foundation (DFG) and Federal Ministry of Education and Research (BMBF) 2003 on: Leibniz Association (WGL) [(b)]	UK Economic and Social Research Council	Statistics Canada	The Australian Department of Families, Community Services and Indigenous Affairs and the Reserve Bank of Australia (for wave 2 in 2002)	Swiss National Science Foundation (mainly), Swiss Federal Statistical Office, and University of Neuchâtel
Design	Indefinite life panel	Indefinite life panel	Indefinite life panel	Overlapping 6-year panels	Indefinite life panel	Indefinite life panel
Year of first interview	1968	1984	1991	1993	2001	1999
Reference population / data collection unit	Heads of family units who have been continuously resident in the USA for at least 2 years.	All private households. All members aged 17 years or over are interviewed.	All private households. All members aged 16 years or over are interviewed.	Private households in the 10 provinces, with the exception of the Indian reserves. All members aged 16 years or over are interviewed. Proxy interviews are accepted.	All private households, excluding those in remote parts of Australia. All members aged 15 years or over are interviewed.	All private households. All members aged 14 years or over are interviewed.
Collection	Waves 1-5 (1968-	Waves 1-14 (1984-	Waves 1-9 (1991-1999)	Since wave 1	Since wave 1 (2001)	Since wave 1 (1999)

Feature	*PSID*	**SOEP**	*BHPS*	*SLID*	*HILDA Survey*	*SHP*
mode	1972) PAPI Since wave 6 (1973) Mainly telephone Since wave 26 (1993) CATI	1997) PAPI Since wave 2 (1985) mixed mode (face-to-face and self-completion) Since wave 15 (1998) Began migrating to CAPI	PAPI plus short self-completion questionnaire. since wave 10 (2000) CAPI Since wave 3 (1993) use short telephone interview as last resort.	(1993) CATI.	PAPI plus self-completion questionnaire. Telephone used as mode of last resort.	CATI
Following rules	Original sample members and their offspring or adopted children. Information is collected for persons who reside with an original sample member, their offspring or adopted children.	Original sample members and their off-spring. From wave 5 (1988) onwards persons who (ever) reside with an original sample member also become permanent sample members.	Original sample members and their off-spring or adopted children. Persons who reside with an original sample member are sample members for that survey wave. Persons who have a child with an original sample member become permanent sample members.		Original sample members and their off-spring or adopted children. Persons who reside with an original sample member are added to the sample for that survey wave. Persons who have a child with an original sample member become permanent sample members.	Original sample members and their off-spring or adopted children. Persons who reside with an original sample member are added to the sample for that survey wave. Persons who have a child with an original sample member become permanent sample members.
Proxy interviews (adult respondents)	Yes – 100 percent. In 1976 and 1985 “wives” were also interviewed	No – 0 percent	Yes – 2 to 4 percent	Yes – about 30 percent	No – 0 percent	Yes – 2-3 percent
Initial responding	4,802 families	5,921 households	5,538 households	15,006 households	7,682 households	5,074 households

Feature	*PSID*	**SOEP**	*BHPS*	*SLID*	*HILDA Survey*	*SHP*
sample size						
Responding sample size in most recent wave	8,002 households (wave 34, 2005)	12,499 households (wave 23, 2006)	8,709 households (wave 15, 2005)	38,776 households (wave 5 of panel 3, wave 2 of panel 4, 2003)	7,139 households (wave 6, 2006)	4,256 households (wave 7, 2005)
Over-sampling / Sample enhancement	Wave 1 (1968) – oversample of low-income households (n=1,872) (2/3 of this sample dropped in 1997). Wave 23 (1990) - Latino supplement (dropped after 1995). Wave 30 (1997) - General immigrant sample top-up.	Wave 1 (1984) - oversample of immigrant households (n=1,393). Wave 7 (1990) - residents of East German supplement (n=2,179 households) Wave 12 (1995) – immigrant refreshment sample Waves 15 (1998) and 17 (2000) – general refreshment samples Wave 19 (2002) - High income households oversample Wave 23 (2006) – general refreshment sample.	Wave 7 (1997) - low-income sample for ECHP - dropped in wave 12 (2002) Wave 9 (1999) - new Scottish and Welsh sub-samples Wave 11 (2001) - new Northern Ireland sub-sample.	Sample based on the Labour Force Survey and hence sample selection probabilities vary across regions (i.e., smaller regions over-sampled).	None.	Wave 6 (2004) – general refreshment sample.
Wave 1 household response rates	76%	West German sample, fully interviewed households = 61% Foreigner sample, fully interviewed households	Partial households = 74% Full households = 69% (includes proxy interviews)		Partial households = 66% Full households = 59%	Partial Households = 49%

Feature	*PSID*	**SOEP**	*BHPS*	*SLID*	*HILDA Survey*	*SHP*
		= 68% East German sample, fully interviewed households = 70% 1998 refresher sample, includes partially interviewed households = 54% 2000 new sample, partial households = 52% 2006 new sample, partial households = 41%	1999 Scottish / Welsh sample, partial households = 63% 2001 Northern Ireland sample, partial households = 69%			
Panel response[c]: Wave 5 Wave 10 Wave 15 Wave 20	 81% 70% 61% 52%	 69% (71%)[d] 53% (55%) 41% (44%) 31% (35%)	 72%[e] 62% n.a.	 Wave 5 rates are: 82% (panel 1) 79% (panel 2) 76% (panel 3)	 74%	 56%
Fieldwork	Data collection contracted out. Management of panel and cleaning of data undertaken in-house.	Data collection and parts of management and processing functions contracted out.	Data collection contracted out. Management of panel and cleaning of data undertaken in-house.	Everything managed in-house.	Data collection, management and processing contracted out.	Data collection contracted out. Management of panel and cleaning of data undertaken in-house.
Data distribution	Freely available from web site.	CD-Rom/DVD. Access restricted to bona fide researchers. Remote access for specific	Deposited in UK Data Archive.	Currently only available via remote access or on-site access at Statistics	CD-Rom. Access restricted to bona fide researchers for specific purpose research.	CD-Rom., Access restricted to bona fide researchers for specific purpose

Feature	*PSID*	**SOEP**	*BHPS*	*SLID*	*HILDA Survey*	*SHP*
		purpose research.		Canada.		research.

Notes

a The PSID's original funding agency was the Office of Economic Opportunity of the United States Department of Commerce. Other organizations that have provided funds to support the PSID include the National Institute on Aging, the National Institute of Child Health and Human Development, the Office of the Assistant Secretary for Planning and Evaluation of the United States Department of Health and Human Services, the Economic Research Service of the United States Department of Agriculture, the United States Department of Housing and Urban Development, and the United States Department of Labor, and the Center on Philanthropy at the Indiana University-Purdue University.

b The German Science Foundation (DFG) and the Leibniz Association (WGL) are financed by the German Federal Government and the Federal States Governments via the Bund-Länder Commission for Educational Planning and Research Promotion. .

c With the exception of the PSID, these response rates are the proportion of respondents in wave 1 that are successfully interviewed at later waves. The figures for the PSID are the proportion of enumerated household members from wave 1 remaining in the sample, as reported in Fitzgerald et al. (1998, Table 1), and thus are not strictly comparable with the figures reported for the other panels.

d Figures in parentheses are for the West German sample (Sample A) only.

e Figures restricted to full interview respondents.

Appendix 1. Variables included in the Cross-National Equivalent File 1980-2005

Label	Data	Variable name
Demographics:		
Age of Individual	B, G, H, P, S, CH	D11101_xxxx
Sex of Individual	B, G, H, P, S, CH	D11102LL
Marital Status of Individual	B, G, H, P, S, CH	D11104_xxxx
Relationship to Household Head	B, G, H, P, S, CH	D11105_xxxx
Number of Persons in Household	B, G, H, P, S, CH	D11106_xxxx
Number of Children in Household	B, G, H, P, S, CH	D11107_xxxx
Education With Respect to High School	G, H, P, S, CH	D11108_xxxx
Number of Years of Education	G, H, P, S, CH	D11109_xxxx
Race of Individual[a]	B, P, S	D11112LL
Employment:		
Annual Work Hours of Individual	B, G, H, P, S, CH	E11101_xxxx
Impute Annual Work Hours of Individual	B, CH	E11201_xxxx
Employment Status of Individual	B, G, H, P, S, CH	E11102_xxxx
Employment Level of Individual	B, G, H, P, S, CH	E11103_xxxx
Primary Activity of Individual	B, G, P, S, CH	E11104_xxxx
Occupation of Individual	B, G, H, P, S, CH	E11105_xxxx
1 Digit Industry Code of Individual	B, G, H, P, S, CH	E11106_xxxx
2 Digit Industry Code of Individual	B, G, H, P, S, CH	E11107_xxxx
Equivalence scale inputs:		
Number HH members age 0-14	B, G, H, P, S, CH	H11101_xxxx
Number HH members age 15-18	B, G, H, P, S, CH	H11102_xxxx
Number HH members age 0-1	B, G, H, P, S, CH	H11103_xxxx
Number HH members age 2-4	B, G, H, P, S, CH	H11104_xxxx
Number HH members age 5-7	B, G, H, P, S, CH	H11105_xxxx
Number HH members age 8-10	B, G, H, P, S, CH	H11106_xxxx
Number HH members age 11-12	B, G, H, P, S, CH	H11107_xxxx
Number HH members age 13-15	B, G, H, P, S, CH	H11108_xxxx
Number HH members age 16-18	B, G, H, P, S, CH	H11109_xxxx
Number HH members age 19+ or 16-18 and indep.	B, G, H, P, S, CH	H11110_xxxx
Indicator - Wife/spouse in HH	B, G, H, P, S, CH	H11112_xxxx
Yearly Income:		
Household Pre-Government Income	B, G, H, P, S, CH	I11101_xxxx
Household Post-Government Income	B, G, H, P, S, CH	I11102_xxxx
Household Labor Income	B, G, H, P, S, CH	I11103_xxxx
Household Asset Income	B, G, H, P, S, CH	I11104_xxxx
Household Imputed Rental Value	B, G, H, P, S, CH	I11105_xxxx
Household Private Transfers	B, G, H, P, S, CH	I11106_xxxx
Household Public Transfers	B, G, H, P, S, CH	I11107_xxxx
Household Social Security Pensions	B, G, P, S, CH	I11108_xxxx
Total Household Taxes	B, G, H, P, S, CH	I11109_xxxx
Individual Labor Earnings	B, G, H, P, S, CH	I11110_xxxx
Household Federal Taxes	G, P	I11111_xxxx
Household Social Security Taxes	B, G, P, CH	I11112_xxxx
Household Post-Government Income (TAXSIM)	P	I11113_xxxx
Total Household Taxes (TAXSIM)	P	I11114_xxxx
Household State Taxes (TAXSIM)	P	I11115_xxxx
Household Federal Taxes (TAXSIM)	P	I11116_xxxx
Household Private Retirement Income	B, G, H, P, S	I11117_xxxx
Household Windfall Income	B, G, H, P, S, CH	I11118_xxxx
Impute Household Pre-Government Income	B, G, H, CH	I11201_xxxx

Impute Household Post-Government Income	B, G, H, CH	I11202_xxxx
Impute Household Labor Income	B, G, H, CH	I11203_xxxx
Impute Household Asset Income	B, G, H, CH	I11204_xxxx
Impute Household Imputed Rental Value	B, G, CH	I11205_xxxx
Impute Household Private Transfers	B, G, H, CH	I11206_xxxx
Impute Household Public Transfers	B, G, CH	I11207_xxxx
Impute Household Social Security Pensions	B, G, CH	I11208_xxxx
Impute Total Household Taxes	G, H, CH	I11209_xxxx
Impute Individual Labor Earnings	B, G, H, CH	I11210_xxxx
Impute Private Retirement Income	B, G, H	I11217_xxxx
Location:		
Area of Residence[b]	B, G, P, S, CH	L11101_xxxx
Region of Residence[70]	B, G, H, CH	L11102_xxxx
Medical/health:		
Whether spent night in hospital in last year	B, G, P, CH	M11101_xxxx
Number of nights (days) spent in hospital	B, G, P, CH	M11102_xxxx
Whether had accident in past year that required hospital	B, G, CH	M11103_xxxx
Frequency of sports or exercise	B, G, P, CH	M11104_xxxx
Have had stroke	B, P	M11105_xxxx
Have or had high blood pressure/hypertension	B, P	M11106_xxxx
Have or had diabetes	B, P	M11107_xxxx
Have or had cancer	B, P	M11108_xxxx
Have or had psychiatric problems	B, P	M11109_xxxx
Have or had arthritis	B, P	M11110_xxxx
Have or had angina or heart condition	B, P	M11111_xxxx
Have or had asthma or breathing difficulties	B, P	M11112_xxxx
Have trouble climbing stairs	B, G, P	M11113_xxxx
Have trouble with bath	B, P	M11114_xxxx
Have trouble dressing	B, G, P	M11115_xxxx
Have trouble getting out of bed	B, G, P	M11116_xxxx
Have trouble shopping	G, P	M11117_xxxx
Have trouble walking	B, P	M11118_xxxx
Have trouble doing housework	B, G, P	M11119_xxxx
Have trouble bending, lifting, stooping	B, P	M11120_xxxx
Health limits vigorous physical activities	B, P	M11121_xxxx
Height (in meters)	G, P, CH	M11122_xxxx
Weight (in kilos)	G, P, CH	M11123_xxxx
Disability Status of Individual	B, G, H, P, S	M11124_xxxx
Subjective Satisfaction with Health	B, G, H, S, CH	M11125_xxxx
Self-Rated Health Status	B, G, H, P, CH	M11126_xxxx
Number of Times Visited Dr. in Past Year	G, CH	M11127_xxxx
Weights:		
Cross-sectional Weight - Respondent Individuals	B, G, H, P, S, CH	W11101_xxxx
Household Weight	B, G, H, P, S, CH	W11102_xxxx
Longitudinal Weight - Respondent Individuals	B, G, H, P, S, CH[71]	W11103_xxxx
Population Factor for W11101$$	B, G, P	W11104_xxxx
Individual Weight - Immigrant Sample	G	W11105_xxxx
Household Weight - Immigrant Sample	G	W11106_xxxx
Cross-sectional Weight - Enumerated Individuals	B, H	W11107_xxxx
Longitudinal Weight - Enumerated Individuals	B, H	W11108_xxxx
Population Factor for W11103$$	B, G, P	W11109_xxxx
Population Factor for W11107$$	B	W11110_xxxx
Population Factor for W11108$$	B	W11111_xxxx

Equivalence Weight Algorithms

Detailed Official U.S. Equivalence Weight

General Official U.S. Equivalence Weight

---

[70] Region of residence is the language region of the interview (German, French, Italian)

[71] W11203 for combined SHP I (original) and SHP II (refreshment) sample.

Official German Equivalence Weight
ELES Equivalence Weight
OECD Equivalence Weight
McClements Equivalence Weight
Other Equivalence Weights

Identifiers:		
Unique Person Number	B, G, H, P, S, CH	X11101LL
Household Identification Number	B, G, H, P, S, CH	X11102_xxxx - X11102_2003
Individual in Household at Survey	B, G, H, P, S	X11103_xxxx - X11103_2003
Oversample Identifier	B, G, P, S	X11104LL
Person in Household Interviewed	B, G, H, CH	X11105_xxxx - X11105_2003
Macro-level Variables:[c]		
Consumer Price Index	B, G, P, S	
Median Pre-government Household Income	B, G, P, S	
Median Post-government Household Income	B, G, P, S	
Purchasing Power Parity for East Germany	G	

*Area of residence is the Region/Metropolitan Area in the BHPS, the Bundesland in the GSOEP, the major city or state in the HILDA, and the US state in the PSID. Province of residence is available on the CNEF SLID files at Statistics Canada.

(B) BHPS: 1991-2004 Survey Years
(G) GSOEP: 1984-2005 Survey Years
(H) HILDA: 2001-2004 Survey Years
(P) PSID: 1980-2003 Survey Years
(S) SLID: 1992-2003 Reference Years
(CH) SHP: 1999-2005 Survey Years

[a]Race in the BHPS and SLID is reported for all sample members. In the PSID, race is coded for any sample member who has ever been a household head or wife.
[b]Area of residence is the Local Authority District of Residence in the BHPS, the *Bundesland* in the GSOEP, the US state in the PSID, the *Kanton* in the SHP. The province of residence is not on the CNEF SLID files on the CD but are available from the CNEF SLID files at Statistics Canada. Local Authority District of Residence data for the BHPS is available by special arrangement with the University of Essex.
[c]Because macro-level variables do not vary across individuals or households, they are only listed in the codebooks for reference purposes.

Appendix 2: Sample sizes for national panels in the CNEF (individuals)

Year	*PSID*	*SOEP*	*BHPS*	*SLID*	*HILDA*	*SHP*
1980	18989	-	-	-	-	
1981	18992	-	-	-	-	
1982	19246	-	-	-	-	
1983	19491	-	-	-	-	
1984	19570	15237	-	-	-	
1985	19787	13747	-	-	-	
1986	19615	13084	-	-	-	
1987	19647	12853	-	-	-	
1988	19687	12253	-	-	-	
1989	19669	11856	-	-	-	
1990	19932	17462	-	-	-	
1991	19962	17094	13780	-	-	
1992	20334	16801	13151	40155	-	
1993	21450	16510	13104	42194	-	
1994	23620	16828	12851	43717	-	
1995	23182	17252	12549	88230	-	
1996	23060	16869	12720	91624	-	
1997	19132	16559	15042	94125	-	
1998	-	18161	14835	139508	-	
1999	19669	17417	21540	94772	-	10437
2000	-	30439	21602	96512	-	9454
2001	20538	27481	26586	141598	19914	8775
2002	-	29280	23435	93680	18295	7648
2003	21277	27553	22559	95792	17691	6944
2004	-	26690	22105		17209	10666
2005	22918	25544	15627		17469	8550
Total observations (Person * Years)	449.767	416.970	261.486	1.061.907	90.578	62.474

Note: These numbers may be different from similar ones found in the documentation of the original survey datasets. For example, the SOEP provides only a 95 per cent version of its data to the CNEF, and the low-income and Latino samples of the PSID are excluded from the CNEF.

Die internationale Einbettung des Sozio-ökonomischen Panels (SOEP) im Rahmen des CNEF[1]

Joachim R. Frick, Stephen P. Jenkins, Dean R. Lillard, Oliver Lipps, and Mark Wooden

1. Hintergrund

In den vergangenen dreißig Jahren haben mehrere Industrienationen zunehmend in Entwicklung, Aufbau und Durchführung von sozialwissenschaftlichen Haushalts- und Personenbefragungen investiert, wobei Längsschnitt- und Panelstudien eine besondere Rolle zukommt. In vielen Fällen von der Wissenschaft vorangetriebenen, unterstützen diese die Erfassung, Beschreibung und Analyse sozioökonomischer Merkmale der Bevölkerung eines Landes und die Beobachtung von Lebensverläufen. Derartige „general purpose social science surveys“ ermöglichen außerdem die Erfassung von Veränderungen des Verhaltens im Zusammenhang mit sozialpolitischen Neuerungen und Veränderungen. Es können darüber hinaus Szenarien zur Abbildung möglicher Verhaltensvariationen unter alternativen sozialpolitischen Maßnahmen entwickelt werden. Diese Daten haben die verhaltenswissenschaftliche Forschung nicht nur auf nationaler Ebene gestärkt, sondern auch länderübergreifende Studien vorangetrieben.

Um nationale Umfragedaten für international vergleichende Analysen nutzen zu können, muss untersucht werden, inwieweit die Daten überhaupt vergleichbar sind bzw. vergleichbar gemacht werden können. Dies erfordert gegebenenfalls umfangreiche methodische Maßnahmen, meist solche zur Anpassung bereits vorhandener Daten im Rahmen der so genannten ex-post Harmonisierung. Da die Befragungen in den einzelnen Ländern mit Hinblick auf nationale Politik und nationale Forschungsziele durchgeführt wurden, sind sie im Allgemeinen nicht von vornherein dafür entworfen worden, für international vergleichbare

[1] Das Cross-National Equivalent File (CNEF) wurde 1991 von der SOEP-Gruppe am Deutschen Institut für Wirtschaftsforschung (DIW Berlin) und der Cornell Universität ins Leben gerufen, damals finanziert vom National Institute of Aging (USA). Inzwischen ist es ein gemeinschaftliches Projekt von sechs Institutionen: Cornell University, SOEP am DIW Berlin, Statistics Canada, Institute for Social and Economic Research (ISER) an der Universität von Essex, Melbourne Institute of Applied Economic and Social Research an der Universität von Melbourne und der Universität von Lausanne (bis 2007 Universität Neuchâtel). Die Autoren danken insbesondere Richard V. Burkhauser, Gaétan Garneau, Robert Schoeni und Gert G. Wagner für hilfreiche Kommentare zu früheren Entwürfen dieses Artikels.

Studien geeignete Daten zu generieren. Ausnahmen bilden hier die Mikrodaten des von Eurostat, dem Statistischen Amt der EU, für die EU-15 Staaten koordinierten Europäischen Haushaltspanels ECHP (European Community Household Panel) und die Kohortenstudie SHARE (Survey of Health, Ageing and Retirement in Europe). Im Rahmen der beim ECHP praktizierten Input-Harmonisierung wurde bereits bei der Erhebung explizit die spätere Verwendung der Daten für international vergleichende Analysen berücksichtigt, indem abgestimmte Erhebungsinstrumente mit möglichst ähnlicher Frageformulierung verwendet wurden. ECHP war jedoch nur zum Teil erfolgreich und wurde nach acht Wellen im Jahre 2001 aufgegeben. SHARE ist erfolgreicher, aber aufgrund seiner Fokussierung auf ältere Kohorten für die Analyse breiterer Bevölkerungsschichten oder gar der Gesamtbevölkerung ungeeignet.[2]

Die meisten Daten müssen jedoch nachträglich harmonisiert werden. Hierbei ist für die Definition von Variablen zur Erfassung äquivalenter Konstrukte und Verhaltensweisen in der international komparativen Forschung erheblicher Aufwand notwendig. Im Zusammenhang mit einfachen Konzepten wie Alter und Geschlecht ist diese Aufgabe vergleichsweise unkompliziert. Erheblich komplexer wird es, wenn es sich um Konzepte handelt, die nur im Kontext von landesspezifischen Institutionen und kulturellen Faktoren verstanden werden können. International vergleichbare Einheiten zur Erfassung von Faktoren wie ökonomische Wohlfahrt, Bildung, Beschäftigung und Gesundheit können nur nachträglich gebildet werden, da die in den jeweiligen Ländern erfassten Daten einen landesspezifischen institutionellen Kontext haben oder sich auf kulturelle Faktoren begründen. Für die vergleichend arbeitenden Forscher ist es daher wichtig, sich mit den Institutionen, Gesetzen und kulturellen Gegebenheiten der verschiedenen Länder vertraut zu machen.

Eine der ersten Studien, die international vergleichbare Daten bereitstellte, war die Luxembourg Income Study (LIS). Seit 1983 harmonisiert LIS Mikrodaten aus Erhebungen aus mehr als dreißig Ländern (vgl. www.lisproject.org und Smeeding/ Jesuit/ Alkemade, 2002), darunter auch für Deutschland auf Basis der Daten des SOEP. Da die Hauptlast der Harmonisierung („Lissification“) bei LIS liegt, wird der Aufwand für den einzelnen Forscher erheblich reduziert.

[2] Burkhauser und Lillard (2005) beschreiben detailliert die Erfolge und Misserfolge bei der Erstellung von sowohl ex-ante als auch ex-post harmonisierten Datensätzen für die international vergleichende Forschung im Allgemeinen; Burkhauser und Lillard (2006) beziehen sich insbesondere auf SHARE.

Obgleich der Umfang der LIS Aktivitäten beeindruckend ist, können damit aber nicht alle Bedürfnisse der internationalen Forschergemeinschaft erfüllt werden. So erhalten externe Wissenschaftler über das Datenfernverarbeitungssystem LISSY nur indirekten Zugang zu den der LIS-Datenbank zugrunde liegenden Mikrodaten, da es sich in einigen Fällen um nicht direkt weitergabefähige Datenbestände der nationalen Statistischen Ämter handelt. Obwohl eine umfangreiche Dokumentation der LIS-Daten vorliegt, hat dieser Ansatz zur Folge, dass Forscher sich mit der vorgegebenen Harmonisierung „zufrieden geben" müssen und deren Einfluss auf substantielle Analysen nicht prüfen können. Zudem liefert LIS nur Querschnittsdaten, so dass an Panel- oder Längsschnittanalysen interessierte Forscher eher wenig von den LIS Daten profitieren können.

Im Folgenden wird ein Projekt vorgestellt, das zwar auf der LIS-Idee aufbaut, dessen Beschränkungen jedoch weitgehend überwindet. Bei diesem Projekt, dem Cross-National Equivalent File (CNEF), handelt es sich um ein gemeinsames Unterfangen von Wissenschaftlern und Vertretern von Institutionen aus sechs verschiedenen Ländern (Stand: Sommer 2008), die jeweils nationale Panel-Erhebungen durchführen. Hierzu gehören die US-amerikanische *Panel Study of Income Dynamics* (PSID), das *Sozio-oekonomische Panel* (SOEP) aus Deutschland, der *British Household Panel Survey* (BHPS) aus Großbritannien, der *Survey of Labour and Income Dynamics* (SLID) aus Kanada, der *Household Income and Labour Dynamics Survey* (HILDA) aus Australien, sowie das *Schweizer Haushaltpanel* (SHP).[3]

Im CNEF werden Daten harmonisiert, die in mindestens zwei der verfügbaren nationalen Datenbestände enthalten sind, die ihrerseits zu den weltweit erfolgreichsten, laufenden sozialwissenschaftlichen Panelstudien gehören. Externe Wissenschaftler haben im Prinzip Zugang sowohl zu den harmonisierten wie auch den Originaldaten, inklusive einer umfangreichen Dokumentation der dem Harmonisierungsprozess zu Grunde liegenden Annahmen und Hinweisen auf Probleme der Vergleichbarkeit.

[3] Das CNEF wird, in enger Zusammenarbeit mit Wissenschaftlern des SOEP am DIW Berlin, des Institute for Social and Economic Research (ISER) an der Universität von Essex, von Statistic Canada in Ottawa, des Survey Research Centers an der Universität von Michigan, des Melbourne Institute of Applied Economic Social Research an der Universität von Melbourne sowie der Universität von Lausanne (bis 2007: Universität Neuchâtel), an der Cornell Universität in Ithaca/NY (USA) organisiert.

Das CNEF unterscheidet sich von anderen Standardisierungs-Projekten nicht nur weil es Daten von laufenden Erhebungen mit einbezieht, sondern auch dadurch, dass die wissenschaftlich arbeitenden Nutzer direkt Einfluss auf die Auswahl neuer CNEF-Variablen sowie deren Harmonisierung ausüben können. Solche Variablen entstehen häufig in Forschungsprojekten, in denen Wissenschaftler international vergleichbar arbeiten. Das dabei erworbene Expertenwissen auf dem jeweiligen Forschungsgebiet beinhaltet auch Kenntnisse von landesspezifischen Institutionen. Dies hat zur Folge, dass die im CNEF enthaltenen harmonisierten Daten das Wissen vieler Wissenschaftler aus verschiedenen Disziplinen vereinen. Da die Forscher ihre harmonisierten Variablen ständig verfeinern, sowie neue Variablen hinzufügen, entwickelt sich Datenbestand des CNEF kontinuierlich weiter.

Einer der großen Vorteile des CNEF liegt im Panel-Charakter der Daten, die es erlauben, eine Person über viele Jahre hinweg zu verfolgen. Dies ermöglicht mit Hilfe geeigneter statistischer Methoden die Analyse von individuellen Verhaltensweisen, die mit wiederholten Querschnittbefragungen nicht beobachtbar sind. Die CNEF-Längsschnittstudien ermöglichen somit nicht nur die Untersuchung des Einflusses von institutionellen? Veränderungen zwischen Ländern, sondern auch über die Zeit. Dies führt zu einem besseren Verständnis des menschlichen Verhaltens. Insbesondere bei lang laufenden Studien wie dem PSID und dem SOEP ermöglichen die nationalen Befragungsdaten die Beobachtung von Familien über mehrere Generationen hinweg. Insofern sind die Daten des CNEF zur Untersuchung von z.B. sozioökonomischem Status und dessen Übertragung von Generation zu Generation, in internationaler wie in interdisziplinärer Perspektive, bestens geeignet.[4]

2. Die Entwicklung des CNEF

Die Geschichte des CNEF begann 1991 mit einer harmonisierten Datenbasis für lediglich zwei Länder, den USA und Deutschland, gefördert durch das National Institute of Aging[5]. Im Jahr 1999 kamen die Daten des BHPS aus Großbritannien und des SLID aus Kanada hinzu. 2007 folgte die HILDA Befragung aus Australien und es begannen die

[4] Siehe hierzu z.B. Butz und Torrey (2006).

[5] Die Principal Investigators dieser Initiative waren Richard V. Burkhauser, zu dieser Zeit an der Maxwell School der Syracuse University, NY (USA) und Gert G. Wagner, Direktor des SOEP am DIW Berlin. Besonderer Dank geht an Richard Hauser, zu dieser Zeit an der Universität Frankfurt/Main, für seine außerordentlich wichtige Unterstützung in der Initialphase dieses Projektes.

Harmonisierungsarbeiten für die Daten der Schweizer Haushaltspanel-Studie SHP, sodass das CNEF heute harmonisierte Datensätze von sechs Ländern beinhaltet.

Der ursprüngliche Fokus des CNEF war der Aufbau einer harmonisierten Datenbasis für die vergleichende Analyse von einkommensbasierter Ungleichheit sowie Einkommensmobilität in den USA und Deutschland mithilfe der repräsentativen Mikrodaten von PSID und SOEP. Viele der frühen vergleichenden Arbeiten zu den Themen ökonomische Wohlfahrt und Einkommensmobilität in den USA und Deutschland in den 1980er und 1990er Jahren basierten auf diesen harmonisierten Daten (vgl. z.B. Burkhauser/Frick/Schwarze, 1997; Burkhauser/Crews-Cutts/Lillard, 1999). Im Rahmen des weiteren, forschungsgetriebenen Ausbaus des CNEF wurden zusätzliche inhaltliche Bereiche berücksichtigt, wobei insbesondere die in den nationalen Befragungen zunehmend umfassendere Abbildung des Bereichs Gesundheit auch für CNEF eine bedeutende Rolle spielt.

Der anhaltend hohe Forschungsbedarf zum Thema *ökonomische Ungleichheit* im Rahmen wohlfahrtsstaatlicher Vergleiche und der Suche nach „best practise" führte fast zwangsläufig zu einer Erweiterung des CNEF um weitere nationale Datensätze. Für den Erfolg des CNEF-Ausbaus war dabei nicht nur wichtig, dass in den zwischenzeitlich neu aufgelegten Panelstudien BHPS und SLID das Einkommen von Individuen und Haushalten einen Befragungsschwerpunkt darstellte, sondern dass beide Befragungen in ihrem Aufbau von den Erfahrungen des PSID und des SOEP beeinflusst waren. Auf Basis dieser erweiterten Datenbasis entstanden eine Reihe von vergleichenden Studien zur Entwicklung von Ungleichheit, Löhnen und (Einkommens-)Mobilität in den USA, Kanada, Großbritannien und Deutschland (vgl. z.B. Jenkins/Schluter, 2003; Jenkins/Schluter/Wagner, 2003; Burkhauser/Giles/Lillard/Schwarze, 2005). Im Laufe der Zeit wurden zusätzliche Variablen harmonisiert. Eine umfangreiche inhaltliche Erweiterung der CNEF-Variablen fand im Jahr 2003 statt, als Variablen zum Thema Gesundheit, die in mindestens zwei der damals teilnehmenden Längsschnittstudien vorlagen, harmonisiert wurden (vgl. Lillard/Burkhauser, 2005). Daten aus der HILDA Befragung kamen im Jahr 2007 hinzu und unterstützen international vergleichende Panel-Analysen zu den Themen Beschäftigungschancen und Erwerbseinkommen von Arbeitnehmern mit Behinderung (vgl. Burkhauser/Schmeiser/Schroeder, 2007). Zum Zeitpunkt der Drucklegung dieses Artikels werden nach entsprechenden Vorarbeiten auch Daten des SHP aus der Schweiz miteinbezogen. Die nächste Version des CNEF wird somit voraussichtlich die folgenden

Daten enthalten: PSID (Wellen 1980-2005), SOEP (1984-2006), BHPS (1991-2005), SLID (1992-2005), SHP (1999-2005), sowie HILDA (2001-2005). Die in den jeweiligen Studien analysierbare Zahl der Beobachtungen auf der Ebene von individuell befragten Erwachsenen und deren minderjährigen Kindern bis 2005 findet sich in Anhang 2. In der Summe aller sechs Befragungen ergeben sich dabei weit über 2,57 Millionen Personen-Jahr Beobachtungen.

3. Aufbau und Inhalt des CNEF

Das Ziel des CNEF ist es, unabhängig von den Erfahrungen der jeweiligen Forscher im Umgang mit Längsschnittdaten, international vergleichende Analysen mit wirtschafts- und sozialwissenschaftlichen Daten zu unterstützen. Zur Erreichung dieses Ziels, welches sowohl Forschung als auch „capacity building“ beinhaltet, stellt das CNEF Mikrodaten mit möglichst benutzerfreundlich definierten Variablen aus verschiedenen Befragungen zusammen. Für jeden nationalen CNEF-Datensatz existiert eine separate Datenbank, in der die entsprechenden Variablen wellen- bzw. kalenderjahrspezifisch abgelegt werden. Diese können dann entweder im Sinne von *stand-alone* Daten oder, wie in den meisten Fällen, in Kombination mit anderen interessierenden Informationen, analysiert werden. So werden z.B. landes- und zeitspezifische Daten zu politischen und ökonomischen Rahmenbedingungen wie Arbeitslosenquoten mit den CNEF-Mikrodaten zusammengeführt. Da die Identifikator-Variablen des CNEF auf Personen- und Haushaltsebene jenen der Originalbefragungen entsprechen, können durch entsprechende Verknüpfungen weitere Mikrodaten aus einer oder mehrerer der ursprünglichen Datenbanken in die CNEF-Datenbank überführt werden, wobei diese als eine Art zentrale Navigationsdatei oder *Masterfile* fungiert.

Der Aufbau des CNEF erleichtert weniger erfahrenen Forschern die Arbeit, da die Variablen in jedem Datensatz mit identischen Variablennamen und – soweit möglich - Wertebezeichnungen versehen sind. Die Namen der Variablen spiegeln ihren Inhalt wider: Der erste Buchstabe bezeichnet die Kategorie der Variablen: demographic (D), employment (E), household composition (H), income (I), weighting (W), sample identifiers (X), location (L), medical & health (M) und macro-level Indikatoren (Y). Die letzten vier Zeichen des Variablennamens zeigen das Erhebungsjahr, aus dem die Variable stammt. Diese parallel aufgebaute Struktur erlaubt es den Benutzern für alle Längschnittstudien (Panels) die gleichen Programme zu verwenden, idealerweise möglichst sogar in einem Durchlauf.

Darüber hinaus erlaubt der spezielle Aufbau des CNEF erfahreneren Forschern schnell und einfach die Algorithmen zu modifizieren, die der Generierung von Variablen zugrunde liegen. Auch weitere Daten können so zur Ergänzung der bestehenden Variablen hinzugefügt werden. Das CNEF-Handbuch beschreibt für jede Variable die für ihre Erstellung verwendeten Algorithmen unter Verwendung der Bezeichnung in den nationalen Originaldaten. Damit Nutzer die vorhandenen CNEF-Daten mit jenen der Originalbefragungen ergänzen können, werden auch die relevanten Identifikatoren für alle Beobachtungseinheiten auf Individual- und Haushaltsebene aus den ursprünglichen Befragungen zur Verfügung gestellt. Dies ermöglicht auch Robustheitsprüfungen der interessierenden Untersuchungsergebnisse bei Modifikation der Definition von Variablen bzw. der für die CNEF-Variablen verwendeten Algorithmen. Schließlich erlaubt es dem Nutzer, gegebenenfalls eine eigene Herangehensweise zur Verbesserung der existierenden Variablen zu entwickeln.

Ergänzend zum oben erwähnten Algorithmus zur Konstruktion von Variablen weist CNEF jeder Variablen einen so genannten *reliability* Wert zu, der auf die Kompatibilität zwischen den Befragungen und damit auf die Qualität der Vergleichbarkeit hinweist. Lautet der Code „1“ wird von hoher Kompatibilität der Variablen ausgegangen, wohingegen eine „4“ anzeigt, dass die Befragungen über keine vergleichbaren Variablen verfügen. Die Vergabe dieser Codes basiert auf den Erfahrungen der CNEF-Datenproduzenten, direkten Vergleichen der verwendeten Befragungsmethoden und der Kenntnis institutioneller Unterschiede zwischen einzelnen Ländern.

Eine wesentliche Innovation des CNEF ist die Konstruktion von (Einkommens)variablen, die weit über das Datenmaterial der ursprünglichen Befragungen hinausgeht. Von besonderer Bedeutung sind dabei Haushaltsjahreseinkommen vor und nach (sozial)staatlicher Umverteilung als Ergebnis der Aggregation einer Reihe von Einkommenskomponenten sowie der Schätzung der Abgabenlast der Haushalte aufgrund von direkten Einkommenssteuern und Sozialversicherungsprämien. Viele dieser Werte könnten ohne erheblichen Aufwand seitens des einzelnen Nutzers nicht errechnet werden, wobei die Simulation der Steuerlast eine besonders große Herausforderung darstellt. Dieser innovative Beitrag des CNEF ermöglicht internationale Vergleiche der verfügbaren Einkommen und der staatlichen Umverteilungseffekte, die in dieser Form mit den Daten der Originalbefragungen bestenfalls nur eingeschränkt möglich sind.

Der nötige Aufwand für die Berechnung des Netto Einkommens variiert dabei von Studie zu Studie. Im Fall von SLID, das auch als nationale Einkommensstichprobe von Statistics Canada dient, sowie dem SHP[6] wird die Steuerlast direkt von den Befragten erfragt. In den anderen Studien müssen diese Werte geschätzt werden. Die Steuer- und Abgabensimulationsprogramme für die BHPS, SOEP und HILDA Befragungen wurden von Forschern an den jeweiligen Instituten entwickelt. Das Programm zur Schätzung der Steuerlast im BHPS (Levy et al., 2006)[7] wurde von Stephen Jenkins und seinen Koautoren an der Universität von Essex geschrieben und wird auch weiterhin von ihm betreut. Das entsprechende Programm des SOEP wurde ursprünglich von Johannes Schwarze (heute Universität Bamberg) entwickelt und wird von Markus Grabka vom DIW Berlin weiter betreut (Schwarze, 1995). Der HILDA Survey bedient sich eines von Bruce Heady an der Universität von Melbourne entwickelten Programms. Im Fall des PSID wurden die Zahlen zur Steuerlast intern geschätzt und der Studie dann bei Veröffentlichung beigefügt. Seit 1993 wird die Steuerlast jedoch mithilfe des Programms TAXSIM (siehe Feenberg/Couts, 1993)[8] des National Bureau of Economic Research von Dean R. Lillard an der Cornell Universität geschätzt.

Der Aufwand zur Berechnung des Nettoeinkommens ist im Falle des SOEP besonders groß, da in Deutschland der Einkommensbegriff eher Monatsbezug hat. Insofern wird im SOEP die relevante Information zum Einkommensbezug im Vorjahr erhoben, indem – getrennt nach mehr als 20 Komponenten – die Zahl der Monate mit Bezug einer gegebenen Komponente zusammen mit dem monatlichen Durchschnittsbetrag erfragt wird. International vergleichbare Angaben zum Jahreseinkommen ergeben sich durch Multiplikation dieser Faktoren und Aggregation der diversen Komponenten. Auf diesem Brutto-Jahreseinkommen aufbauend generiert das Steuer-Simulationsprogramm unter Berücksichtigung aller relevanten Parameter (wie z.B. Splitting, Freibeträge, Steuerprogression) eine geschätzte Steuer- und Sozialversicherungsabgabenlast für alle SOEP Haushalte.

---

[6] Im SHP müssen allerdings Sozialversicherungsabgaben auf das Arbeitseinkommen von den SHP Datenproduzenten geschätzt werden.
[7] Datensätze zum Nettoeinkommen können direkt vom UK Data Archive abgerufen werden (http://www.data-archive.ac.uk/findingData/snDescription.asp?sn=3909/mrdoc/pdf/3909userguide.pdf).
[8] Butrica und Burkhauser (1997) bieten eine detaillierte Besprechung der Steuer Kalkulations-Algorithmen des NBER und des PSID. Darüber hinaus vergleichen sie die PSID Schätzungen, die mithilfe von TAXSIM entstanden sind, mit jenen aus den Jahren 1980-1992.

Ziel für alle CNEF-Datenbestände ist die Konstruktion eines für international vergleichende Wohlfahrtsanalysen relevanten (Vor-)Jahreseinkommens nach Umverteilung durch Steuern und Transfers, vereinfachend als „*household post-government income*“ bezeichnet.[9] Dieses ergibt sich aus der Summe von Erwerbseinkommen aus selbständiger und abhängiger Beschäftigung (inklusive eventueller Einmalzahlungen wie Weihnachts- und Urlaubsgeld, Gratifikationen, Boni, etc.), Kapitalerträgen, Einnahmen aus Vermietung und Verpachtung, privaten Transfers, öffentlichen Transfers, sowie allen anderen monetären Zuflüssen aller Mitglieder eines Haushaltes, abzüglich der Einkommensteuer und Sozialversicherungsabgaben. Darüber hinaus steht in einer separaten Variablen auch eine der wichtigsten nicht-monetären Einkommenskomponenten privater Haushalte, der fiktive Mietwert selbstgenutzten Wohneigentums („*imputed rent*“), zur Verfügung. Alle Variablen auf der Haushaltsebene werden jeweils den einzelnen Mitgliedern eines Haushaltes zugeordnet.

Anhang 1 listet für alle aktuell im CNEF enthaltenen Variablen den jeweiligen Namen, eine Kurzbeschreibung und die Verfügbarkeit in den nationalen Datenbasen. Darüber hinaus enthalten die Code-Bücher des CNEF Makroinformationen für jedes Land, wie beispielsweise den jährlichen Lebenshaltungskostenindex. Da diese Daten für alle Befragten eines Landes zu einem gegebenen Zeitpunkt konstant sind, werden sie nur im Code-Buch aufgeführt. Anhang 2 zeigt den Stichprobenumfang in den nationalen CNEF-Datensätzen.

4. Vergleichende Darstellung der nationalen Panelstudien im CNEF

Alle sechs im CNEF enthaltenen Längsschnittstudien erfassen unter anderem Informationen zu Haushaltsstruktur, Einkommen, Erwerbstätigkeit, und demographischen Merkmalen. Die Vergleichbarkeit der Mikrodaten wird dabei sowohl von Unterschieden zwischen den verschiedenen Befragungen – zum Beispiel bezüglich der Art und Weise der Erhebung – als auch von intertemporalen Variationen innerhalb einer gegebenen Befragung beeinflusst.

---

[9] Im Allgemeinen richtet sich die Definition der Einkommensvariablen nach den Empfehlungen der „Canberra Group on Household Income Measurement“ (Canberra Group, 2001). Aufgrund von „item-non-response“ fehlende Einkommenswerte werden in den CNEF-Daten unter Zuhilfenahme verfügbarer Längsschnittdaten imputiert. Aufbauend auf einer vergleichenden Analyse von CNEF-Daten für Deutschland, Australien und UK zum Einfluss von Imputation auf Ergebnisse zur Einkommensverteilung und –mobilität argumentieren Frick und Grabka (2007) für eine stärkere Harmonisierung der Imputationstechniken in international vergleichend aufgebauten Datenbasen.

Um einen Eindruck der generellen Vergleichbarkeit der Mikrodaten der sechs nationalen Panels zu vermitteln, zeigt Übersicht 1 eine synoptische Darstellung zentraler Merkmale rund um Erhebungsmethode, Design und Umfang der diversen Stichproben, Weiterverfolgungsregeln, Teilnahmequoten, Feldarbeit, etc.. Mit Ausnahme des SLID verfolgen alle CNEF-Befragungen die Mitglieder der Haushalte der Startwelle der jeweiligen Befragung inklusive deren Nachkommen[10]. Die Panels unterscheiden sich jedoch bezüglich der Definition der zu befragenden Mitglieder eines Haushaltes. So werden zwar in BHPS, SOEP, HILDA und dem SHP alle Erwachsenen eines Haushaltes befragt, jedoch variiert das relevante Befragungsalter zwischen 14 und 17 Jahren. Alle Befragungen erheben Informationen zu Personen, die in einen bestehenden Stichproben-Haushalt einziehen. BHPS, SOEP, HILDA und SHP erlangen diese Informationen auf direktem Wege, da sie jährlich alle erwachsenen Mitglieder des Haushaltes befragen. Im PSID wird immer nur ein Mitglied eines Haushaltes, im Sinne eines Vorstandes oder einer Bezugsperson, befragt, während SLID in großem Maße Proxy-Interviews durchführt. Darüber hinaus unterscheidet sich die SLID-Stichprobe als rotierendes Panel grundsätzlich von den anderen Befragungen, wobei sich die Untersuchungspopulation eines gegebenen Jahres jeweils aus zwei über sechs Jahre laufenden Längsschnittstudien ergibt, die sich über drei Jahre hinweg überschneiden.

Fünf der vorliegenden Befragungen haben im Laufe der Zeit ihre Methoden zur Datenerfassung verändert. Die ursprünglichen Umfragen basierten auf interviewerbasierten, persönlichen Interviews, in deren Verlauf der Interviewer die Angaben der Befragten in einen Fragebogen aufnahm („paper and pencil interviewing“ (PAPI)). In den 90er Jahren stiegen die meisten Befragungen auf computergestützte Methoden um. Die wichtigste Unterscheidung in der Art und Weise der Durchführung computergestützter Interviews bezieht sich darauf, ob diese persönlich („face to face“) oder per Telefon durchgeführt werden. BHPS, SOEP und der HILDA Survey werden überwiegend interviewergestützt durchgeführt. BHPS und SOEP bedienen sich dabei jedoch immer häufiger eines Laptops („computer assisted personal interviewing“ (CAPI)). In SOEP und HILDA findet sich auch eine parallele Anwendung verschiedener Methoden, wobei im SOEP Selbstausfüller („self-completion“) eine immer größere Rolle spielen. HILDA setzt verstärkt Telefon-Interviews ein, da – bedingt durch

[10] In der SLID-Studie werden nur die ursprünglichen Mitglieder eines Haushaltes über einen Zeitraum von maximal sechs Jahren weiterverfolgt. Nachkommen dieser Haushalte werden nicht berücksichtigt, sobald sie den gemeinsamen Haushalt verlassen. Insofern sind diese nur im Querschnitt von Bedeutung, im Längsschnitt ist ihr Gewicht gleich Null. Gemäß des für diese Studie geltenden Rotationsprinzips beginnt alle drei Jahre ein neues Panel, welches somit zu einem gegebenen Zeitpunkt (=Welle) jeweils etwa die halbe Stichprobe repräsentiert.

Umzüge seit der Stichprobenziehung – die regionale Clusterung der zu befragenden Haushaltsadressen abnimmt, was zu entsprechend höheren (Anreise-)Kosten der Interviewer führt. Fast sieben Prozent der von HILDA in der 6. Welle durchgeführten Interviews wurden telefonisch abgewickelt. PSID ging 1973 von interviewergestützten („face-to-face"), schriftlichen („paper and pencil interviews") zu telefonischen Interviews über und wechselte schließlich 1993 zu computergestützten Telefon-Interviews (computer-assisted telephone interviewing (CATI)). Sowohl SLID (seit 1992) als auch SHP (seit 1999) verwenden von Anfang an ein CATI-System.

Der Befragungszeitraum, während dem die Befragungen innerhalb eines jeweiligen Jahres durchgeführt werden, variiert zwischen den Studien.[11] Je nach Art und Zielsetzung einer Analyse sind somit saisonale Effekte als Erklärung für internationale Unterschiede nicht auszuschließen. Die Datenerhebung in SOEP und PSID erfolgt primär in den ersten vier Monaten des Jahres, während sich BHPS auf die Herbstmonate konzentriert. Im Fall von HILDA findet das Gros der Befragungen zwischen September und Dezember (also im australischen „Frühsommer") statt, für SHP sind es die Wintermonate September bis Februar. Zum Teil sind diese Unterschiede zurückzuführen auf nationale Unterschiede in der Definition des fiskalischen bzw. des Haushaltsjahres.

Unterschiede zwischen den sechs Befragungen gibt es im Zusammenhang mit der Teilnahmewahrscheinlichkeit sowohl bezüglich der Beteiligung in Welle 1 als auch - in etwas geringerem Maße - beim Ausfallverhalten im weiteren Verlauf der Längsschnittstudie („Panel Attrition"). Je nach Definition beträgt die durchschnittliche Teilnahme-Quote der sechs Befragungen in der Startwelle circa 70 Prozent. Der Anteil vollständig realisierter Haushalte (also Haushalte, in denen alle Befragungspersonen auch erfolgreich befragt werden konnten) variiert von rund 50 % im Fall des SHP[12], über 59 % im HILDA Survey und bis zu 76 % in PSID. Die entsprechenden Werte für BHPS und SOEP liegen bei respektive 69 % und 65 %[13]. Da bei der PSID-Erhebungsmethode nur die Beteiligung eines einzigen Haushaltsmitgliedes nötig ist, schneiden BHPS und SOEP auch im direkten Vergleich mit der

[11] In ihren Anfängen haben alle sechs Befragungen die Daten jährlich erfasst. PSID wechselte jedoch 1997 von einem jährlichen zu einem zwei-jährlichen Rhythmus.

[12] Zu beachten ist hierbei, dass aufgrund der im SHP angewandten CATI Methode jeder nicht erfolgreich kontaktierte Haushalt als Ausfall gilt, was qualitätsneutrale Ausfälle einschließt.

[13] Die Teilnahmequoten in der ersten Welle der beiden ursprünglichen Teilstichproben des SOEP betrugen 61 % bei den „West-Deutschen" (Sample A) und 68 % im „Ausländer"-Sample B.

US-Studie recht gut ab. In jüngeren Stichproben ist ein allgemein rückläufiger Trend der Teilnahmewahrscheinlichkeit in der Startwelle erkennbar. Dies gilt sowohl im internationalen Vergleich als auch, wie am Beispiel der Auffrischungsstichproben des SOEP erkennbar, innerhalb der einzelnen Länder.

Der Anteil der Panelausfälle stabilisiert sich in den meisten Längsschnittstudien nach einigen Wellen auf recht niedrigem Niveau (circa 4 Prozent oder weniger). Dies trifft auch auf die meisten der im CNEF vertretenen Panels zu. Der Anteil der Personen aus der Startwelle, die bis zur 5. Welle weiterhin teilnehmen, bewegt sich bei SOEP, BHPS und HILDA zwischen 71 % und 74 %. Im SHP ist die Teilnahmewahrscheinlichkeit bis Welle 5 mit lediglich 56 % deutlich niedriger, während die entsprechenden Quoten bei SLID und PSID mit mehr als 80 % überdurchschnittlich hoch ausfallen. Diese höheren Teilnahme-Quoten liegen zumindest teilweise daran, dass PSID nur jeweils ein Haushaltsmitglied befragt bzw. die SLID-Erhebung den Mitgliedern eines Haushaltes erlaubt, stellvertretend füreinander zu antworten (circa 30 Prozent der Antworten erfolgen im Rahmen solcher Proxy-Interviews). Die Zahl der Befragten mit ununterbrochener Teilnahme an einer Längsschnittstudie („balanced panel") schrumpft grundsätzlich mit jedem zusätzlichen Beobachtungsjahr aufgrund demographischer Ereignisse wie Tod und Auswanderung und – oft selektivem –Ausfallverhalten (*attrition*). Diese negativen Entwicklungen können zumindest im Hinblick auf die Stichprobengröße bei Querschnittuntersuchungen durch die Geburt neuer Haushaltsmitglieder und Zuzüge in bereits teilnehmende Stichprobenhaushalte ausgeglichen werden.

Die Befragungen unterscheiden sich in Bezug auf Erweiterungen der Stichprobe und die Anwendung von „oversampling", also bewusst überrepräsentierenden Teilstichproben von Untersuchungseinheiten mit eher selten vorkommenden Charakteristika. Nachdem Kritik an der mangelnden Repräsentation von Migranten in den Daten der PSID laut geworden war, wurde 1990 eine Stichprobe von Lateinamerikanern („Latino-Sample") hinzugenommen, die jedoch später wieder aufgegeben wurde, sowie eine nicht herkunftslandspezifische Zuwandererstichprobe, die bis heute besteht. Besonderes Merkmal des SOEP ist eine kontinuierliche Erweiterung der zu befragenden Population durch eine Reihe von repräsentativen Auffrischungs- und Ergänzungsstichproben (diese fanden bisher in den Jahren 1998, 2000 und 2006 statt). Darüber erfolgte im Juni 1990, nur wenige Monate nach dem Mauerfall und noch vor Inkrafttreten der Währungs-, Wirtschafts- und Sozialunion, die Ausdehnung des SOEP-Erhebungsgebietes auf Ostdeutschland. Auch das over-sampling bestimmter Bevölkerungsgruppen mit hoher Relevanz für sozialstrukturelle Analysen ist ein Charakteristikum des SOEP. Bei solchen Teilgruppen handelt es sich insbesondere um Personen in Haushalten mit Migrationshintergrund (in den Wellen 1984 und 1995) und einkommensstarke Haushalte (im Jahr 2002). Auch BHPS hat im Rahmen der Integration des UK-ECHP-Samples eine Teilstichprobe von Haushalte mit niedrigem Einkommen aufgenommen, aber später wieder aufgegeben. Für die Befragungen HILDA und SLID ist das Thema der Wiederauffrischung der Stichprobe von etwas geringerer Bedeutung, da SLID sich auf die Überlappung relativ kurz angelegter Längsschnittstudien beschränkt und der HILDA Survey noch verhältnismäßig jung ist. Das SHP ist zwar ebenfalls noch relativ jung, reagierte aber bereits im Jahre 2004 mit einer repräsentativen Auffrischungsstichprobe auf die überdurchschnittlichen Verluste durch Panel Attrition.

Es gibt diverse gute Gründe für repräsentative Stichprobenergänzungen insbesondere in lang laufenden Längsschnittstudien. Solche Erweiterungen dienen nicht nur zur Vergrößerung der Stichprobe bzw. zur Stabilisierung der Fallzahlen im Querschnitt, sondern sind auch Basis für die empirische Untersuchung von Befragungs- oder Paneleffekten in länger laufenden Stichproben (vgl. z.B. Frick et al., 2006). Darüber hinaus können solche Auffrischungsstichproben demographische Veränderungen in der Grundgesamtheit (aufgrund von Immigration nach dem Zeitpunkt der Stichprobenziehung) nachbilden und somit entsprechende Verzerrungen der bereits vorliegenden Stichproben ausgleichen.

Die Ziehung solcher Ergänzungsstichproben ist somit neben dem Saldo der „natürlichen“ Stichprobenentwicklung (Verluste aufgrund demographisch bedingter Ausfälle und Verweigerungen bzw. Gewinne aufgrund von Abspaltungen aus bestehenden Stichproben-Haushalten) die entscheidende Einflussgröße für die Abweichung der befragungsspezifischen Stichprobengröße in der aktuellsten Welle im Vergleich zur jeweiligen Startwelle. Dies kann am Beispiel des SOEP anschaulich gemacht werden. Im Rahmen der ersten Welle im Jahre 1984 wurden knapp 6 000 Haushaltsinterviews (Samples A und B) durchgeführt. Bedingt durch die Ergänzungen um sechs weitere Sub-Stichproben (Samples C bis H) sowie das Entstehen „neuer“ Haushalte, die durch regionale Mobilität aus den ursprünglichen Haushalten hervorgingen, konnten im Jahr 2006 rund 12 500 Haushalte interviewt werden.

Schließlich unterscheiden sich die Studien auch recht deutlich bezüglich der Organisation und Durchführung von Feldarbeit und Datenproduktion. Da SLID die zentrale

Einkommensstichprobe des nationalen statistischen Amtes (Statistics Canada) ist, wird der gesamte Prozess der Datenerfassung, -prüfung und -bereinigung intern betreut. Tendenziell ähnlich ist es bei PSID, wo wissenschaftliche Leitung und Datenbereinigung in ein und demselben wissenschaftlichen Institut vereint sind. Im Falle der anderen Studien erfolgt die Datenerhebung extern durch private Firmen. Codierung und Aufbereitung der Daten für die weitere wissenschaftliche Analyse erfolgt in den meisten der CNEF-Studien in-house. Zwar sind beim SOEP weite Teile der Aufbereitung ebenfalls an das Erhebungsinstitut Infratest Sozialforschung ausgegliedert, jedoch die Generierung nutzerfreundlicher Variablen im Quer- und Längsschnitt, die Imputation fehlender Werte sowie Hochrechung und Gewichtung finden fast ausschließlich beim SOEP selbst statt.

Die Auswahl der hier dargestellten Unterschiede und Gemeinsamkeiten dieser sechs Studien ist offensichtlich fokussiert auf die direkt für die CNEF relevanten Aspekte. Weitere Details zu den diversen Studien finden sich in Frick et al. (2007) sowie auf den entsprechenden Web-Seiten[14]:

- The PSID (http://www.psidonline.isr.umich.edu/)
- The SOEP (http://www.diw.de/gsoep)
- The BHPS (http://www.iser.essex.ac.uk/ulsc/bhps/doc/)
- The SLID (http://www.statcan.ca/start.html)
- The HILDA Survey (http://melbourneinstitute.com/hilda)
- The SHP (http://www.swisspanel.ch)

Zukünftige Entwicklungen dieser und neuer nationaler Befragungen werden in erster Linie die Informations- und Analysebedürfnisse der nationalen politischen Entscheidungsträger und Forschungsstrategien reflektieren. Somit bleibt die Notwendigkeit einer kontinuierlichen Harmonisierung auch der zukünftig noch zu erhebenden nationalen Daten für dieses international vergleichende Projekt erhalten.

5. Ausblick und Zugang zu den CNEF-Daten

Die Daten des CNEF erlauben Forschern mit Interesse an sozio-ökonomischen Fragestellungen, international vergleichende Analysen im Quer- und Längsschnitt für

[14] Im Literaturverzeichnis finden sich Hinweise für vertiefende Darstellungen der einzelnen Befragungen.

Australien, Kanada, Deutschland, Großbritannien, die Schweiz und die USA durchzuführen. Im Gegensatz zu anderen Querschnittsdatenbanken enthält das CNEF nicht nur detaillierte Auskunft darüber, wie die vorliegenden Variablen generiert wurden, sondern unterstützt Wissenschaftler auch bei entsprechenden Modifikationen und Erweiterungen, da die CNEF-Daten mit den Original-Befragungsdaten verknüpfbar sind. Die bisher auf Basis des CNEF vorgelegten Analysen bieten insofern auch Anwendungsbeispiele für zukünftige Weiterentwicklungen im Sinne einer forschungsgetriebenen Verbesserung des Datenangebotes. Die Prozesse zur Harmonisierung bereits bestehender Variablen werden dabei kontinuierlich überprüft und gegebenenfalls optimiert. Derzeit enthält das CNEF zwar nur einen Bruchteil der Originaldaten der nationalen Panels. Laufende Forschungsarbeiten der internationalen Nutzergemeinschaft, die kontinuierlich neue Forschungsbereiche aufgreifen führen aber auch zur Generierung von mit großer Sorgfalt befragungsübergreifend definierten Daten. Das jüngste Beispiel für diese Art der Erschließung neuer Themenfelder ist der Bereich Gesundheit[15]. Gleichzeitig wird die verbesserte Zusammenarbeit zwischen Datenproduzenten und Analysten, die derzeit gemeinsam an der nachträglichen Harmonisierung der vorliegenden Survey Daten arbeiten, auch die zukünftige ex-ante Harmonisierung neuer Bereiche unterstützen. Dies wird wiederum zur Verbesserung der internationalen Vergleichbarkeit der Mikro-Daten beitragen und so auch die Qualität der Forschungsergebnisse steigern.

Der Zugang zu den CNEF-Daten ist abhängig von nationalen Datenschutzbestimmungen. Da PSID-Daten grundsätzlich öffentlich zugänglich sind, werden die PSID-CNEF-Datensätze auf der CNEF-Webseite ebenfalls der Öffentlichkeit zugänglich gemacht. Der Zugang zu BHPS-CNEF, SOEP-CNEF, HILDA-CNEF und SHP-CNEF-Daten ist nur mit der Erlaubnis des jeweiligen nationalen Datenproduzenten möglich[16]. Diese Erlaubnis muss dann wiederum, per e-mail oder Fax, an das CNEF-Büro an der Cornell Universität geschickt werden. Erst dann wird die CNEF-CD zugestellt. Eine Ausnahme bilden die SLID Daten, die sich aufgrund der kanadischen Datenschutzbestimmungen nicht auf der CNEF-CD befinden. Registrierte CNEF-Nutzer haben die Möglichkeit, im Rahmen einer Datenfernverarbeitung ihre Analyse-

[15] Als zukünftige Erweiterungen werden derzeit subjektive Indikatoren wie die „Allgemeine Lebenszufriedenheit“ diskutiert sowie zusätzliche „nicht-monetäre“ Einkommenskomponenten zur Ergänzung der bereits verfügbaren Messung des fiktiven Mietwerts für selbstnutzende Eigentümer. Diese Informationen werden in der Variable „Imputed Rental Value of owner-occupied housing“ (Variable I11105_xxxx) zur Verfügung gestellt.

[16] Es handelt sich hierbei jedoch um eine relativ einfache und schnell abgewickelte Formalität. Eine detaillierte Beschreibung der relevanten Schritte findet sich auf der CNEF-Webseite der Cornell Universität.

Programme an Statistics Canada zu senden. Diese werden dort nach datenschutzrechtlicher Prüfung ausgeführt und die entsprechenden Ergebnisse an den Forscher zurück gesandt.[17].

Die einmalige Anmeldegebühr für die Daten des CNEF beträgt 125 US Dollar, die an die Cornell Universität gezahlt wird. Für weitere Informationen zum Datenzugang konsultieren Sie bitte die CNEF-Webseite unter
http://www.human.cornell.edu/che/PAM/Research/Centers-Programs/German-Panel/cnef.cfm
oder senden Sie eine e-mail an cnef@cornell.edu

Literatur

Burkhauser, R.V. / *Crews-Cutt* A. / *Lillard* D.R. (1999) How Older People in the United States and Germany Fared in the Growth Years of the 1980s: A Cross-Sectional versus a Longitudinal View. Journal of Gerontology, September, Vol. 54B, No. 5, S279-S290.

Burkhauser, R. V., *Frick*, J. R. and *Schwarze*, J. (1997): A Comparison of Alternative Measures of Economic Well-Being for Germany and the United States. Review of Income and Wealth, 43(2), 153-171.

Burkhauser, R.V. / *Giles*, P. / *Lillard*, D.R., / *Schwarze*, J. (2005): Until Death Do Us Part: An Analysis of the Economic Well-Being of Widows in Four Countries. Journal of Gerontology: Social Sciences. Vol. 60B No. 5, S238-S246.

Burkhauser, R.V. / *Lillard*, D.R. (2005): The Contribution and Potential of Data Harmonization for Cross-National Comparative Research, Journal of Comparative Policy Research 7, 313-330.

Burkhauser, R.V. / *Lillard*, D.R. (2007): The Expanded Cross-National Equivalent File: HILDA joins its International Peers, The Australian Economic Review 40, 208-215.

Burkhauser, R. V. / *Schmeiser*, M. D. / *Schroeder*, M. 2007. The Employment and Economic Well Being of Working-Age Men with Disabilities: Comparing Outcomes in Australia, Germany, and Great Britain with the United States. Paper presented at the HILDA Survey Research Conference 2007, July 19-20, 2007 University of Melbourne

[17] Im Rahmen kürzlich geänderter Bestimmungen ist der Zugang zu den SLID-Daten nun auch im Rahmen eines persönlichen Forschungsaufenthaltes in einem der Research Data Centres (RDC) in Kanada möglich. Dieses Netzwerk umfasst vierzehn Forschungsdatenzentren, sechs Zweigstellen und das staatliche Forschungszentrum in Ottawa.

Butrica, B.A. / *Burkhauser*, R. V. (1997): Estimating Tax Burdens in the PSID Using the TAXSIM Model, Aging Studies Program Project Paper No. 12, Center for Policy Research, The Maxwell School, Syracuse University, Syracuse NY.

Butz, W.B. / *Torrey*, B.B. (2006): Some frontiers in Social Sciences, Science 312 (30 June 2006), 1898-1900

Canberra Group (2001): Expert Group on Household Income Statistics: Final Report and Recommendations, Ottawa.

Feenberg, D. / *Coutts*, E. (1993): An Introduction to the TAXSIM Model, The Journal of Policy Analysis and Management 12, 189-194.

Frick, J. R., / *Jenkins*, S.P., / *Lillard*, D. R., *Lipps*, O. and *Wooden*, M. (2007): The Cross-National Equivalent File (CNEF) and its Member Country Household Panel Studies. Schmoller's Jahrbuch - Journal of Applied Social Science Studies. 127 (4): 627-654.

Frick, J. R. / *Goebel*, J. / *Schechtman*, E. / *Wagner*, G. G. / *Yitzhaki*, S. (2006): Using Analysis of Gini (ANoGi) for detecting whether two sub-samples represent the same universe: The German Socio-Economic Panel Study (SOEP) Experience, Sociological Methods & Research 34 (4), 427-468

Frick, J. R. / *Grabka*, M.M. (2007): Item Non-Response and Imputation of Annual Labor Income in Panel Surveys from a Cross-National Perspective. IZA Discussion Paper No. 3043, September 2007, Bonn: IZA. http://ftp.iza.org/dp3043.pdf

Haisken-DeNew, J. P. / *Frick*, J.R. (2005): Desktop Companion to the German Socio-Economic Panel Study (GSOEP), Version 8.0 – Update to Wave 21, DIW Berlin.

Headey, B.W. (2003): How Best to Impute Taxes and Measure Public Transfers? HILDA Discussion Paper Series no. 2/03, Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

Jenkins, S.P. / *Schluter*, C. (2003): Why are Child Poverty Rates Higher in Britain than in Germany? A Longitudinal Perspective, Journal of Human Resources 38, 441-465.

Jenkins, S.P. / *Schluter*, C. / *Wagner*, G. (2003): The Dynamics of Child Poverty: Britain and Germany Compared, Journal of Comparative Family Studies 34, 337-353.

Levy, H. / *Zantomio*, F. / *Sutherland*, H. / *Jenkins*, S.P. (2006): Documentation for Derived Current and Annual Net Household Income Variables, BHPS Waves 1-14. [Retrieved from: http://www.data-archive.ac.uk/doc/3909/mrdoc/pdf/3909userguide.pdf]

Lillard, D. / *Burkhauser*, R.V. (2006): Evaluation of the Cross-National Comparability of the Survey of Health, Ageing, and Retirement in Europe, the Health and Retirement Study, and the English Longitudinal Study of Ageing, Report prepared for the National Institutes of Aging, Washington DC.

Lillard, D / Wagner, G.G. (2006); The Value Added of Biomarkers in Household Panel Studies. Data Documentation No. 14, Berlin: DIW Berlin.

Schwarze, J. (1995): Simulating the Federal Income and Social Security Tax Payments of German Households Using Survey Data, Cross-National Studies in Aging Program Project Paper No. 19, Center for Policy Research, The Maxwell School, Syracuse University, NY.

Smeeding, T.M. / *Jesuit*, D. K. / *Alkemade*, P (2002): The LIS / LES Project Databank: Introduction and Overview. Schmollers Jahrbuch (Journal of Applied Social Science Studies) 122, 497-717.

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

Watson, N. (2006): Options for a Top-up Sample to the HILDA Survey, Paper presented at the ACSPRI Social Science Methodology Conference, University of Sydney, 10-13 *December.*

Zimmermann E. / *Tillmann*, R. (eds) (2004): Vivre en Suisse 1999-2000 - Leben in der Schweiz 1999-2000. Une année dans la vie des ménages et familles en Suisse. Ein Jahr im Leben der Schweizer Familien und Haushalte (Vol. 3. Population, Famille et Société), Peter Lang Verlag AG, Berne.

Übersicht 1: Synopse zentraler Merkmale der Panelstudien im CNEF

Merkmal	*PSID*	**SOEP**	*BHPS*	*SLID*	*HILDA Survey*	*SHP*
Institut	Institute for Social Research, University of Michigan	SOEP; Deutsches Institut für Wirtschaftsforschung (DIW Berlin)	Institute for Social and Economic Research, University of Essex	Statistics Canada	Melbourne Institute of Applied Economic and Social Research, University of Melbourne	Schweizer Haushalt-Panel, Universität Neuchâtel (seit 2008: Forschungszentrum Sozialwissenschaften (FORS), Universität Lausanne)
Finanzierung	National Science Foundation, National Institute of Health, sowie einige andere Organisationen[(a)]	1984 bis 2002: Deutsche Forschungsgemeinschaft (DFG); Bundesministerium für Bildung und Forschung (BMBF) Seit 2003: Wissenschafts-gemeinschaft Gottfried Wilhelm Leibniz (WGL)[(b)]	UK Economic and Social Research Council	Statistics Canada	Australian Government Department of Families, Community Services and Indigenous Affairs. Reserve Bank of Australia (für Welle 2 im Jahr 2002)	Schweizerischer Nationalfonds
Studiendesign	Unbegrenzt laufende Längsschnittstudie	Unbegrenzt laufende Längsschnittstudie	Unbegrenzt laufende Längsschnittstudie	Überlappende Längsschnittstudien von jeweils sechsjähriger Dauer	Unbegrenzt laufende Längsschnittstudie	Unbegrenzt laufende Längsschnittstudie
Jahr des ersten Interviews	1968	1984	1991	1993	2001	1999
Referenz-Population / Erhebungs-einheit	Haushaltsvorstände, deren Haushalt seit mindestens zwei Jahren ohne Unterbrechung in den USA lebt.	Alle privaten Haushalte. Alle Mitglieder ab dem Kalenderjahr, in dem sie 17 Jahre alt werden.	Alle privaten Haushalte. Alle Mitglieder im Alter von 16 und mehr Jahren	Private Haushalte der zehn Provinzen, mit Ausnahme der Indianerreservate. Alle Mitglieder im Alter von 16 und mehr Jahren. Proxy-Interviews zulässig.	Alle privaten Haushalte, mit Ausnahme derer in sehr abgeschiedener Lage. Alle Mitglieder im Alter von 15 und mehr Jahren.	Alle privaten Haushalte. Alle Mitglieder im Alter von 14 und mehr Jahren.
Erhebungs-	Wellen 1-5 (1968-	Wellen 1-14 (1984-1997)	Wellen 1-9 (1991-1999)	Seit Welle 1 (1993)	Seit Welle 1 (2001) PAPI	Seit Welle 1 (1999)

Merkmal	*PSID*	**SOEP**	*BHPS*	*SLID*	*HILDA Survey*	*SHP*
methode	1972) PAPI. Seit Welle 6 (1973) hauptsächlich Telefon. Seit Welle 26 (1993) CATI.	PAPI. Seit Welle 2 (1985) Methoden-Mix (face-to-face und Selbstausfüller). Seit Welle 15 (1998) Übergang zu CAPI.	PAPI sowie ein kurzer selbst auszufüllender Fragebogen. Seit Welle 10 (2000) CAPI. Seit Welle 3 (1993) kurzes Telefon-Interview als letzte Möglichkeit.	CATI.	sowie ein kurzer selbst auszufüllender Fragebogen. Kurzes Telefon-Interview als letzte Möglichkeit.	CATI.
Weiter-verfolgungs-regeln	Mitglieder der Stichprobe in Welle 1 und deren Nachkommen und adoptierten Kinder. Informationen über Personen, die mit einem der ursprünglichen Mitglieder der Stichprobe oder mit deren Nachkommen oder adoptierten Kindern zusammenleben.	Mitglieder der Stichprobe in Welle 1 und deren Nachkommen und adoptierten Kinder. Seit Welle 5 (1988) werden Personen, die mit einem ursprünglichen Stichprobenmitglied zusammengewohnt haben, im Falle eines Umzugs innerhalb Deutschlands ebenfalls weiterverfolgt.	Mitglieder der Stichprobe in Welle 1 und deren Nachkommen und adoptierten Kinder. Personen, die mit einem ursprünglichen Mitglied zusammenwohnen werden zu Mitgliedern der aktuellen Welle. Personen, die ein Kind mit einem ursprünglichen Mitglied haben, werden permanente Mitglieder.	Mitglieder der Stichprobe in Welle 1.	Mitglieder der Stichprobe in Welle 1 und deren Nachkommen und adoptierten Kinder Personen, die mit einem ursprünglichen Mitglied zusammenwohnen werden zu Mitgliedern der aktuellen Welle. Personen, die ein Kind mit einem ursprünglichen Mitglied haben, werden permanente Mitglieder.	Mitglieder der Stichprobe in Welle 1 und deren Nachkommen und adoptierten Kinder Personen, die mit einem ursprünglichen Mitglied zusammenwohnen werden zu Mitgliedern der aktuellen Welle. Personen, die ein Kind mit einem ursprünglichen Mitglied haben, werden permanente Mitglieder.
Proxy-Interviews (Erwachsene)	Ja—zu 100 Prozent. 1976 und 1985 wurden auch Partner befragt.	Nein—keine	Ja—zwischen 2 und 4 Prozent	Ja—circa 30 Prozent	Nein—keine	Ja—2 bis 3 Prozent
Zahl der Interviews in Startwelle	4 802 Familien	5 921 Haushalte	5 538 Haushalte	15 006 Haushalte	7 682 Haushalte	5 074 Haushalte
Zahl der HH-Interviews in der jüngsten Welle	8 002 Haushalte (Welle 34, 2005)	11 689 Haushalte (Welle 24, 2007)	8 709 Haushalte (Welle 15, 2005)	38 776 Haushalte (Welle 5 / 3. Panel, Welle 2 / 4. Panel, in 2003)	7 139 Haushalte (Welle 6, 2006)	4 256 Haushalte (Welle 7, 2005)

Merkmal	*PSID*	**SOEP**	*BHPS*	*SLID*	*HILDA Survey*	*SHP*
Over-sampling (Überrepräsentation) / Zusatzstichprobe	Welle 1 (1968) – Oversampling einkommensschwacher Haushalte (*n*=1 872). Ca 2/3 dieser Stichprobe wurden ab 1997 nicht weiterverfolgt. Welle 23 (1990) – „Latino“ Sample (nur bis 1995 erhoben). Welle 30 (1997) – Zuwanderer top-up Sample	Welle 1 (1984) - Oversampling von Migranten-Haushalten (*n*=1 393). Welle 7 (1990) – Repräsentative Stichprobe Ostdeutschland (*n*=2 179 Haushalte) Welle 12 (1995) – Oversampling „Zuwanderer seit 1984“ Wellen 15 (1998), 17 (2000) und 23 (2006) – Repräsentative Ergänzungsstichproben Welle 19 (2002) – Zusatzstichprobe „Einkommensstarke Haushalte“	Welle 7 (1997) – Niedrigeinkommens Stichprobe für ECHP –ab Welle 12 (2002) nicht weiterverfolgt. Welle 9 (1999) – Stichproben für Schottland und Wales Welle 11 (2001) – Stichprobe für Nord-Irland	Stichprobe basiert auf dem Labour Force Survey, wobei kleinere Regionen überrepräsentiert werden	Keine.	Welle 6 (2004) – Repräsentative Ergänzungs- und Auffrischungs-stichprobe
Welle 1 Teilnahme-Quote in % aller Haushalte	76%	Jeweils vollständig befragte Haushalte: 1984 Sample A: = 61% 1984 Sample B: = 68% 1990 Sample C: = 70% Jeweils inkl. teilweise realisierte Haushalte 1998 Sample E: = 54% 2000 Sample F = 52% 2006 Sample H = 41%	Vollständig realisierte Haushalte = 69% (inkl. Proxy Interviews) Inkl. teilweise realisierte Haushalte = 74% 1999 Sample Schottland/ Wales, inkl. teilweise realisierte Haushalte = 63% 2001 Sample Nord-Irland, inkl. teilweise realisierte Haushalte = 69%	93%	Vollständig realisierte Haushalte = 59% Inkl. teilweise realisierte Haushalte = 66%	Inkl. teilweise realisierte Haushalte = 49%
Teilnahme-Quote [(c)]:				Teilnahmequote in Welle 5:		

Merkmal	*PSID*	**SOEP**	*BHPS*	*SLID*	*HILDA Survey*	*SHP*
Welle 5 Welle 10 Welle 15 Welle 20	81% 70% 61% 52%	69% (71%)[(d)] 53% (55%) 41% (44%) 31% (35%)	72%[(e)] 62% - -	82% (Panel 1) 79% (Panel 2) 76% (Panel 3)	74% - - -	56% - - -
Feldarbeit	Datenerfassung extern. Panel-Management und Datenbereinigung intern.	Datenerfassung, Teile des Managements sowie der Datenbereinigung extern. Weitergehende Datenprüfung- und Generierung intern.	Datenerfassung extern. Management der Studie und Datenbereinigung intern.	Alles intern.	Datenerfassung, Management und Datenverarbeitung extern.	Datenerfassung extern. Management der Studie und Datenbereinigung intern.
Datenweitergabe	Frei zugänglich (online verfügbar).	DVD. Zugang nur für wissenschaftliche Forschungszwecke. Zusätzlich Fernzugang für zweckgebundene Forschung möglich.	Im UK Data-Archive hinterlegt	Derzeit nur über Fernzugang oder Vor-Ort Zugang bei Statistics Canada.	CD-Rom. Zugang nur für wissenschaftliche Forschungszwecke im Rahmen eines definierten Forschungsprojektes.	CD-Rom. Zugang nur für wissenschaftliche Forschungszwecke im Rahmen eines definierten Forschungsprojektes.

Fußnoten:

a Die ursprüngliche Finanzierung der PSID erfolgte durch das Office of Economic Opportunity of the United States Department of Commerce. Außerdem haben zur Finanzierung des PSID beigetragen: National Institute on Aging, National Institute of Child Health and Human Development, Offices of the Assistant Secretary for Planning and Evaluation of the United States Department of Health and Human Services, Economic Research Service of the United States Department of Agriculture, United States Department of Housing and Urban Development, United States Department of Labor, Center on Philanthropy at the Indiana University-Purdue University.

b Die Deutsche Forschungsgemeinschaft (DFG) und die Wissenschaftsgemeinschaft Gottfried Wilhelm Leibniz (WGL) werden finanziert durch die Bundesregierung und die Länder-Regierungen über die Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung.

c Mit Ausnahme der Angaben für die PSID zeigen diese Teilnahme-Quoten den Anteil der Befragten aus Welle 1, die auch in späteren Wellen erfolgreich befragt wurden. Die Zahlen für die PSID beziehen sich auf die Anzahl der Haushaltsmitglieder aus Welle 1, die auch in späteren Wellen noch im Sample enthalten sind (vgl. Fitzgerald et al., 1998 Tabelle 1). Dies hat zur Folge, dass die Angaben für die PSID in diesem Zusammenhang nur eingeschränkt mit jenen der anderen Studien vergleichbar sind.

d Die Angaben in Klammern gelten nur für die westdeutsche Stichprobe A.

e Diese Angaben gelten nur für Befragte mit vollständig absolviertem Interview.

Anhang 1. Variablen im Cross-National Equivalent File (CNEF) 1980-2005

Label	Data	Variable name
Demographie (Demographics)**:**		
Age of Individual	B, G, H, P, S, CH	D11101_xxxx
Sex of Individual	B, G, H, P, S, CH	D11102LL
Marital Status of Individual	B, G, H, P, S, CH	D11104_xxxx
Relationship to Household Head	B, G, H, P, S, CH	D11105_xxxx
Number of Persons in Household	B, G, H, P, S, CH	D11106_xxxx
Number of Children in Household	B, G, H, P, S, CH	D11107_xxxx
Education With Respect to High School	G, H, P, S, CH	D11108_xxxx
Number of Years of Education	G, H, P, S, CH	D11109_xxxx
Race of Individual[a]	B, P, S	D11112LL
Erwerbstätigkeit (Employment)**:**		
Annual Work Hours of Individual	B, G, H, P, S, CH	E11101_xxxx
Impute Annual Work Hours of Individual	B, CH	E11201_xxxx
Employment Status of Individual	B, G, H, P, S, CH	E11102_xxxx
Employment Level of Individual	B, G, H, P, S, CH	E11103_xxxx
Primary Activity of Individual	B, G, P, S, CH	E11104_xxxx
Occupation of Individual	B, G, H, P, S, CH	E11105_xxxx
1 Digit Industry Code of Individual	B, G, H, P, S, CH	E11106_xxxx
2 Digit Industry Code of Individual	B, G, H, P, S, CH	E11107_xxxx
Input zur Berechnung von Äquivalenzskalen (Equivalence Scale inputs)**:**		
Number HH members age 0-14	B, G, H, P, S, CH	H11101_xxxx
Number HH members age 15-18	B, G, H, P, S, CH	H11102_xxxx
Number HH members age 0-1	B, G, H, P, S, CH	H11103_xxxx
Number HH members age 2-4	B, G, H, P, S, CH	H11104_xxxx
Number HH members age 5-7	B, G, H, P, S, CH	H11105_xxxx
Number HH members age 8-10	B, G, H, P, S, CH	H11106_xxxx
Number HH members age 11-12	B, G, H, P, S, CH	H11107_xxxx
Number HH members age 13-15	B, G, H, P, S, CH	H11108_xxxx
Number HH members age 16-18	B, G, H, P, S, CH	H11109_xxxx
Number HH members age 19+ or 16-18 and indep.	B, G, H, P, S, CH	H11110_xxxx
Indicator - Wife/spouse in HH	B, G, H, P, S, CH	H11112_xxxx
(Vor-)Jahreseinkommen (Yearly Income)**:**		
Household Pre-Government Income	B, G, H, P, S, CH	I11101_xxxx
Household Post-Government Income	B, G, H, P, S, CH	I11102_xxxx
Household Labor Income	B, G, H, P, S, CH	I11103_xxxx
Household Asset Income	B, G, H, P, S, CH	I11104_xxxx
Household Imputed Rental Value	B, G, H, P, S, CH	I11105_xxxx
Household Private Transfers	B, G, H, P, S, CH	I11106_xxxx
Household Public Transfers	B, G, H, P, S, CH	I11107_xxxx
Household Social Security Pensions	B, G, P, S, CH	I11108_xxxx
Total Household Taxes	B, G, H, P, S, CH	I11109_xxxx
Individual Labor Earnings	B, G, H, P, S, CH	I11110_xxxx
Household Federal Taxes	G, P	I11111_xxxx
Household Social Security Taxes	B, G, P, CH	I11112_xxxx
Household Post-Government Income (TAXSIM)	P	I11113_xxxx
Total Household Taxes (TAXSIM)	P	I11114_xxxx
Household State Taxes (TAXSIM)	P	I11115_xxxx
Household Federal Taxes (TAXSIM)	P	I11116_xxxx
Household Private Retirement Income	B, G, H, P, S	I11117_xxxx
Household Windfall Income	B, G, H, P, S, CH	I11118_xxxx
Impute Household Pre-Government Income	B, G, H, CH	I11201_xxxx

Impute Household Post-Government Income	B, G, H, CH	I11202_xxxx
Impute Household Labor Income	B, G, H, CH	I11203_xxxx
Impute Household Asset Income	B, G, H, CH	I11204_xxxx
Impute Household Imputed Rental Value	B, G, CH	I11205_xxxx
Impute Household Private Transfers	B, G, H, CH	I11206_xxxx
Impute Household Public Transfers	B, G, CH	I11207_xxxx
Impute Household Social Security Pensions	B, G, CH	I11208_xxxx
Impute Total Household Taxes	G, H, CH	I11209_xxxx
Impute Individual Labor Earnings	B, G, H, CH	I11210_xxxx
Impute Private Retirement Income	B, G, H	I11217_xxxx
Regionalinformation (Location):		
Area of Residence [b]	B, G, P, S, CH	L11101_xxxx
Region of Residence[89]	B, G, H, CH	L11102_xxxx
Gesundheit (Medical/health)**:**		
Whether spent night in hospital in last year	B, G, P, CH	M11101_xxxx
Number of nights (days) spent in hospital	B, G, P, CH	M11102_xxxx
Whether had accident in past year that required hospital	B, G, CH	M11103_xxxx
Frequency of sports or exercise	B, G, P, CH	M11104_xxxx
Have had stroke	B, P	M11105_xxxx
Have or had high blood pressure/hypertension	B, P	M11106_xxxx
Have or had diabetes	B, P	M11107_xxxx
Have or had cancer	B, P	M11108_xxxx
Have or had psychiatric problems	B, P	M11109_xxxx
Have or had arthritis	B, P	M11110_xxxx
Have or had angina or heart condition	B, P	M11111_xxxx
Have or had asthma or breathing difficulties	B, P	M11112_xxxx
Have trouble climbing stairs	B, G, P	M11113_xxxx
Have trouble with bath	B, P	M11114_xxxx
Have trouble dressing	B, G, P	M11115_xxxx
Have trouble getting out of bed	B, G, P	M11116_xxxx
Have trouble shopping	G, P	M11117_xxxx
Have trouble walking	B, P	M11118_xxxx
Have trouble doing housework	B, G, P	M11119_xxxx
Have trouble bending, lifting, stooping	B, P	M11120_xxxx
Health limits vigorous physical activities	B, P	M11121_xxxx
Height (in meters)	G, P, CH	M11122_xxxx
Weight (in kilos)	G, P, CH	M11123_xxxx
Disability Status of Individual	B, G, H, P, S	M11124_xxxx
Subjective Satisfaction with Health	B, G, H, S, CH	M11125_xxxx
Self-Rated Health Status	B, G, H, P, CH	M11126_xxxx
Number of Times Visited Dr. in Past Year	G, CH	M11127_xxxx
Hochrechnung und Gewichtung (Weights):		
Cross-sectional Weight - Respondent Individuals	B, G, H, P, S, CH	W11101_xxxx
Household Weight	B, G, H, P, S, CH	W11102_xxxx
Longitudinal Weight - Respondent Individuals	B, G, H, P, S, CH[90]	W11103_xxxx
Population Factor for W11101_xxxx	B, G, P	W11104_xxxx
Individual Weight - Immigrant Sample	G	W11105_xxxx
Household Weight - Immigrant Sample	G	W11106_xxxx
Cross-sectional Weight - Enumerated Individuals	B, H	W11107_xxxx
Longitudinal Weight - Enumerated Individuals	B, H	W11108_xxxx
Population Factor for W11103_xxxx	B, G, P	W11109_xxxx
Population Factor for W11107_xxxx	B	W11110_xxxx

[89] In the SHP region of residence gives the language region of the interview (German, French, Italian).
[90] W11203 for combined SHP I (original) and SHP II (refreshment) sample.

Population Factor for W11108_xxxx	B	W11111_xxxx
Algorithmen zur Ableitung von Äquivalenzskalen (Equivalence Weight Algorithms):		
Detailed Official U.S. Equivalence Weight		
General Official U.S. Equivalence Weight		
Official German Equivalence Weight		
ELES Equivalence Weight		
OECD Equivalence Weight		
McClements Equivalence Weight		
Other Equivalence Weights		
Identifikatoren (Identifiers)**:**		
Unique Person Number	B, G, H, P, S, CH	X11101LL
Household Identification Number	B, G, H, P, S, CH	X11102_xxxx
Individual in Household at Survey	B, G, H, P, S	X11103_xxxx
Oversample Identifier	B, G, P, S	X11104LL
Person in Household Interviewed	B, G, H, CH	X11105_xxxx
Makro-Indikatoren (Macro-level Variables):[c]		
Consumer Price Index	B, G, P, S	
Median Pre-government Household Income	B, G, P, S	
Median Post-government Household Income	B, G, P, S	
Purchasing Power Parity for East Germany	G	

(B) BHPS: Erhebungsjahre 1991-2004
(G) SOEP: Erhebungsjahre 1984-2005
(H) HILDA: Erhebungsjahre 2001-2004
(P) PSID: Erhebungsjahre 1980-2003
(S) SLID: Erhebungsjahre 1992-2005
(CH) SHP: Erhebungsjahre 1999-2005

[a] Die Variable "Race" wird in BHPS und SLID für alle Stichprobenmitglieder erhoben. In den PSID-Daten trifft dies nur für solche Personen zu, die in mindestens einer Welle Haushaltsvorstand oder Partner waren.
[b] "Area of residence" entspricht in der BHPS dem *Local Authority District of Residence*, dem *Bundesland* in den SOEP-Daten, dem *US-Bundesstaat* in der PSID, dem *Kanton* im SHP. Diese Information ist in den CNEF-Daten des SLID auf der CD-Rom nicht enthalten, aber im Rahmen eines Forschungsaufenthalts bei Statistics Canada verfügbar. Zugang zu Informationen des *Local Authority District of Residence* in den BHPS-Daten ist im Rahmen spezieller Arrangements mit der University of Essex möglich.
[c] Da diese zeitpunktspezifischen Indikatoren auf Makro-Ebene für alle Individuen oder Haushalte eines Erhebungsjahres gelten, sind diese lediglich in den CNEF-Handbüchern gelistet.

Anhang 2: Fallzahlen der nationalen Panelstudien im CNEF (Individualdaten)

Jahr	*PSID*	*SOEP*	*BHPS*	*SLID*	*HILDA*	*SHP*
1980	18989	-	-	-	-	
1981	18992	-	-	-	-	
1982	19246	-	-	-	-	
1983	19491	-	-	-	-	
1984	19570	15237	-	-	-	
1985	19787	13747	-	-	-	
1986	19615	13084	-	-	-	
1987	19647	12853	-	-	-	
1988	19687	12253	-	-	-	
1989	19669	11856	-	-	-	
1990	19932	17462	-	-	-	
1991	19962	17094	13780	-	-	
1992	20334	16801	13151	40155	-	
1993	21450	16510	13104	42194	-	
1994	23620	16828	12851	43717	-	
1995	23182	17252	12549	88230	-	
1996	23060	16869	12720	91624	-	
1997	19132	16559	15042	94125	-	
1998	-	18161	14835	139508	-	
1999	19669	17417	21540	94772	-	10437
2000	-	30439	21602	96512	-	9454
2001	20538	27481	26586	141598	19914	8775
2002	-	29280	23435	93680	18295	7648
2003	21277	27553	22559	95792	17691	6944
2004	-	26690	22105		17209	10666
2005	22918	25544	15627		17469	8550
Insgesamt (Personen * Wellen)	449.767	416.970	261.486	1.061.907	90.578	62.474

Hinweis: Diese Fallzahlen können von jenen in der jeweiligen nationalen Dokumentation abweichen. So sind z.B. im Format des PSID-CNEF das Niedrigeinkommens- und das Latino-Sample ausgeschlossen und der SOEP-CNEF-Datensatz basiert aus Datenschutzgründen lediglich auf einer 95 % Zufallsstichprobe der Daten der ersten Welle.

Income Imputation in the Swiss Household Panel 1999-2007

Oliver Lipps[1]

This paper describes the methods used and the steps taken to impute missing income values in the Swiss Household Panel Survey (SHP). Missing values that result from both item- and unit-nonresponse are imputed. We impute income on the individual level, distinguishing between several income components.

The imputed item- and unit nonresponse income distributions are compared with the distributions of the validly reported cases. The level of imputed income from *employment* resulting from *item*-nonresponse is similar to that of the validly reported. Other imputed income components from *item*-nonresponse are generally slightly, imputed income from *unit*-nonresponse considerably higher than that from the validly reported cases. This result shows that imputing missing income may avoid biased level estimates. Income variations of the valid cases and the imputed cases are not different.

Keywords: item-nonresponse, unit-nonresponse, imputation, income

JEL-classification: C81, D31, I32

[1] FORS, c/o University of Lausanne, Bâtiment Vidy, CH – 1015 Lausanne
oliver.lipps@fors.unil.ch
Acknowledgment: This study has been realized using the data collected in the "Living in Switzerland" project, conducted by the Swiss Household Panel (SHP), which is based at the Swiss Centre of Expertise in the Social Sciences FORS, University of Lausanne. The project is financed by the Swiss National Science Foundation. I thank two colleagues for valuable comments on a former version of this paper.

Missing Income and Imputation in the SHP in brief

Item nonresponse occurs in surveys if respondents are not able or not willing to give a valid answer on survey questions. Mostly, difficult or sensitive questions such as income questions are concerned. In some cases, also technical errors might be responsible. Another source of nonresponse are individuals who do not give an interview at all (unit nonresponse). Both forms of nonresponse occur in both cross-sectional and panel surveys; in the latter in all or only some waves.

To impute missing income data in the Swiss Household Panel (SHP), we generally use both cross-sectional and longitudinal methods for all income variables.[2] For individuals with a never validly reported income component, a cross-sectional method must be used first: the income is "initialized" using a stochastic regression. We start by using all available relevant covariates, and impute the missing value in the first wave to be imputed with all covariates validly given. If one of the covariates is missing in all waves, we stepwise drop covariates according to significance. The last initialization step, which involves only a few missing values, is a median imputation. This procedure is repeated using a reversed order of waves, i.e., from the most recent to the first wave ("last-first") to be imputed. In case of a discrepancy between the first-last and the last-first initialized value, the value that is imputed in an earlier regression step is used. If the imputation step is the same, one of the two is randomly selected.

Once the income component is initialized or validly reported in at least one wave by all income eligible individuals, we impute missing income data in all other waves using longitudinal imputation methods. Similar methods to those performed in the German Socio-Economic Panel (SOEP) (Frick and Grabka 2004), or the Household, Income and Labour Dynamics in Australia (HILDA) Panel Survey (Starick and Watson 2007) are used, e.g. by also giving preference to the Little & Su (L&S) imputation technique. Unlike the SOEP, which does not distinguish class variables to match recipients and possible donors at all, the HILDA uses age classes, and we use education as class variable.

We impute missing income values for all income components of all individuals, who report to having received income from this source without giving the amount or a guesstimate. In addition we impute missing income values asked in a proxy interview. If the likelihood is high that a *unit*-nonresponding individual earns income from a specific source, this is also imputed. Concluding an income from a specific source is possible, because in the SHP socio-demographic information such as sex, age, education, or especially occupational

[2] Frick and Grabka (2004) give an overview of the commonly used single imputation techniques (p. 6 ff.).

status of all household members is available from the household grid questionnaire. Being listed in the household grid is a necessary condition to be eligible for imputation.

Motivation for Item and Unit missing Income Imputation in the SHP

An obvious motivation to impute missing income data is that not doing so leads to a loss of prediction power due to listwise deletion of cases. Also, because of its membership in the Cross National Equivalent File (CNEF[3]), the Swiss Household Panel (SHP) must deliver imputed item-nonresponse income variables to the CNEF (Frick et al. 2007, Lipps and Kuhn 2009). The major motivation however is that using only validly reported income may bias both population level and variation estimates and model results.

Although weights are designed to help correct for unit-nonresponse, they do nothing in correcting for the bias due to item-nonresponse. However item-missing values are not missing completely at random. This makes imputation of item-missing income values necessary. In the SOEP, "ignoring cases with item-nonresponse tends to underestimate income levels as well as variance …. Additionally, in line with findings in the literature, item-nonresponse on income appears to be selective with respect to both tails of the income distribution, especially at the upper end" (Frick and Grabka 2004:20). Therefore, imputing missing income cases may give more realistic income level and variation measures. In longitudinal analyses, income mobility seems to be underestimated using only validly reported cases (Frick and Grabka 2007).

With the exception of labor income, all (imputed) income variables in the SHP delivered to the CNEF are aggregated on the household level. This renders the imputation of missing *units* (individuals) important to avoid underestimation of these aggregated sums. (Partial household) unit nonresponse is a problem particularly in centralized telephone household surveys such as the SHP and mostly affects individuals other than household reference persons (Lipps 2009). Frick et al. (2009) suggest the following measures to deal with unit-nonresponse:

[3] Apart from the SHP, current members of the CNEF are the SOEP, the HILDA, the U.S. Panel Study of Income Dynamics (PSID), the British Household Panel Survey (BHPS), and the Canadian Survey of Labour and Income Dynamics (SLID).

"(a) Ignoring the fact that a household member (and its income information) is missing, thus assuming the non-responding individual's income is zero ...

(b) Adjusting the calculation of the equivalence scale by ignoring the person's contribution to household income as well as to household needs, thus in principle ignoring the person's existence…

(c) Eliminating all households observed with [nonresponding individuals] …, thus assuming that these households are missing completely at random" (p. 2).

(d) Imputing unit-nonresponding income values.

Frick et al. (2009) find that applying (a) to (c) results in "a systematic downward bias in level and development of income inequality and relative poverty whereas income mobility will be overstated" (p. 31). They conclude that imputation of various components instead of only adjusting the income measure may be considered advantageous.

Income Components imputed

All income components to be imputed are collected in the individual questionnaire in the SHP. Some income concepts and therefore income questions and calculation algorithms have changed since the start of the SHP in 1999 (Kuhn 2008) and are fully comparable only since 2002. Since we impute and construct all income components from 1999 on, using different algorithms, care must be taken when comparing income across waves until 2001, and from a wave before 2002 with one after 2001.

Imputed income variables comprise the following income sources (annual amounted), constructed from the original income variables asked (Kuhn 2008):[4]

(1) Income from employment: net ("i$$empyn")[5]

(2) Income from independent work: net ("i$$indyn")

(3) (old age) pension: annual income ("i$$oasiy")

(4) Invalidity pension: ("i$$aiy")

(5) Income from pension insurance: ("i$$peny")

(6) Income from unemployment fund: ("i$$uney")

(7) Income from social assistance: ("i$$wely")

(8) Grants, scholarship: ("i$$gray")

---

[4] In the waves before 2002, we impute other (aggregated) income variables, e.g., total working income (wyn). Due to non comparability across the waves analysed here, they are not listed.

[5] In the SHP notation, "i" is the first letter of income variable names; "$$" denotes the survey year (from 99 on).

(9) Income from other institutions, organizations: ("i$$insy")

(10) Income from family allowances: ("i$$famy")[6]

(11) Income from people in private households (outside the household): ("i$$pnhy")

(12) Yearly income from other sources: ("i$$osy")

Codes to be imputed are[7]

-8 (other error)

-7 (filter error)[8]

-5 (irregular, difficult to say)

-1 (does not know)

-2 (does not want to say)

In the SHP 21,732 individuals in 8,529 households were ever listed between 1999 and 2007 (including children), of whom 18,320 ever responded, either by completing an own individual questionnaire, or by means of a proxy. In the following Table 1, we list the number[9] of missing values by nonresponse category (item-, or unit-), the number of individuals that need initialization, the number of possible donors (i.e., those with validly reported nonzero income), and a variation (standard-deviation/mean) characteristic for the income variables. Missing values and donors are summed over the respective waves. Note that wave specific item and unit nonresponse is exclusive. We also calculate pearson correlations of income with age-group and education (compare Table 1 in Starick and Watson (2007)), to find income discriminating variables, available from the household grid. These characteristics help to find the most suitable method for imputation. Note that due to comparability, of the 9 waves from 1999-2007 only descriptive statistics from 2002 until 2007 are listed. Unit-nonrespondents are assumed to earn income according to their occupational status and number of children, both available from the household grid. E.g. if the unit-nonrespondent is actively occupied, s/he is supposed to earn positive wages, which are imputed by distinguishing part or full time employment. Similarly, while for (old age) retired people the components of old age social security pensions (oasiy and peny) are imputed, for "other" retired people invalidity pensions (aiy) are imputed. Unemployed individuals are

[6] Family allowances are asked separately only from 2004 on (Kuhn 2008).

[7] -4 denotes „no income from the respective income source.

[8] does not occur.

[9] Note that we do not list the percentages because absolute numbers of donors and recipients provide better insight about the feasibility of the imputation procedures.

attributed unemployment assistance. Child allowances are allocated (to households) according to the number of children.

The number of recipients is comparatively high for income from social assistance (wely). “Unit-nonrespondents” in this category include those who refuse to give information on any source of income and for whom it is not clear from the available information which income source they touch. To impute income for these individuals, rather than to assign a fixed minimum income, we use donors who earned “income from social assistance”.[10]

	Missings		of which:	Potential Donors			
Income source:	N Item NR	N Unit NR	N to be initialized	Number donors	Std-dev. / Mean	Corr with age[11]	Corr with Educat[12]
employment	1689	7,744	2,806	23,561	.84	.24	.31
indep. work	862	0	325	3,769	2.02	.11	.12
old age pension	516	1,581	627	6,209	.46	.10	.02
invalidity pension	145	234	164	1,080	.66	-.07	.10
pension insurance	485	1,537	763	3,859	1.34	-.06	.14
unemployment fund	73	212	186	880	.94	.29	.14
social assistance	52	678	428	417	1.11	-.02	-.03
grants, scholarship	74	0	46	415	1.12	.17	.12
other institutions	113	0	91	1211	2.60	.08	.04
family allowances[13]	329	1,754	845	3757	.77	.07	.01
private transfer (ext.)	304	0	171	3420	2.61	.22	.16
other sources	622	0	448	3,498	4.93	.01	.05

Table 1: Nonresponse characteristics of the income variables in the SHP 2002-2007. Descriptive statistics (shaded right) are averaged over waves.

If an income variable is imputed for both nonresponse components, we find that more missing cases stem from unit nonresponding individuals than from item nonresponding cases.

[10] We are aware that this procedure is based on very strong assumptions. In addition, this artificially blows up the unit nonrespondents in this income category.
[11] Age classes are in 10 year groups.
[12] Education measured in three (about equally sized) levels.
[13] Variable available from 2004 on.

Imputation Methods

In this section, the different *longitudinal* methods and the stochastic *cross-sectional* regression used to impute missing income values are described. The longitudinal imputation methods L&S, its extended variant, and the simple carryover method are described first. If the income component is never validly reported, longitudinal methods fail to provide a positive imputation value (Frick and Grabka 2004). In these cases, cross-sectional imputation methods must be used first. We describe the cross-sectional regression method used to "initialize" the income component. As is usual, we assume that the income missing mechanism responsible is MAR (missing at random). This means, the missing data are at random once controlled for observed variables. All imputation procedures are programmed in STATA©.

Little & Su Method

The L&S imputation technique, also known as the "row and column" imputation procedure (Frick and Grabka 2004), considers longitudinal as well as cross-sectional information in the imputation process. The imputed value is the result of a combination of a row effect, a column effect and a residual effect. The column (year) effects are given by $c_j = \frac{\overline{Y_j}}{\overline{\overline{Y}}}$, where j = 1, ..., m [number of years], $\overline{Y_j}$ is the sample mean income for year j, and $\overline{Y}$ is the mean of $\overline{Y_j}$ over all j. The column effect c_j can be interpreted as the inflation factor in year j. The row (person) effects, $r_i = \frac{\sum \frac{Y_{ij}}{c_j}}{m_i}$, are computed for each sample member i. Y_{ij} is the income for individual i in year j and m_i is the number of recorded waves. r_i corresponds to i's mean expected income Sorting cases by r_i and matching the incomplete case i with information from the nearest complete case, say l (the donor), yields the imputed value $\tilde{Y}_{ij} = [r_i] * [c_j] * \left[\frac{Y_{lj}}{r_l * c_j}\right]$. The three terms in brackets represent the row, column, and residual effects. The first two terms estimate the predicted mean, and the last term is the stochastic component of the imputation from the matched case. Again, it must be noted that this approach fails to provide a positive imputation value if only cross-section information is available for a given individual.

Extended Little & Su Method

The extended L&S technique with imputation class (Starick and Watson 2007) distinguishes donors and recipients by taking into account common characteristics. Since donors and recipients should have similar characteristics that are associated with the variables being imputed, we calculate the correlation between age-group and income component, and education and income component, see the last two columns in Table 1. Unlike Starick and Watson (2007), we use education for the extended L&S technique. This is because not only does age have a high correlation with some income components like unemployment benefits, but even more so does education, e.g., with income from employment. We can thus expect more similarities between donors and recipients by looking for donors within the same education group.

Carryover Procedure

If reported information from another wave is available, the closest reported value is imputed without modification by the carryover method.[14] Note that Starick and Watson (2007) use only the wave before the missing as imputation candidates ("last value carried forward"). In our version, we start with the missing value's next wave, and proceed with the previous wave, if the value from the next wave is not valid. We use values from more distant waves if the components from closer ones are all missing or otherwise not applicable. Unlike Starick and Watson (2007), we do not use the random carryover method that draws one of two possible neighboring values at random.

Imputation of Individuals without Income Information: cross-sectional "initialization"

All imputation methods described above require that the income component is validly reported in at least one wave. If it is missing in all waves, it needs to be "initialized" first. This is done by means of a cross-sectional stochastic regression based imputation technique[15]. We use as many covariates as possible for the initial regressions, and drop covariates subsequently. Generally, we use similar covariates as Grabka and Frick (2003) to impute the different income components. In each regression step, we regress the income component on

[14] We will consider an inflation factor in the next program version, since e.g., old age pensions increase at a comparable rate.
[15] We use the procedure "uvis" in STATA.

all covariates using the first wave to be imputed, and proceed using the next wave, until the value is imputed or still missing values make a reduction of the number of covariates necessary.

Specifically, we proceed as follows, separately for each income component for each eligible individual:

1. Check, if income component needs unit-nonresponse imputation. If yes, also include unit-nonresponding cases to the imputation dataset.[16]
2. Check, if income component is validly reported in at least one wave. If yes, use the appropriate longitudinal imputation method ((extended) L&S, Carryover). Also use the longitudinal imputation once a value is "initialized", that means, imputed in one wave according to steps 3.
3. If income component is never validly reported, initialize. This means:
   - Check possible covariates for the regression imputation. Include also other income components that are (already) available[17].
   - Regress on the whole set of relevant covariates using the reported cases, starting with the first wave to be imputed. If no covariate is missing, use the regression based predicted value as initialization.
   - If any covariate is missing in the first wave, use the second wave to be imputed, etc., until the last wave to be imputed.
   - If any covariate is missing in all waves, drop covariates according to significance, and start again with the first wave. Proceed with increased wave/ dropped covariates until there are any missing values in the last covariate(s).
   - If all significant covariates contain missing values, use a median imputation in the final step.
   - Repeat the whole "initialization" procedure starting with the last wave to be imputed, until the first wave.
   - To decide whether the initialized value from the first ("left to right") or the second ("right to left") regression imputation procedure is used, check which procedure finds a valid value in an earlier regression step. This is the finally initialized value. If both procedures deliver a valid value at the same step, randomly select one of the two values.

[16] The dataset that underlies the imputation has as many records as individuals, and stores wave-specific income variables and covariates in columns.
[17] This requires carefully analyzing the optimal order of income imputation.

Which longitudinal Method for which variable?

Starick and Watson (2007) report from a simulation study, that for cross-sectional estimates, carryover methods often perform the best, but perform poorly on the distributional accuracy of change between waves. The L&S method usually provides a reasonable compromise between the accuracy of level estimates versus estimates of change, particularly for respondents. Where there is a reasonably good correlation between the imputation class variable (age or education) used in the L&S method and the variable being imputed, the L&S variant that uses imputation class performs better than the basic L&S. However, when the imputation class variable is only weakly associated with the variable to be imputed, the basic L&S method performs better, especially when the donor pool is small. They find in addition that the carryover methods are more likely to understate change and overstate correlation between waves. With respect to *cross-survey* robustness of results from different imputation methods, Frick and Grabka (2007) compare imputed values from the SOEP, the HILDA, and the British Household Panel Survey (BHPS). While for the SOEP and the HILDA the L&S imputation method was traditionally used, for the BHPS both cross-sectional methods (traditionally used to impute BHPS missing income values) was tested against the L&S (new) method. They find that using L&S technique also for the BHPS produces remarkably similar (structural) results. Giving priority to the longitudinal L&S method is certainly in line with the harmonization efforts put forward by Frick and Grabka.

Following Starick's and Watson's (2007) recommendations, we use the longitudinal method listed in Table 2 to impute item- or unit- missing income values.

Income component Income source:	Longitudinal Imputation Method	Unit-Nonrresponse Imputation
employment: net (empyn)	Ex. L&S (Education)	yes
independent work: net (indyn)	Ex. L&S (Education)	no
(old age) Pension: annual income (oasiy)	Carryover	yes
invalidity pension (aiy)	Carryover	yes
pension insurance (peny)	Ex. L&S (Education)	yes
unemployment fund (uney)	L&S	yes
social assistance (wely)	L&S	yes

grants, scholarship (gray)	L&S	no
institutions, organizations (insy)	L&S	no
family allowances (famy)	Carryover	yes
people from other private households (pnhy)	Ex. L&S(Education)	no
other sources (osy)	L&S	no

Table 2: Longitudinal Imputation method and unit-nonresponse Imputation used.

Income Distributions with and without imputed Values

In this section, we compare the imputed income variables, distinguishing the three disjoint missingness classes:

- Validly reported income values
- Imputed item-nonresponse values
- Imputed unit-nonresponse values

We generally drop the upper 1% percentile from the data (from each wave for the income from employment variable; from the pooled waves for all other variables). In the kernel density estimates graphs, the curves of the income variable densities by missingness class are drawn. In the tables, we list the medians, the standard deviations and the sample sizes, before we graph them to facilitate interpretation.

Income from employment (empyn)

The sample size of the (imputed) income from employment (empyn) allows for analyses separated by wave.

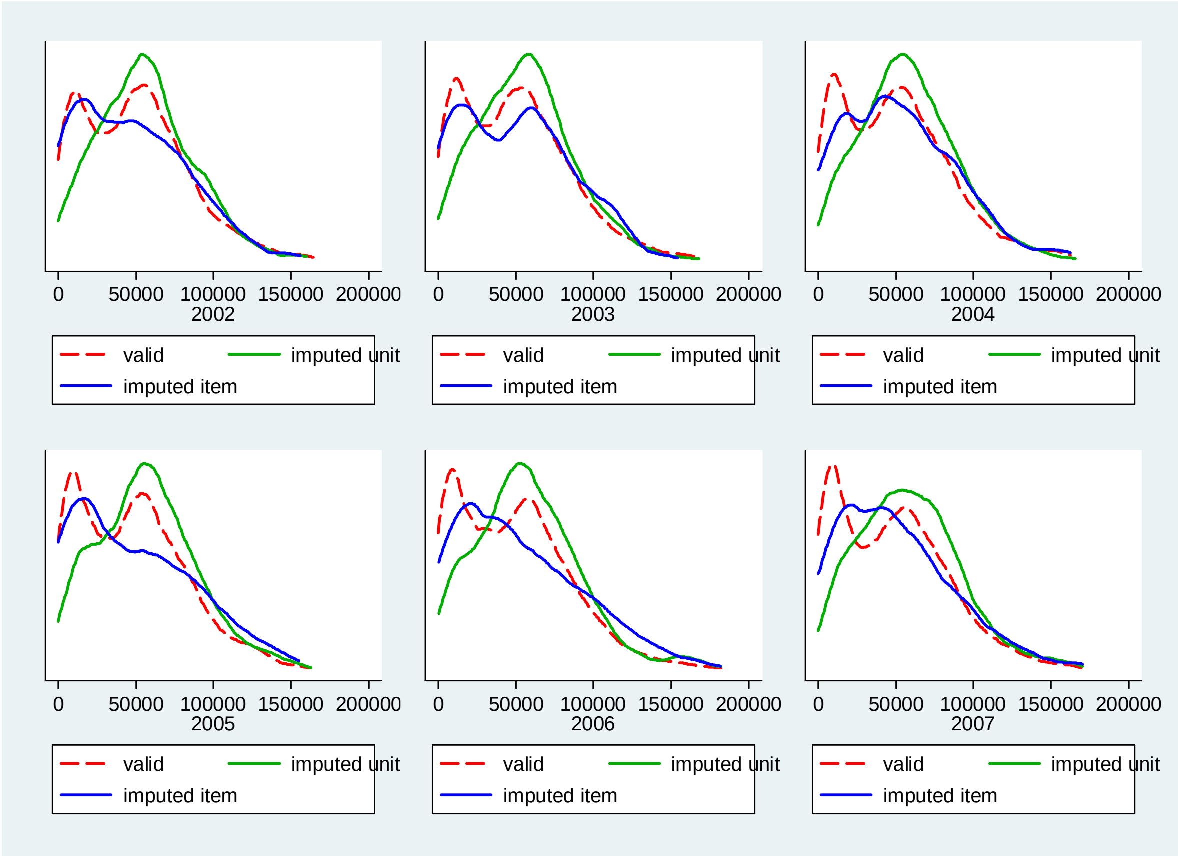


Figure 1: Densities of income from employment in the SHP 2002-2007, by missingness class and year.

While the validly reported income from employment is bimodal in all waves, this is not necessarily the case for the imputed item and never for the imputed unit responding incomes. Imputed income curves from the item-nonrespondents, although close to that from the validly reporting cases, is somewhat "smoothened", and slightly underrepresent the lowest income groups. The imputed unit-nonrespondent´s incomes generally have a smaller variance. This is due to a comparatively strong underrepresentation of the lower income groups.

Table 3 depicts the medians, standard deviations and sample sizes of the income from employment, by missingness class and wave.

	1. valid reported			2. imputed item nonresp.			3. imputed unit nonresp.		
	Median	Std.dev	N	Median	Std.dev	N	Median	Std.dev	N
2002	48,600	33,883	3,297	43,071	34,903	293	54,688	30,585	1,063
2003	48,000	34,541	3,076	48,576	35,470	189	55,889	30,914	876
2004	47,840	34,068	4,710	49,348	35,201	446	55,898	30,244	1,853

2005	47,390	34,829	3,978	43,664	39,358	247	55,832	32,558	1,426
2006	47,390	36,118	3,998	45,478	40,833	245	55,396	34,289	1,246
2007	46,800	35,757	4,274	46,020	37,197	242	57,238	33,144	1,210

Table 3: Median and Standard Deviation: income from employment (in Swiss Franks) SHP 2002-2007.

To facilitate interpretation, we graph the medians and the standard deviations:

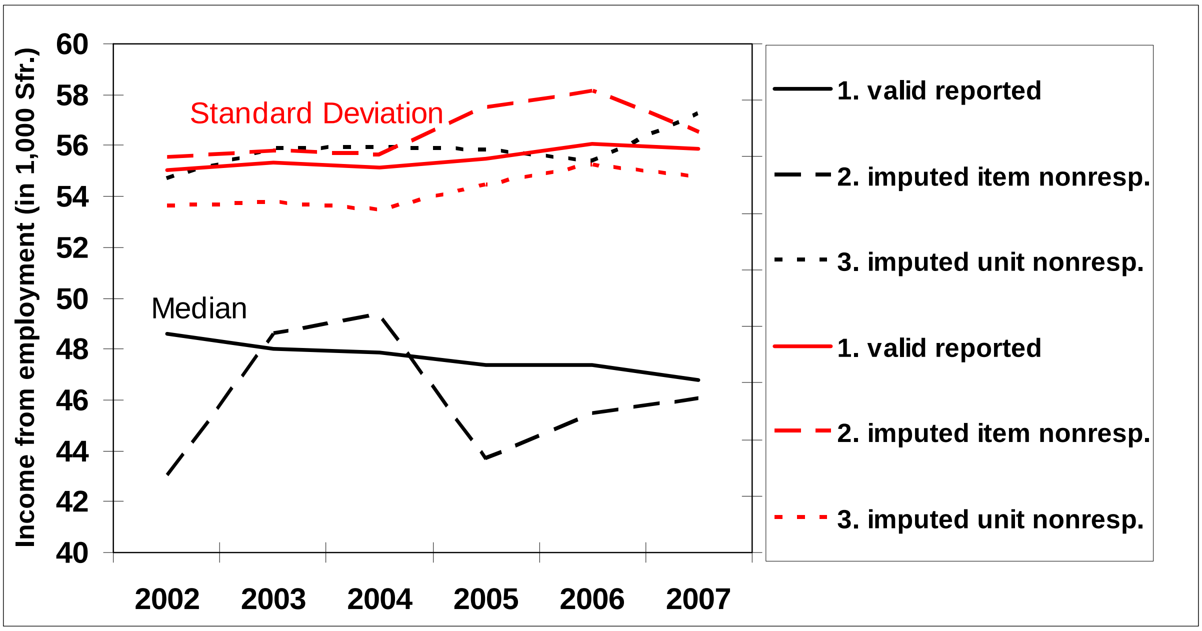


Figure 2: Median (black) and Standard Deviation (red) for income from employment, SHP 2002-2007.

Expectedly from Figure 1 and Table 3, while the level (black: median) of the validly reported incomes and the item-nonresponse imputed incomes are similar, we find a higher level among the imputed unit-nonresponding individuals in all years. The variation (red: standard deviation) across the missingness classes is the same over the years.

Income from other sources with imputed unit-nonresponse

In addition to income from employment, unit-nonresponding individuals are imputed for the following income sources:

- Income from invalidity pension (aiy)
- Income from unemployment fund (uney)
- Income from social assistance (wely)

- Income from old age pension (oasiy)
- Income from pension insurance (peny)
- Income from family allowances (famy)

Similarly to income from employment, we first depict the graph with the density curves of these income components by missingness class in Figure 3. Note that we now pool the data over all waves between 2002 and 2007.

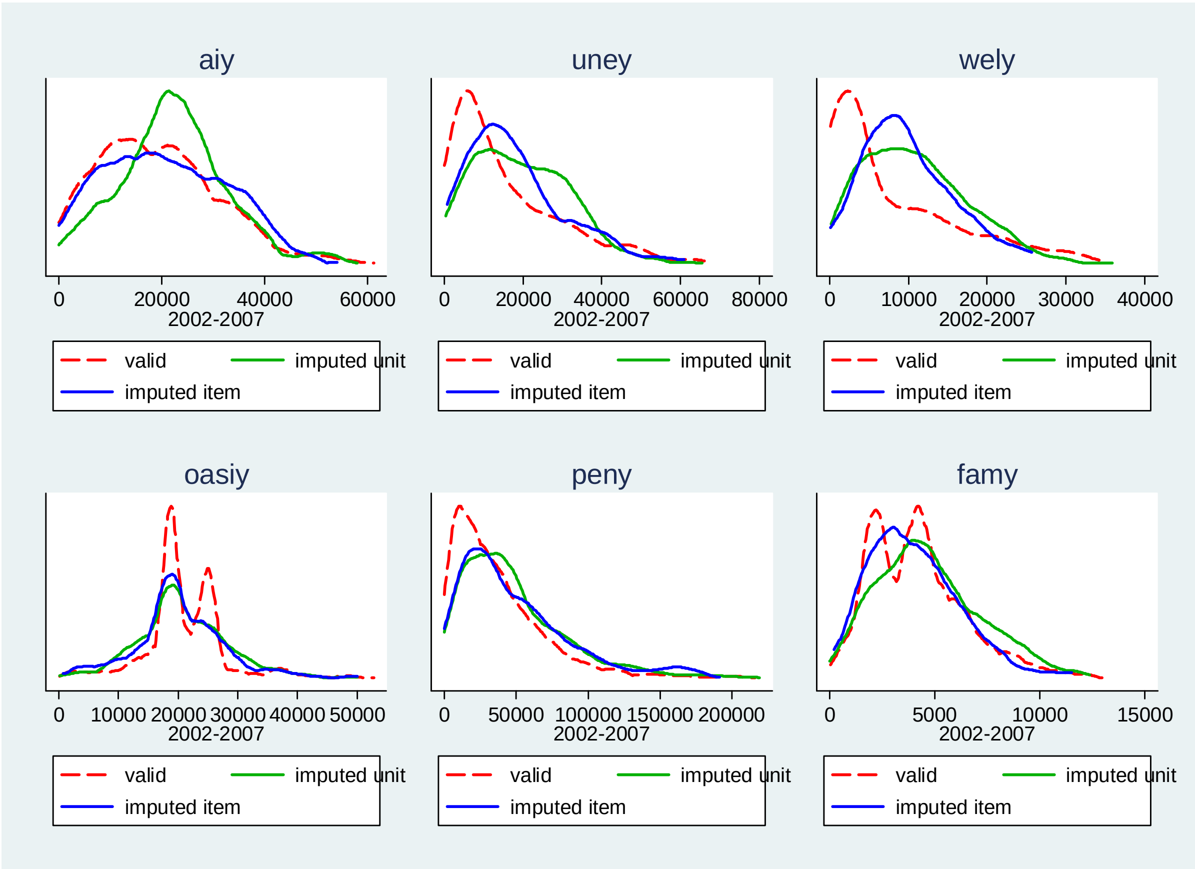


Figure 3: Densities of unit-imputed income components other than from employment in the SHP 2002-2007 (Family allowances: 2004-2007), by missingness class.

(Wave pooled) income curves from different missingness classes have different shapes. While the item and unit imputed income curves are rather similar with the exception of those from invalidity pensions (aiy), the validly reported income curve coincides surprisingly well with the imputed item-nonresponse curve for this component. Again, there is the tendency that low income cases are underrepresented in the imputed income variables.

	1. valid reported			2. imputed item nonresp.			3. imputed unit nonresp.		
	Median	Std.dev	N	Median	Std.dev	N	Median	Std.dev	N
aiy	18,000	11,889	1,069	19,856	11,791	144	22,105	10,412	232
uney	11,060	14,909	870	16,307	13,295	73	18,656	12,900	211
wely	4.200	8,374	412	8,459	5,735	51	10,018	6,532	673
oasiy	19,625	6,784	6,146	19,200	7,862	513	19,683	7,438	1,571
peny	28,800	32,692	3,813	37,285	38,600	478	40,849	36,269	1,533
famy	4,080	2,337	3,724	3,840	2,343	325	4,357	2,513	1,737

Table 4: Median and Standard Deviation: unit-imputed income components other than from employment (in Swiss Franks) SHP 2002-2007.

Figure 4 graphs the measures for the components:[18]

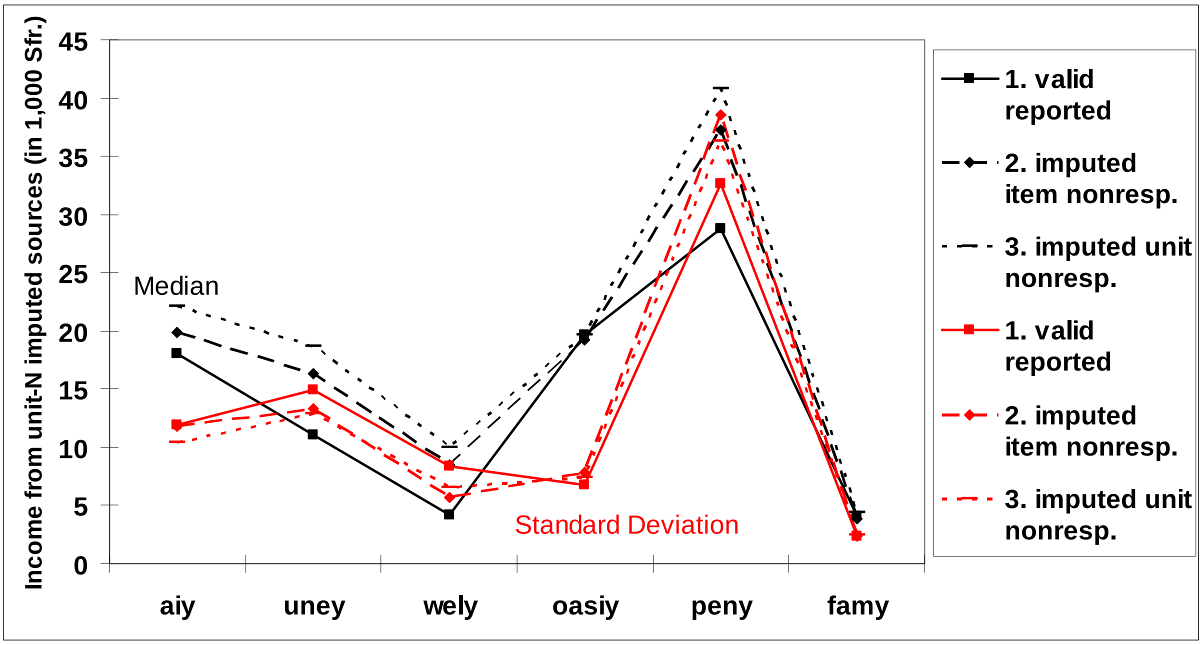


Figure 4: Median (black) and Standard Deviation (red), for unit-nonresponse imputed income components other than from employment, SHP 2002-2007.

Only the two measures for income from old age pensions (oasiy) and from family allowances (famy) are the same for all missingness classes, due to their small variations. For all other median values, the following order is respected: validly reported lowest, imputed item-nonresponse second, and imputed unit-nonresponse highest.

[18] Note that the faint lines connecting the income sources are for better readability.

Income from sources without imputed unit-nonresponse

For the following income variables, unit-nonresponse is not imputed:

- Income from independent work (indyn)
- Income from Grants, scholarship (gray)
- Income from other institutions or organizations (insy)
- Income from people in private households (outside the household) (pnhy)
- Income from other sources: (osy)

In Figure 5, we depict kernel density estimates for these income sources, again dropping the upper 1% percentile.

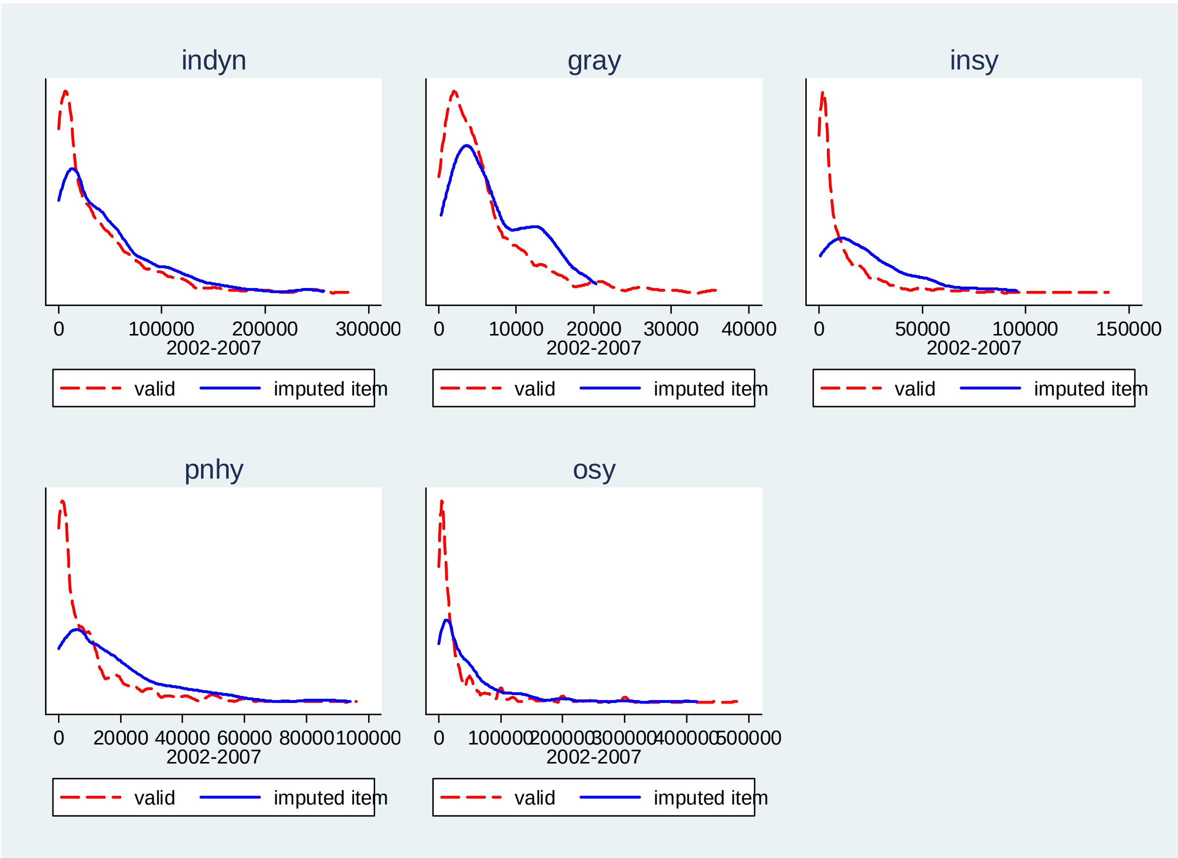


Figure 5: Densities of income from sources that are not unit nonresponse imputed in the SHP 2002-2007 (aggregated), by missingness class.

Again, low income cases are underrepresented in the imputed item income variables, relative to the validly reported cases.

	1. valid reported			2. imputed item nonresp.		
	Median	Std.dev	N	Median	Std.dev	N
indyn	24,000	46,638	3,728	35,127	48,479	859
gray	4,000	6,020	411	5,173	5,114	74
insy	4,720	19,130	1,201	18,185	21,313	113
pnhy	5,000	15,660	3,385	12,310	17,389	304
osy	12,000	62,702	3,469	29,983	59,950	615

Table 5: Median and Standard Deviation: not unit-imputed income components in the SHP 2002-2007 (Family allowances: 2004-2007), by missingness class (in Swiss Franks).

Figure 6 depicts the level and variation differences across the missingness classes for each component:

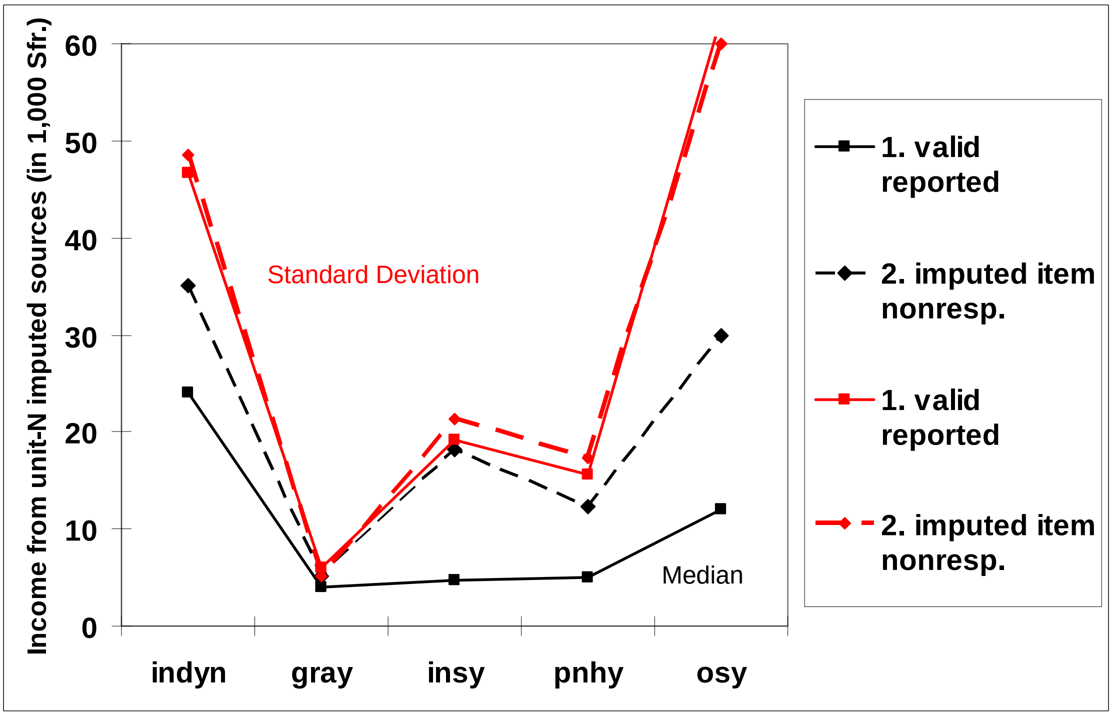


Figure 7: Median (black) and Standard Deviation (red), for not unit-nonresponse imputed income components, SHP 2002-2007.

Also with respect to the income variables that are not unit-nonresponse imputed, with the exception of income from grants (gray), the imputed item-nonresponse cases always have a much higher level than the validly reported cases. The variances are again about the same.

Summary

In the paper we describe the methods used to impute different income components in the Swiss Household Panel, and compare the results of the imputed cases with the validly reported values. We use a two stage approach: for each individual entitled to earn income from a specific income component, we first make sure that positive income is available in at least one wave; i.e., validly reported. If this is not the case, we use a cross-sectional stochastic regression-based imputation procedure to "initialize" the income component for all eligible individuals in one wave. Given the income component is present in at least one wave, we use appropriate longitudinal imputation procedures. Depending on the component, we use the stochastic Little & Su method (standard or extended version, by education), or the simple carryover method.

To check which consequences the imputation has on cross-sectional measures, we analyze level and variation of each income component, distinguished by the following missingness classes: validly reported, imputed value from item-nonresponse, and imputed value from unit-nonresponse. It turns out that while the *levels* of the imputed item-nonresponses are generally higher than the levels of the validly reported values, the imputed unit-nonresponses are considerably higher for almost all income components. The *variations* of the imputed values are mostly the same as those of the validly reported values. These findings prove the need to impute missings of both item- and unit nonresponding income values. The first is necessary to avoid bias from underestimated levels because item-missing cases (that have a higher income) are ignored. The second is necessary to avoid underestimated household income when income from all household members is aggregated.

Especially neglecting missing income from unit-nonresponse is a problem. Frick et al. (2009) show that – among the possible treatments of unit-nonresponse – unlike imputing income from unit-nonresponding cases, ignoring partial unit nonresponse, adjusting equivalence scales, or deleting partial unit-nonresponding households or individuals is not an option. Further research using data from the SHP could go beyond pure level or variation measures but compare income equality and mobility aspects that result by either imputing unit-nonresponding household members or just using reported income to adjust for the missing information.

References

Frick, J., S. Jenkins, D. Lillard, O. Lipps and M. Wooden (2007) The Cross-National Equivalent File (CNEF) and its Member Country Household Panel Studies. *Schmollers Jahrbuch* **4/2007**

Frick, J. and M. Grabka (2004) Missing Income Data in Panel Surveys: Incidence, Imputation and its Impact on the Income distribution. *DIW Discussion Papers* **376**, Berlin

Frick, J. and M. Grabka (2007) Item Non-Response and Imputation of Annual Labor Income in Panel Surveys from a Cross-National Perspective. *IZA Discussion Paper* **3043**

Grabka, M. and J. Frick (2003) Imputation of Item Non-Response on Income Questions in the SOEP. *DIW Research Notes* **29**, Berlin

Frick, J., M. Grabka and O. Groh-Samberg (2009) Imputation of Annual Income in Household Panel Surveys with partially non-Responding Households. Paper presented at the second ESRA conference, 29/6-3/7, Warsaw

Kuhn, U. (2008) Collection, construction and plausibility checks of income data in the Swiss Household Panel. Swiss Household Panel Working Paper 1_2008, Lausanne

Lipps, O. (2009) Attrition of Households and Individuals in Panel Surveys. *SOEPpapers* **164**

Lipps, O. and U. Kuhn (2009) Codebook for CNEF variables in the SHP (1999 - 2007). *Swiss Household Panel Working Paper* **5_09**, Lausanne

Starick, R. and N. Watson (2007) Evaluation of Alternative Imputation Methods: The HILDA Experience. HILDA, June 2006

Watson, N. (2004) Income and Wealth Imputation for Waves 1 and 2. *HILDA Project Technical Paper Series* **3/04**, University of Melbourne

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, individual attrition, household attrition, attrition bias, reference person, household head.

1. Introduction

Attrition is an important quality criterion in large household panel surveys (Zabel 1998, Watson and Wooden 2004, Behr et al. 2005, Gramlich 2007, 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

[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.

unrealistic it is not entirely impossible. Peracchi (2002) for example, reports that between 2 and 6 % of 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 (2007) 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 on the address (GSOEP, SHP, ECHP), or use the former reference person as an anchor if all members of a household move. 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

higher share of female reference persons in the SHP also results from 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. The household socio-demographic characteristics 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, e.g. 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., 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.

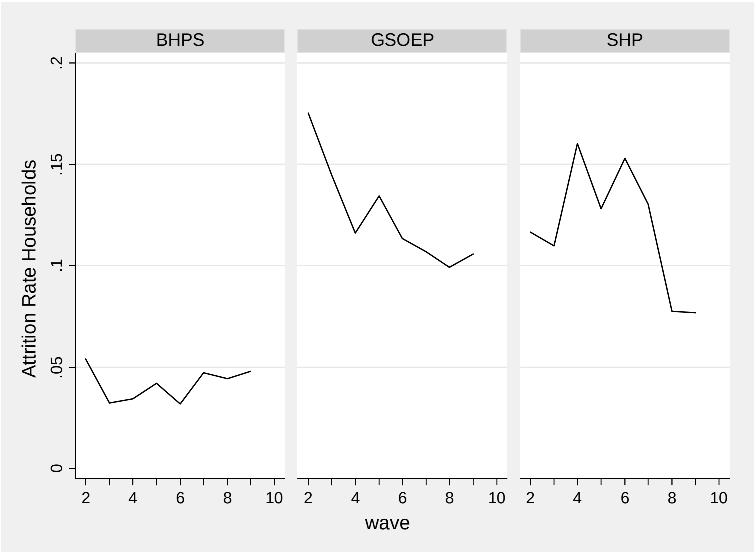


Figure 1: Attrition Rates of first wave Households over transitions wave n->n+1 (n=2,…9), by survey.

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:

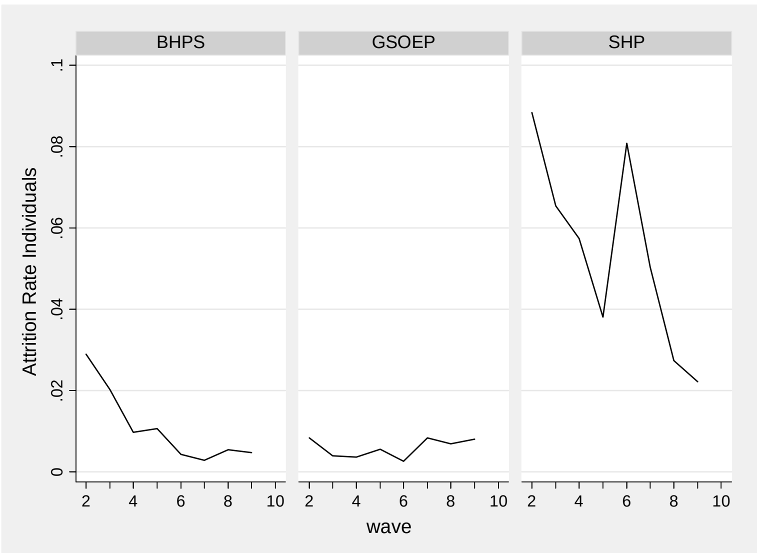


Figure 2: Attrition of first wave Respondents until wave 9, by survey.

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]	Households	Individuals
Reason for dropping-out [%]	(N=1,854)	(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 (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.

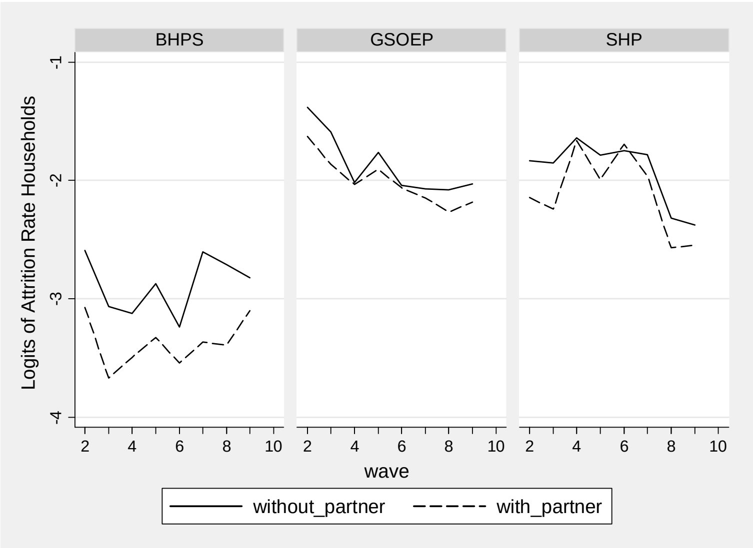


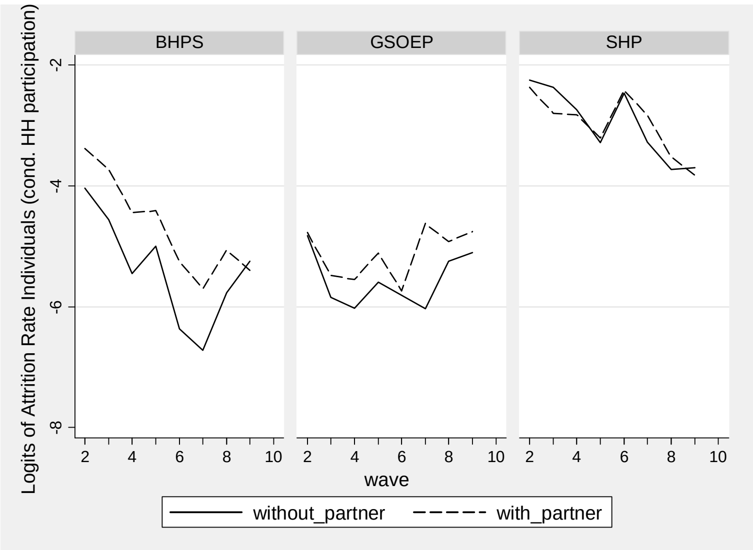
Figure 3: Logits Attrition Rates of first wave Households over transitions wave n->n+1 (n=2,…9), by survey.

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 (movers) and older (refusers) 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. There is a comparatively strong attrition selection 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.[7]



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 (wave effect controlled): Individual Attrition of Individuals conditional on Household Participation. **=significant at 1%-level, *=significant at 5%-level, -=not included.

[7] The variable "working" cannot be used because of co-linearity with the attrition variable in too many cases.

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 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.

[8] Unfortunately she did not distinguish between the reference person and his/her partner.

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 (2007) Analyse der Panelausfälle im Sozio-Oekonomischen Panel SOEP (Analysis of Panel Attrition in the German Socio Economic Panel Study GSOEP). Diplomarbeit (Master thesis), University of Konstanz

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

Infratest (2002) Verbleibstudie bei Panelausfällen im SOEP. Infratest Sozialforschung München, Dec 3, 2008 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

Attrition in the Swiss Household Panel

Oliver Lipps[#]

Swiss Household Panel, Neuchâtel

In this paper, we analyse magnitude and possible selectivity of attrition in first wave respondents in the Swiss Household Panel (SHP), from wave two (2000) through wave seven (2005). After comparing attrition of first wave respondents with that of other panel surveys, we proceed to model selectivity of attrition in two steps: we first build separate wave-to-wave models, and second a longitudinal all-wave model. The latter model includes wave interaction effects. The first models allow for tracing of selectivity development, i.e. whether an initial selectivity might compensate or cumulates over time, the second to assessing the effects of the covariates in a specific wave, controlling for the base attrition effect. In particular it allows for the analysis of consequences due to discrete fieldwork events.

Our results support the findings in the literature: attritors are in general the younger people and the males, foreigners, the socially and politically "excluded", i.e. those who show little social and political interest and participation, those who are mostly dissatisfied with various aspects in their life, and those who live in households with high unit nonresponse, and who exhibit a worse reporting behaviour. This pattern is rather cumulative than compensating over panel waves. Excessive attrition in two waves presumably caused by two discrete events in the panel is not particularly selective. Still existing variation in selective attrition is worth to be further explored.

Keywords: Attrition, Panel Nonresponse, cumulative effects.

Introduction

The major purpose of a household panel survey, in which individuals in households are surveyed repeatedly over waves, is to represent the real dynamics in the sample population. If

[#] This work is supported by the Swiss National Science Foundation. I wish to thank my colleagues at the SHP for their valuable comments. Eric Graf suggested conducting an in-depth correlation analysis, which led to the finally used aggregated health and satisfaction parameters. I would like to thank three anonymous reviewers for helpful comments. Any responsibility for the analysis or interpretations presented in this paper is however with the author.

individuals other than non-sample cases drop out of the panel ("attrition"[1]), this has at first obvious consequences on the longitudinal sample size. However, other than making analyses impossible due to cell sizes becoming too small after some waves, or merely producing higher standard errors in descriptive statistics, a selective attrition may in addition lead to wrong conclusions of important measures under consideration. For instance, in a recent analysis, Stocké and Stark (2005) show with data from the Eurobarometer[2] that due to listwise deletion of individuals because of item nonresponse the share of persons going to a vote is around 9% higher than including the whole sample. Because the mechanisms leading to item nonresponse in a cross-sectional survey or to attrition in a panel survey might originate from similar factors (Loosveldt, Pickery and Billiet, 2002, and the literature review below), we generally suspect a selective attrition in longitudinal surveys. E.g. in the British Household Panel Survey (BHPS), several refusal conversion techniques are being applied, showing that converted refusals have characteristics which are partly distinct from easier to convince respondents (Burton et al. 2004). Thus the characteristics of the members of a panel might well change after a longer time period.

Attrition is usually modelled and predicted with the help of standard socio demographic variables collected in a former wave. E.g., in the German GSOEP the size of the community a respondent lives is significant for the odds of a successful contact (Spiess and Kroh 2005). In the European Household Panel (ECHP) the individual longitudinal panel response can be explained to a good extent by the *socio-demographic variables* age, employment status (i.e., full-time vs. not), and partnership (Nicholetti and Perrachi 2004). However, it is important to note that the socio-demographics are "fallible: they are correlates, not causes of the survey participatory behaviour" (Groves and Couper 1996, p.81). This is emphasised also by Stoop (2005), who specify these *causes* for (non)cooperation: "social isolation, social participation, …, interest in societal well-being, doing voluntary work, political interest and knowledge, …, electoral participation, the type of sponsor, and attitudes towards surveys" (p. 126). Therefore, if available, variables measuring political interest and social participation (e.g. Pickery, Loosveldt and Carton 2001), and item nonresponse (INR) on difficult (Loosveldt, Pickery and Billiet 2002) or sensitive (Schräpler 2004) questions to include motivational factors are usually used in analyses on attrition. Groves et al. give the theoretical concepts to explain unit

[1] With the term attrition we refer to all drop outs of a panel survey, i.e. refusals (non-cooperation) and non contacts of all *interview eligible* individuals (i.e. all who continue to be part of the sample: all who did not decease, are not being institutionalised, or for whom a valid reason for a proxy interview is given; see for the latter in the case of the SHP http://www.swisspanel.ch/project/participation/index.php?lang=en&pid=53).
[2] http://ec.europa.eu/public_opinion/index_en.htm

nonresponse in surveys (2004, pp. 176): “The theoretical perspectives that are most commonly applied to survey participation include

- “*opportunity costs*” (which is) based on the notions that busy persons disproportionately refuse to be interviewed because the costs of spending time away from any other pursuits is more burdensome that for others …
- “*social isolation*”, which influence persons at the high and low ends of the socio-economic spectrum to refuse survey requests …
- “*topic interests*” (and motivation), which fuel hypotheses about how the interested disproportionately responding …
- “*oversurveying*” that suggests fatigue from survey requests.” (p. 176)

In addition to the cross-sectional factors, there is in addition the longitudinal aspect that the reasons to attrite from a panel need not be constant: De Keulenaer (2004) analysing attrition in the Panel Study of Belgian Households, finds “that the effects of SES variables decrease with additional wave …, while the effects of the variables describing the *interview situation* increase …” (p. 5).

In the literature on attrition evidence is reported for all issues but to opportunity costs. As to *opportunity cost* related variables, one might e.g. think of interview time. In the US-Panel Survey on Income Dynamics (PSID), although proved to be negatively correlated with panel attrition, interview length has found not to be causally related to attrition (Zabel 1998). In the Household, Income and Labour Dynamics in Australia (HILDA) panel survey, a non significant relation exists between attrition and interview time (Watson and Wooden 2004). The latter state in this regard that “Interview time, … is a product of instrument length, respondent interest in the survey, and respondent difficulty with the questions. Consequently, we expect the interview length to comprise a mix of respondents, some of whom found the interview very difficult and others who enjoyed the experience.” (p. 302). Moreover, different interview length in the same survey is largely a matter of filter complexity: those who find themselves in more filters, usually have a specific socio-demographic status. This applies especially to individuals who are in the labour force, and have therefore to answer a number of work related questions. It can thus be expected that interview length correlates with the socio-demographic status.

As to *social isolation* factors, Watson and Wooden (2004), using the HILDA panel survey, find that attritors after one wave are more likely to have reported lower life satisfaction levels, are more likely renter of a flat rather than owner of a house or a flat. Also housing conditions have been shown to be significant in terms of attrition in other large panel surveys (Watson and Wooden 2004, Fitzgerald et al. 1998, Zabel 1998, Gray et al. 1996). However, in the HILDA survey, when controlling interview situation in the first wave, these indicators largely lose explanatory power.

As bad health prevents often from active social participation, we subsume *health* conditions under this category: Gray et al. (1996) analysing attrition in a survey of health and life style in Great Britain, state that “those characteristics which were found to be related to attrition … are smokers, the less sporting and those who did not feel a part of the community” (p. 171). However, there was “little or no relationship between the health and psychological variables and a person’s conscious decision to participate …” (p. 182).

With respect to *topic interests* and *motivational* factors, in cross-sectional analyses little political interest and social participation and many “don’t know” answers are shown to be strongly correlated with little political knowledge and weakly pronounced attitudes of the individuals as well as a general negative attitude toward surveys (Stocké and Stark 2005). Such (non)response behaviour can be considered as resulting from a lack of cognitive efforts and disinterest toward the survey (Pickery, Loosveldt and Carton 2001, Loosveldt, Pickery and Billiet 2002, Schräpler 2004, Stocké and Stark 2005). The within household response rate is a good indicator for household specific motivation as according to Watson and Wooden (2004), in the HILDA and in the BHPS surveys, “in line with the results … for the BHPS, we see that coming from a partially responding household is a major risk factor for non-participation at the next wave” (p. 302).

Loosveldt and Carton (1997), analysing the respondent’s decision to participate in the second wave of a Flemish election panel survey, find that the ability to provide an answer during the first interview plays a crucial role. Respondents who have problems to answer the questions and are difficult to interview are more likely to refuse to be interviewed in the second wave. The respondent’s knowledge about the surveyed aspects and reasons for a survey is therefore crucial for panel participation. This correlates with the experience of a “pleasant” interview during the first panel wave, which seems to be an important issue for further panel stay (Loosveldt, Pickery and Billiet 2002).

As to the answers given to subjective categorical questions, Pickery and Loosveldt (2004) view the non-occurrence of at least some extreme category answers as proving a low interview quality, because “One can expect that even respondents without a pronounced opinion will use the extreme response options now and then, especially when different scales are considered. If they do not, they probably do not expend the effort required. … this is a form of satisficing ....” (p. 79). Similar results can be concluded with respect to the use of the midpoint category (see e.g. Tourangeau and Rasinski 1988). Scherpenzeel (2002) considers an excessive use of middle category answers as a proof of a low motivation to conduct the survey.

Regarding *panel specific aspects*, in the US Panel Study of Income Dynamics (PSID), the cross sectional representativeness seems to remain “roughly intact” (Fitzgerald et al. 1998, p. 251). However there is evidence that attrition is correlated with higher levels of “turnover and stability in earnings, marital status, and geographical mobility” (op. cit., p. 296). Also Watson and Wooden (2004) find more attritors among those in the HILDA panel who change their address more often.

With few exceptions, in the literature, only two panel waves are used in order to analyse panel attrition.[3] As we are especially interested in the stability of the attrition variables across waves, we model the successive transitions between waves, but do also built a longitudinal all-wave model.

Specifically the article is organised as follows: First, we present the Swiss Household Panel (SHP), the data used for the analysis. To get a better impression on magnitude and selectivity of nonresponse and attrition in the SHP we describe the first wave nonresponse process, before we compare the distributions of the first wave respondents and its stayer subset after five waves with those in the BHPS and the ECHP. Next the modelling variables used are described in more detail. In a first modelling approach, these variables are used in a year to year attrition analysis. Here we will in particular investigate which variables are significant and whether the significant variables are cumulative rather than compensative as regards to attrition. Next a longitudinal model is built and analysed. Most importantly, wave interaction terms will show whether certain covariates deviate from the base attrition model in

[3] One exception is Rendtel (2002), who uses econometric models to analyse whether attrition is missing at random.

specific panel waves. The results are discussed before the last section concludes. Note that we do not consider interviewer effects on nonresponse in this work.[4]

Data: the Swiss Household Panel

In this work we use data from the Swiss Household Panel (SHP), an ongoing, nationwide, yearly conducted, centralised CATI panel survey on the Swiss residential population. Questions are about household composition and socio-demographics, health, well being and attitudes, politics, social networks, and economics. The SHP started in 1999 with slightly more than 5000 households. In the SHP, in each year first the household composition together with the relationships between all household members, and the basic socio-demography is asked of the household reference person in the grid questionnaire. The household reference person is an adult who is sufficiently knowledgeable of the household characteristics, including the household finances. The grid questionnaire completion takes three to ten minutes, depending on household size and complexity of relationships. Then, a household related questionnaire is to be completed (about 10 minutes), again by the reference person. After this information is given, each household member from the age of 14 on has to complete his/her own individual questionnaire (about 35 minutes).

First wave nonresponse

In Switzerland, survey analysts and researchers face comparatively high nonresponse rates[5], and subsequent attrition in the case of a panel survey. In the first wave (1999), questionnaire response in the SHP is as follows, in the different steps (MIS-Trend 2000):

1. out of the 14,174 (gross) addresses drawn from the national telephone register and called by the interviewers, 1,025 were no valid telephone numbers (fax etc.). This leaves 13,149 net addresses.

[4] An attrition analysis using the SHP together with interviewer characteristics can be found in Lipps (2006b).
[5] See for a cross-country comparison of the response rates of e.g. the European Social Survey (ESS) Stoop 2005. Switzerland ranks at the very bottom within all countries involved in the ESS. A probable reason for the high nonresponse rate might be over-surveying in Switzerland, see Budowski and Scherpenzeel (2005) for the special case of the SHP. Note that similar to Nicoletti and Peracchi (2005), if we talk about nonresponse in the first wave, we distinguish between noncontact and refusal.

2. out of the 13,149 net addresses, 1,065 could not be reached (i.e. 8.1% non contact rate).
3. out of the 12.084 contacted households, 2,712 (22.5%) are non-sample cases (i.e. business lines, language problems, etc.), 2,309 of the remaining 9,366 households (24.7%) refused to complete the grid questionnaire.
4. out of the 7,057 households who completed the grid questionnaire, 1,062 (15.0%) refused to complete the household main questionnaire.
5. out of the remaining 5,995 households, in 921 households (15.4%) all individuals refused to complete their individual questionnaire.

This leaves us with a household net response rate of 48.6% (5,074 "completed households of 10,437 sample households), i.e. the grid and household questionnaires are completed and at least one household member filed his/her individual questionnaire.

On the individual level, according to the screening results from the household grid questionnaire, there are 10,293 individuals living in the 5,074 participating households. Of these, 921 (9.0%) are non-sample cases (language problems, illness, etc.). Of the remaining 9,372 persons, 1,573 (16.8%) refused, leaving a sample of 7,799 respondents.

Because, apart from the geographical region, there is no information on the gross sample, sample selectivity can in principle only be calculated using information of the households who completed the grid questionnaire. Based on the screened households, it can be shown that foreign households are underrepresented. Within households, males, younger individuals, and again persons with foreign nationality answer to the survey in a worse way (Cornali and Vonlanthen 2001).

In all it can be assumed that due to nonresponse, in the first SHP wave especially foreign individuals are underrepresented to a quite strong degree.

Attrition in the SHP compared with other panel surveys

Despite various measures to motivate panel participants (Budowski and Scherpenzeel 2005), the SHP faces a relatively high attrition of around 17% per year. This figure is higher than in other well established large (mostly CAPI) panels like the German SOEP (e.g. Kroh and Spiess 2005), the British BHPS (e.g. Burton et al. 2004) or the US PSID[6]. However we

[6] See the various articles in the Journal of Human Resources 33 (2), Special Issue: Attrition in Longitudinal Surveys, 1998

talk about attrition in a comparatively restrictive longitudinal manner: we only include persons who already answered the individual questionnaire in the first wave, and thus – other than sometimes done in other panels – do not include any new entrants into the panel[7]. Nevertheless, e.g. the first sample in the SOEP faces a longitudinal attrition of 10% during the transition from wave 1 (in 1984) to wave 2, 7% from wave 2 to wave 3, 5% from wave three to wave 4, and declines to around 2% in the long term (Kroh and Spiess 2005, p. 21). Similarly, the PSID has an attrition of 12% from wave 1 (in 1968) to wave 2, and about 3% in the years to follow (Fitzgerald et al. 1998, p. 254). Also in the BHPS, after an attrition of longitudinal first wave (in 1991) respondents of around 12%, attrition reduced to less than 5% from wave three on (Burton et al. 2004, p. 4).

We first compare some important distributions of the SHP total original and five waves stayers sample, with those from the BHPS (Lynn et al. 2006). The BHPS started in 1991 and was an example for the SHP with respect to design and content. However the BHPS interviews are conducted face-to-face. In order to have comparable samples, we compare the total 1999 sample persons with those who validly report during all five waves between 1999 and 2003 in the SHP. Also we drop all individually who become (known) nonsample cases within this time period. We have the following sample sizes:

Sample	SHP	BHPS
# Respondents eligible in all waves until wave 5	7654	10264
# Participants first wave, with all waves until wave 5	3891	7246
First 5 wave response rate [%]	58.8	70.6

The attrition in the SHP is almost 12 % points higher than in the BHPS.

Sex	SHP all	SHP 5w	BHPS all	BHPS 5w
Proportion of men [%]	43.6	42.3	47.7	46.2

With about 1.4 % point differences between the men's share in the total and the stayer sample, the differences are about the same in both surveys.

Age [%]	SHP all	SHP 5w	BHPS all	BHPS 5w
14 (BHPS: 16) - 24	15.2	10.4	15.9	14.6
25-34	19.0	17.8	19.1	20.0
35-44	23.2	25.7	17.5	19.1

[7] These concern new household members who are therefore not original or longitudinal panel members (Naud and Latouche 2001).

45-54	18.9	20.9	14.5	15.0
55-64	12.2	13.7	12.9	13.3
65-74	8.1	8.5	12.1	11.9
75+	3.4	3.0	8.0	6.1

The differences with respect to (starting) age groups are considerable between the SHP 1999 stayer and total sample on one hand, and the BHPS 5 wave respondents and total sample on the other. In the SHP the selection is mostly due to the youngest age class. This may be an indicator of problems with tracing of movers by the survey agency.

Marital Status [%]	SHP all	SHP 5w	BHPS all	BHPS 5w
Divorced or separated	6.7	6.7	5.7	5.8
Living as couple	7.4	6.6	6.3	6.4
Married	57.8	63.4	58.3	61.0
Never married	23.7	19.1	20.7	18.6
Widowed	4.5	4.3	9.1	8.2

The differences in the marital status between the 1999 five waves stayer sample and the 1999 total sample in the SHP on one hand, and the 1991 total sample and those, who responded all waves until wave five in the BHPS on the other, while higher in the SHP, are not tremendous.

Education [%]	SHP all	SHP 5w	BHPS all	BHPS 5w
High (Degree)	31.2	34.6	23.3	25.8
Middle (Level, Other)	47.5	47.3	41.4	42.2
Low/ No Qualification	21.3	18.1	34.7	31.4

The SHP and BHPS differences in percentage points between the total and stayer samples with respect to education are about the same.

Household Size [%]	SHP all	SHP 5w	BHPS all	BHPS 5w
1 PHH	17.4	16.8	13.7	13.1
2 PHH	30.0	29.5	33.4	33.7
3 PHH	15.5	14.6	20.4	20.4
4+ PHH	37.2	39.1	32.5	32.8

Compared to the BHPS, larger households are slightly overrepresented in the SHP after five waves.

Household Income: Quintiles	SHP all	SHP 5w	BHPS all	BHPS 5w
Lowest	20.0	16.7	20.0	17.5
2	20.0	19.8	20.0	19.2
3	20.0	20.7	20.0	20.8
4	20.0	21.1	20.0	20.9

Highest	20.0	21.4	20.0	21.6

As to income quintiles, the SHP distribution is not very different from that of the BHPS.

General Health	SHP all	SHP 5w	BHPS all	BHPS 5w
Very well	33.5	34.0	28.2	29.2
Well	50.4	51.7	45.0	45.7
Average	13.8	12.7	18.6	17.7
Bad	2.0	1.6	6.2	5.6
Very bad	0.2	0	2.0	1.7

Both distribution differences are similar.

Qualitatively similar results to those of the BHPS are reported from other panel surveys (e.g. the US PSID Fitzgerald et al. 1998). Generally, it appears that – given the much higher attrition in the SHP – higher attrition in the SHP does not automatically mean higher (cross sectional) selectivity. One exception is the higher drop out of younger persons.

In the following, we compare the most important socio-demographic attrition ratios (proportion in total sample / proportion in stayer sample after five waves) with those of the ECHP. For the latter, the attrition ratios from the middle 80 percentile countries are shown, omitting the lower and upper 10%. The ratios of the first and the last waves are depicted, where last wave means between third and fifth wave, depending on ECHP country.

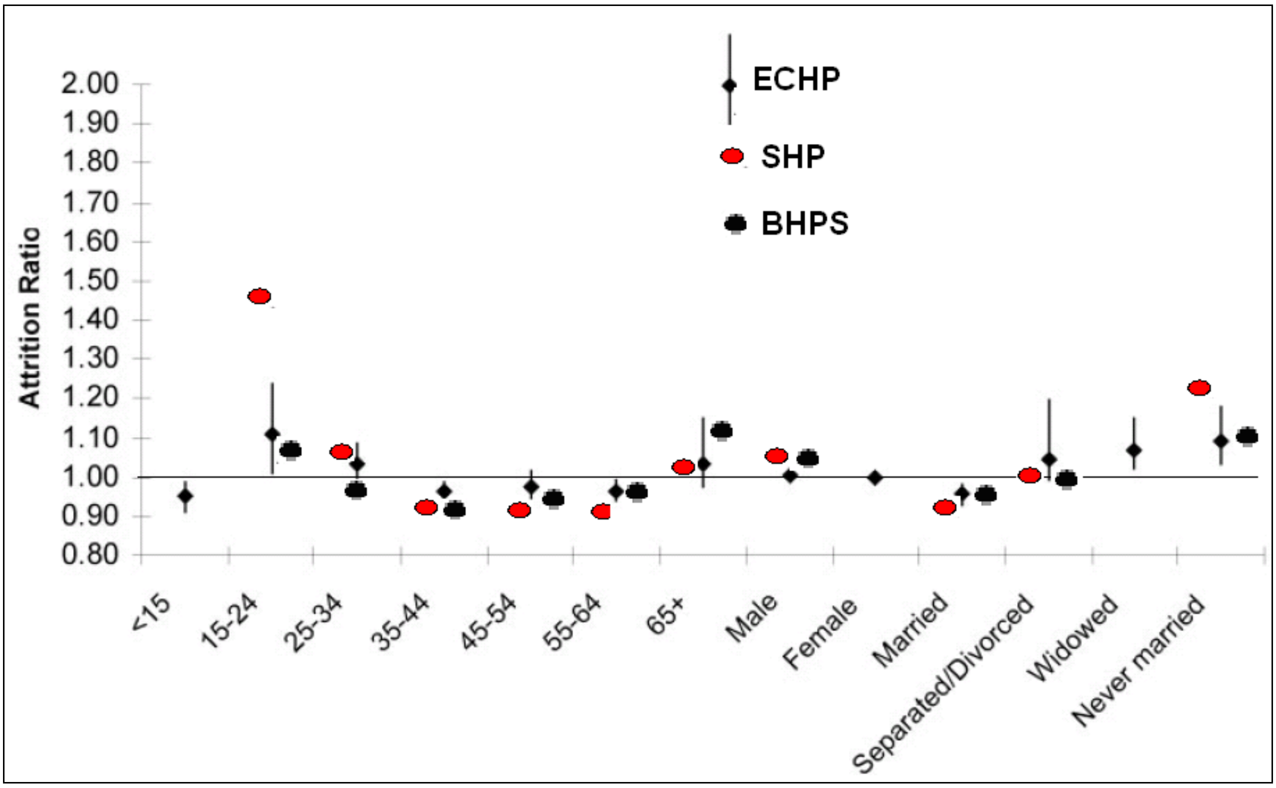


Figure 1: Attrition ratios (proportion in total sample / proportion in stayer sample after 5 waves) in the European Community Household Panel, Swiss Household Panel, and British Household Panel Survey. Figure taken from: Watson (2003), Figure 1.

Again, the problem to keep younger people in the sample becomes apparent.

Panel Survival in the SHP

The longitudinal sample in the SHP, which is still interview eligible in the next wave, started in 1999 with 7769 individuals. In 2000, this sample size drops to 6333, to 5719 in 2001, to 4874 in 2002, to 4332 in 2003, 3592 individuals in 2004, and finally ends up with 2999 in 2005. We depict the SHP wave specific number of participating individuals together with the attrition rates in **Figure 2**.

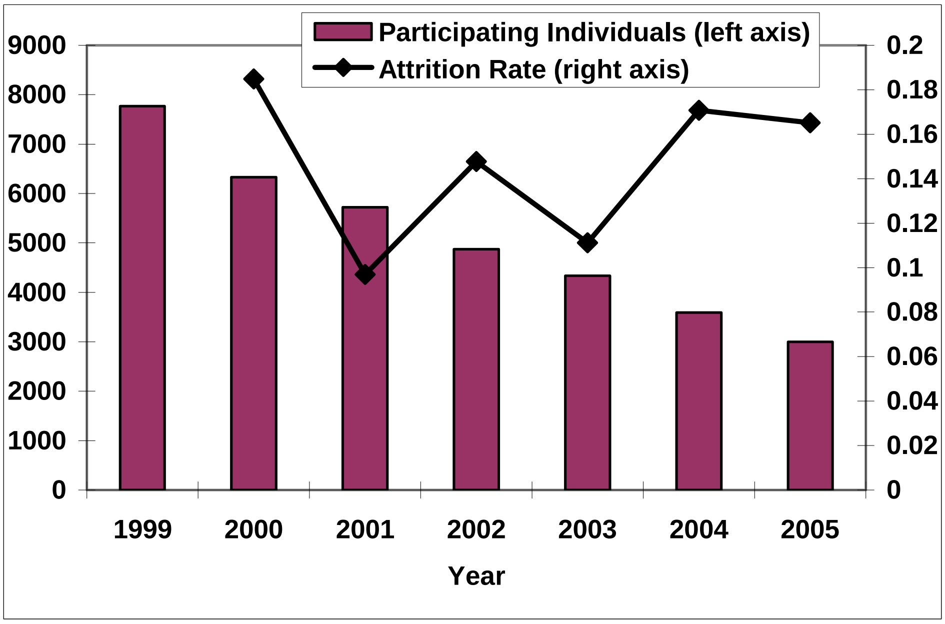


Figure 2: Number of participating eligible first Wave Individuals and Attrition Rates in the Swiss Household Panel, wave 2 (2000) through wave 7 (2005).

The main problem of the SHP is that the attrition rate does not decline: Although the attrition in the second wave is comparatively high, it is not outrageous, especially taking into account an "oversurveying" effect prevalent in Switzerland (Zimmermann and Joye 2003) with harmful effects on response rates and presumably attrition.

We like to assess the consequences of the two major panel specific "events" in the history of the SHP:

- The biographical questionnaire, a P&P self completion sent to the respondents in May 2002 (Budowski and Scherpenzeel 2005). A small part of the sample was used for the 2001 pretest, but this concerned only around one sixth of the whole sample.
- To notify the respondents of a duration of the panel survey of five years, before the first wave started. This was the time horizon of the initial project funding, which the respondents were deliberately told (Budowski and Scherpenzeel 2005).

Although the 2001 pretest of the biographical questionnaire did not show any adverse effects toward the SHP CATI response rate in the next wave (Scherpenzeel et al. 2002), we are now in a position to analyse effects biographical main questionnaire. In terms of the "five wave announcement", we expect a negative influence on those who were difficult to convince to take part in the first five waves, and who might have an argument to drop out in the sixth wave. Remarks noted by the interviewers during the household grid questionnaire during the sixth wave like “In the beginning it has been told that the survey lasts five years and I already took part for five waves but have enough now” supports this hypothesis.

Apart from the expected high attrition in the second wave (2000) we find consequently two peaks in **Figure 2**, in 2002 and in 2004. These coincide with these “events” in the SHP. We therefore find the expected higher attrition rates supposedly due to the biographical questionnaire and the “five wave announcement”. The all wave model to be discussed later will give more insight in the covariate influence on attrition in these particular two waves.

In the following attrition analysis we combine all kinds of possible drop-outs, i.e. due to refusals, non contacts, or others (De Keulenaer 2004). The reason for this decision is first that in a CATI survey like the SHP, from the second wave on, non-contact is a minor problem compared to refusal at least regarding its magnitude. This is also due to the easier respondent tracing by the survey agency based on the information from previous waves. Secondly, a true refusal may easily be confused with a noncontact: this may be because the respondent is alerted by the advance letter announcing the survey call some days before and therefore does not answer the phone once the assumed number appears on the display. Moreover, it happens sometimes that a respondent Y disowns another eligible member X in the household, telling that X is not available, on X’s behalf. What is usually coded noncontact is a true refusal. In addition, a metadata comparison of the characteristics of those who refuse and those who did not continue to respond due to other reasons shows that both groups are not very different with the exception of age and household size (Gray et al. 1996, Table 1). We nevertheless include covariates which can be expected to be strongly related to noncontact rather than refusal, such like age, or whether the household expects to move in the next 12 months, if available. With respect to the former, we already realised that young adults drop out of the SHP to a higher extent which points to problems of tracing moving households.

Independent Variables

We model attrition with the help of a bundle of *socio-demographic* variables on one hand, including the household composition (number of adults 18 years or older, number of children under the age of 18), sex, age[8] in seven groups (14-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70+), whether the individual is married, education on a binary scale (median cut), and finally the employment status, i.e. full time or part time employed, unemployed, or out of the labour force.

We analyse the influence of having Swiss citizenship, and the interview language. The latter variable is also dichotomised, as within the three SHP survey languages German, French and Italian, about 73% of the Swiss residents speak German as first language (Swiss Federal Statistical Office 2005). We further distinguish between the seven Swiss regions “Lake of Geneva”, “Middleland”, “North-West Switzerland”, “East Switzerland”, “Central Switzerland”, “Zurich”, and “Ticino”. In addition we investigate the size of the community in which the respondent lives, measured by a binary variable which amounts to 1 if the respondent lives in a large or middle sized community (“Grosszentrum” or “Mittelzentrum”), and 0 else.

Also whether the household moved since the last panel wave is taken into account. Moving is an important predictor of the odds to make contact with a household. However it is a priori not clear whether moving has an influence on response, once contact is established. In the HILDA survey at least, the likelihood of obtaining an interview is independent of moving. In order to assess whether the survey institute successfully tracks households, the intention to move is taken into account in the attrition model as well.

In order to measure panel motivation using information of the degree of *social isolation*, we include house ownership as this variable was found to discriminate grid response (Naud and Latouche 2001). In order to approach disposable income – a direct use of income is not possible because the missing values are not (yet) imputed in the SHP, a variable which counts various potential bad states of the dwelling (6 at maximum) as well as flat or house ownership are used.

Next we investigate a bundle of satisfaction variables; all satisfaction variables are equally measured on an 11pt scale, where 0 means completely unsatisfied, and 10 means completely satisfied:

- satisfaction with the financial situation of the household
- satisfaction with living together with the other household members or with living alone, respectively

[8] Age is measured in 1999 and kept constant afterwards, i.e. we measure initial age.

- satisfaction with the activities in the free time
- satisfaction with the amount of leisure time

These satisfaction measures are standardised and then combined into an aggregate parameter.[9]

Health reasons are very often stated when people argue why they do not like to continue to take part. In order to avoid collinearity, we combine the following (standardised) health related variables into one aggregate parameter ("health problems") which is the mean of the following variables:

- the degree to which the self rated health is bad
- whether there were at least one day during the last years at which one encountered health problems
- whether there are impediments in everyday activities due to health problems
- satisfaction with health
- whether medications are needed
- the degree to which the person feels anxious in terms of her health

The extent to which health has improved during the last 12 months is not added to this aggregated health parameter, but kept as a single *health change* measure.

As to *topic interest and motivational* factors, we investigate a series of variables which are a measure of societal and political interest and participation. We consider in more detail the variables, which are also originally measured on an 11pt scale

- interest in politics
- in degree to which the respondent has trust in the federal government (compare Stocké and Stark 2005)
- the self assessment of his or her political influence
- and the number of times one would go to ten possible votes.

These variables enter a common parameter "political interest".

In order to first distinguish the rather theoretical parameter political interest from actual or potential political participation and because a correlation analysis shows that the variables between the two groups correlate only rather weakly, we define a second aggregated political parameter "political participation", which is the mean of the following standardised variables:

[9] Worth mentioning is that because of the aggregation, we encounter only few missing values due to item nonresponse: a variable containing a missing value is simply not considered for the individual concerned for the mean value calculation. The mean parameter consists then of fewer variables.

- the extent one is willing to take part at demonstrations
- the degree one would take part at strikes
- the extent one would participate at boycotts

We furthermore include the political left-right political orientation. Also this variables is standardised. We further consider whether the individual is active in voluntary work and does sports at least once a weak.

As *survey related motivation variables*, we include the household response rate of the preceding wave. Furthermore, we introduce the dummy "reference person" from the last year in multi-adult households, as Lipps (2006b) shows that response propensities of former reference persons are much higher than of other persons in the household.

Next, we include variables measuring the interview quality of the preceding wave. The first variable indicates whether at least one of the 23 subjective attitude questions present in all waves has not been validly answered (subjective questions noanswer), the number of extreme and middle category answers to these questions, as well as not providing enough information for the data editor to be able to calculate total individual income.

As to the extreme categories, we suspect that not only an underuse, but also an overuse of extreme categories may be a form of satisficing and therefore include the number of all extremes in the data. The same is done for the midpoint (cat 5) category.

In order to avoid deletion of records due to missing values other than the validly coded INR categories "don't know" or "no answer", the (few) missing values are recoded to the respective modal value.

Single Waves Models

The single wave models aim to getting an idea which person groups do especially attrite in which wave. In each single wave model, we include all longitudinal persons who give a valid interview in the first of the two transition waves under consideration and are still interview eligible in the second wave. As we are especially interested in the identification of even weakly significant attrition variables, we build logistic forward regression models with a comparatively small inclusion level equal to .01. As mentioned above, we consider whether the household intents to move within the next twelve months in these single wave models. This variable is not available in the years 2000 and 2002. To interpret the coefficients and their significance notice that they are not comparable due to the different sample sizes of the models.

The results of the forward regression models are depicted in the left columns of Table 1 and Table 2. To facilitate an overview, we denote in Table 1 only the significant variables, with a plus sign if the variable correlates positively with attrition in the wave considered[10], or a minus sign if the variable correlates negatively with attrition in this wave[11]. The full coefficients for the significant variables at the 1% level are listed in Table 2.

We first compare significant variables in the different variable groups possibly susceptible for attrition. Evidently there are large differences: within *socio-demography* especially moving factors and (young) age are crucial. Nationalities other than Swiss show higher attrition rates, but only in the second wave. Spatial issues thus are not of major importance as long as they do not measure the main language regions. In the *social isolation* variable group, only satisfaction is important, while housing does not play a role at all. *Topic interest and motivation,* and *survey status* play a major role in all waves, as well as *survey status*. That health problems are of minor importance in *each* wave is surprising, because health problems are often a main reason for individual refusal, as far as reasons for dropping out are given in the grid questionnaire.

A second immediate notion is that if an attrition variable is significant in several waves, attrition is always cumulative rather than compensating. This is consistent with findings in the literature (e.g. Gray et al. 1996). In addition the magnitudes are rather stable over time, as can be seen in Table 2. Thirdly we notice that the pseudo R^2 statistics range between .05 are thus not tremendously high. This shows that, although we included many covariates trying to explain attrition variation across different variable groups derived from theory, there are still other reasons responsible.

In the following, we identify and interpret the selective effects in the single transition models.

As to the *socio-demographic* variables, there are no particularly surprising results. The number of adults in a household does not discriminate attrition until the year 2004. In this

[10] "+" if significant at the 1% level, "++"if significant at the 1‰ level.
[11] "-" if significant at the 1% level, "--"if significant at the 1‰ level.

year, persons living in households with two (or more) adults attrite way above average. This may be due to the above stated “five wave announcement”. Households with several adults may be more sensitive to such announcements, perhaps more so because they face a higher response burden due to having to answer several individual questionnaires.

For the variable number of children in the household on the other hand, we find negative effects as regards to attrition in the first and the third transition. An explanation is that households with children have a more stable lifestyle, and are much easier to be found and contacted. After four waves however, the remaining households in the panel and the remaining family households seems to be in “balance” insofar, as there is no difference in the attrition rate from wave five on.

The same stability argument applies for the middle aged groups between 50 and 59 years of age, and for the married persons; for the former there is a negative attrition effect in the second and especially the fourth wave, which further cumulates in the sixth wave. Very severely however, the adolescent individuals and particularly the younger adults attrite to a very high degree constantly from the second wave on. With respect to age distribution, this may lead to an “overaging” of the panel sample in later panel waves. This high attrition is an alarming signal regarding tracing of moving respondents, as moving is more prevalent in this age group. This finding is strengthened by the very high attrition rate of those who intend to move within the next 12 months. As can be seen in Table 2, the odds to attrite of the “willing to move” people do not decrease over waves; i.e., tracing efforts undertaken are not increasingly efficient.

As to the older cohort aged 70 and older, we encounter a higher attrition in the second wave which does not continue later. This is surprising, as due to increasingly bad health, deceases, institutionalisation, or simply a too high response burden, one might expect a constantly higher attrition in this age group. Their higher stability in life seems to play a higher role than a higher “natural” transition probability towards ineligibility.

Also unanticipated is that education or being full time employed or jobless is never significant. This may be a result from a conflict between the higher interest in the topic by higher educated or full time employed people and their tougher time budget. That the unemployed do not show a higher attrition rate could not be expected from the literature. In fact the attrition is as high as for the reference group, which consists of those who are not in the labour force. In this respect, both interest in the survey and response burden seem to be important factors. The contrary may hold for the part time employed: they show a smaller attrition in the third and especially in the sixth wave, as response burden may play a minor role for this group.

Investigating nationality, Swiss citizens attrite to a lesser rate, as can be expected from other panel surveys. Swiss-German speaking people show a high risk to attrite only in wave five. This partially complements the slightly higher attrition rate of the East-Swiss who all belong to the German part of Switzerland in the third wave. Second, it may result from an institution effect: the "five wave announcement" may have been communicated more explicitly in the German speaking CATI centre compared to the French speaking centre.

Not surprisingly, the intention to move decreases the odds of a contact and therefore increases the probability of attrition. Households who moved since the last interview show a significantly higher participation rate: this is probably due to the fact that the propensity to move right again is smaller for these households.

Concerning *social isolation* factors, housing variables, which can be used as weak proxies for wealth (ownership) and income (state of housing), do not play a role at all. This comes as a surprise. Similarly, neither health problems nor health improvements largely affect attrition in the subsequent wave. This fits well to the non-increasing attrition of the elderly after the second wave and shows that health per se does not affect attrition. On the other hand, the aggregated satisfaction parameter is significant through the first waves.

The most interesting effects stem from the variables measuring *topic interest and motivation*: expectedly, being interested in politics and potentially participating in political or societal activities largely decrease the odds to attrite. This is partially also true for involvement in voluntary work. Having a left rather than a conservative political orientation on the other hand has no effects. Overall, the lowest attrition can be expected from those with a high political interest. This pattern holds virtually during all panel waves.

The timing of the influence is also interesting: those who are involved in voluntary work exhibit a decreased attrition only from wave four on, whereas individuals who participate in politics show a particularly low attrition in the second and third waves. Perhaps the latter do act more immediately and to a stronger degree.

Regarding the data quality characteristics of the interviews preceding the wave under investigation, especially those who use many extreme categories answers to the subjective questions have highly positive odds to attrite from wave three on. In fact, this intraindividually rather stable variable[12] seems to be an excellent measure to assess attrition in the next wave, and does not correlate high with the other variables considered. A possible interpretation for the strong effect on attrition is that giving a high number of extreme categories answers is an indicator of little substantive interest in the survey. A similar albeit much weaker result holds for those who give a high number of middle category answers. The latter effect is however prevalent only in the fifth panel wave.

Not giving a valid answer to at least one of the subjective questions has no significant effect at all; however a not valid answer to one of the income questions has strong positive attrition effects in the second wave. This should have positive consequences on the income nonresponse in subsequent waves.

Finally a high within household response rate leads to a constantly highly significant lower attrition in the next wave. This confirms that other household members' disposition in a preceding panel wave has strong impacts on the own participation behaviour (Lipps 2006).

Similarly one might wonder why the reference person of the previous wave, who has to answer the household grid and preferably the household questionnaire, attrites to a smaller extent than other persons in the household. On one hand, these tasks speak in favour of the hypotheses that her commitment to the panel is from the beginning stronger than those of other individuals. However, all characteristics are measured in the year before the possible attrition under investigation occurs: It may equally be true that the previous year's reference person suffered such a high response burden due to not only having completed the grid and household questionnaire, but also her own individual questionnaire that she is more likely to attrite. Nevertheless we find a cumulative negative attrition of the reference person, so the first hypothesis is confirmed.

We note in addition that the 2002 and the 2004 ("panel events") models differ only in some minor respects from the other models: as to the "five year announcement", in 2004 there is by and large a higher attrition especially by those who live together with other adults, a lower for part time employed and again persons engaged in voluntary work. A possible interpretation may be that especially those with a higher household specific survey burden have been especially bothered by continuing the panel despite the "five year announcement". However these findings should be interpreted carefully.

[12] The correlation coefficient between two waves never falls below the value of .5.

All Waves Model

In the second modelling step we are especially interested in testing for duration dependence, i.e. all things equal, we try to identify person groups for whom the base attrition rate differs and, in addition, whether this base rate is different in single waves or even shows a monotonous development over time. Second we like to identify potential effects of the two panel "events" biographic survey" and "five wave announcement" in wave four (2002) and wave six (2004), respectively, in the context of *all* panel waves.

In the SHP the decision is made to keep all persons in the sample who temporary refuse to take part, i.e., refusal only in one wave. Once an individual refuses for two consecutive waves, (s)he is excluded from further contact attempts. Because we like to keep all interview eligible individuals including the temporary attritors in the analysis, we cannot apply a true survival model.[13] Nevertheless, it is important to control for the clustering of waves within individuals when duration dependence is tested[14] (Zabel 1998). Therefore we apply a two-level random effects model, thus treating the single individual residuals as random variables. This controlling constitutes the main difference vis-à-vis the single wave models, where significances depend on the sample sizes, which vary considerably between waves. Note in addition that we cannot estimate a random effect forward all wave model. Thus in this stage all variables are included, not only the significant ones. This may have implication on co-varying variables.

The qualitative results of the all wave model are depicted in the right columns of Table 1, the odds ratios of all coefficients in the right columns of Table 2.

Looking at the modelling results at the right columns of Table 1, it becomes clear that only few of the variables considered are significant in terms of a wave specific deviation from the base attrition rate. This means that only few person characteristics which discriminate attrition change over waves. In addition the significance of these "deviance variables" reaches the 1‰ level only in the case of the satisfaction measure in wave six. This reflects the stability of the

---

[13] See Lipps (2006a) for an attrition analysis of the SHP using only individuals with monotonous participation patterns.
[14] I.e., all individuals contribute to the same extent, irrespective of their panel participation duration.

base attrition rate over the waves, i.e. time is cumulative rather than compensating. E.g. people who are engaged in voluntary work attrite to a lesser extent constantly in every wave. That the high initial attrition does not significantly decrease over time can be seen by the insignificant wave specific effects.

There are some wave specific peculiarities vis-à-vis the results of the single-wave models. E.g., being married or part time employed has transition specific effects in the single wave models. However these effects completely vanish in the all wave model. These variables are captured in other now included variables, see the above remark about co-varying independent variables. Conversely, having health problems, then without effects, do significantly increase the odds of attrition in the all wave model. Overall however, in the all wave model, the base effect is “stabilised”.

As we are especially interested in the effects of the panel events “biographical questionnaire” and “five wave announcement”, we check the 2002 and 2004 columns in more detail. As to the fourth wave attrition effects, we identify a comparatively higher attrition by those who show a higher political participation, or are reference persons. Both effects affect the highly negative base attrition with respect to these characteristics in a positive way. This shows that person groups, who generally show smaller attrition rates, are particularly deterred by the additional burden. Regarding the “five year announcement” effects on attrition in 2004, we find a highly significant increased attrition rate for those who are more satisfied with various aspects in their lives. Again this particular wave effect compensate for the small base attrition rate of more satisfied individuals. These persons may have also attrited to a higher rate due to the higher response burden, because of their tougher disposable time budget.

Overall we find that individuals who generally show a higher panel loyalty attrite to a higher extent due to the “events” considered. This may prove their sensitivity toward additional survey burdens whose possible effects on future attrition rates should be carefully analysed before they are implemented.

Summary and Conclusions

In this article we analyse attrition in the Swiss Household Panel (SHP) from wave two (2000) through wave seven (2005). We only include individuals, who already completed the individual questionnaire in the first wave, and are still interview eligible in the wave under consideration.

First, comparing attrition in the SHP with that in the British Household Panel and the European Community Household Panel, we show that although attrition is comparatively high in the SHP, it is not particularly selective with respect to important socio-demographic or – economic variables. However the problem to keep younger individuals in the sample is a challenge for the SHP – the more that this is the person group which moves and forms new households, thereby maintaining the (cross-sectional) sample size in the panel survey.

We use the characteristics of the respondent in the preceding wave as covariates to model independent logistic forward regression transition models in a first step. We find that the younger people, those without children in the households, those who intend to move, foreigners, those living in households with interview refusing participants or showing little interest and (potential) participation in politics and society, those dissatisfied with various aspects in life, and those who are not reference person in their household attrite to a higher degree than their respective counterparts. Only very weak effects are due to spatial aspects and the housing situation of the respondents, as well as their health status or physical activities exerted.

These findings are in line with attrition analysis results from other European Panel surveys (Buck et al. 2003, Watson 2003) and confirms the "social exclusion" theory (Groves and Couper 1998, Groves et al. 2004, Stoop 2005). In addition we find strong effects from interview quality characteristics in that those who exhibit an extreme answering behaviour (overuse of extreme categories answers) or tend to refuse to answer income questions also attrite to a higher rate. An assumed satisficing behaviour (overuse of middle category answers) or proving little cognitive effort invested in the answers (not answering the subjective questions) is not necessarily related to a higher attrition. The effects are in all models cumulative rather than compensative.

In the second step an all wave random effects model with wave interaction terms is estimated. We find that the base model is rather similar to a synopsis of the single wave models, and that the effects of the wave interactions are minor. In particular, although attrition in the waves with the two specific panel "events"[15] is higher, there are neither significantly increased attrition effects from the two wave dummies nor a particularly increased attrition by special person groups. The small deviations from the base attrition in the two "event" years

[15] First the biographic survey in 2002 and second the assumed end of the panel due to communicating to the respondents that the panel is initially funded for five years at the start of the survey in 1999.

are comparatively small compensatory effects: those who generally show a smaller attrition are positively affected and vice versa. In general however, the mostly insignificant wave interaction effects show that the panel participants are affected to more or less a similar degree.

Our models have a comparatively small explanation power. This shows that other than easily measurable factors used here from completed household or individual interviews also play an important role with respect to attrition behaviour. The quality of the contacts of interviewers with respondents prior to an interview can give some hints for future research: E.g., while the respondent socio-demography is significant for the outcome of the first contact with an interviewer, Groves and Couper show that it looses its predictive power for those requiring more than one contact to obtain a final disposition (1996, p.74, 1998, p.255). First analyses to better understand the process leading to attrition are currently under way (Lipps 2007), who uses call data from the SHP.

In order to reduce the high attrition the SHP, a couple of measures were taken for the 2006 wave (MIS-Trend 2007), partially based on the findings in this article: most importantly, an incentive experiment has been implemented among panel households[16] in order to increase motivation. In addition, unlike the rule adopted in the SHP not to ask households who refused to participate for two consecutive waves, households, who did not answer during the 2004 and the 2005 waves, have been approached all the same. Finally, the notion of completed households has been changed[17]. First results show that these measures decreased attrition in the 2006 wave to a considerable extent.

References

Budowski, M. and Scherpenzeel, A. (2005) Encouraging and Maintaining Participation in Household Surveys: The Case of the Swiss Household Panel. *ZUMA Nachrichten* **56**, pp. 10-36.

[16] A random quarter of households received 1.) no incentives (control group), 2.) stamps unconditionally, 3.) participation at a lottery or 4.) a donation to be donated to a charity institution. 3.) and 4.) applied on the individual level after completion of the individual interview (conditional). Also incentives for interviewers have been implemented depending on collective response rate achievements.

[17] Before the 2006 wave, a household has been considered „complete“ if the household grid and at least one individual interview was filled. From wave 2006 on, completeness means that *all* interview eligible household members complete their interview. The idea was to increase cooperation among those who are not reference persons in the household.

Burton, J., Laurie, H. and Lynn, P. (2004) The Long-Term Effectiveness of Refusal Conversion Procedures on Longitudinal Surveys. *ISER Working Papers* 2004-11.

Cornali, A. and Vonlanthen, C. (2001) Description de la pondération de l'échantillon du Panel Suisse de ménages. *Swiss Household Panel Working Paper* **4_01**, Neuchâtel.

De Keulenaer (2004) Socio-Economic Status Bias in Survey Nonresponse: An Analysis of the Panel Study of Belgian Households (PSBH). *Proceedings of the 57th WAPOR Annual Conference*, May 12-13, Phoenix, Arizona.

Fitzgerald, J., Gottschalk, P. and Moffitt, R. (1998) An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics. *J. Human Resources* ***33*** *(2), Special Issue: Attrition in Longitudinal Surveys*, (Spring, 1998), pp. 251-299.

Gray, R., Campanelli, P., Deepchand, K. and Prescott-Clarke, P. (1996) Exploring Survey Nonresponse: the Effect of Attrition on a Follow-up of the 1984-85 Health and Life Style Survey. *The Statistician* **45** (2), pp. 163-183.

Groves, R. and Couper, M. (1996) Contact-Level Influences on Cooperation in Face-to-Face Surveys. *J Official Statistics*, **12** (1): 63-83.

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

Groves, R.M., Fowler, F.J., Couper, M.P., Lepkowski, J.M., Singer, E. and Tourangeau, R. (2004). *Survey Methodology.* Wiley Series in Survey Methodology.

Kroh, M. and Spiess, M. (2005) Documentation of Sample Sizes and Panel Attrition in the German Socio Economic Panel (SOEP) 1984-2004. *Data Documentation* **6**, DIW Berlin.

Lipps, O. (2006) Analysis of Panel Participation in Couples using Interviewer Characteristics and the Partner's Behaviour. *Swiss Household Panel Working Paper* **3_06**, Neuchâtel.

Lipps, O. (2006a) Attrition and Motivation in the Swiss Household Panel. *Swiss Household Panel Working Paper* **2_06**, Neuchâtel.

Lipps, O. (2006b) Attrition of and within Households in the Swiss Household Panel – Roles of the Reference Person, Other Persons, and the Interviewer in the Course of Time. *Swiss Household Panel Working Paper* **6_06**, Neuchâtel.

Lipps, O. (2007) Cooperation in Centralised CATI Household Panel Surveys – A Contact-based Multilevel Analysis to Examine Interviewer, Respondent, and Fieldwork Process Effects. Submitted.

Loosveldt, G. and Carton, A. (1997) Evaluation of Nonresponse in the Belgian Election Panel Study '91-'95. *Proceedings of the Survey Research Methods Section, American Statistical Association*.

Loosveldt, G., Pickery, J. and Billiet, J. (2002) Item Nonresponse as a Predictor of Unit Nonresponse in a Panel Survey. *Journal of Official Statistics* **18** (4), pp. 545-557.

Lynn, Peter, Nick Buck, Jonathan Burton and Heather Laurie (2006) Quality Profile: British Household Panel Survey, Version 2, Waves 1 to 13, 1991-2003. *Institute for Social and Economic Research*, University of Essex.

MIS-Trend (2000) Statistiques finales. Unpublished Field Report of the SHP wave 1 (1999) data collection, Lausanne.

MIS-Trend (2007) Panel Suisse de Ménages. Rapport technique réalisé pour Le Panel Suisse de Ménages – Neuchâtel. Unpublished Field Report, Lausanne.

Naud, J.-F. and Latouche, M. (2001) Weighting of the Swiss Household Panel. Wave 2 Detailed Description. *Swiss Household Panel Working Paper* **8_01**, Neuchâtel.

Nicoletti, C. and Peracchi, F. (2005) Survey response and survey characteristics: Micro-level evidence from the European Community Household Panel. *Journal of the Royal Statistical Society: Series A* **168** (4), pp. 763–781.

Pickery, J. and Loosveldt, G. (2002) A Multilevel Multinomial Analysis of Interviewer Effects on Various Components of Unit Nonresponse. *Quality and Quantity* **36**, pp. 427-437.

Pickery, J. and Loosveldt G. (2004) A Simultaneous Analysis of Interviewer Effects on Various Data Quality Indicators. *Journal of Official Statistics*, **20** (1), pp. 77–89.

Pickery, J., Loosveldt, G. and Carton, A. (2001) The Effects of Interviewer and Respondent Characteristics on Response Behaviour in Panel Surveys. *Sociological Methods & Research* **29** (4) (May 2001), pp. 509-523.

Rendtel, U. (2004) Attrition in Household Panels: A Survey. *Chintex* (The Change from Input Harmonisation to Ex-post Harmonisation in National Samples of the European Community Household Panel – Implications on Data Quality) *Working Paper* **22**.

Scherpenzeel, A. (2002) Interviewer Effects in data from the Swiss Household Panel. *Swiss Household Panel Working Paper* **2_06**, Neuchâtel.

Scherpenzeel, A., Zimmermann, E., Budowski, M., Tillmann, R., Wernli, B. und Gabadinho, A. (2002) Experimental pre-test of the biographical questionnaire. *Swiss Household Panel Working Paper* **8_02**, Neuchâtel.

Schräpler, J. (2004) Respondent Behavior in Panel Studies. A Case Study for Income Nonresponse by MEans of the German Socio Economic Panel (SOEP). *Sociological Methods and Research* **33** (1), pp. 118-156.

Stocké, V. and Stark, S. (2005) Stichprobenverzerrung durch Item-Nonresponse in der International vergleichenden Politikwissenschaft (with English Summary). *Sonderforschungsbereich 504*, Working Paper No. **05_43**, University of Mannheim, Germany.

Stoop, I. (2005) The Hunt for the last Respondent: nonresponse in Sample Surveys. PhD Thesis, Utrecht University.

Swiss Federal Statistical Office (2005) Eidgenössische Volkszählung 2000 – Sprachenlandschaft in der Schweiz. Neuchâtel, April 2005.

Tourangeau R. and Rasinski K.A: (1988) Cognitive Processes underlying Context Effects in Attitude Measurements. Psychological Bulletin **103** (3), pp. 299-314.

Watson, D. (2003) Sample Attrition between Wavers 1 and 5 in the European Community Household Panel. *European Sociological Review* **19** (4), pp. 361-378.

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

Zabel, J.E (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. *J. Human Resources 33 (2), Special Issue: Attrition in Longitudinal Surveys*, (Spring, 1998), pp. 479-506.

Zimmermann, E and Joye, D. (2003) *Item Nonresponse: Ignorance or Refusal?* Paper presented for the 2003 Workshop "Item Non-response and data quality in large social surveys", October 2003, Basel, Switzerland.

	Single Wave Models (single transitions)						All Wave Model (indiv. clustering controlled)					
	2000	**2001**	**2002**	**2003**	**2004**	**2005**	**base**	**+2001**	**+2002**	**+2003**	**+2004**	**+2005**
Wave effect (only all wave model)							Ref.					
SOCIO-DEMOGRAPHY												
Number of adults in household					++							
Number of kids in household	--		--				--					
Male	++						++					
age14-19	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Reference category					
age20-29	++	++	+									
age30-39						++	--					
age40-49							--					
age50-59	-		--		-		--					
age60-69			-				--					
age70-	+											
Married		--	-									
Education high												
Full time employed												
Part time employed		-			--							
Unemployed												
Swiss Citizen	--						--					
Language Swiss German				++						+		
Lake Geneva	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Reference category					
Middleland								Only base effect				
North-West CH								Only base effect				
Zurich								Only base effect				
East-CH		+						Only base effect				
Central CH								Only base effect				
Ticino								Only base effect				
Lives in Urban Centre												
Household moved within last 12 months	--	--	--	--	--	--	--					
Household has intention to move within next 12 months	.	++	.	++	++	++	.	.	.	.	.	.
SOCIAL ISOLATION												
House bad												
House owner												
Satisfaction with various aspects	--	-	-				--				++	
HEALTH												
Health Problems							++	-				
Health improved during last 12 months												
Do_sports				--						-		
TOPIC INTEREST AND MOTIVATION												
Engaged in voluntary work			-		-		-					
(Potential) Political Participation	--	--					--		+			
Political Left orientation							-					
Political interest	--		--	--	--	--	--					
Subjective questions noanswer												
Number of extreme categories		++	++	+			++					
Number of middle categories				+								
Income noanswer	++						++					
Response rate within household	--	--	--	--	--	--	--					
SURVEY STATUS OF INDIVIDUAL												
Person is Reference Person (only more-adult-HH)	--	--		--	-	-	--		+			
N	7769	6333	5719	4874	4332	3592	N=32619, ρ=0.20 (Intra-cluster coeff.)					
MCFadden Pseudo R^2	0.068	0.064	0.057	0.053	0.045	0.055	-					
LR chi^2	506	331	309	212	209	176	Wald chi^2 (Deg. Freedom=180)					1485

Table 1: "Qualitative" influence on Attrition: "+"=positive, 1% significance level, ++"=positive, 1‰ significance level, "-"=negative, 1% significance level, --"=negative, 1‰ significance level. "."= n.a. Single wave models: forward regression.

All wave model: only significant (1%) effects indicated.	Single Wave Models (single transitions)						All Wave Model (indiv. clustering controlled)					
	2000	2001	2002	2003	2004	2005	base	+2001	+2002	+2003	+2004	+2005
Wave effect (only all wave model)							Ref.	.83	.88	.38	1.17	.97
SOCIO-DEMOGRAPHY												
Number of adults in household					1.19		1.05	.94	.97	1.02	1.13	1.02
Number of kids in household	.89		.84				.85	1.01	.98	1.06	1.15	1.01
Male	1.25						1.19					
age14-19	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Reference category					
age20-29	1.57	1.74	1.38				.96					
age30-39						.64	.57					
age40-49							.54					
age50-59	.78		.67		.75		.42					
age60-69			.65				.46					
age70-	1.48						.75					
Married		.72	.77				.99	.78	.86	.87	1.01	.96
Education high							.95	.94	.88	.91	.94	1.03
Full time employed							1.09	1.06	1.15	1.15	.98	1.11
Part time employed		.74			.71		1.1	.82	1.08	1.06	.71	.9
Unemployed							1.61	1	.69	1.27	1.23	.76
Swiss Citizen	.65						.61	1.15	1.27	1.36	1.23	1.66
Language Swiss German				1.58			1.23	1.14	.84	1.41	1.08	.73
Lake Geneva	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Reference category					
Middleland							.97	Only base effect				
North-West CH							.95	Only base effect				
Zurich							.90	Only base effect				
East-CH		1.40					1.05	Only base effect				
Central CH							.87	Only base effect				
Ticino							.95	Only base effect				
Lives in Urban Centre							.98	1.29	.98	.99	.98	.82
Household moved within last 12 months	.39	.38	.26	.21	.20	.18	.33	1.45	.77	.83	.71	.66
Household has intention to move within next 12 months	.	1.28	.	1.39	1.26	1.32	.	.	.	.	.	.
SOCIAL ISOLATION												
House bad							.98	1.03	1.06	1.06	.96	1.07
House owner							.83	1.25	1.11	1.07	.95	1.1
Satisfaction with various aspects	.80	.82	.79				.75	1.06	1.06	1.12	1.44	1.23
HEALTH												
Health Problems							1.31	.71	.95	.95	.77	.92
Health improved during last 12 months							.99	1.01	1.01	1.04	1.05	.99
Do_sports				.74			1.07	.99	.99	.74	.8	.98
TOPIC INTEREST AND MOTIVATION												
Engaged in voluntary work			.77		.78		.86	1.01	.88	.89	.88	1.06
(Potential) Political Participation	.85	.84				.82	.84	1.04	1.19	1.14	1.12	.96
Political Left orientation							.89	1.05	1.16	1.04	1.16	1.06
Political interest	.77		.82	.79	.75	.74	.82	1.07	1.01	1.04	.9	.9
Subjective questions noanswer							.97	1.15	1.21	1.21	1.08	1.33
Number of extreme categories		1.05	1.05	1.03			1.03	1.01	1.03	1.01	.96	1
Number of middle categories				1.07			1.01	1.01	.98	1.05	1.03	.97
Income noanswer	1.35						1.42	.84	.79	.79	.83	.8
Response rate within household	.14	.14	.23	.22	.30	.20	.14	.84	1.2	1.55	1.68	1.16
SURVEY STATUS OF INDIVIDUAL												
Person is Reference Person (only more-adult-HH)	.56	.66		.63	.77	.72	.6	1.17	1.52	1.22	1.32	1.26
N	7769	6333	5719	4874	4332	3592	N=32619, ρ=.20 (Intra-cluster coeff.)					
MCFadden Pseudo R^2	.068	.064	.057	.053	.045	.055	-					
LR chi^2	506	331	309	212	209	176	Wald chi^2 (Deg. Freedom=180)					1485

Table 2: « Quantitative » influence on Attrition: odds ratios. "."= n.a. Single wave models: only significant (1%) effects from forward regression model included.

Effects of different Incentives on Attrition and Fieldwork Effort in Telephone Household Panel Surveys

Oliver Lipps*

Keywords: attrition, bias, incentive effects, cost, telephone, individual level, household level

Little is known about sample behavior and fieldwork effects of different incentives introduced in a household panel survey. This is especially true for telephone surveys. In a randomized experiment, the Swiss Household Panel implemented one prepaid and two promised non-monetary incentives in the range of 10 to 15 Swiss Francs (7-10 €), plus a no incentive control group. The aim of the paper is to compare effects of these incentives especially on cooperation, but also on sample selection and fieldwork effort, separated by the household and the subsequent individual level.

We find small positive cooperation effects of the prepaid incentive on both the household and the individual level especially in larger households. Sample composition is affected to a very minor extent. Finally, incentives tend to save fieldwork time and partially the number of contacts needed on the individual level.

* Swiss Foundation for Research in Social Sciences (FORS)
c/o University of Lausanne, Vidy, CH – 1015 Lausanne, oliver.lipps@fors.unil.ch.
Acknowledgments: Support by the Swiss National Science Foundation is gratefully acknowledged.

Introduction

In many household panel surveys, there is a two-stage sequential response procedure (e.g. Frick et al., 2007): First the household reference person is asked to provide information on the household in the household grid questionnaire. Only after all interview eligible individuals are known, they can be asked to complete the survey. In addition to the household grid (and consequently all individual questionnaires), individual questionnaires are at risk to be refused especially by individuals other than the household reference person. This holds especially true in telephone surveys (Lipps, 2009): unlike in face-to-face surveys, usually not all eligible household members can be interviewed at once. To reduce attrition, measures must be effective on all sample members. First, it is crucial to convince the household reference person, who can be considered as a gatekeeper, to cooperate. Next, all household members must be motivated to answer the survey. Using different incentives can possibly accommodate this sequential survey design. Up to now, often the same incentives are used for all sample members. An examination of different incentives that are designed to reduce attrition on the household and the individual level separately is however still lacking in the literature. This research tries to tackle this problem. In addition, we consider effects of different incentives on sample selection and fieldwork effort on both levels.

The article is organized as follows: first we review the literature on effects of incentives in both cross-sectional and longitudinal surveys. Next, we set up the hypotheses on the effects which we expect from an incentive experiment, which is described together with the data used. We analyze attrition and sample selection effects, and effects on fieldwork effort, distinguished by grid and individual level. We summarize and conclude in the last chapter.

Theoretic Considerations

Incentives are used to encourage and motivate survey sample members to see their participation as being valued and help the interviewer through the process of reciprocity and social interaction. Several theories have been proposed and tested to explain how incentives affect the respondent's decision to participate in surveys (see for an overview e.g. Ryu, Couper, & Marans, 2005). For example, the *economic* exchange theory considers incentives as a compensation for the respondent's time and efforts invested to complete the survey. The *social* exchange theory (Dillman, 2000) explains that small prepaid incentives work in many cases by demonstrating trust that the potential respondent will answer the survey. Closely related is the principle of *reciprocity*, i.e. the norm that people should help those who helped them. An explanation why different individuals react differently on incentives is provided by the *leverage-salience* theory (Groves, Singer, & Corning, 2000). Individuals perceive and value the same survey attribute (e.g. topic, sponsor, or incentive) in different ways. This thus has different effects on the decision to participate (Groves & Couper, 1998).

Evidence in Cross-sectional Surveys

Generally, monetary incentives are more effective than gifts, and prepaid incentives result in significantly higher response rates compared to incentives that are conditional on participation (Singer et al., 1999; Singer, 2002). In a series of experiments on the monthly conducted random digit dial (RDD) telephone U.S. Survey of Consumer Attitudes (SCA), Singer, van Hoewyk and Maher (2000) find that prepaid incentives enclosed with advance letters have positive effects on response in telephone surveys, especially among those who had initially refused to participate. Such positive effects need not be true with promised incentives, which Becker, Imhof & Mehlkop (2007) test in mail surveys Also, prepaid incentives may be able to reduce the number of calls on sample members after having established contact. James (1997) and Curtin, Singer and Presser (2007) show that while incentives reduce the number of calls needed in face-to-face and RDD surveys, the amount paid may have different effects on response. Brick et al. (2006) also report positive response effects from prepaid cash incentives in an RDD survey, but also diminishing effects per amount of incentive. Finally, prepaid incentives do not seem to affect the sample composition.

Also forms the incentive takes, plays a role: For example, Lengacher et al. (1995) find in a face-to-face panel survey among older individuals that compared with a cash payment or a gift, charitable giving tends to increase subsequent wave response rates among those who were already cooperative respondents. The conclusion is that charitable giving can be viewed as a proxy for altruistic activities more generally, including seeing participating in social surveys as a civic duty (Laurie & Lynn, 2009). Lottery incentives are not uncommon in household surveys, probably because they allow for easy administration (e.g. Hansen, 2006). According to Singer (2002), “lotteries function as promised cash incentives” (p. 6) that can be “coded as a monetary incentive with a value equivalent to the prize divided by the number of subjects in that experimental condition” (Singer et al., 1999, p. 221). Simmons and Wilmot (2004) and Hansen (2006) however, consider lotteries as a nonmonetary incentive. Research suggests that lotteries are less effective than prepaid incentives in stimulating survey response (Singer & Kulka, 2000) but might be more attractive to individuals who are less risk averse

(Holt & Laury, 2002). Stamps sent with the advance letter have a monetary value and have properties similar to phone cards, which are often used as incentives (Teisl et al., 2006).

Evidence in Panel Surveys

While there are many studies on incentives effects in cross-sectional surveys, only few studies exist in longitudinal surveys. This is on one hand surprising because to keep respondents in the sample is essential for the survival of panel surveys. On the other, it may be dangerous to experiment with panel sample members, who have participated since many waves. It was only recently that Laurie & Lynn (2009) extensively reviewed the use of incentives in longitudinal surveys. They report that while attrition can be reduced, incentives do not affect sample composition. Jäckle & Lynn (2008) find similar results from an incentive experiment in a UK panel survey of young people. It is shown that positive effects on retention rates are larger for unconditional than conditional incentives.

Several sources suggest positive effects even when small incentives are given (e.g., Laurie & Lynn, 2009). Although attrition may already have left a sample which is essentially fairly cooperative, the British Household Panel Survey had good experiences with a small increase of a prepaid voucher from £7 to £10 per respondent, even after 14 waves of data collection. This was especially the case for those who did not respond in the wave before the incentive increase, being thus an effective strategy for reluctant respondents. Similarly, Martin, Abreu & Winters (2001), introducing incentives in waves 8 and 9 of the US Survey of Income and Program Participation (SIPP), show that these are able to improve refusal conversion rates.

Cost Issues

Incentives do not only incur costs, but might also reduce fieldwork efforts. For example, Singer, van Hoewyk & Maher (2000) find that a $5 incentive included with an advance letter significantly reduced both the number of calls in a telephone survey to work a case and the number of interim refusals. For the face-to-face SIPP and the US Health and Retirement Survey (HRS), incentives reduced the number of calls that interviewers needed to make at wave 1 (cited in Laurie & Lynn, 2009). While Brick et al. (2006) do not observe strong differences between the incentive experiment groups in the RDD administered US National Household Education Survey (NHES), Curtin, Singer & Presser (2007) find moderately negative incentive effects on the number of calls in the RDD SCA, and Rodgers (2002) considerable cost savings from the reduced number of calls in the face-to-face administered HRS. Haggerty et al. (2000) finds an even “dramatically reduced” (p. 1272) number of calls per completed case in small business surveys.

Hypotheses

We expect the following effects:

1. Providing an unconditional incentive to the household is effective as a “door-opener” also for the skeptical households. Compared to the situation without incentive, we expect:

a.) higher household grid completion rates[1], especially in samples that are introduced later.

b.) no difference with respect to the household sample composition.[2]

Because of the announcement in the advance letter, we expect smaller but also positive effects from incentives that are conditional on individual participation. This is because the reference person might not want to exclude other household members (and him/herself) from benefitting from the conditional incentive by not completing the household grid.

2. Once the household grid is completed, incentives conditional on individual participation are expected to motivate all eligible household members to answer the individual questionnaires. Relative to the situation without individual incentives, we expect

a.) higher individual completion rates, especially in samples that are introduced later.

b.) different individual sample composition effects depending on the conditional incentive used. The unconditional incentive is expected to have a positive effect also on individual participation that might however be smaller than that from the conditional incentives.

3. Relative to the situation without incentives, we expect reduced fieldwork efforts from all incentives on both

a.) the grid and

b.) the individual level.

Incentive Experiment

In 2006, the Swiss Household Panel (SHP) implemented an experiment which used three incentive groups plus a no incentive group as a control. It was decided not to use cash, because the incentive should not be misunderstood as a monetary compensation for the time

[1] Throughout this paper we use the AAPOR (2008) definitions COOP1 for the cooperation rate, REF1 for the refusal rate, and RR1 for the response rate.

[2] Although the longitudinal unit in household panel surveys is the single individual, in order to analyze household level attrition, it is necessary to follow the household. We let the household in which the last year’s household reference person lives represent the household. In case a household splits, we follow the household where the reference person from the previous wave resides. Note that the in the SHP, the household reference person remains the same after a wave with a probability of almost 90%. In face-to-face household panel surveys a change of the household reference person is even less probable between waves (Lipps 2009).

and efforts invested by the sample members, but rather as a token of appreciation. Due to the relatively small budget, the incentives were not to exceed a certain value. Furthermore a large experimental sample was planned to be able to analyze effects for subgroups, if necessary. Of course, decision in favor of a higher value of the incentives or a larger sample size was a trade-off. Easy administration was an additional condition. All other communications with panel households already conducted in former waves are maintained, for example a newsletter with survey results, and advance letters (now with different content, according to the incentive delivered) sent to each household. A random quarter of all households that answered in the previous wave were divided into one of the following groups:

- *No incentives* (control group). No incentives were mentioned in the advance letter. Here, the content of the letter was similar to that of the previous years.
- *Stamps.* These households were sent 12 stamps at 1 Swiss Francs (about 8.00 €) unconditionally with the advance letter. The stamps were printed with the SHP logo.
- *Lottery.* These households were told in the advance letter that each respondent to the individual questionnaire will participate in a lottery with three prizes: 1.) 5,000 Swiss Francs 2.) 3,000 Swiss Francs, and 3.) 2,000 Swiss Francs. The monetary value per respondent amounts to 5.55 €.[3]
- *Donation.* Households in this group are told in the advance letter that each respondent to complete the individual questionnaire may donate 10 Swiss Francs (about 6.67 €) to a charity, to be selected from a list at the end of the interview.

Data

The SHP is a centralized CATI panel survey aiming to observe social change, in particular the dynamics of changing living conditions in Switzerland. The SHP survey started in 1999 with slightly more than 5000 randomly selected households. Each year, the household reference person is asked to report the current household composition together with basic socio-demographic characteristics in the household grid. Completion of the grid questionnaire takes two to ten minutes, depending on household size and complexity of relationships. Then, the household questionnaire is to be completed (about 10 minutes), again by the reference

---

[3] Each group contains about 1,000 households with each around 1.5 eligible respondents. Assuming a response rate of 80% (1,200 respondents), the "monetary value for each participant amounts to 8.3 Swiss Francs (5,55 €).

person. Finally, each household member from the age of 14 on has to complete his/her own individual questionnaire (about 35 minutes), including the reference person. To keep the sample size at a reasonable level, a refreshment sample has been added in 2004. Like the original sample from 1999, also the refreshment sample is representative of the Swiss residential population. The original and the refreshment sample each contain about 2,500 successfully interviewed households in 2005. There is no different treatment of the samples. In particular, neither the members of the original sample nor those of the refreshment sample obtained incentives before the experiment, other than the usual SHP newsletter. Until wave seven (2005), the attrition rate on the grid level amounts to between 11% (in 2000, 2001), 13% (in 2003, 2005), and even slightly more than 15% (2002, 2004) (Lipps, 2009). As for individual attrition conditional on household grid completion, the rates amount to between 4% (2003), 5% (2005), 6% (2001, 2002), to 8% (2000, 2004). Concerning bias, the young[4] and the old, male headed, not working, and smaller households tend to attrite to a higher extent. Selection due to individual attrition is effective towards the middle-aged, thus aggravating bias already from household attrition, and those living with a partner. With respect to socio economic and attitudinal variables, foreigners, the socially and politically excluded, those who are mostly dissatisfied with various aspects in their life, and those who exhibit a worse reporting behaviour in previous waves show higher attrition (Lipps, 2007).

In addition to the nonrespondents from the 2005 wave, households whose address was not known or who sent back the newsletter between the 2005 and the 2006 wave are excluded from the experiment. This is predominantly because of the risk that in such households the advance letters notifying them of the incentive are not read. By using this procedure, there are only very few (2.2 %) refusing individuals from the previous wave in the sample. We exclude previous refusers for the analyses. In addition, because the reasons driving noncontacts and refusals are different (Groves & Couper, 1998), and incentives are more effective to reduce refusals rather than not contacted sample members (Singer & Kulka, 2000), we do not take into account sample members that could not be contacted in 2006.

In Table 1, the sample randomization results, distinguished by original and refreshment sample, is depicted.

< Table 1 here >

[4] Household socio-demographic characteristics are represented by the household reference person (Lipps, 2009).

Results[5]

Household Level

Table 2, Table 3, and Table 4 contain the 2006 grid response rates, separated by incentive group and household size, for the total, the original, and the refreshment sample, respectively.

< Table 2 here >

< Table 3 here >

< Table 4 here >

Using Fisher's exact test, we find that in both the total and the refreshment sample, the response rate of the stamps group in all and in 3+ person households is significantly higher than in the donation group. In addition the stamps group outperforms the control group in case of 2-person households in the original sample.[6]

Hypothesis 1a (higher grid response rates with conditional incentive) cannot be rejected. Higher cooperation does however not hold for the conditional incentives.

Next, we test for different sample composition effects, using the sample with a completed reference person individual questionnaire in 2005. The dependent variable equals 1 if the grid questionnaire in 2006 is completed, 0 else. We first include (the few) independent variables that were shown to be affected by incentives in the literature, like sample maturity, social inclusion (employment status, married, education, health, satisfaction with life, political interest, participation) and respondent behavior assessment during the interview by the interviewer. Age and sex are added because some selection effects have been shown in surveys that observe special age groups only. All independent variables included affect attrition in the SHP (Lipps, 2007; Voorpostel 2010) or in other panel surveys (Laurie & Lynn, 2009; Ryu, Couper, & Marans, 2006; Voogt & Saris, 2003).

Individual level variables are taken from the 2005 reference person questionnaires:

- original sample (first asked in 1999) vs. refreshment sample (first asked in 2004)

[5] Throughout the paper we will use the .05 (=5%) significance level.
[6] Note that there are also significant differences between cells from the original sample (table not shown). Due to a too small cell size, however, we do not comment on them here.

- number of children under the age of 18 in the household
- full time employment of reference person (2005)
- reference person male (2005)
- reference person married
- reference person age
- reference person age squared
- education level of reference person (8 ordinal degrees in 2005)
- health status (1 (very bad) -5 (very good))
- satisfaction with life (0 (very bad) -10 (very good))
- political interest (0 (absolutely not) -10 (a great deal))
- participation in clubs or groups
- Interviewer assessment: respondent behaves friendly (1 (hostile) – 4 (friendly)
- Interviewer assessment: respondent understands questions (1 (poor) – 3 (good))
- Interviewer assessment: respondent difficult to get (1 (vey difficult) – 4 (easy))
- Interviewer assessment: respondent repeats in next wave (1 (no) – 4 (absolutely))

Using multinomial logit models with the incentive as dependent variable, the coefficients of each incentive group with the control group as base are listed in Table 5, .

< Table 5 here >

Two coefficients are significant in the donation-control comparison model: households from the original sample and those who are more difficult to be convinced to participate show a slightly (5% level) higher grid participation. The first group of households tend to attrite to a lesser, the latter to a higher extent in the SHP (Lipps, 2007). Based on the high number of variables (16) entered in the models, it could be expected that a small number of variables are significant. Therefore we do not consider the sample composition of the incentive and the control group to be different.

Hypothesis 1b (no different sample composition effects) cannot be rejected.

Finally, we check if the incentive has an influence on the effort necessary to work a case until the final response status is determined. We compare the number of calls that result in a noncontact, the number of actual contacts, and the total number of days it takes from the first until the last call. To control the effect from unobserved household effects, we calculate the difference between the 2006 (experiment) and the 2005 values (Δ) for each household.

Similar to the analysis of selection effects, we use multinomial logit models with the incentive as dependent, and the 2006-2005 fieldwork effort differences as independent variable. For comparison reasons, we exclude households that do not cooperate in 2006.[7] For completeness, we depict the fieldwork effort measures from 2005 in the upper part of Table 6 in addition. For example, it took on average 2.8 contacts to work a household grid in the control group.

< Table 6 here >

None of the within household differences in any of the incentive groups turn out to be significant when compared with the control group.

Hypothesis 3a (reduced fieldwork efforts on the grid level) must be rejected.

Individual Level

Because grid questionnaire completion is necessary before individuals can be contacted, individual response behavior is analyzed conditional on grid completion. We find the following final statuses in 2006, distinguished by incentive group, as well as for the maturity of the sample:[8]

< Table 7 here >

< Table 8 here >

< Table 9 here >

In the combined sample, the stamps group outperforms (5%) all other groups for the larger households and for all households. In the original sample, cooperation of the stamps group in all and in 3+ person households is significantly (5%) higher than that of the control group. In addition, the control group outperforms both conditional incentive groups if all households are combined. Finally, in the refreshment sample, cooperation of the stamps group in all and in 3+ person households is significantly higher than that of the lottery group only. Similar to the situation in the original sample, the control group outperforms the lottery group if all households are combined.

---

[7] Ultimately non-cooperating households need much more calls and fieldwork time. Because the sample is limited to the 2005 respondents, we must therefore drop the 2006 nonrespondents.

[8] The strong differences by household size are due to the fact that reference persons generally cooperate once they completed the grid and the likelihood to be reference person decreases with household size.

Hypothesis 2a (higher individual completion rates) again holds for the prepaid incentive only, and must be rejected for the individual conditional incentives.

.

Also for individuals, we analyze sample composition effects, using the same independent variables as in the case of households, plus the binary variable reference person status. Like in the analysis of household selection effects, we compare the coefficients of each incentive group with those of the control group, using multinomial logit models with the incentive as dependent variable.

< Table 10 here >

In the stamps group, we find that the number of children in the household, dissatisfaction with life, and participation in clubs or groups is associate with a slightly (5%) higher attrition relative to the control group. In the lottery group, there are no selection effects compared to the control group. Members of the donations group exhibit smaller (1%) attrition among the original sample, and slightly higher among the middle aged[9]. The effects are not consistent, as for example in the stamps group, we expected a reduction of the attrition among those who usually tend to attrite to a higher extent, like the childless or those who do not participate in clubs. The same interpretation holds for the selection effects in the donations group. We therefore do not consider the (few and mostly on the 5% level only) significant selection effects as causal effects due to the incentives.

Hypothesis 2b (different composition effects) must be rejected.

Also on the individual level, we analyze fieldwork effort effects, using the same methods than for households.

< Table 11 here >

Compared to the control group, there is a significantly (1%) smaller number of fieldwork days necessary in all incentive groups to work an individual, compared with the control group. In the lottery group, the number of contacts is also significantly smaller..

Hypothesis 3b (reduced fieldwork efforts on the individual level) cannot be rejected.

[9] Note from Table 10 that the age relation to higher completion is u-shaped

Summary and Conclusion

The Experiment

Based on an incentive experiment, this study analyzes attrition in the CATI Swiss Household Panel (SHP) survey both on the household (1st response stage) and the subsequent individual (2nd response stage) level, using three incentive groups and one control group. The prepaid household incentives (stamps) amount to a monetary value of 12 Swiss Francs (about 8 €), the individual expected value of the conditional lottery incentive to 8.33 Swiss Francs, and the conditional charity donation incentive to 10 Swiss Francs, for each participating individual. On both the household and the individual level, we expect motivation to participate to be higher if incentives are offered, with different composition effects due to incentives only on the individual level. Due to increased panel loyalty, we expect these effects to be higher in the sample that was introduced later, then in its 3rd wave, compared to the original sample, then in its 8th wave. Finally we hypothesize some cost saving effects until a final response status is obtained.

Findings and Limitations

With respect to completing the household grid, while the prepaid incentive outperforms the control in 2-person households in the original sample only, members of one of the conditional incentive groups (donations) cooperate worse than those of the prepaid group in all households and in 3+ person households in both the refreshment and the combined sample. Relative to the control group, there are no sample composition effects in any of the incentive groups. Finally, there is no reduction of fieldwork effort needed to work a case.

Given the household grid is filled, as for individual questionnaire completion, there is also a higher participation due to the unconditional incentive, and outperforms all other incentive groups. In the original sample, the unconditional incentive outperforms the control group only, in the refreshment sample, one conditional incentive group (lottery). Also interesting, the control group members show a higher participation than both conditional incentive groups in the case of the original sample, and one conditional incentive group (lottery) in the case of the refreshment sample. Also on the level of the individual, although there are some weak sample composition effects, they are not systematically related to what could be expected from the literature. Finally as for cost saving effects, relative to the control group, all incentive groups appear to significantly shorten the fieldwork duration.

A limitation is the slight difference in monetary equivalents of the incentives per respondent. We however believe that they are nonetheless comparable. Overall, it may be that the size of the incentives was too small to induce stronger and more systematic effects in a panel, where the newest sample is already in its third wave.

A Consequence for the SHP and general Suggestions

In the literature there is evidence that incentives may be effective to increase response especially among sample members with low response propensities, for example during refusal conversion (e.g. Singer, Groves, & Corning, 1999; Lengacher et al., 1995). Rodgers finds that "the greatest cost-benefit ratio would likely have been achieved by offering the higher incentive to households in which there was non-response at the previous wave" (2002, p. 2933). Acknowledging this, and given the results of the incentive experiment described in this paper, the SHP decided to send a 50 Swiss Francs voucher with the advance letter to the households, who refused in the previous wave, in the 2007 wave. A logit model of grid completion results in a significantly higher response, when compared with a roughly similar sample from the previous (2006) wave. Unfortunately, however, it is not possible to make these samples fully comparable in this nonexperimental setting.

To draw a conclusion, it is very likely that while effects from small or mid valued incentives are quickly decreasing with the maturity of the panel, more reluctant sample members probably remain sensitive to higher monetary incentives. Generally, we suggest considering the value and the form of incentives to be introduce in a mature panel very carefully. With respect to the latter, if incentives are conditional on participation, our findings show that they could backfire in reduced cooperation behaviors.

References

AAPOR (The American Association for Public Opinion Research) (2008). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 5th edition*. Lenexa, Kansas: AAPOR.

Becker, R., Imhof, R., & Mehlkop, G. (2007). Effects of prepaid monetary Incentives on the Return of Mail Survey and selfreporting about Delinquency (German with English summary). *methoden – daten – analysen,* **1** (2): 131-59.

Brick, J., Hagedorn, M., Montaquila, M., Brock Roth, S., & Chapman, C. (2006). *Impact of Monetary Incentives and Mailing Procedures: An Experiment in a Federally Sponsored Telephone Survey (NCES 2006-066). Methodology Report.* U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Curtin, R., Singer, E., & Presser, S. (2007). Incentives in Random Digit Dial Telephone Surveys: A Replication and Extension. *Journal of Official Statistics* **23** (1): 91-105.

Dillman, D. (2000). *Mail and Internet Surveys: The Tailored Design Method.* New York: John Wiley & Sons.

Frick, J., Jenkins, S., Lillard, D., Lipps, O., & Wooden, M. (2007). The Cross-National Equivalent File (CNEF) and its Member Country Household Panel Studies. *Journal of Applied Social Science Studies* (Schmollers Jahrbuch) **127** (4): 627-654.

Groves, R., & Couper, M. (1998). *Nonresponse in Household Interview Surveys.* New York: Wiley & Sons.

Groves, R., Singer, E., & Corning, A. (2000). Leverage-Salience Theory of Survey Participation – Description and an Illustration. *Public Opinion Quarterly,* **64**: 299-308.

Haggerty, C., Grigorian, K., Harter, R., & Wolken, J. (2000). The 1998 Survey of Small Business Finances: Sampling and Level of Effort associated with gaining Cooperation from Minority-owned Businesses. In: Board of Governors, the Federal Reserve System, *Proceedings of the Second International Conference on Establishment Surveys* (pp. 1268-1273). Buffalo, NY.

Hansen, K. (2006). The Effects of Incentives, Interview Length, and Interviewer Characteristics on Response Rates in a CATI-Study. *International Journal of Public Opinion Research,* **19** (1): 112-121.

Holt, C., & Laury, S. (2002). Risk Aversion and Incentive Effects. *American Economic Review* **92** (5): 1644-1655.

Jäckle, A., & Lynn, P. (2008). Respondent Incentives in a Multi-Mode Panel Survey: Cumulative Effects on Nonresponse and Bias. *Survey Methodology,* **34** (1): 105-117.

James, T. (1997). Results of wave 1 incentive experiment in the 1996 survey of income and program participation. In: American Statistical Association, *Proceedings of the Survey Research Methods Section* (pp. 834-839)*,* Alexandria, VA.

Laurie, H. (2007). The Effect of Increasing Financial Incentives in a Panel Survey: an experiment on the British Household Panel Survey, Wave 14. *Working Paper 2007-05 of the Institute for Social and Economic Research*, Colchester: University of Essex.

Laurie, H. & Lynn, P. (2009). The use of respondent incentives on longitudinal surveys. In Lynn, P. *Methodology of Longitudinal Surveys*, Colchester: John Wiley & Sons.

Lengacher, J., Sullivan, C., Couper, M., & Groves, R. (1995). *Once Reluctant, Always Reluctant? Effects of Differential Incentives on Later Survey Participation in a Longitudinal Study*. Survey Research Centre, University of Michigan.

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

Lipps, O. (2009). Attrition of Households and Individuals in Panel Surveys. *SOEPpapers,* 164.

Martin, E., Abreu, D., & Winters, F. (2001). Money and Motive: Effects of incentives on Panel Attrition in the Survey of Income and Program Participation. *Journal of Official Statistics,* **17** (2): 267-84.

Rodgers, W. (2002). Size of Incentive Effects in a Longitudinal Study. In: American Association for Public Research: *Strengthening Our Community - Section on Survey Research Methods,* (pp. 2930-2935)*,* Alexandria, VA.

Ryu, E., Couper, M., & Marans, R. (2005). Survey Incentives: Cash vs. In-Kind, Face-to-Face vs. Mail, Response Rate vs. Nonresponse Error. *International Journal of Public Opinion Research,* **18** (1): 89-106.

Simmons, E., & Wilmot, A. (2004). Incentive Payments on Social Surveys: a Literature Review. *Survey Methodology Bulletin* **53** (1): 1-11.

Singer, E. (2002). The Use of Incentives to reduce Nonresponse in Household Surveys, *Survey Methodology Program Working Paper*, *051*, Ann Arbor: Institute for Social Research, University of Michigan.

Singer, E., Van Hoewyk, J, Gebler, N., Raghunathan, T., & McGonagle, K. (1999) The Effects of Incentives on Response Rates in Interviewer-Mediated Surveys. *Journal of Official Statistics,* **15** (2*):* 217-30.

Singer, E., Groves, R., & Corning, A. (1999). Differential incentives: beliefs about practices, perceptions of equity, and effects on survey participation. *Public Opinion Quarterly,* **63** (2): 251-260.

Singer, E., & Kulka, R. (2000). Paying respondents for survey participation. *Survey Methodology Program Working Paper, 092*, Ann Arbor: Institute for Social Research, University of Michigan.

Singer, E., Van Hoewyk, J., & Maher, M. (2000). Experiments with Incentives in Telephone Surveys. *Public Opinion Quarterly* **64**, (2): 171-188.

Teisl, M., Roe, B., & Vayda, M. (2006). Incentive Effects on Response Rates, Data Quality, and Survey Administration Costs. *International Journal of Public Opinion Research,* **18** (3): 364-373.

Voogt, R., & Saris, W. (2003). To Participate or Not to Participate: The Link between Survey Participation, Electoral Participation, and Political Interest. *Political Analysis,* **11**: 164-179.

Voorpostel, M. (2010). Attrition patterns in the Swiss Household Panel by demographic characteristics and social involvement. Appears in: Lipps, O., R. Tillmann, U. Kuhn & D. Lillard (eds.): Longitudinal Analysis in Switzerland. Special Issue of the *Swiss Journal of Sociology* **2**/2010

Sample Randomization		
[N Households], [Column %]	SHP original Sample (started 1999)	SHP Refreshment Sample (started 2004)
Control	622 (25.2)	468 (25.2)
Stamps	613 (24.8)	475 (25.5)
Lottery	597 (24.2)	475 (25.5)
Donations	637 (25.8)	442 (23.8)
All Households	2,469 (100.0)	1,860 (100.0)

Table 1: Randomization of Sample for Incentive Experiment (Households with 2005 completed Grid).

Response Rate Grid 2006: all Households				
[%]	1 Pers.Household	2 Pers.Household	3+Pers.Household	All Households
Control	88.1 (1.9), N=294	87.7 (1.7), N=367	91.1 (1.4), N=429	89.2 (.9), N=1090
Stamps	87.8 (1.9), N=303	91.4 (1.6), N=339	92.4 (1.3), N=446	90.8 (.9), N=1088
Lottery	89.3 (1.9), N=272	87.6 (1.7), N=372	90.4 (1.4), N=428	89.2 (.9), N=1072
Donations	84.7 (2.2), N=281	87.1 (1.8), N=363	88.0 (1.6), N=435	86.8 (1.0), N=1079

Table 2: Response Rates Grid 2006 in all Incentive (+Control) Groups (in Brackets: Standard Errors). Note: Cooperation is higher (5%) in the stamps than in the donation group in all and in 3+ Person households.

Response Rate Grid 2006: Original Sample				
[%]	1 Pers.Household	2 Pers.Household	3+Pers.Household	All Households
Control	92.1 (2.1), N=164	90.6 (2.0), N=203	94.9 (1.4), N=255	92.8 (1.0), N=622
Stamps	91.7 (2.2), N=157	96.4 (1.3), N=197	94.6 (1.4), N=259	94.5 (.9), N=613
Lottery	89.8 (2.5), N=147	94.1 (1.6), N=205	94.3 (1.5), N=245	93.1 (1.0), N=597
Donations	87.3 (2.7), N=157	93.4 (1.7), N=212	93.7 (1.5), N=268	92.0 (1.1), N=637

Table 3: Response Rates Grid 2006 in all Incentive (+Control) Groups (in Brackets: Standard Errors). Note: Cooperation is higher (5%) in the stamps than in the control group in 2 person households.

Response Rate Grid 2006: Refreshment Sample				
[%]	1 Pers.Household	2 Pers.Household	3+Pers.Household	All Households
Control	83.1 (3.3), N=130	84.1 (2.9), N=164	85.6 (2.7), N=174	84.4 (1.7), N=468
Stamps	83.6 (3.1), N=146	84.5 (3.0), N=142	89.3 (2.3), N=187	86.1 (1.6), N=475
Lottery	88.8 (2.8), N=125	79.6 (3.1), N=167	85.2 (2.6), N=183	84.2 (1.7), N=475
Donations	81.4 (3.5), N=124	78.1 (3.4), N=151	79.0 (3.2), N=167	79.4 (1.9), N=442

Table 4: Response Rates Grid 2006 in all Incentive (+Control) Groups (in Brackets: Standard Errors). Note: Cooperation is higher (5%) in the stamps than in the donation group in all and in 3+ person households.

Grid Completion in 2006, relative to Control Group			
[N=3,853]	**Stamps**	**Lottery**	**Donations**
Original Sample	.04 (.10)	.04 (.10)	.22 (.10)*
number of children	-.05 (.06)	.02 (.06)	-.01 (.06)
full time employment	.15 (.13)	-.02 (.13)	.01 (.13)
male	-.05 (.12)	.09 (.12)	.01 (.12)
married	.04 (.11)	-.06 (.11)	.01 (.11)

age	-.00 (.02)	-.02 (.02)	-.03 (.02)
age squared	.00 (.00)	.00 (.00)	.00 (.00)
education	.00 (.02)	-.01 (.01)	.01 (.02)
health	.01 (.08)	.02 (.08)	.02 (.08)
satisfaction life	.05 (.03)	.00 (.03)	.02 (.03)
political interest	-.01 (.02)	-.02 (.02)	.00 (.02)
participation	-.08 (.10)	-.03 (.10)	.04 (.10)
Is friendly	-.15 (.15)	.20 (.15)	-.18 (.15)
Understands questions	.18 (.14)	-.16 (.13)	.05 (.14)
Is difficult Case	.04 (.19)	.30 (.19)	.44 (.19)*
Will repeat wave	-.06 (.09)	-.02 (.10)	.08 (.10)

Table 5: Multinomial Logit Coefficients by Incentive (Base: Control Group). Data: Households with 2005 responding Reference Persons.*(**) significant on 5(1)% Level.

Fieldwork Effort 2005			
	Mean Number Noncontacts	Mean Number Contacts	Mean Number Days
Control [n=972]	6.2 (.4)	2.8 (.1)	18.9 (.9)
Stamps [n=988]	6.8 (.5)	2.7 (.1)	18.1 (.8)
Lottery [n=956]	6.8 (.6)	2.7 (.1)	19.7 (.9)
Donations [n=937]	7.0 (.6)	2.9 (.1)	20.3 (1.0)
Δ Fieldwork Effort 2006-2005, relative to Control Group			
[N=3,853]	**Stamps**	**Lottery**	**Donations**
Nr. Noncontacts	-.000 (.002)	-.002 (.002)	-.003 (.002)
Nr. Contacts	.007 (.014)	.002 (.014)	.007 (.014)
Nr. Days	-.001 (.001)	-.001 (.001)	-.000 (.001)

Table 6: Fieldwork Effort: 2005 (upper), Multinomial Logit Coefficients by Incentive (Base: Control Group) (lower). Data: Households with 2006 completed Grid. *(**) significant on 5(1)% Level (only Differences).

Response Rate Individuals 2006				
[%]	1 Pers.Household	2 Pers.Household	3+Pers.Household	All Household
Control	96.9 (1.1), N=258	83.2 (1.6), N=554	78.1 (1.3), N=991	82.4 (.9), N=1803
Stamps	97.0 (1.1), N=264	85.3 (1.5), N=538	81.9 (1.2), N=996	85.2 (.8), N=1798
Lottery	97.1 (.1.1), N=242	82.2 (1.6), N=557	77.1 (1.4), N=945	81.5 (.9), N=1744
Donations	96.2 (1.2), N=237	83.6 (1.6), N=555	78.2 (1.3), N=976	82.3 (.9), N=1768

Table 7: Response Rates Individuals 2006 in all Incentive (+Control) Groups (in Brackets: Standard Errors). Note: Cooperation is higher (5%) in the stamps than in all other incentive groups in all and in 3+ person households.

Response Rate Individuals 2006: Original Sample				
[%]	1 Pers.Household	2 Pers.Household	3+Pers.Household	All Household
Control	96.7 (1.5), N=150	87.1 (1.9), N=302	81.1 (1.6), N=576	85.1 (1.1), N=1028
Stamps	99.3 (.7), N=142	90.1 (1.7), N=312	86.6 (1.4), N=566	89.4 (1.0), N=1020
Lottery	99.2 (.8), N=131	88.2 (1.8), N=314	83.5 (1.6), N=546	87.1 (1.1), N=991
Donations	97.1 (1.5), N=136	87.8 (1.8), N=337	82.4 (1.5), N=631	85.9 (1.0), N=1104

Table 8: Response Rates Individuals 2006 in all Incentive (+Control) Groups (in Brackets: Standard Errors). Note: Cooperation is higher (5%) in the stamps than in the control group in all and in 3+ person

households. In addition, the control group outperforms (5%) the lottery and the donation groups if all households are combined.

Response Rate Individuals 2006: Refreshment Sample				
[%]	1 Pers.Household	2 Pers.Household	3+Pers.Household	All Household
Control	97.2 (1.6), N=108	78.6 (2.6), N=252	74.0 (2.2), N=415	78.8 (1.5), N=775
Stamps	94.2 (2.1), N=122	78.8 (2.7), N=226	75.8 (2.1), N=430	79.6 (1.4), N=778
Lottery	94.6 (2.2), N=111	74.5 (2.8), N=243	68.4 (2.3), N=399	74.2 (1.6), N=753
Donations	95.0 (2.2), N=101	77.1 (2.9), N=218	70.4 (2.5), N=345	76.4 (1.7), N=664

Table 9: Response Rates Individuals 2006 in all Incentive (+Control) Groups (in Brackets: Standard Errors). Note: Cooperation is higher (5%) in the stamps than in the lottery group in all and in 3+ person households. In addition, the control group outperforms (5%) the lottery group if all households are combined.

Individual Completion in 2006			
[N=5,130]	**Stamps**	**Lottery**	**Donations**
Reference Person 06	.01 (.09)	.11 (.09)	.13 (.09)
Original Sample	.02 (.08)	.07 (.08)	.28 (.08)**
number of children	-.10 (.04)*	.05 (.04)	.00 (.04)
full time employment	.19 (.10)	.06 (.11)	.08 (.11)
male	-.13 (.09)	-.06 (.09)	-.14 (.09)
married	-.01 (.10)	-.07 (.10)	.07 (.10)
age	-.00 (.01)	.00 (.01)	-.03 (.01)*
age squared	.00 (.00)	.00 (.00)	.0003 (.0001)*
education	.00 (.01)	-.02 (.02)	.00 (.01)
health	-.04 (.06)	.01 (.07)	.01 (.06)
satisfaction life	.06 (.03)*	.03 (.03)	.01 (.03)
political interest	-.01 (.02)	-.01 (.02)	.00 (.02)
participation	-.19 (.08)*	-.04 (.08)	-.02 (.08)
Is friendly	-.08 (.12)	.00 (.12)	-.09 (.12)
Understands questions	.06 (.11)	-.09 (.11)	.12 (.12)
Is difficult Case	.13 (.16)	-.03 (.16)	.20 (.15)
Will repeat wave	.02 (.06)	-.04 (.08)	.01 (.08)

Table 10: Multinomial Logit Coefficients of Incentives (Base: Control Group). Data: 2005 Individual Completion.*(**) significant on 5(1)% Level.

Fieldwork Effort 2005			
	Mean Number Noncontacts	Mean Number Contacts	Mean Number Days
Control [n=1,326]	5.2 (.4)	2.7 (.1)	13.8 (.6)
Stamps [n=1,402]	5.4 (.4)	2.7 (.1)	14.6 (.6)
Lottery [n=1,269]	5.4 (.4)	2.9 (.1)	15.0 (.6)
Donations [n=1,300]	4.8 (.3)	2.7 (.1)	14.7 (.6)
Δ Fieldwork Effort 2006-2005, relative to Control Group			
[N=5,952]	**Stamps**	**Lottery**	**Donations**

Nr. Noncontacts	-.002 (.002)	-.003 (.002)	-.002 (.002)
Nr. Contacts	-.015 (.010)	-.028 (.001)**	-.019 (.010)
Nr. Days	-.004 (.001)**	-.006 (.002)**	-.005 (.002)**

Table 11: Fieldwork Effort: 2005 (upper), Multinomial Logit Coefficients by Incentive (Base: Control Group) (lower). Data: Individuals with 2006 completed Questionnaire. *(**) significant on 5(1)% Level (only Differences).

A Note on Interviewer Performance Measures in Centralised CATI Surveys

OLIVER LIPPS*

Keywords: non-response, interviewer performance, contact level, multiple membership, cross-classified multilevel, random assignment.

Interviewer performance with respect to convincing sample members to participate in surveys is an important dimension of survey quality. However, unlike in CAPI surveys where each sample case "belongs" to one interviewer, there are hardly any good measures of interview performance for centralised CATI surveys, where even single contacts are assigned to interviewers at random. If more than one interviewer works one sample case, it is not clear how to attribute success or failure to the interviewers involved. In this article, we propose two correlated methods to measure interviewer *contact* performance in centralised CATI surveys. Their modelling must take complex multilevel clustering effects, which need not be hierarchical, into account. Results are consistent with findings from CAPI data modelling, and we find that when comparing effects with a direct ("naïve") measure of interviewer contact results, interviewer random effects are realistic only in one measure.

* Swiss Foundation for Research in Social Sciences (FORS), University of Lausanne, Vidy, CH – 1015 Lausanne, oliver.lipps@fors.unil.ch. Support by the Swiss National Science Foundation is gratefully acknowledged.

Introduction

In CAPI surveys, interviewers usually work all contacts on a sample member until the latter is either ready to complete the interview, refuses, or leaves the interviewer with a pending appointment. In the case of a CAPI survey, the assignments of the sample member's contacts to interviewers can therefore be schematised as follows:

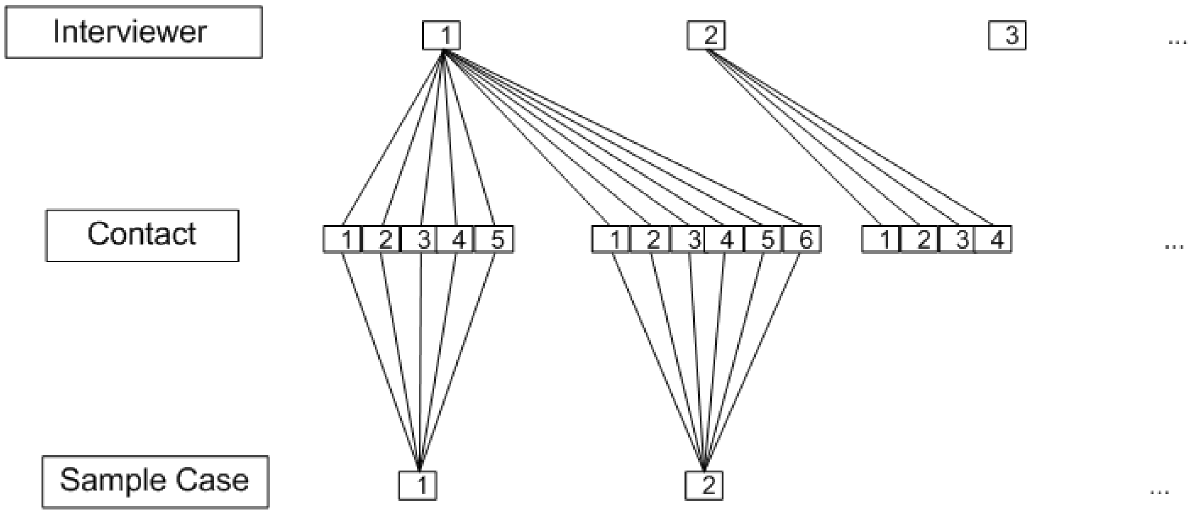


Figure 1: Interviewer-Sample Case Assignments via Contacts in CAPI Surveys.

Here, it is straightforward to measure interviewer performance in convincing sample members to participate in the survey, simply by calculating the mean number of *finally* participating sample cases worked by the interviewer. Methodologically, the only problem is a possible confusion of area and interviewer effects, because interviewers may obtain more or less "difficult" areas.[1]

In centralised CATI surveys, separation of these effects is guaranteed by the randomised sample case – interviewer - contact assignment:

[1] In the British Household Panel Survey (BHPS) wave 2, an interpenetrated sample experiment has been performed on a subsample in order to be able to separate interviewer and area effects (O'Muircheartaigh and Campanelli 1999).

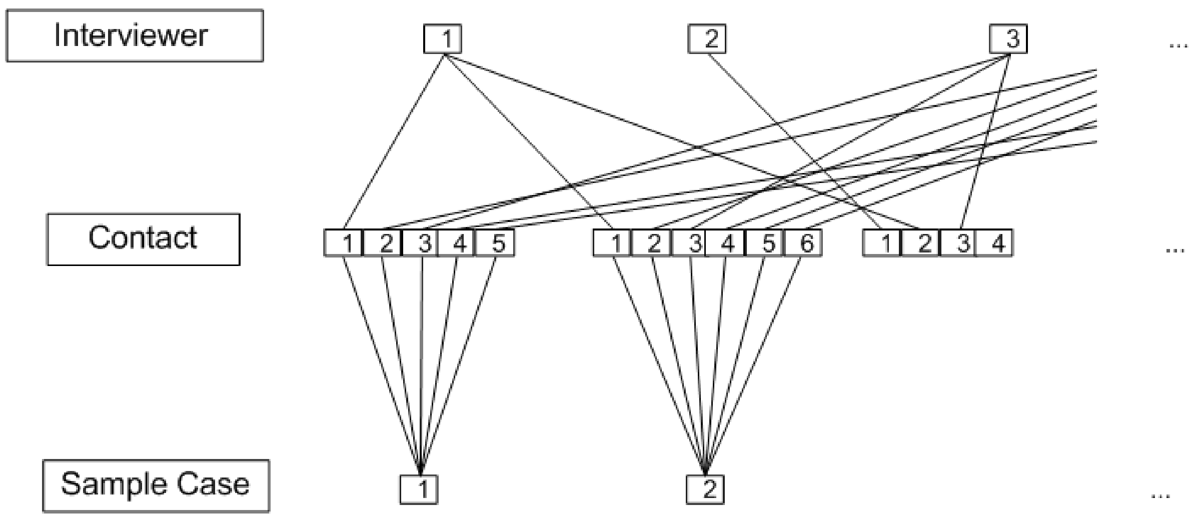


Figure 2: Interviewer-Sample Case Assignments via Contacts in Surveys with Random Assignment.

Here, although the problem of interviewer-area confusion is usually resolved (unless, inter alia, interviewers are used according to the dialect spoken in an area), it is not obvious how to measure interviewer performance. Most existing approaches focus on single interviewer-sample member contact results, where generally only cooperation rates based on first contacts are retained (e.g. Mayer and O'Brien 2001). The reason is "… to avoid contaminating the measure with the performance of a previous interviewer" (Durand 2005, p. 763). This is in line with analysis by Groves and Couper (1998, p. 256), who conclude that for the later contacts, the attributes of the prior contacts are the most important indicators of cooperation likelihood. However, if a final disposition is not achieved after the first contact, as in refusal conversion cases or if appointments are made, this approach is not applicable. Recent models therefore assign bonus or malus points to *transitions* (Durand 2005) achieved; i.e., they assess single contact results dependent on the previous contact result of the sample case. In addition, single contact results are directly assessed, with the result of the previous contact controlled for in regression models (Lipps 2007a).

These measures suffer from various problems:

- Arbitrariness of "point" assignment according to call or contact achievement. Durand, for example, attributes one credit point for a completed interview from a previous appointment (2005, p. 766). However, this procedure is not very convincing for appointments with a fixed date and time. No special interviewer performance is required to conduct a standardised interview at a fixed date. The achievement is rather to convince the sample member to fix a date and time for an interview. However no points are attributed for this achievement. Moreover, the degree of bindingness of appointments may vary widely. Lipps (2007a) shows that there is a substantial difference in the probability of finally completing a case, depending on whether a *vague* or a *fixed*[2] appointment has been made in the Swiss Household Panel Survey. In addition, the *order* of the contact on a sample case plays a role: after a first fixed appointment with the target person has been agreed, 88% of all household interviews are finally completed. This probability of final participation decreases continuously with the number of fixed appointments with the target person: to 84% after the second, 82% after the third, 76% after the fourth, 67% after the fifth fixed appointment, etc. A similar picture emerges after vague appointments with the target person: after the first vague appointment, 71% of the households are finally completed, 66% after the second, 64% after the third, 57% after the fourth, 51% after the fifth, etc. This example shows the difficulty of assigning credit points after a certain contact result.

- Clustering effects of contacts for sample members with interviewers are not considered. For example, in Durand's analysis (2005), a multilevel model for growth (with a random effect of time) with interviewers as second level is used, in order to assess interviewer learning effects on performance over time. However, sample member clustering effects can be assumed to be much higher but are not taken into account. There is also no consideration made for the complex cross-clustering of contacts within interviewers and sample cases (see Figure 2).

Performance Measures for Centralised CATI Surveys

"Cooperation" performance

We define the first performance index by referring the contacts to the specific contact *target,* namely the survey participation of the sample case. Therefore, a straightforward way to measure interviewer performance is to define *final* participation of the sample member as binary performance index, and relate it to the assigned contacts. I.e., all contacts done on a participating case would be assigned a "1" if the sample member finally participates, "0" if not.

[2] Fixed means with a fixed date and time for the interview.

The first measure interprets each contact of the interviewers involved in working the sample case as one partial contribution to the participation of the sample member. We call this approach “cooperation”, because the target is to convince the sample case as a complete unit to participate, irrespective of the outcome of single contacts.

“Refusal Avoidance” performance

To make things more complicated, survey research theory might suggest a different measuring approach. Groves’ and Couper’s concept of “maintaining interaction” with sample members (1998, p. 243 ff.), which has already proved successful in CAPI surveys, might form the basis of such an approach. This concept is based on the strategy of avoiding termination of the interaction during initial contacts (p. 249), because “the odds of success are increased with the continuation of the conversation” (p. 250). If this is accepted as interviewer guidance to be strictly adhered to in terms of a single contact, it would mean that the interviewer tries to minimise the odds of a “no” rather than to maximise the odds of a “yes”. Thus, a high interview performance in a randomly assigned setting could also mean that the interviewer does *not obtain a refusal* from the sample member. This concept includes the ability of “stepping back” (e.g. Hox et al. 1999, p. 193) as one possible interviewer tactic to adequately react to initially reluctant individuals.

In the second measure we thus understand high interviewer contact performance as not obtaining a refusal. Accordingly, we call this approach “refusal avoidance” (Groves and Couper 1998, Mayer and O’Brien 2001).

Control: “Optimisation” performance

Finally, we use a direct or “naïve” interviewer performance indicator as a control measure. The idea is that interviewers usually try to optimise single contact outcomes per se. As Sonnentag and Frese state: “… because teams are composed of individuals, team processes and team performance cannot be completely understood and improved without taking individual performance into account” (2002, p. 17).

In order to approximate the performance of a single contact, we calculate the mean probability of sample case cooperation by contact result (see Lipps 2007a). The trivial values are contacts resulting in a completed interview (optimisation performance=1) and refusals (optimisation performance=0). Concerning the more interesting intermediate contact results, we distinguish between fixed appointments with the target person and an agreed contact date and time, fixed appointment with another person and an agreed contact date and time, and vague appointments. Because in this setting the interviewer tries to optimise his or her single contact outcome, we call this approach "optimisation". Because of the averaging over all sample cases, there is concern that the optimisation index - much more than the other two indices - contains only a small part of the true interviewer effects.

Before we model the three performance indices, the distribution of the optimisation index is depicted in the table below. We use data described later (SHP/SILC 2005/2006), separated by regular and refusal conversion fieldwork phases. For example, after a vague appointment with a target person has been agreed, 59 percent of these sample cases are finally completed. For the sake of completeness, we also depict the trivial cases, completion and refusal.

(N=39,207 Contacts, 2005/2006 SHP/SILC) Contact Result	N contacts	% contacts	Sample Case Mean Cooperation Probability
Completed Interview	10,200	26	1
Refusal	2,686	7	0
Fixed Appointment with target person	6,375	16	.86
Vague Appointment with target person	16,762	43	.59
Fixed Appointment made with another person	3.171	8	.65
All Contacts	39,194	100	.70

Table 1: Sample Case Cooperation Probability by Contact Result. Data: SHP/SILC 2005/2006 Sample. Regular Fieldwork Phase.

Table 2 contains the respective figures for the refusal conversion phase:

(N=39,207 Contacts, 2005/2006 SHP/SILC) Contact Result	N contacts	% contacts	Sample Case Mean Cooperation Probability
Completed Interview	888	15	1
Refusal	2,142	36	0
Fixed Appointment with target person	562	9	.70
Vague Appointment with target person	1,929	33	.37
Fixed Appointment made by another person	398	7	.37
All Contacts	5,919	100	.36

Table 2: Sample Case Cooperation Probability by Contact Result. Data: SHP/SILC 2005/2006 Sample. Refusal Conversion Phase.

Because the optimisation index is measured as probabilities, we treat this variable as binomially distributed later in the models (Browne 2005).

Relationship of performance indices

Because the three performance indicators all intend to measure the same thing – interviewer performance – we expect them to be positively correlated. Even more than this, some contacts result in the same index value by definition: if an interviewer performs well on the cooperation index, s/he necessarily avoids a refusal. Similarly, a refusal implies a cooperation of 0. On the other hand, a refusal avoidance other than a cooperation may not necessarily mean a cooperation of 1, because another interviewer may still obtain a refusal by the sample member.

Table 3 lists the correlation matrix between the three performance indices, averaged over each interviewer.

Correlation Coefficient (N Interviewers) Significance Level	Regular Fieldwork Phase (N=202)			Refusal Conversion Phase (N=69)		
Modelling Approach/Fieldwork Phase	Coop	RA	Opt	Coop	RA	Opt
Cooperation (Coop): regular phase	1					
Refusal Avoidance (RA): regular phase	.58 .000	1				
Optimisation (Opt): regular phase	.74 .000	.76 .000	1			
Cooperation (Coop): refusal conv. phase	.18 .136	.05 .691	.16 .195	1		
Refusal Avoidance (RA): refusal conv. phase	.14 .269	.28 .020	.21 .085	.73 .000	1	
Optimisation (Opt): refusal conv. phase	.22 .073	.21 .085	.23 .060	.89 .000	.86 .000	1

Table 3: Correlation of Performance Indices. Data: SHP/SILC 2005/2006 sample.

As expected, we find positive correlations between the performance indices, with different degrees of correlation. Generally, high and highly significant correlations exist within either the regular or the refusal conversion fieldwork phase. For example, interviewers who perform well on the cooperation index during the regular phase, also do so on the refusal avoidance index during the same phase (R=.58). Correlations are not so high across different fieldwork phases: for example, the performance of interviewers on the refusal avoidance index during the regular phase has an insignificant (R=.05) correlation with performance on the cooperation index during the refusal conversion phase. Note however that the correlations across phases refer only to the subsample of interviewers who conduct refusal conversion contacts, with a supposedly higher performance also during the regular phase.[3]

[3] The appointment of interviewers to conduct refusal conversion contacts is up to the survey agency. Exact selection methods are not known.

Also interesting are the correlations across the two fieldwork phases on the same index: their magnitudes range from an insignificant R=.18 (cooperation) to a significant R=.28 (refusal avoidance). This means that interviewers tend to perform slightly better during the refusal conversion fieldwork phase if they already did so during the regular phase. This holds despite the positive interviewer selection for the refusal conversion phase mentioned above.

Modelling Interviewer Performance

In the modelling step, we are interested in the magnitude of fixed and random effects on interviewer and sample case level, using the three performance indices as dependent variables. Given that the indices all aim to describe the same thing, we would expect that the coefficients are similar.

Previous research has used CAPI data to analyse interviewer effects on sample member participation (Hox et al. 1991, Groves and Couper 1996, 1998, Japec 2005). This research has shown that considerable interviewer effects on survey cooperation exist, so we would expect that interviewer random effects would be significant for our measures. However, due to construction, it is probable that a large portion of true interviewer variance on sample case participation is not captured by our measures. This is likely to hold especially for the optimisation index. As to fixed interviewer effects, it is usually hard to identify significant variables (Groves and Couper 1998, Pickery et al. 2001, Japec 2005, Lipps 2007a). If significant at all, main effects of interviewers are likely to be weak (Groves and Couper 1998). The most important effects of interviewers on cooperation seem to be training and experience (Snijkers et al. 1999, Hox and de Leeuw 2002). Groves and Couper state that “most of the acculturation process of producing effective interviewers occurs during training on the job” (1998, p. 195).[4] Although the turnover in CATI is relatively high, even relatively short experience should have an impact. This can be expected because “performance initially increases with increasing time spent in a specific job and later reaches a plateau (Sonnentag and Frese 2002). Therefore we use interviewer experience measuring covariates and survey related indicators in order to model the three modelling approaches.

For each of the three modelling variables, we build three subsequent models: first an intercept only model, which allows for calculating the variance portions on the level of the

[4] Japec (2005), however, reports findings that do not show a positive relationship between interviewer experience and response rates.

sample cases and the interviewers. In addition this model yields a baseline deviance statistic, which can be used to assess the model improvement by including fixed effects. In a second step, we include sample characteristics variables explaining the part of the total variance due to panel and sample cohort membership effects, which serve as controls. In the third step, we include fieldwork and interviewer experience characteristics, along with outcome characteristics of the previous contact for the optimisation model. It is the portion of the interviewer variance reduction between the second and the third step in the different models, and the coefficients of the covariates entering the third step, which we are especially interested in. The interviewer experience variables include whether the interviewer is already in his/her second panel year, and the number of contacts s/he already worked during the fieldwork period. We control the difficulty of accessing sample members measured by the number of the contact on the sample case (optimisation) and the total number of contacts on a sample case until final disposition (all indicators), the working shift at which the contact takes place, and the elapsed number of days in the fieldwork period. In addition, we are interested in the question of whether it is advisable to have the same interviewer conduct subsequent contacts. Rendtel et al. (2004) report highly positive response effects from interviewer continuity between waves for the European Community Household Panel. However, Campanelli and O'Muircheartaigh (1999) did not find such effects in a subsample of the BHPS.

In the cooperation index models we use the sample member's cooperation behaviour outcome as a constant dichotomous variable over all contacts on this sample member within one fieldwork phase.[5] In the parlance of multilevel modelling, we have a non-hierarchical multiple membership setting (e.g. Fielding and Goldstein 2006): each lowest level unit (sample case) is a member of possibly more than one higher level unit (interviewer). The (single) outcome on one sample member thus has contributions from possibly more than one interviewer. Interviewer related effects can be conceptualised as weighted contributions of the interviewers working on that sample case. We set the weights according to the effort necessary to work the case and the suspected effect of the interviewer on the case: the n^{th} contact on a sample case is given a weight of 1/n. We thus take the increased difficulty of sample cases requiring more contacts to be finalised into account. To estimate the fixed and random coefficients of the multiple membership models, we use the Markov Chain Monte

[5] Equal to 1 if the sample case finally cooperates, otherwise 0.

Carlo (MCMC) estimation technique (Browne 2005), which is implemented in the MLWin Software.[6] If, as in the second or third modelling approaches, single contact results are to be analysed, cross-classified multilevel models are the modelling of choice (e.g. Fielding and Goldstein 2006). Here, contacts are clustered in sample cases, but sample cases are not clustered in interviewers (see Figure 2). Finally, the cooperation and the refusal avoidance indices are modelled as logistically distributed, with the optimisation index as a binomially distributed variable.

Data

We use call (process) data from two "multi-purpose" household panel surveys, conducted in Switzerland during the years 2005 and 2006. More specifically we use data from:

1. the Swiss Household Panel (SHP), an ongoing, nationwide, yearly conducted centralised CATI panel survey, which started in 1999 with slightly more than 5000 households;
2. the Swiss pilot of the Europe-wide Survey on Income and Living Conditions (SILC).

In each year, both surveys first ask the household composition together with the relationships between all household members, and the basic socio-demography of the household reference person in the grid questionnaire. Preferably, the household reference person should be the same individual across years. If, however, the previous year's reference person is not available, another adult person in the household who is knowledgeable enough about the household can replace him/her. The grid questionnaire takes three to ten minutes to complete, depending on household size and complexity of relationships. After filling the grid, a household related questionnaire is to be completed (about 10 minutes), again by the reference person. After the household related information is given, each household member from the age of 14 years on has to complete his/her own individual questionnaire (about 35 minutes). We restrict our analysis to the first step, i.e. the household grid level response, leaving aside the subsequent household and individual questionnaire responses.

Due to high attrition of former respondents (Lipps 2007), the SHP recruited a refreshment sample in 2004, representative of the Swiss residential population. For the Swiss SILC pilot,

---

[6] http://www.cmm.bristol.ac.uk/MLwiN/index.shtml

the first wave was conducted in 2004 in parallel to the SHP, by the same survey agency, also using CATI mode, with a partial overlap of the interviewers involved. The questionnaires of the SILC and the SHP are almost the same with the grid and household questionnaires almost completely, and around 60% of the questions of the individual questionnaire being identical. A random half of the pilot SILC households sampled and first interviewed in 2004 was asked to take part a second time in the subsequent year. Also in 2005 a new, smaller SILC sample was drawn and interviewed. The main difference between the two surveys – from the sample members' point of view – is twofold:[7]

1. the SHP sample members are informed about the structure, but not the exact duration of the survey. According to funds available, they are told that the survey will go on at least for another two years.
2. the sponsors of the SHP are the Swiss National Science Foundation and the University of Neuchâtel, which are both research institutions. By contrast, the Swiss Federal Statistical Office acts as both organiser and sponsor of the SILC survey. The SILC can therefore primarily considered to be government based.

Each year, after the regular fieldwork phase is finalised, an attempt is made to convince the sample members who refused to answer the survey to complete it during the refusal conversion phase. Generally, all refusals at the first stage are re-contacted unless a written refusal is sent to the Swiss Household Panel, or the centre's survey manager considers re-contacting to be hopeless.

The number of contacts on a household until final disposition (cooperation or refusal) is in principle not limited in both survey stages, but it is also at the discretion of each centre's[8] survey manager to decide not to make further attempts to contact a household. Thus some households remain "unworked" in the sense that either they cannot be contacted or that a vague or fixed appointment is still pending. The latter can be considered a (soft) refusal. These are, however, very rare cases; in the data used the maximum number of contacts in order to work a household grid is 70 during the regular fieldwork phase and 28 during refusal conversion.

[7] See Graf and Tillmann (2005) for details.
[8] The interviews are conducted from two centres: Berne, mainly responsible for the Swiss-German speaking area, and Lausanne, mainly responsible for the French and Italian speaking parts of Switzerland.

To summarise the “pre-field” variables in the model, we distinguish the following samples and survey years. First in the survey year 2005:

- the original SHP sample, then in its seventh wave (SHP I)
- the SHP refreshment sample, then in its second wave (SHP II)
- the original SILC sample, then in its second (and last) wave (SILC I).
- the SILC refreshment sample, then in its first (and last) wave (SILC II).

and in the survey year 2006:

- the original SHP sample, then in its eighth wave (SHP I)
- the SHP refreshment sample, then in its third wave (SHP II)

Because we expect both different random and fixed effects for the regular and for the refusal conversion fieldwork phase, we build separate models. Interviewers who conduct less than ten contacts during a respective fieldwork phase are omitted from the analysis. During the regular fieldwork phases, 39,194 contacts were made on 8,745 households by a total of 202 interviewers; during the refusal conversion phase, 5,919 contacts were made on 2,509 households by 69 interviewers. We can assume that interviewers who are appointed to conduct refusal conversion attempts are those who had already proved good performance with the SHP/SILC responding households during the regular phases.

Modelling Results

The results of the MCMC estimated multiple membership and the cross-classified multilevel regression models of the three interviewer performance measures are listed in Table 6 and Table 7. We discuss the modelling results of the first two performance indicators, and use the results of the optimisation indicator primarily for comparison purposes. Looking at the deviance statistics development, we realise immediately that both the models of the regular and the refusal conversion fieldwork phase improve significantly when the two covariate blocks (“prefield” and “postfield” variables) are added.[9] This effect is especially

[9] The difference of the deviance (= -2* Log Likelihood) statistics is approximately chi^2 distributed with the number of additional variables as a degree of freedom. Note that the likelihood estimate is only approximate for discrete models.

strong in the refusal avoidance model during the refusal conversion phase after the inclusion of the postfield variables block.

The first independent variable ("Swiss German Part") distinguishes the two interview centres with the language regions. As to the sample considered, and as expected, contacts in the original SHP sample (seventh/eight wave) show the highest performance, and contacts in the SILC II sample (first wave) the worst. This is due to the much longer panel membership ("panelisation") of the SHP I survey members. There are some differences between the SHP II (second/third wave) and the SILC I (second wave) samples; however, it is not the case that one of these samples performs better on both indices. This shows that the fact that the sample member knows about the structure of the survey (SHP II), or the kind of sponsor does not significantly affect contact performance. The survey year variable coefficients emphasise the importance of panelisation effects on contact performance.

It is the third models (post-field) that we are mostly interested in. In all models contact performance significantly worsens with fieldwork time. This is to be expected since the more difficult cases are usually reached later and they take longer to be worked.

Contact time of day is more important during refusal conversion; a contact during the evening shift is in general less successful, while contacting a household on afternoons has positive effects on refusal avoidance. Evening contacts affect refusal avoidance in a negative way during the regular phase. We speculate that the effects of time of interview on performance are a consequence both of reaching differently predisposed households at certain times and of the different performance quality of interviewers working the different shifts. We test this hypothesis by including the time of interview in the pre-field models, and compare interviewer and household random effects with those from the pre-field models. Surprisingly, at least in the regular phase, only the interviewer random effects decrease, while the sample case random effects remain the same. This means that the effects from different times of day are entirely due to the different performance of the interviewers working the different shifts.

The total number of contacts on a household during the regular fieldwork phase has a highly significant negative effect on contact cooperation results, and a highly significant positive effect on refusal avoidance results. This latter finding holds especially for the refusal conversion phase, and is in line with the "maintaining interaction" concept. It is probably the case that some interviewers might have followed the "stepping back" strategy. The negative effect on the cooperation indicator is most probably due to the higher difficulty to convince cases who are reluctant and thus require more contacts.

Using the same interviewer for the next contact on a sample case has no effect during the regular phase, and a positive effect during refusal conversion. It is probably not until the more problematic refusal conversion phase that respondents begin to have confidence in the interviewers given the few possible tools of communication available over the phone.

The last three variables in the third variable block measure interviewer experience made during the two panel waves considered (contact number and second year) and the total workload (total number contacts). They have rather small effects both in the regular and the refusal conversion fieldwork phases. While the contact performance slightly improves with each contact, the effect of panel experience is not consistent. Also a high total workload does not necessarily pay off.

As to the interviewer random effects, they are quite substantial in all models. We find a strong decrease of the interviewer cooperation performance random effect after the inclusion of the post-field variables during the regular phase. Probably a large portion of interviewer variance stems from the fieldwork time s/he is employed: interviewers working later are more likely to be contacting more difficult cases. Regarding refusal avoidance, the fieldwork progress and the number of contacts on a household have opposite effects on performance. Therefore, one cannot definitely say that fieldwork progress is positively correlated with a higher refusal rate of contacts. A correlation analysis confirms this: while the correlation coefficient between the number of days of fieldwork and the refusal of a contact amounts to a positive value of .09 (significant on 1%), the correlation with the cooperation index is a high negative value of -.34 (significant on 1%).

We try to further decrease the unexplained interviewer model variance by the inclusion of variables collected with the help of a paper and pencil interviewer questionnaire. This questionnaire contains, amongst other things, interviewer socio-demography and socio-economy, job satisfaction, variables on attitudes towards trying to convince or persuade a sample member to participate (de Leeuw et al. 1998), job motivation (Sonnentag and Frese 2002), perceived burden and to what degree one is able to adapt to people or situations (Japec 2005). None of these variables proved significant in the (fieldwork variables) controlled models, neither during the regular nor during the refusal conversion fieldwork period. This finding reinforces previous results that interviewer main effects do not have an impact on their performance at convincing sample members to participate in surveys.

Application Example 1: Residual Analysis of Interviewer Performance

Similarly to the work in Pickery and Loosveldt (2004), we are able to identify exceptional interviewers in a residual analysis. For the survey agency this might be an appropriate tool to assess interviewer performance in an equitable way. For example, if it turns out that an interviewer performs badly before post-fieldwork quantities are controlled, and better *after* controlling for these, it can be concluded that his/her fieldwork assignment might have produced bad fieldwork results. For example, in the cooperation model during the regular fieldwork phase, we find the following residual plots of interviewer performance after controlling for the pre-field variables (left) and after the inclusion of all variables (right):

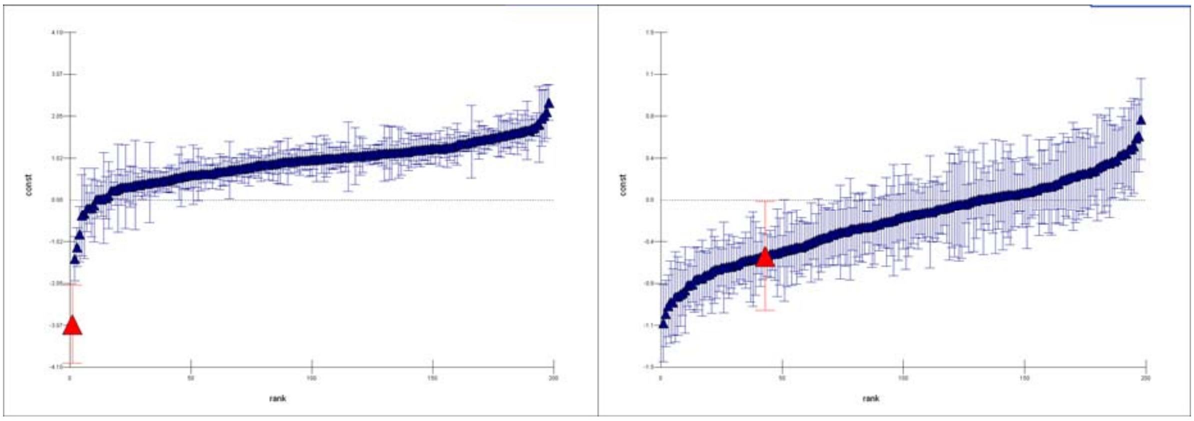


Figure 3: Residual of Interviewers in Cooperation Model, Regular Fieldwork Phase. Vertical Lines Standard Deviations. Negative Outlier Highlighted in Pre-field Model (left), Post-field Model (right).

The highlighted interviewer shows a relatively bad performance in the left panel. However, controlled for the fieldwork variables, the outlier problem almost vanishes. The reason for the highlighted interviewer to have performed so badly was his/her late fieldwork period with a difficult sample to be worked: while on average interviewers worked 43.3 (s.e.=.22) days after the fieldwork started, the interviewer concerned has a value of 143.4 (s.e.=6.0). Also the number of contacts on the households contacted by this interviewer is comparatively high. It is very likely that this interviewer joined the fieldwork staff quite late and had a high workload, and thus only obtained hard to convince households. Of course this special case is quite easy to detect and has only illustrative purposes. More sophisticated reasons might be responsible for a bad (or good) interviewer performance. The instrument described above can nevertheless help to find a reason for under/over performance using the different performance indices.

Application Example 2: Intermediate Contact Results

In this example, we consider intermediate contact results, i.e. vague or fixed appointments, as regards to the probability of completing a household, averaged by interviewer and contact result. We are interested in the question of whether interviewers achieving appointments X on a finally successfully administered household, are also successful with appointments Y. In addition, we would like to answer the question of whether interviewers who are successful with appointments X during the regular fieldwork phase, are also successful with appointments X during the refusal conversion. As above, X and Y may be fixed appointments with the target person, vague appointments with the target person, or appointments with another person in the household.

Here, we use the cooperation index. However, we do not model in a multilevel way but use the interviewer specific weighted[10] means of the household cooperation, distinguished by intermediate contact result (see Table 1 and Table 2). We calculate simple correlation coefficients between the mean household cooperation, averaged for each interviewer's intermediate contact results. In addition, we depict the number of interviewers having obtained the corresponding contact results, and the significance level of the correlation coefficient.

Correlation Coefficient N (Interviewers) Significance Level	Regular Fieldwork Phase (RE)			Refusal Conversion Phase (RF)		
Contact Result	FT	VT	FO	FT	VT	FO
Fixed Appointment with Target Person (FT)	1 200			1 64		
Vague Appointment with Target Person (VT)	.28 200 .01	1 202		.24 62 .06	1 67	
Fixed Appointment with Other Person (FO)	.09 45 .56	-.04 162 .61	1 163	.09 45 .56	.30 47 .04	1 48

[10] Similarly to the weights in the multiple membership multilevel models we use the inverse of the contact number on the household.

Table 4: Correlations of Interviewer Specific Mean Household Cooperation after Appointments. Regular and Refusal Conversion Stage Separated.

Interviewers who obtained a vague appointment during the regular phase, and "whose" households finally cooperate, tend to also be successful with fixed appointments with the target person (corr = .28). There are no further significant correlations during the regular phase. To the contrary, there is even a negative, albeit insignificant, correlation between vague appointments and fixed appointments with household members other than the target person.

There are two correlations worth mentioning for the refusal conversion phase (when there are a much smaller number of interviewers): finally successful vague appointments positively correlate with finally successful fixed appointments made both with the target person and with other persons.

To summarise, interviewers who obtained a fixed or vague appointment on a finally cooperating household, are not necessarily also successful with other appointment types. In addition, final "successes" on appointments work differently during the regular and the refusal conversion fieldwork phase.

In the following we depict the correlations which result after the same appointment type across the fieldwork phases:

Phase	Correlation Coefficient N (Interviewers) Significance Level	Regular Fieldwork Phase (RE)		
		FT	VT	FO
Refu-sal Conv. Phase	Fixed Appointment with Target Person (FT)	.06 64 .64		
	Vague Appointment with Target Person (VT)		.16 67 .19	
	Fixed Appointment with Other Person (FO)			.28 48 .06

Table 5: Correlations of Interviewer-Specific Mean Household Cooperation after Appointments, across Regular and Refusal Conversion Stages.

Regarding cooperation of households with an intermediate contact result across fieldwork phases, there are positive correlations, of which only one is significant. Interviewers who are (un)successful with households after obtaining appointments with other persons during the regular phase, are also rather (un)successful after the same contact result during the refusal conversion phase.

Summary and Conclusion

The article has investigated the question of how to measure interviewer performance and effects as to sample case participation in CATI surveys, in which sample cases are not completely assigned to single interviewers, but where several interviewers work the same phone number, through a randomised allocation of contacts. Existing approaches mostly focus on the first contact, whose result implies the highest interviewer effect. This approach makes the investigation of the performance of interviewers impossible, who work later contacts on a sample case without a final disposition after the first contact. Others measure single contact outcomes.

Both approaches suffer from the problem of how to assign values to intermediate results (mostly appointments), i.e. contact results other than a completed interview or a refusal. What is it worth if the interviewer obtains, for example, a fixed appointment with an agreed date and time, rather than a vague appointment? How should one take the outcome of a possible previous contact into account? In addition, both existing approaches do not take into consideration the complex clustering of contacts within sample cases within interviewers, which might or might not be hierarchical. To model interviewer performance effects makes complex multilevel models necessary.

In this article we propose and model two interviewer performance measures for centralised CATI surveys, built on existing theories of cooperation in CAPI surveys, in which each sample case “belongs” to one interviewer:

1. the “cooperation” index measures the binary outcome of the sample member; all contacts on a sample case after which the treated sample member *finally* cooperates are given a value of 1, and of 0 if s/he does not *finally* cooperate. The idea behind this measure is that it is not so much the individual contact outcome which is decisive, but that interviewers who work the sample case follow the common target “cooperation” of the sample member.

2. the “refusal avoidance” index measure is derived from the well known theory elaborated by Groves and Couper (1996, 1998), with the strategy of maintaining interaction with the (reluctant) sample case and of avoiding refusals, rather than trying to push a sample case and

to risk a final refusal. Binary success in this context is defined as 1 if the contact outcome is not a refusal.

In addition, we define and model a "naïve" interviewer performance measure, which is a direct conversion of the contact result into a real number:

3. the "optimisation" index: this measure directly assesses the contact result by calculating the rate of *finally* cooperating sample members, by contact result. Trivial contact results are cooperation (=1) and refusal (=0), but it can be shown by means of contact data that fixed appointments result in a higher mean number of finally cooperating sample members than vague appointments. The idea is that each interviewer tries to optimise the outcome result of the contact as to finally try to convince the sample member to participate.

In the empirical part of the paper, we model the three indices using data from two waves of two Swiss general panel surveys, distinguished by regular and refusal conversion fieldwork phases. Due to the complex clustering structure, we model the cooperation index using multiple membership multilevel models, and the refusal avoidance and the optimisation indices using cross-classified multilevel models. It first turns out that the interviewer effects during refusal conversion measured by the optimisation index are rather small. This is probably caused by defining the index as the contact results averaged over all sample cases. Therefore this index is not suitable to measure interviewer effects.

Second, we find for the two remaining indices that both fixed and random effects differ; while sample effects are comparable, fieldwork effects are sometimes quite different. However the effects are mostly consistent with the underlying theoretical concepts, e.g. "maintaining interaction" or "stepping back". We are able to substantially reduce interviewer variance by adding fieldwork variables, especially in the models which use regular fieldwork data. Importantly, we show the importance of controlling for fieldwork time in order to assess interviewer performance. This is most important when analysing the cooperation performance index.

The different results obtained for the two indices call for a more sophisticated treatment of how interview performance and effects should be measured and modelled in centralised CATI surveys, possibly also considering special survey characteristics and performance targets. A tentative application of the indices considered might be tried here: The refusal avoidance performance measure could be used in surveys in which it is of crucial importance to have as many sample members as possible turned into respondents. Examples are panel surveys, whose long-term existence depends crucially on a low attrition of the sample members. The cooperation performance measure could be used in any other random sample survey, in which one important target is to maximise response rate, and where teamwork rather than single contact results are to be improved.

The proposed measures still need to be evaluated on other surveys. The next step could be to conduct experiments in which the measures are tested in varying survey specific conditions.

References

Browne, W. (2005) *MCMC Estimation in MLwiN. Version 2.0.* Centre for Multilevel Modelling, University of Bristol

Campanelli, P. and O'Muircheartaigh C. (1999) Interviewers, Interviewer Continuity, and Panel Survey Nonresponse. *Quality & Quantity* **33**: 59-76

De Leeuw, E., Hox. J., Snijkers, G., and De Heer, W. (1998) Interviewer Opinions, Attitudes and Strategies Regarding Survey Participation and Their Effect on Response. *ZUMA Nachrichten Spezial*, August 1998: 239-248

Durand, C. (2005) Measuring Interviewer Performance in Telephone Surveys. *Quality & Quantity* **39**: 763–778

Fielding, A. and Goldstein, H. (2006) Cross-classified and Multiple Membership Structures in Multilevel Models: An Introduction and Review. *Research Report* **RR791**, department for education and skills, University of Birmingham

Graf and Tillmann (2005) Comparaison du déroulement des enquêtes PSMI, PSMII et SILC (2004-2005). Working Paper **4_05**, *Swiss Household Panel*, Neuchâtel

Groves, R., Cialdini, R. and Couper, M. (1992) Understanding the Decision to participate in a Survey. *Public Opinion Quarterly*, **56** (4): 475-93

Groves, R. and Couper, M. (1996) Contact-Level Influences on Cooperation in Face-to-Face Surveys. *J Official Statistics*, **12** (1): 63-83

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

Hox, J. and de Leeuw, E. (2002) *The Influence of Interviewers' Attitude and Behaviour on Household Survey Nonresponse: An international Comparison.* In R. Groves, D. Dillman, J. Eltinge and R. Little (eds.): Survey Nonresponse. New York: Wiley

Hox, J., de Leeuw, E. and Kreft, I. (1991) *The Effect of Interviewer and Respondent Characteristics on the Quality of Survey Data: A Multilevel Model*. In: Biemer, Groves, Lyberg, Mathiowetz and Sudman (eds.): Measurement Errors in Surveys. New York: Wiley

Hox, J., de Leeuw, E. and Snijkers, G. (1998, August). Fighting Nonresponse in Telephone Interviews; Successful Interviewer Tactics. *ZUMA Nachrichten Spezial*: 173-185

Japec, L. (2005) *Quality issues in interviewer surveys: some contributions.* Department of Statistics, Stockholm University. PhD Thesis

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

Lipps, O. (2007a) Cooperation in centralised CATI Household Panel Surveys – a Contact based multilevel Analysis to examine Interviewer, Respondent, and Fieldwork Process Effects. Mimeo, University of Neuchâtel

Mayer, T. and O'Brien, E. (2001) Interviewer refusal-aversion training to increase survey participation. American Statistical Association, Alexandria VA

O'Muircheartaigh C. and Campanelli, P. (1999) A Multilevel Exploration of the Role of Interviewers in Survey Non-Response. *J. Royal Statistical Society A* **162** (3): 437-446

Pickery, J. and Loosveldt, G. (2004) A Simultaneous Analysis of Interviewer Effects on Various Data Quality Indicators with Identification of Exceptional Interviewers. *J Official Statistics*, **20** (1): 77-89

Pickery, J., Loosveldt, G. and Carton, A. (2001) The Effects of Interviewer and Respondent Characteristics on Response Behaviour in Panel Surveys. *Sociological Methods & Research* **29** (4): 509-523

Rasbash, J., Steele, F., Browne, W. and Prosser, B. (2004) A User's Guide to MLwiN. Version 2.0. Centre for Multilevel Modelling, University of Bristol

Rendtel, U., Behr, A., Bellgardt, E., Neukirch, T., Pyy-Martikainen, M., Sisto, J., Lehtonen, R., Harms, T., Basic, E. and Marek, I. (2004) Report on Panel Effects. *CHINTEX Working Paper* **16**, European Commission

Snijders, T. and Bosker, R. (1999) *Multilevel analysis*. Newbury Park, California: Sage

Snijkers, G., Hox, J. and de Leeuw, E. (1999) Interviewer's tactics for Fighting Survey Nonresponse. *J. Official Statistics*, **15** (2): 185-198

Sonnentag, S. and Frese, M. (2002) *Performance Concepts and Performance Theory.* In: Psychological Management of Individual Performance. S. Sonnentag (ed.), Wiley & Sons

Regular Fieldwork Phase	**Binary Result for Household (1=Completed): "Cooperation" Multiple Membership Model**			**Binary Result for Contact (1=No Refusal): "Refusal Avoid." Cross classified Multilevel**			**Cooperation Probability for HH (Binomial): "Optimisation" Cross classified Multilevel**		
Model	**Only Intercept**	**+ Pre-Field**	**+ Post-Field**	**Only Intercept**	**+ Pre-Field**	**+ Post-Field**	**Only Intercept**	**+ Pre-Field**	**+ Post-Field**
N (Contacts on Households)	39,194								
N (Interviewers)	202								
Intercept	**.297**	**1.041**	**2.418**	**2.679**	**2.910**	**3.174**	**2.676**	**2.941**	**2.075**
+ Pre-Field Variables									
Swiss German Part		(.059)	.410		.155	(-.091)		(.129)	(.015)
SHP I sample		Base	Base		Base	Base		Base	Base
SHP II sample		**-.809**	**-.701**		**-.607**	**-.638**		**-.617**	**-.714**
SILC I sample		**-1.091**	**-.710**		-.625	-.447		-.641	-.650
SILC II sample		**-1.731**	**-1.525**		**-1.267**	**-1.213**		**-1.277**	**-1.199**
Survey Year 2006		.087	-.166		.544	.387		.534	.611
+ Post-Field Variables									
Number of Day of Fieldwork			**-.014**			**-.023**			**-.018**
Contact Time of Day: 9 am – 1 pm			Base			Base			Base
Contact Time of Day: 1 pm – 5 pm			(-.055)			(-.111)			(-.017)
Contact Time of Day: 5 pm – 10 pm			(.020)			-.308			(.056)
Household Contact Number									**-3.165**
Total Number of Contacts on Household			**-.065**			**.190**			**3.164**
Same Interviewer as in previous Contact			(.025)			(.146)			.210
Interviewer Contact number on Household			.001			.002			.003
Total Number of Contacts of Interviewer			(.000)			.001			-
Interviewer second year at SHP/SILC (only 2006)			.239			(-.025)			(-.090)
Previous Contact: none (fresh sample line)									Base
Previous Contact: fixed Appointment									**1.900**
Previous Contact: vague Appointment									(.047)
Previous Contact: Appointment by other Person									(.095)
Random Intercept *Sample member*	3.290[153]	3.290	3.290	3.290	3.290	3.290	3.290	3.290	3.290
Random Intercept *Interviewer*	2.251	1.433	.274	.509	.439	.375	.506	.443	.241
Deviance (MCMC) Statistic	45,362	43,526	39,750	18,586	18,145	16,349	18,597	18,146	11,235

Table 6: Household Grid Completion, MCMC Estimates, Regular Fieldwork Phase. All Coefficients "significant" (at least twice their Standard Error). All Interviewer Random Effects at least three Times their Standard Error.[154] Bold: at least 10x their s.e.. in Brackets: not significant Effects.

[153] In logit models the variance at the lowest level can be interpreted as the area under the logistic curve ($\pi^2/3 \sim 3.29$); see Snijders and Bosker (1999).
[154] See Fielding and Goldstein (2006): " … more than 3 times their standard errors. As such if it were desired to refer them to the appropriate test null distribution they would all be significantly different from zero beyond the 1% level." (p. 30).

Refusal Conversion Fieldwork Phase	**Binary Result for Household (1=Completed): "Cooperation" Multiple Membership Model**			**Binary Result for Contact (1=No Refusal): "Refusal Avoid." Cross classified Multilevel**			**Cooperation Probability for HH (Binomial): "Optimisation" Cross classified Multilevel**		
Model	**Only Intercept**	**+ Pre-Field**	**+ Post-Field**	**Only Intercept**	**+ Pre-Field**	**+ Post-Field**	**Only Intercept**	**+ Pre-Field**	**+ Post-Field**
N (Contacts on Households)	5,919								
N (Interviewers)	69								
Intercept	-.166	-.276	1.054	.593	.252	(.363)	**-1.163**	-.945	-.273
+ Pre-Field Variables									
Swiss German Part		(.199)	(.071)		.383	(.003)		(.171)	(.190)
SHP I sample		Base	Base		Base	Base		Base	Base
SHP II sample		-.512	-.490		-.183	-.302		-.546	-.409
SILC I sample		-.319	-.436		-.392	-.755		-.836	-.783
SILC II sample		-.395	-.571		-.655	-.892		-.815	-.745
Survey Year 2006		.339	(.215)		.484	.566		(.096)	(.311)
+ Post-Field Variables									
Number of Day of Fieldwork			-.012			**-.016**			-.005
Contact Time of Day: 9 am – 1 pm			Base			Base			Base
Contact Time of Day: 1 pm – 5 pm			(.175)			.922			(.210)
Contact Time of Day: 5 pm – 10 pm			-.563			-.561			(-.113)
Household Contact Number									(.023)
Total Number of contacts on Household			.027			**.773**			-.105
Same Interviewer as in previous contact			.361			.492			.260
Interviewer Contact number on Household			.001			.002			(.000)
Total Number of Contacts of Interviewer			(.000)			-.001			-
Interviewer second year at SHP/SILC (only 2006)			(.049)			-.511			(-.355)
Previous Contact: none (fresh sample line)									Base
Previous Contact: fixed Appointment									**1.885**
Previous Contact: vague Appointment									(.096)
Previous Contact: Appointment by other Person									(-.022)
Previous Contact: (soft) Refusal									-.347
Random Intercept *Sample member*	3.290	3.290	3.290	3.290	3.290	3.290	3.290	3.290	3.290
Random Intercept *Interviewer*	.856	.515	.520	.524	.306	.306	.271	.263	.289
Deviance (MCMC) Statistic	7,496	7,440	7,266	7,293	7,239	5,306	6,386	6,304	5,816

Table 7: Household Grid Completion, MCMC Estimates, Refusal Conversion Fieldwork Phase. All Coefficients "significant" (at least twice their Standard Error). Bold: at least 10x their s.e.. in Brackets: not significant Effects.

Cooperation in Centralised CATI Household Panel Surveys

– A Contact-based Multilevel Analysis to Examine Interviewer, Respondent, and Fieldwork Process Effects

Oliver Lipps [155]

In this research, we analyse the contact-specific mean of the final cooperation probability, distinguishing on the one hand between contacts with household reference persons and on the other hand with other eligible household members, between first and later contacts. Data comes from two Swiss Household Panel surveys.

The interviewer-specific variance is higher for first contacts, especially in the case of the reference person. For later contacts with the reference person, the contact-specific variance dominates. This means that interaction effects and situational factors are decisive. The contact number has negative effects on the performance of contacts with the reference person, positive in the case of other persons. Also time elapsed since the previous contact has negative effects in the case of reference persons. The result of the previous contact has strong effects, especially in the case of the reference person. These findings call for a quick completion of the household grid questionnaire, assigning the best interviewers to conducting the first contact.

While obtaining refusals has negative effects, obtaining other contact results has only weak effects on the interviewer's next contact outcome. Using the same interviewer for contacts has no positive effects.

Key words: Cross-classified; call data; random interviewer-respondent assignment.

---

[155] Swiss Foundation for Research in Social Sciences (FORS), University of Lausanne, Vidy, CH – 1015 Lausanne, oliver.lipps@fors.unil.ch.
Acknowledgments: Support by the Swiss National Science Foundation is gratefully acknowledged. I thank four anonymous reviewers for valuable comments on an earlier version of the article.

1. Introduction

In centralised telephone surveys, it is usually difficult to measure cooperation effects of respondents and interviewers because the same interviewer typically does not conduct all calls with a sample member (Lipps 2008). In the Swiss Household Panel (SHP), for instance, the interviewer might call the telephone number of a household that other interviewers have already contacted. It may be the case that a fixed appointment was agreed on, or the reference person showed some reluctance during a former contact but agreed to be called later. Thus many interviewers are possibly involved in the completion of a single household grid or a single individual questionnaire. The choice of the telephone number is performed completely at random from the pool of still uncompleted numbers at a given time (interviewer shift). This assignment allows for the separation of the effects of interviewers, respondents, and contacts, on contact outcomes in a randomised setting, thus effectively achieving an interpenetrated design.

A schematic relationship of this random assignment is depicted in Figure 4:

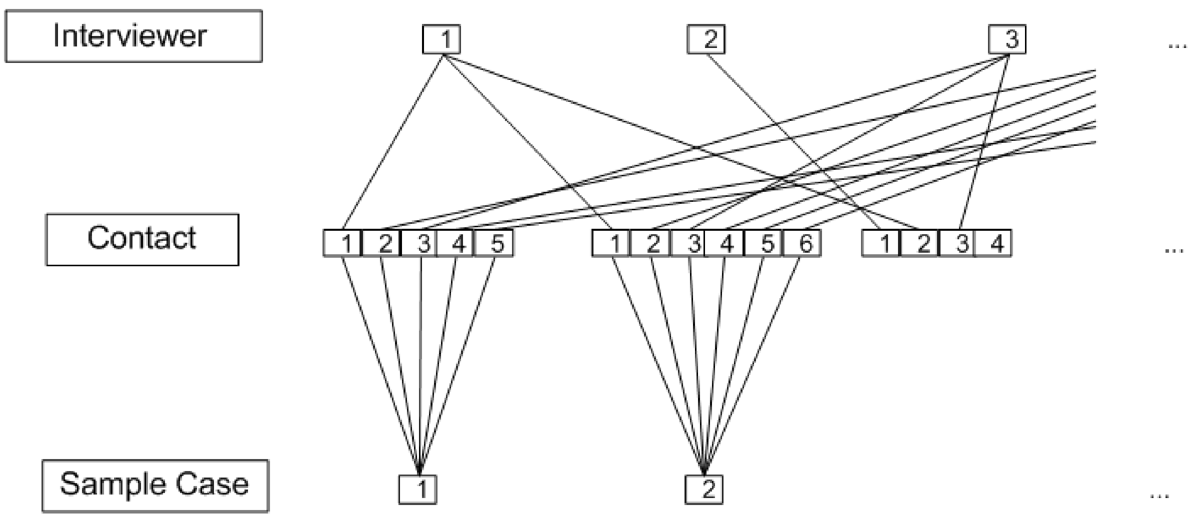


Figure 4: Interviewer-Sample Case Assignments via Contacts in Surveys with Random Assignment

The primary aim of the current article is to promote a better understanding of different effects of the actors[156] involved in the response cooperation process in a centralised CATI panel survey. The actors are the interviewers conducting the contacts, the household reference persons asked to complete the household grid questionnaire, and "other persons"[157] asked to complete their individual questionnaires. To analyse and disentangle interviewer and respondent effects, we have to investigate the contact level. Furthermore, we seek to identify covariates which are able to explain the variance on the appropriate level. This might shed light on appropriate measures to be taken in order to improve the calling procedure.

The article is organised as follows. First, we present an appropriate model of survey cooperation, also previous findings of interviewer, respondent and fieldwork characteristic effects. Next, we introduce the data and the modelling approach used, before discussing the model results. We conclude with recommendations for fieldwork organisation, namely how to assign interviewers to contacts in an efficient way.

2. Models of Cooperation

While the respondent socio-demography is significant for the outcome of the first contact with an interviewer, Groves and Couper show that it loses its predictive power for those requiring more than one contact to obtain a final disposition. The reasons are twofold: first, socio-demographic variables with higher cooperation propensities fall out of the sample due to the omission of first-contact respondents; second, for later contacts it is rather the attributes of the prior contacts which are important indicators of the cooperation likelihood (1996, p. 74). Important to note is that the socio-demographics are "fallible: they are correlates, not causes

[156] In accordance with the language of multilevel modelling we will subsequently talk about levels when appropriate.
[157] In the sequel, "other" persons are taken to be interview-eligible household members other than reference persons.

of the survey participatory behaviour" (p. 81). This is also emphasised by Stoop (2005), who specifies causes of (non)cooperation: "social isolation, social participation, …, interest in societal well-being, doing voluntary work, political interest and knowledge, …, electoral participation, the type of sponsor, and attitudes towards surveys" (p. 126).

For interviewers, as it is likely that "most of the acculturation process of producing effective interviewers occurs during training and on the job" (Groves and Couper 1998, p. 195), training, and experience seem to be important characteristics (Snijkers, Hox, and de Leeuw 1999; Hox and de Leeuw 2002). Inteviewers' attitudes towards the importance of their work and their expectations regarding difficulty in gaining cooperation seem to be significantly related to response rates (Singer, Frankel, and Glassman 1983; Lehtonen 1996). In an interviewer survey, Groves and Couper (1998, p. 209) find that positive interviewer expectations are associated with higher response rates. Also greater perceived authority and legitimacy of the sponsor might play a role (p. 206 and Cialdini 1984; Groves et al. 1992).

Groves and Couper built a theory of cooperation behaviour which heavily relies on the characteristics of the interaction between respondent and interviewer (1996, 1998). It is not so much fixed interviewer characteristics which determine the outcome of contacts. Stoop, reviewing the determinants, states that the interaction "depends on survey characteristics and fieldwork design, the social environment, characteristics of the interviewer and individual and household characteristics" (2005, pp. 55f.). Groves and Couper's concept of "maintaining interaction" (1998, pp. 37 ff.) is based on the strategy of avoiding a termination of the interaction during initial contacts. This concept includes the ability of "stepping back" (e.g. Hox, De Leeuw, and Snijkers 1998, p. 174) as one possible interviewer tactic to adequately react to initially reluctant individuals.

3. Previous Findings

Respondent effects on cooperation in panel surveys are analysed mostly for CAPI surveys. For example, while experiencing a “pleasant” interview during the first panel wave seems to be an important factor with regard to continuation (Loosveldt, Pickery, and Billiet 2002), respondents who complain about survey burden or express an intention to quit the survey are, in fact, more likely to drop out (Martin, Abreau, and Winters 2001). Campanelli and O’Muircheartaigh (2002) find that the interviewer’s subjective rating of the respondent’s cooperation in the previous wave is a good predictor for nonresponse in a future wave. Nonresponse in the following wave can be predicted with the aid of standard socio-demographic variables collected in a former wave, plus political interest and social participation measures in order to include motivational factors (e.g., Pickery, Loosveldt, and Carton 2001).

Also interviewers are an important determinant of cooperation. In face-to-face panel surveys, in order to build up confidence and trust to reduce attrition, often the same interviewer is used for the same household over many years (Schräpler 2001; Campanelli and O’Muircheartaigh 1999). Positive interviewer continuity effects are reported by Buck et al. (2003), using the German Socio-Economic Panel (SOEP) and the British Household Panel Survey (BHPS), both conducted face-to-face. However, it is not clear whether confidence between interviewer and respondent is improved because the interviewer visits the same households year after year, or because the interviewer revisits just those households with whom confidence could be successfully established. The latter seems to be the case in the U.S. Health and Retirement Survey (HRS[158]), where the field supervisors reassign interviewers to successfully interviewed households (Hill and Willis 2001). In order to analyse the trust hypothesis and to disentangle area from interviewer effects in a CAPI survey,

[158] The HRS samples only individuals 50 or more years old. In the HRS the first interview is done face-to-face; in the biennial follow-ups, mostly the telephone is used.

an interpenetrated[159] design was used in a subsample of the second wave of the British Household Panel Survey by Campanelli and O'Muircheartaigh (1999). They find considerable area and interviewer random effects, without significance of the easily measurable interviewer characteristics of the conventional sex-age type. More interestingly, "there was significant variation in the effectiveness of an interviewer continuity strategy among individual level refusals. This variability, however, could not be explained by the measurable characteristics of individuals, households, or areas [or interviewers]" (p. 73). They conclude that "interviewer continuity per se does not affect response rates directly" (Campanelli and O'Muircheartaigh 2002, p. 143). The remainder of the BHPS nevertheless shows interviewer continuity effects, which suggests that without experimental control one could come to the wrong conclusion.

As to the magnitude of interviewer effects in cross-sectional surveys, Hox, de Leeuw and Kreft (1991) calculate ρ=.02 with an insignificant variance component for interviewers in a mixed mode (telephone and face-to-face) small controlled field experiment. None of the interviewer variables are significant. The authors admit that the interviewer sample was perhaps simply too homogeneous, at least after receiving thorough training and using a detailed script to persuade respondents to participate. Pickery, Loosveldt, and Carton (2001), using the second wave of the face-to-face Belgian Election panel survey, find an interviewer intraclass correlation coefficient ρ of ρ=.044 (p. 517, Table 3). Surprisingly, the effect of the interviewer from the first wave on the refusals in the second wave is stronger than the effect of the interviewer who had to actually convince the respondent to cooperate. None of the interviewer socio-demographic or experience variables is significant. Pickery, Loosveldt, and Carton (2001) wonder whether the interviewer variability is in reality geographical variability. However, similar to O'Muircheartaigh and Campanelli (1999), they find only a small

[159] Random interviewer-respondent assignment, see Mahalanobis (1946).

geographical variability. Japec (2005) reports an interviewer intraclass correlation coefficient of ρ=.027 in the Swedish part of the 2002 face-to-face European Social Survey (ESS) for the response rate, and ρ=.048 for the refusal rate. She does not find a positive relationship between interviewer experience and response rates.

Although the number of possible stimuli in telephone surveys is smaller than in face-to-face interviews, interviewer effects can still be expected. For example, voice characteristics and speech patterns seem to play a role (Oksenberg and Cannell 1988). Interviewers may not follow directions or have different argumentation skills (Stokes and Yeh 1988; Snijkers, Hox, and de Leeuw 1999), or face different kinds and magnitudes of (perceived) burden (Japec 2005). Also existing research using data from the telephone SHP confirms this: although there are considerable interviewer effects on survey cooperation, it is not possible to substantially reduce them using available interviewer socio-demographic, attitudes, or satisfaction variables (Lipps 2006).

4. Data

We use call[160] data from the SHP, a nationwide, yearly, centralised CATI panel survey. The SHP started in 1999 with slightly more than 5,000 randomly selected households. Every year, the household reference person is required to first complete the household roster in the grid questionnaire. After the completion of the grid questionnaire, a household-related questionnaire is to be completed. Once all individuals in the household are enumerated, each household member from the age of 14 on has to complete his/her own individual questionnaire. The SHP recruited a refreshment sample in 2004, also representative of the Swiss residential population. The same year, the first wave of the Swiss pilot of the Survey on Income and Living Conditions in Switzerland (CH-SILC) was conducted in parallel to the

[160] The term call will be used for any contact attempt, whether someone was contacted or not (Stoop 2005, p. 139).

SHP, by the same survey agency. The fieldwork design for both surveys was the same, the questionnaires almost identical. Half of the pilot SILC households were surveyed a second time in the subsequent year 2005. In the wave analysed here (2005) three samples are therefore available:

- the original panel members then in their seventh wave (SHP Wave 7)
- the refreshment sample members, then in their second wave (SHP Wave 2)
- the Swiss SILC sample members, then in their second wave (SILC Wave 2).

Attrition analyses regarding the SHP confirm that, similar to those with other surveys (Groves and Couper 1998; Stoop 2005), the socially isolated drop out to a greater extent (Lipps 2006, 2007). In addition there is evidence of problems in tracking households which have moved. Attrition in the SHP occurs predominantly at two stages within the household survey process:

- when the household reference person is asked to complete the household grid.
- once the household grid is completed, when eligible individuals *other than* household reference persons are asked to complete their individual questionnaires.

In the models we included only individuals who completed their individual questionnaires in the preceding wave (2004).

5. Modelling Approach and Variables

From the considerations above, it becomes clear that first contacts with households are different from later contacts. First and later contacts therefore need separate analyses. We suspect different effects from the survey, socio-demography and attitudes, and the previous call history as between the reference person and other persons. As a result, we distinguish between contacts with the household reference person and other interview-eligible persons.

As dependent variable, we use the mean final cooperation probability, distinguished by reference persons and others, and for the first and the later contacts, respectively. We distinguish the following contact results:

Refusal (incl. broken appointment)
Vague appointment
Fixed appointment made by another person
Fixed appointment
Completed Interview

Table 8: Considered outcomes of the contacts**.**

In the lower right part of Table 9 in the appendix, the mean probabilities of a final cooperation depending on the contact result are depicted. For example, any "vague appointment" with a reference person made in the first contact leads to a final grid questionnaire completion of the current case with a probability of 71%. For other persons, also contacted the first time, the same contact outcome has an overall individual questionnaire completion probability of 81%. Because the dependent variables are probabilities we use poisson (count) models, with a log link.

Lipps (2008) uses cross-classified[161] models with a similarly defined dependent variable, but only considers contacts with reference persons and does not distinguish between first and later contacts. With respect to the fieldwork effects, he finds negative effects from later fieldwork times in the SHP. This is a typical "late case" effect.

Similar to the analysis by Pickery, Loosveldt, and Carton (2001), we first examine whether the interviewer from the 2004 Wave has an effect on the first contact result in 2005 (Table 10). Both models include first the former interviewer of the individual questionnaire (2004;

[161] See for an instructive introduction Fielding and Goldstein (2006).

upper part), then the current first contact interviewer as second level (lower part). Only the current interviewer has effects on both grid (intraclass correlation coefficient 4.7%) and other person first contact cooperation (intraclass correlation coefficient 2.0%). Contrary to the findings of Pickery, Loosveldt, and Carton (2001), the former interviewer does not have an effect on current cooperation in the telephone SHP. We thus decide to only include the current interviewer in the final first contact models. We also test the interviewer random effects against an interviewer fixed effects model, using the two interviewer experience covariates which are significant in the (respondent random effect only) model (Table 11). A Hausman test results in Prob>chi^2=.27; it is therefore safe to use the random effects model. Note that in Table 11 only variables which are significant in at least one model are listed. To estimate the models, we use the default setting implemented in the MLwiN software: the first-order Taylor approximated MQL method. Departing from this default caused nonconvergence of many models. Due to severe underdispersion, however, we relax the assumption of a poisson distribution by allowing for an extra-distributional parameter.

We build up the final models step by step, including covariates from different categories:

1. Variance components model: this model includes only the intercept. In order to separate interviewer and respondent and contact effects in the later contacts models, we build cross-classified multilevel models first, with the first level the contact, and the second levels the crossed respondents and interviewer (Rasbash et al. 2004; also Figure 4). We include as many covariates as interviewers, with random coefficients for each interviewer. All variances are constrained to have the same value.

   In the models with the inclusion of substantive covariates, we drop the interviewer random effects and end up with standard underdispersed poisson models in the case of the first contact and with hierarchical two-level models in the case of later contacts.

This simplification is due to convergence problems and to a desire not to overburden the models.

2. + Survey/ Survey Phase: here we consider whether the contact is a refusal conversion[162] attempt and we add the two survey dummies (SHP Wave 7 and SILC Wave 2). The SHP Wave 7 individuals can be expected to show a stronger panel commitment, because uninterested individuals of this (original) sample may have already refused during the previous waves. The difference between the SILC and the SHP sample is that the latter know that they are subject to a longer survey duration. Moreover the sponsors of the SHP and the SILC surveys are different: the SHP is mainly funded by the Swiss National Science Foundation, the SILC is a Eurostat project, run by the Swiss Statistical Federal Office in Switzerland. Here a scientific institution is contrasted with a federal authority, with the latter supposed to exert a higher authority.
3. + Socio-Demography: here we add all relevant individual or household characteristics already shown to be significant for attrition in the SHP (Lipps 2007). The political interest score is a combination of satisfaction with various life domains, standardised with mean 0 and standard deviation 1. We include the 2004 interviewer assessment of the likelihood that the respondent will participate in future waves, from 0="most probably not" to 3= "definitely."[163]
4. + Call history: here we add the contact number and the result of the previous contact (dropped in the case of a first contact). The latter variable is a dummy for "appointment fixed" (vs. vague appointment). We include information on whether the contact is (incidentally) made by the interviewer who interviewed the respondent in

[162] Apart from a selection of experienced interviewers who obtained additional training for the refusal conversion phase, no special design changes were made for the refusal conversion.
[163] This question is not asked in the SILC and was therefore imputed by the mean value from the SHP.

the previous year (in the case of the first contact) or did the previous contact (in the case of a later contact). We also include the number of calls already done on the household, and, in the case of a later contact, the number of days elapsed since the last contact. The interviewer within-wave learning experience variables are the number of vague or fixed appointments thus far and the number of completed interviews and refusals. These are controlled for the total number of contacts of the interviewer.

6. Modelling Results

6.1. First Contact

In the first contact variance components models, around 4.7%[164] (grid) and 2.0% (other persons) of the total variance of the contact-specific completion probability stems from the interviewers. This figure is in line with the literature. The deviance statistics[165] for both the grid and the other person models strongly increase once we drop the interviewer random intercept. Even the last grid model (+ call history) has a much higher deviance statistic than the variance components model. This shows the relevance of the interviewer variation to grid cooperation.

The survey/survey phase variables have rather strong fixed effects. Not surprisingly, all respondents in the refusal conversion phase show much lower cooperation. Also as expected, seventh-wave SHP respondents exhibit a substantially better cooperation than second-wave respondents both of the SHP and especially the SILC. This is probably mainly due to a distinct scepticism towards the European Union in Switzerland; further half of the SILC households received a written questionnaire asking income details in between their first and second survey wave. Dropping these households, the first contact outcome of the SILC sample is the same as that of the SHP wave two sample (not shown).

[164] =.020/(.020+.408)

[165] The difference of the deviances of two nested models is approximately chi^2 distributed with the number of additional variables as degree of freedom. Note that the likelihood estimate is only approximative for binomial and poisson models.

The previous year's within-household response rate has an expected strong positive effect on the contact performance, as well as some socio-demographic respondent characteristics. Neither age, nor sex, nor language (German/French-speaking part of Switzerland) plays a role. In line with the social exclusion concept, multi-adult households, those individuals with a greater political interest and higher education cooperate better. This holds, however, only for reference persons. It seems that other persons' cooperation is to some degree determined by the household reference person rather than their own characteristics. In fact, only other persons whose household grid is completed are asked to participate. A positive assessment of future participation by the 2004 interviewer has positive effects on the first contact results.

The call history only weakly affects the first contact results. Using the same interviewer as in the wave before has no effect. This could be expected as there are inconsistent effects already in face-to-face panel surveys, which offer more interviewer stimuli. More interestingly, what the interviewer has experienced before this first contact has some effect: the number of refusals already experienced has a proportionately negative influence; this effect is greater for reference persons. While the number of completed interviews has no effect for grid respondents, the number of fixed appointments seems to slightly positively affect first contact results.

6.2. Later Contacts

Roughly 26% (32%) of other persons (reference persons) considered here are only contacted twice; 17% (18%) have three contacts, 12% (12%) have four contacts, 8% (8%) five contacts etc. The maximum number of contacts amounts to 178 (58). Other persons thus are contacted more often, with a mean number of 9.2 (5.5) contacts.

For the variance components, we find that in the grid model only 1% of the total variance is due to the interviewer, 21% is due to the respondent. The rest is contact-specific variance, *within* respondent and interviewer. For the other person contact performance, the corresponding model does not converge. As for the random intercepts in the subsequent models, the grid contact effect is three to four times as high as the grid respondent effect. Regarding the other persons' random effect, respondents and contacts have about the same magnitude. This shows that the interviewer-respondent interaction quality is of particular importance for the reference person in the grid completion.

Interestingly, after inclusion of the survey and the survey phase, grid respondents in the refusal conversion phase perform only slightly worse. SHP Wave 7 respondents again show better cooperation. The effects of the socio-demographic variables are similar to those of the first contact models. While political interest is now significant for other persons, the effect of the previous wave interviewer assessment is reduced. Education is no more significant for reference persons.

After inclusion of the call history, the reference person level variance decreases to almost zero. This is mostly caused by the inclusion of the previous contact result. Other persons' contact performance is only weakly affected. Using the interviewer from the previous contact has a negative effect on the current contact with other persons. We can confirm findings from Groves and Heeringa (2006), who report a negative effect from the number of prior calls, and – with respect to the grid response – especially from the time elapsed since the last contact. Similar to the effects on the first contact result, the interviewer's negative learning experiences also apply to the later contact models: we find similar effects, though with smaller magnitudes, with respect to the grid respondent.

7. Conclusion

In order to learn more about respondent, interviewer, and fieldwork process effects on cooperation in centralised CATI household surveys, we examine the degree of cooperation on the contact level. We use data from the Swiss Household Panel (SHP) and the second wave of the Swiss part of the Survey on Income and Living Conditions (CH-SILC) pilot study, both conducted by the same fieldwork agency. The interviewer-respondent interaction is completely random in the surveys considered, thus achieving an effective interpenetrated design. We consider those response stages, which are most "critical" with respect to attrition: the household reference person when asked to complete the household grid questionnaire, and eligible individuals in the household "other" than the reference person when asked to complete their individual questionnaire. We distinguish between first and later contacts.

Using a multilevel modelling approach, we find that the interviewer effects are highest in the first contact models, especially when contacting the reference person. In later contacts, the interviewer share of the total variation is almost negligible. Generally, contact performance in the refusal conversion phase is much worse; however, only slightly negative for reference persons in later contacts.

Socio-demography and the last wave within-household response rate are more important for the reference persons' than for other persons' contact performance. Contrary to previous research, these variables are still important in later contacts with the result of the previous contact controlled for. In these later contacts, reference persons' performance increases with the number of the contact, while it decreases for other persons. Interestingly, the result of the previous contact is much more important for reference persons' than for other persons' cooperation.

We conclude that especially first contacts with reference persons should preferably have a favourable result (with a completion of the grid questionnaire or at least a fixed interview appointment) while with other persons the principle of maintaining interaction appears more important. It would probably be a good idea to let only the best interviewers do the first household contacts.

Using the same interviewer from previous contacts has no positive effects. The interviewer experience within the survey plays a role: while positive experiences like obtaining an interview do not improve the performance of future contacts, it worsens with negative experiences like obtaining a refusal. This "frustration" effect is especially pronounced in first contacts with reference persons. Also this speaks in favour of an assignment of the best interviewers to conduct the first contact with a household.

8. References

Buck, N., Nicoletti, C., McCulloch, A., and Burton, J. (2003). Report on Attrition Analysis and Item-Nonresponse. CHINTEX Working Paper #16, ISER, University of Essex.

Campanelli, P. and O'Muircheartaigh C. (1999). Interviewers, Interviewer Continuity, and Panel Survey Nonresponse. Quality & Quantity 33, 59-76.

Campanelli, P. and O'Muircheartaigh C. (2002). The Importance of Experimental Control in Testing the Impact of Interviewer Continuity on Panel Survey Nonresponse. Quality & Quantity 36, 129-144.

Cialdini, R.B. (1984). Influence: The New Psychology of Modern Persuasion. New York: Quill.

Fielding, A. and Goldstein, H. (2006). Cross-classified and Multiple Membership Structures in Multilevel Models: An Introduction and Review. Research Report RR791, Department for Education and Skills, University of Birmingham.

Groves, R. and Heeringa, S. (2006). Responsive Design for Household Surveys: Tools for Actively Controlling Survey Errors and Costs. Journal of the Royal Statistical Society, Series A, 169, 439-458.

Groves, R., Cialdini, R., and Couper, M. (1992). Understanding the Decision to Participate in a Survey. Public Opinion Quarterly, 56, 475-493.

Groves, R. and Couper, M. (1996). Contact-Level Influences on Cooperation in Face-to-Face Surveys. Journal of Official Statistics, 12, 63-83.

Groves, R. and Couper, M. (1998). Nonresponse in Household Interview Surveys. New York: Wiley.

Hill, D. and Willis, R. (2001). Reducing Panel Attrition. A Search for effective Policy Instruments. Journal of Human Resources, 36, 416-438.

Hox, J., de Leeuw, E., and Kreft, I. (1991). The Effect of Interviewer and Respondent Characteristics on the Quality of Survey Data: A Multilevel Model. In Biemer, Groves, Lyberg, Mathiowetz and Sudman (eds.): Measurement Errors in Surveys. New York: Wiley.

Hox, J., De Leeuw, E., and Snijkers, G. (1998). Fighting Nonresponse in Telephone Interviews; Successful Interviewer Tactics. ZUMA Nachrichten Spezial (August), 173-185.

Japec, L. (2005). Quality Issues in Interviewer Surveys: Some Contributions. Department of Statistics, Stockholm University. PhD Thesis.

Lehtonen, R. (1996). Interviewer Attitudes and Unit Nonresponse in Two Different Interviewing Schemes. In Laakonen, S., (ed). International Perspectives on Nonresponse. Proceeding of the Sixth International Workshop on Household Survey Nonresponse. Statistics Finland, Helsinki, 130–140.

Lipps, O. (2006). Attrition of and within Households in the Swiss Household Panel – Roles of the Reference Person, Other Persons, and the Interviewer in the Course of Time. Discussion Paper 6-06, Swiss Household Panel, Neuchâtel.

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

Lipps, O. (2008). A Note on Interviewer Performance Measures in Centralised CATI Surveys. Survey Research Methods, 2, 61-73.

Loosveldt, G., Pickery, J., and Billiet, J. (2002). Item Nonresponse as a Predictor of Unit Nonresponse in a Panel Survey. Journal of Official Statistics, 18, 545-557.

Mahalanobis, P. (1946). Recent Experiments in Statistical Sampling in the Indian Statistical Institute. Journal of the Royal Statistical Society, 109, 325-370.

Martin, E., Abreau, D., and Winters, F. (2001). Money and Motive: Effects of Incentives on Panel Attrition in the Survey of Income and Program Participation. Journal of Official Statistics, 17, 249-266.

Oksenberg, L. and Cannell, C. (1988). Effects of Interviewer Vocal Characteristics on Nonresponse. In Groves, R., Biemer, P., Lyberg, L., Massey, J., Nicholls II, W., and Waksberg, J.: Telephone Survey Methodology. New York: Wiley.

O'Muircheartaigh, C. and Campanelli, P. (1999). A Multilevel Exploration of the Role of Interviewers in Survey Non-Response. Journal of the Royal Statistical Society, Series A, 162, 437-446.

Pickery, J., Loosveldt, G., and Carton, A. (2001). The Effects of Interviewer and Respondent Characteristics on Response Behaviour in Panel Surveys. Sociological Methods & Research, 29, 509-523.

Rasbash, J., Steele, F., Browne, W., and Prosser, B. (2004). A User's Guide to MLwiN. Version 2.0. Centre for Multilevel Modelling, University of Bristol.

Singer, E., Frankel, M., and Glassman, M (1983). The Effect of Interviewer Characteristics and Expectations on Response. Public Opinion Quarterly, 47, 68-83.

Snijders, T. and Bosker, R. (1999). Multilevel Analysis. Newbury Park, California: Sage.

Snijkers, G., Hox, J., and de Leeuw, E. (1999). Interviewer's Tactics for Fighting Survey Nonresponse. Journal of Official Statistics, 15, 185-198.

Stokes, L. and Yeh, M.-Y. (1988). Searching for Causes of Interviewer Effects in Telephone Surveys. In Groves, R., Biemer, P., Lyberg, L., Massey, J., Nicholls II, W. and Waksberg, J.: Telephone Survey Methodology. New York: Wiley.

Stoop, I. (2005). The Hunt for the Last Respondent: Nonresponse in Sample Surveys. PhD Thesis, Utrecht University.

Received March 2007

Revised July 2009

Appendix: Modelling Results

Samples: SHP Wave 2, SHP Wave 7, SILC Pilot Wave				
Respondents who have been validly interviewed in				
Number of households **called**	6,422			
Number of households **contacted**	6,343			
Number of **calls**	144,093			
Thereof:				
Phone not answered	76.0 %			
Fixed appointment with person concerned	3.1 %			
Vague appointment with person concerned	9.8 %			
Fixed appointment made by another person	2.2 %			
Refusal (incl. broken appointment)	3.0 %			
Completed interview	5.5 %			
Other (nonsample calls, etc.)	.4 %			
First contacts	Other person	Grid	Final cooperation of respondent:	
Number	3,318	7,279		
Thereof:			Other person	Grid
Refusal (incl. broken appointment)	9.9 %	14.4 %	0 %	0 %
Vague appointment	40.1 %	31.2 %	81 %	71 %
Fixed appointment made by another person	20.8 %	7.3 %	88 %	72 %
Fixed appointment	14.0 %	22.0 %	95 %	90 %
Completed interview	14.4 %	25.2 %	100 %	100 %
Later contacts:			Final cooperation of respondent:	
Number	13,475	9,876		
Thereof:			Other person	Grid
Refusal (incl. broken appointment)	6.3 %	21.6 %	0 %	0 %
Vague appointment	47.3 %	41.1 %	74 %	59 %
Fixed appointment made by another person	7.9 %	8.3 %	81 %	64 %
Fixed appointment	13.0 %	7.2 %	92 %	84 %
Completed interview	25.6 %	21.8 %	100 %	100 %

Table 9: Descriptive Statistics 2005: Households, Calls, and Contacts.

Grid questionnaire	Coefficient
N (first contacts)	7,279
N (Interviewers 2004)	162
Fixed effects	
Intercept	**-.320**
Interviewer 2004 intraclass corr. Coeff. (rho)	.011
N (Interviewers 2005)	152
Fixed effects	
Intercept	**-.321**
Interviewer 2005 intraclass corr. Coeff. (rho)	.047

Other persons	Coefficient
N (first contacts)	3,318
N (Interviewers 2004)	155
Fixed effects	
Intercept	**-.234**
Interviewer 2004 intraclass corr. Coeff. (rho)	.000
N (Interviewers 2005)	153
Fixed effects	
Intercept	**-.234**
Interviewer 2005 intraclass corr. Coeff. (rho)	.020

Table 10: Completion Probability, Poisson Regressions (log link) with Underdispersion, first contact with Respondent (Grid or other Individual), Interviewer: second level, Respondent: first level.

1st contact	**Variance components model**				**d**		**+ Socio – Demography**		**+ Call history**	
N (first contacts)	Grid:7,279/ Other person: 3,318									
	Grid	**Other**	**Grid**	**Other**	**Grid**	**Other**	**Grid**	**Other**	**Grid**	**Other**
Intercept	**-.321**	**-.230**	**-.324**	**-.234**	**-.268**	**-.231**	**-.577**	**-.594**	**-.575**	-.513
+ Survey										
Refusal conversion					**-.833**	**-.839**	**-.805**	**-.806**	**-.677**	**-.818**
SHP wave 7					.075	.064	.069	.061	.060	.050
SHP wave 2					Base	Base	Base	Base	Base	Base
SILC wave 2					-.048	(-.029)	-.072	-.048	(-.028)	-.051
+ Response propensity/ Socio-demography										
Within HH response rate previous year							**.239**	.317	**.240**	.278
Political interest score							.013		.013	
Number of adults in household							.018		.018	
Higher education							.023		.023	
"Respondent takes part in next wave" (Iwer)							.027	.026	.027	.024
+ Call history										
Same interviewer as in previous interview										
Number of calls already to household										-.001
Interviewer number of fixed appointments									.003	
Interviewer number of refusals									-.007	-.001
Random intercept *Interviewer* σ^2	.020	.006	-	-	-	-	-	-	-	-
Random intercept *Respondent* σ^2	**.408**[1]	**.290**	**.461**	**.306**	**.457**	**.303**	**.454**	**.303**	**.451**	**.299**
(Under) Dispersion factor	.124	.088	.140	.093	.139	.092	.138	.092	.137	.091
Deviance statistic	3372	685	4006	744	3686	671	3604	663	3549	646

Later contacts	**Variance components model**[2]	**+ Survey/ survey phase**	**+ Socio – demography**	**+ Call history**
N (later contacts)	Grid: 13,475/ Other person: 9,876			

[1] In binomial or poisson models the variance at the lowest level is constrained to the area under the logistic curve ($\pi^2/3 \sim 3.29$); see Snijders and Bosker (1999). Due to underdispersion, we have a variance of .408= .124 (dispersion factor) * 3.29 (constraint).
[2] Cross-classified multilevel for first two models (not converging for other persons), multilevel hierarchical clustering structure (omitting interviewer random intercept) in subsequent models.

	Grid	Other	Grid	Other	Grid	Other	Gri	Other	Grid	Other
Intercept	**-.291**	(n.c.)	**-.287**	**-.235**	**-.389**	**-.274**	**-**	**-.631**	**-.481**	**-.562**
+ Survey										
Refusal conversion					-.138	**-.668**	-	**-.613**	**-.205**	**-.606**
SHP wave 7					**.120**	.095	**.116**	.094	.063	.078
SHP wave 2					Base	Base	Base	Base	Base	Base
SILC wave 2					(-.008)	.036	(-	(.031)	-.022	(.031)
+ Response propensity/ Socio-demography										
Within HH response rate previous year							**.226**	.378	**.126**	.323
Political interest score							.034	.029	.019	.027
Higher education										
Number of adults in household							.014		.008	
"Respondent takes part in next wave" (Iwer)							.016		(.008)	
+ Call history										
Contact number 2									Base	Base
Contact number 3									(-.003)	.024
Contact number 4									-.022	.040
Contact number 5+									-.043	.045
Status of last Contact (fixed vs. vague apptmt)									**.210**	.025
Same Interviewer as in previous contact										-.025
Number of calls already to household									-.001	-.001
Days since last contact									**-.003**	
Interviewer Number of fixed appointments										.001
Interviewer Number of refusals									-.001	-.001
Random intercept *Interviewer* σ^2	.003	(n.c.)	-	-	-	-	-	-	-	-
Random intercept *Respondent* σ^2	**.051**[3]	(n.c.)	**.054**	**.104**	**.045**	**.107**	**.041**	**.107**	**.000**	**.105**
Random intercept *Contact* σ^2	**.181**	(n.c.)	**.181**	**.095**	**.191**	**.099**	**.197**	**.099**	**.253**	**.105**
(Under) Dispersion factor	.055	(n.c.)	.055	.029	.058	.030	.060	.030	.077	.032
Deviance statistic	-529	(n.c.)	-383	-	-380	-3631	-478	-	-1330	-

Table 11: Completion Probability, Poisson Regressions (log link) with Underdispersion. All Coefficients "significant" (at least twice their Standard Error). Bold: at least 10x their s.e., - : not applicable/ not considered. Not listed: not significant. In brackets: not significant. Deviance statistics for poisson models are approximative.

Cross-sectional and longitudinal Interviewer and Respondent Survey Quality Effects in a CATI Panel

OLIVER LIPPS[#]

Especially in panel surveys, respondent attrition, respondent learning, and interviewer experience effects play a crucial role with respect to data quality. We examine three interview survey quality indicators in the same survey in a cross sectional as well as in a longitudinal way. In the cross sectional analysis we compare data quality in the mature original sample with that in a refreshment sample, surveyed in the same wave. Because in the same wave an interviewer survey was conducted, collecting attitudes on their socio demography, survey attitudes and burden measures, we are able to consider interviewer fixed effects as well. The longitudinal analysis gives more insight in the respondent learning effects with respect to the quality indicators investigated by considering the very same respondents across waves.

The Swiss Household Panel, a CATI survey representative of the Swiss residential population, forms an ideal modelling database: the interviewer – respondent assignment is random, both within and across waves. This design avoids possible confusion with other effects stemming from a non-random assignment of interviewers, e.g. area effects or effects from assigning the best interviewers to the hard cases. In order to separate interviewer, respondent and wave effects, we build cross-classified multilevel models.

Key words: Survey data quality, centralised CATI, random assignment, longitudinal effects, cross-classified multilevel, interviewer effects, respondent effects.

[#] Swiss Household Panel, 4, Espace de l'Europe, P.O. Box 688, CH - 2002 Neuchâtel oliver.lipps@unine.ch
Support by the Swiss National Science Foundation is gratefully acknowledged.

Introduction

Quite recently the « *Bulletin de Méthodologie Sociologique* » published an extensive literature overview on survey interviewers, with the aim to encourage research and contributions on the subject (van Meter 2005). Our paper extends existing work in at least two dimensions:

- It is based on data which uses pure random interviewer-respondent assignments
- It analyses interviewer effects with respect to *time* as an additional dimension, i.e., it examines several panel data waves on respondents and interviewers in a longitudinal way

Because interviewers are usually not randomly assigned to areas in face-to-face surveys, interviewer and area effects may easily be confused in CAPI panel surveys. E.g. Schräpler (2001), analysing income nonresponse using the German Socio Economic Panel (GSOEP), applies multilevel models to separate interviewer and respondent effects, but admits that possibly interviewer and/or area effects are confused (p. 150) with the design used by the GSOEP. A non randomised interviewer-respondent assignment makes the analysis both of respondent and interviewer effects and a clear separation of them impossible. Therefore, in the second wave of the British Household Panel Study (CAPI), a subsample was surveyed using a special "interpenetrated" design (Mahalanobis 1946), in order to be able to separate interviewer and area effects with respect to noncontact and refusal (O'Muircheartaigh and Campanelli 1999), and to interviewer continuity on nonresponse (Campanelli and O'Muircheartaigh 1999, 2002). A further difficulty arises in panel surveys like the GSOEP because the same interviewer is preferably assigned to the same respondents after waves. The reason is that building up confidence between respondent and interviewer may have positive effects on data quality and response. E.g. Spiess and Kroh (2005) confirm this for the GSOEP and Laurie et al. (1999) for the British Household Panel Survey (BHPS), however in the Australian HILDA panel, using the same interviewer after one wave has no effects on attrition (Watson and Wooden 2004).

In centralised telephone surveys with a completely random interviewer-respondent assignment, an interpenetrated design is realised by definition. However also in centralised telephone surveys interviewers are often assigned to respondents in a non-random way. E.g. in the US Health and Retirement Study, in which most interviews are being conducted by telephone, "supervisors tend to assign the best interviewers … to the most difficult respondents." (Hill and Willis 2001, p. 426). In addition, often local dialects or knowledge of the area are tried to be matched by assigning interviewers, who live(d) in the district or are knowledgeable of the area, to the respective respondents.

In panel surveys, it is especially important to examine effects of experience and burden of interviewers on respondent behaviour, without the confusion of a non-random interviewer-respondent assignment. To our knowledge a separation and analysis of interviewer, respondent and wave[1] effects in a panel survey are hitherto lacking in the literature. One reason is the mentioned non randomness of the assignments, which makes appropriate (non experimental) panel surveys very rare. Another reason is the rather complex cross-classified data structure (see Figure 1) which necessitates sophisticated modelling tools.

This article seeks to add more insight in interview quality effects in a centralised CATI panel survey, over several waves. In a first step we analyse one panel cross section during which an interviewer survey has been conducted in addition to the CATI survey. In this wave, in addition to the longitudinal respondents, a representative refreshment sample has been

[1] The number of waves in a panel survey can be considered as survey experience of (longitudinal) respondents and (longitudinal) interviewers. In reality, however, most interviewers stay with the panel only for few waves, at least in centralised telephone panel surveys.

drawn and interviewed for the first time. Thanks to the refreshment sample, we are able to control effects from attrition and aging. We like to analyse whether the respondents who are new to the panel exhibit different survey quality behaviours, controlled for possible interviewer effects. In the longitudinal analysis step, for which no interviewer characteristics are available, we are particularly interested in learning effects of respondents and experience and survey burden effects of interviewers, respectively. Also panel habitualisation effects are investigated. We also examine the effect of incidentally interviewing the same respondent after one wave, and effects of incidentally matching interviewers and respondents by age, sex, and education.

Respondent and Interviewer survey quality effects

Reasons for survey respondents not to give valid answers might be extremely different (Shoemaker 2002). A simple reason may be that the correct answer is not known (Frick and Grabka 2005), or an inadequate comprehension of the intent of the question or failure to retrieve adequate information (Groves et al. 2004), but also the sensitivity of questions and privacy or confidentiality reasons might play a role (Mayer 2001). The latter might even be partially dependent on field conditions like familiarity with the interviewer: Mensch and Kandel (1988) e.g. find that the number of prior interviewing contacts depresses drug use reporting. The respondent's cognitive trade off of giving or not a valid answer may lead to a socially desired, untrue answer or to complete nonresponse. If respondents provide acceptable answers without expending the cognitive effort necessary to give the *accurate* response, this results in "satisficing" behaviour (Krosnick 1991, Pickery and Loosveldt 1998, 2001).

Individual behaviour can be explained by means of the rational choice theory, according to which respondents of a survey behave depending on the perceived consequences. The choice depends on features of the situation with regard to own preferences (Esser 1993). E.g. an individual answers a survey if "the act of participation is expected to bring rewards that exceed the cost of participation" (Hill and Willis 2001, p. 418).[2] A critique is that for low cost decisions like answers to a survey question no complicated cost benefit calculations are made. Rather, habits are chosen, which are automatic dispositions adopted from former behaviour and experiences, and prove successful in everyday problems (Esser 1993). A slightly more reflected behaviour is to typify the situation and to specify a *frame* which contains a dominant goal characteristic of the specific situation, learnt from previous situations. The frame thus determines the actions to take in the specific situation (Esser 1993). In an interviewer survey for example, if there are effects from specific interviewer characteristics which cause special respondent reactions, specific frames are activated by the respondents.

Interviewers are a possible source of error, too, both in face-to-face and in telephone surveys. The extreme form of interviewer effect is to fabricate interviews (Schnell 1991, Diekmann 2005). More prevalent, however, are not closely followed scripts or coding procedures. Peneff (1988), analysing 45 survey interviewers for the INSEE national French statistics institute, finds discrepancies between researcher's norms concerning standardisation neutrality, and identical wording, and how questions are asked and responses constructed.

In short, the interviewer influence may affect both variation and localisation measures of the responses (Japec 2005, paper I, p. 3):

- The interviewer can influence the respondent to pay less attention to the questions
- The interviewer can influence the respondent in a certain direction

Van der Zouwen and Dijkstra (1988) show that there are basically four types of inadequate interviewer behaviour, which cause biases:

- The adaptation of questions for the interviewee by the interviewer

[2] A nice illustration of the response decision of individuals who value different aspects with different weights is given in Groves et al. (2004, p. 177).

- The interviewer's inattentiveness
- The interviewer's choice of a response
- The hinting by the interviewer to clarify the interviewee's response (see also Smit 1993, Smit, Dijkstra and van der Zouwen 1997)

On the other hand, van der Zouwen (2006) pleads for not to "over-standardize" the interviewer behaviour, leaving the interviewer some room for "repairing" respondent inadequate answers or misunderstandings. This behaviour may cause different answers depending on the interviewer's knowledge of concepts and definitions, which can be "in one of the following four cognitive states:

- Available: the concepts and definitions can be retrieved with minimal effort
- Accessible: the concepts and definitions can be retrieved with some effort
- Generatable: the concepts and definitions are not exactly known, but may be guessed using other information in memory from previous surveys
- Not available: the requested concepts and definitions are not known"

(Japec 2006, p. 34). For example the interviewer might recall a similar problem from an earlier interview and adopt the definition taken there; others might choose to record a "don't know" answer (Japec 2006).

Van der Zouwen and de Leeuw (1989) find in addition small but significant effects of the method of data collection on the quality of survey data. They report three "mode characteristics" or intervening variables which explain these mode effects:

- The "persuasion power" of a mode
- The complexity of the task for the respondent
- The degree of the control over the question-answer process

Although the number and magnitude of possible stimuli in telephone surveys are smaller than in face-to-face interviews, interviewer effects still have to be expected. E.g. voice characteristics and speech patterns seem to play a role (Oksenberg and Cannell 1988), interviewers may not follow directions or have different argumentation skills (Stokes and Yeh 1988, Snijkers et al. 1999), or face different kinds and magnitudes of (perceived) interview burden (Japec 2005). The latter stems from different workloads, a poorly planned survey administration, a lack of positive feedback and of clearly defined expectations, of getting many refusals, of trying to persuade reluctant respondents, of poorly designed questions, or of feeling that the survey topic is not important to society (Japec 2005, paper II, p. 15 f.). A "direct effect [on response quality] is if an interviewer finds it difficult to ask a respondent a sensitive question and chooses to tell the respondent that he or she does not have to answer the question" (p. 27).

Groves and Magilvy (1986) find that interviewer effects largely vary between surveys and try to reduce this instability by cumulating results over many surveys. For the central telephone surveys conducted by the University of Michigan Survey Research Center (SRC) they report a rather small interviewer intraclass correlation coefficient of around .01, which is smaller than those typically found in CAPI surveys (also Groves 1989, cited in Groves et al. 2004, p. 277). With respect to relevant covariates identified to vary with interviewer effects, Singer et al. (1983) report from an early study on the basis of a CATI survey that interviewer expectations, age and their assignment size have effects on the cooperation rates. Hox et al. (1991) are among the first to analyse interviewer effects using (adequate) multilevel models. From a literature review they confirm the interviewer effect magnitude reported earlier by Groves and Magilvy (1986): while "in well-conducted face to face interviews the intraclass correlation typically clusters around .02; in controlled telephone surveys it averages below .01" (Hox et al., 1991, p. 440), Hox et al. report an intraclass correlation of .01 for item nonresponse and .06 for response bias (acquiescence) in a small experimental controlled mixed mode study. They find "very few interviewer variables that explain significant interviewer variance" (p. 457) and conclude that while the interviewer effects were generally small, they were derived from data "employing a thoroughly pilot tested questionnaire, and interviewers who were well trained … , closely supervised, and provided with scripts for difficult situations. In large-scale surveys the field conditions may be less optimal, and differences between … interviewers may be larger" (p. 458).

In the more recent literature, in order to analyse interviewer effects on item-nonresponse using the multilevel modelling technique, mostly CAPI data has been used. Pickery and Loosveldt (1998, 2001 and 2004), analysing CAPI data from Belgium, find "interviewer effects [on data quality measures], but were not able to explain them in terms of [available] interviewer variables" (1998, p. 43). As to the extent of the effects, they report comparatively high intraclass correlation coefficients of .21 for income (non)responses, .16 for "no opinion" responses, .18 for "don't know" responses, and .04 for extreme response categories (2004, p. 83). Concerning dependent survey quality variables susceptible to interviewer effects, Pickery and Loosveldt conclude that "question difficulty and scope of the interviewer task might explain the size of the interviewer effects on item nonresponse" (2001, p. 337, see also Brick et al. 1995, for CATI surveys). Schräpler (2001) analyses the German SOEP which also mostly uses CAPI techniques, separately for several waves. He reports comparatively high interviewer intraclass correlations with respect to stereotypical response styles amounting to between .29 and .41, and to even between .66 and .71 with respect to item nonresponse to gross income questions. Furthermore, he finds respondent correlations of stereotypical response styles which amount to around .15 between waves. Schräpler (2001) concludes that these 'small' values suggest "that this respondent behaviour is not a stable personality trait over time, but a temporary habit caused by a motivational deficit" (p. 10). As to income nonresponse, he reports higher nonresponse for both female interviewers and female respondents, and that the income refusal rate is a quadratic function of the respondent's age in the SOEP. The between wave income nonresponse correlation is higher than .4 (p. 15).

Because in CAPI surveys the interviewer has a much stronger – physical - presence to the interviewee than in CATI, which works only through acoustic stimuli, telephone survey data should produce more conservative interviewer effects than face to face surveys (Pannekoek 1988). On the other hand, CATI survey based estimates are especially sensitive to interviewer effects because each interviewer usually performs many more interviews than in CAPI. This might dramatically increase the design effect (deff), which depends linearly on the sample assignment size, see Groves et al. (2004, p. 276). In addition, CATI interviewers tend to be a more homogenous population group as to their socio-demography and attitudes (Groves and Couper 1998, and Scherpenzeel 2005 for the Swiss Household Panel)

Heeb and Gmel 2001, using a CATI interviewed sample of 2,746 individuals conducted by 39 interviewers, analyse interviewer effects on alcohol consumption in Switzerland. They report an interviewer intraclass correlation coefficient of .023. This leads to a design effect deff=1.89. Andersen and Olsen (2002) analyse the Danish National Birth Cohort Study 1997-1999 which contains 12,910 CATI interviews carried out by 34 interviewers. They find little evidence for interviewer effects on answers to questions concerning smoking and alcohol consumption in the first trimester of pregnancy. Neither the interviewers' personal habits, nor their attitudes toward smoking, their alcohol consumption during pregnancy or their education, age, or parity correlate with the answers obtained. The authors admit that "training of the interviewers and continuous supervision may have contributed to this finding." (p. 95, see also Billiet and Loosveldt (1988) as to the importance of interviewer training to difficult to administered questions). Lipps (2005) analyses various item nonresponse measures simultaneously using the fifth (2003) wave of the CATI Swiss Household Panel data. Although he finds comparatively high interviewer effects, only few interviewer fixed characteristics are significant. As a consequence no clear picture of favorable (easily measurable) interviewer characteristics can be drawn.

To summarise there seem to be interviewer effects on survey quality and item nonresponse aspects, especially in face to face surveys, in terms of difficult to ask or to code questions, and in surveys with less controlled interviewers and a laxer script handling by interviewers. In CAPI panel surveys like the GSOEP, we encounter rather high correlations of response quality measures between adjacent waves. The question remains on the magnitude of interviewer and respondent variation in a longitudinal context in a CATI panel survey.

Data and dependent variables

The Swiss Household Panel (SHP) is a yearly conducted centralised CATI panel survey which started in 1999 with slightly more than 5,000 households, representative for the Swiss residential population. Questions are about household composition and socio-demographics, health, well being and attitudes, politics, social networks, and economics. Because of the survey design with a randomised interviewer-respondent assignment, we are able to disentangle interviewer, respondent and wave effects. For the longitudinal analysis, we use those respondents from the 2000 (2^{nd}) through the 2005 (7^{th}) wave, who are successfully interviewed throughout all six waves. These 2,733 respondents are interviewed by 237 interviewers over the six waves. The majority of the 237 interviewers involved stay in the panel for only one wave, as can be seen from Table 1:

Duration: number of waves	Number of Interviewers	Percent
1	132	55.7
2	64	27.0
3	23	9.7
4	13	5.5
5	3	1.3
6	2	.8

Table 1: Swiss Household Panel Interviewer stay number of waves, between 2000 and 2005

The interviewer-wave-respondent association can be schematised as follows:

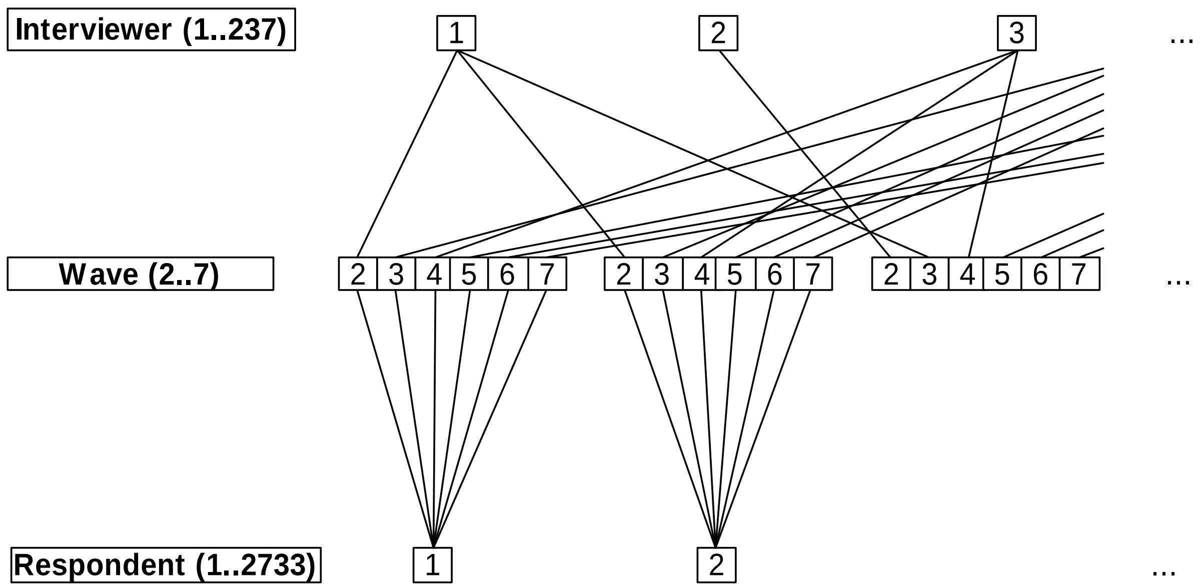


Figure 1: Pattern of interviewer-wave-respondent association in the Swiss Household Panel

All 2,733 longitudinal respondents are interviewed in all six waves 2000 through 2005. In the figure, interviewer 1 interviews both in wave 2 and wave 3 (i.e. he/she belongs to the 64 interviewers who interview in two waves, see Table 1), interviewer 2 interviews only in wave 2, interviewer 3 only in wave 4, etc.

In 2004 (wave 6), the SHP interviewed a refreshment sample in addition to the "old" panel households, equally representative of the Swiss residential population. There are only slightly more "old" panel households, then in their sixth wave, than "new" panel households in 2004. After 2000 and 2003, the SHP conducted an additional interviewer survey in 2004, and - anticipating interviewer specific differences for the "old" and the "new" sample - asked the interviewers in addition about their difficulties to convince new and old respondents to participate separately. At large, the 2004 SHP contains 8067 completed individual CATI interviews, conducted by 114 interviewers. The 2004 interviewer questionnaire was filled out and sent back by 87 interviewers. After excluding the not usable questionnaires and the interviewers who conducted 5 or less individual interviews, a sample of 81 interviewers having completed 6796 individual CATI interviews remains for the cross-sectional analysis.

Of the 81 interviewers analysed, only 19 (23 %) interviewers are male; 59 (73%) have German as their mother tongue; 22 interviewers (23%) French. Three interviewers (4 %) are also able to conduct interviews in Italian These proportions are about in line with the languages spoken by the Swiss population. 34 interviewers (41%) are still in education; accordingly, 61 (75%) are younger than 25 years. Most interviewers find it easy to ask the questions. Slightly more interviewers state higher difficulties to convince the new rather than the old households to participate. 67 (83%) of the interviewers claim that they would report their income, similarly, 45 (56%) would take part at a comparable survey.

In order to operationalise the dependent variables to examine data quality measures like giving a social desirable answer or to satisfice, we partly follow Jäckle et al. (2006), in that they also did not test the social desirability connotations of the items and use similar quality indicators of satisficing and giving socially desirable answers available in the SHP data, taking account of the Swiss context.[3]

[3] See for an example of the development of an appropriate scale for measuring social desirability Winkler, Kroh and Spiess (2006).

In particular we use for the social desirability index four questions from the politics module of the SHP individual questionnaire. These comprise the extent to which a person is interested in politics, participates in federal polls, whether one agrees in that women are generally penalised in society, and whether one is in favour of measures against these conditions. On a scale from 0 to 10, where a higher value means higher accordance, in order to quantify our social desirability indicator, we calculate the number of categories between 8 and 10. Because these measures can be assumed to correlate with social status, the respondent covariates shown in the modelling results in Table 3 are control variables with respect to the social desirability variable, rather than substantial *explanations* of this measure. This applies particularly for the education variable.

For the satisficing index to be analysed, we use 22 satisfaction and attitude questions from the modules health, work, family and social networks, and politics. According to Krosnick (1991), we consider answering “don’t know” or “no answer” a form of satisficing, because the respondent is not motivated to expend the mental effort necessary to generate a substantive answer (see also Pickery and Loosveldt 2001). We also consider the extreme category response propensities (0 or 10) to answer to the 22 questions.[4] The latter are known to be a strong predictor for unit-nonresponse in the next wave in the SHP (Lipps 2007). Lastly, we include income nonresponse, as a binary survey quality variable. Income nonresponse equals 1 if based on the information given by the respondent the total yearly personal income cannot be calculated (Gabadinho and Budowski 2002).

[4] E.g. Pickery and Loosveldt (2004) view the non-occurrence of at least one extreme answer category as proving a low interview quality. However we have the impression that a heavy use of extremes documents a certain form of satisficing. This assessment is the result of listening to a number of individual CATI interviews.

In the following Table 2, we depict the percentage of occurrence of the indicators under consideration, in the cross-sectional sample 2004 (including the refreshment sample) and the longitudinal sample, respectively:

Variable	Statistics	Mean			Standard Deviation		
	sample	2004		2000-2005	2004		2000-2005
		"old"	refresh	longitud.	"old"	refresh	longitud.
Social Desirability		.290	.279	.386	.221	.220	.274
Noanswer		.008	.014	.009	.030	.043	.033
Extreme Categories		.210	.281	.204	.189	.206	.185
Income Nonresponse		.048	.075	.064	.213	.263	.245

Table 2: Probability of Occurrence of the Quality Measures Indicators considered for the cross sectional Sample in 2004 (N=6796) and the longitudinal sample in 2000-2005 (N=16,398 observations, 2,733 individuals).

It becomes clear that the members of the "old" sample in 2004 and especially the longitudinal sample answer in a more socially desired way, give fewer noanswers and extreme answer categories, and do more often report their income. In the following, we do not further consider the noanswer index, because its occurrence does not seem to be a problem.[5]

Models

As we deal here with clustered hierarchical data, multilevel methods are chosen for modelling. They are now often applied to survey data, in which respondents (1st level) are clustered within interviewers (2nd level). Unlike the purely hierarchical three level analysis e.g. applied by Schräpler (2001) with wave as lowest level, in our longitudinal models, a cross-classified structure is necessary. This is because though respondents are clustered in interviewers and waves are clustered in respondents, respondents over waves are not clustered in interviewers. The schematic interviewer-wave-respondent association can be seen in Figure 1 above.

[5] With 1% of all 22 attitude questions not answered, each respondent has a mean number of .2 not answered questions.

According to the distribution of the variables, we model social desirability and extreme category use as poisson[6] distributed variables with a log link, and the income nonresponse variable as a logistically distributed binary variable. Due to underdispersion, we relax the standard binomial variance assumption of all indexes in all but the income nonresponse in the cross-sectional full (see below) model.

In order to construct the cross-classified longitudinal models, we first build two-level models with second level the respondents and first level the wave. In order to find the covariates of the full (intermediate) two level models, we proceed similar to Hox et al. (1991, p. 445): based on a theoretically meaningful choice (Hox et al. 1991, Pickery and Loosveldt 2001, Japec 2005), we include the significant wave dummies and select promising respondent level variables by backward OLS regression (.05 as criterion). Then, the interviewer specific variables are added and tested, again using a backward regression. The 2004 wave hierarchical cross-sectional model is built in the same way. Then these variables are entered in the two-level (interviewer-respondent in the case of the cross-sectional, respondent-wave in the case of the longitudinal) models. Unlike Hox et al., we do not test random effects of variables other than the intercept, in the two-level models, in order not to overburden the longitudinal models.

Based on the converged two-level longitudinal models, a (pseudo) third level with 1s for all observations is constructed in the longitudinal models. We define dummies for each of the interviewers, which have random coefficients on the third level. All variances on the third level are constrained to have the same value. Because of the small interviewer effects in the longitudinal variance components models, and because some of the three-level models did not converge with covariates other that the intercept, we decide not to include the interviewer level in the respective *full* longitudinal models, which contain the significant covariates.

[6] See for poisson models in a multilevel context with interviewers as second level Pickery and Loosveldt (1998).

The estimated fixed and random effects for each quality indicator for the cross-sectional and the longitudinal models are depicted in the right half of Table 3 and are divided in two parts: first the variance components models, i.e. the fixed intercept and the random intercept coefficients on all levels considered ("VC"=Variance Components") are listed. The random coefficients allow for the calculation of the variance share for each level considered. The interviewer variance proportion is depicted in the third last line of the variance components models. Secondly, the results of the "full" models which contain the fixed coefficients, together with the random parts are shown.

Modelling Results

Cross sectional 2-level models

For the 2004 cross-sectional models, after having identified the set of significant covariates, we first test whether the structure of the models for the "new" and the "old" panel members is different. We thus build two series of OLS regression models for each quality indicator: one with "new" as a dummy variable, and one using the interactions of "new" with all other significant covariates. As it turns out, the fit of the interaction models is only moderately better, but at the expense of a number of degrees of freedom (results not listed). A loglikelihood test shows that the models do not improve significantly. For the sake of parsimony we decide not to include interaction terms.

In the left of Table 3 the estimated fixed and random parts of the three cross sectional variance components models are depicted. We find a very small interviewer specific proportion of the total variance of less than 1% for the social desirability model, a medium proportion of around 5% for the extreme categories model, and a high proportion of 10% for the income nonresponse model. All three models improve with the covariates included, as can be seen from the deviance statistics.[7] Most interestingly, the sample dummy ("old" sample) is never significant. That is, controlled for the significant respondent covariates (most important age and education), the "mature" panel members do not tend to give more socially desired answers, less extreme category responses, or less income nonresponse answers than those from the refreshment sample. This means that although there is attrition selectivity due to which predominantly the politically uninterested and socially excluded respondents dropped out during the waves 2000 through 2004 (Lipps 2007), survey quality measured by a combination of these variables is not different between the two samples. As to income nonresponse at least this is surprising because Schräpler (2001) reports that income nonresponse measures seem to be a predictor for unit nonresponse in the next wave, and Loosveldt, Pickery and Billiet (2002) find this for more general item nonresponse measures.

Concerning the other significant respondent variables, the German speaking Swiss give fewer socially desired and extreme category answers. With respect to education, it could be expected that the higher educated people give more socially desired answers, according to the definition of this indicator. Similarly they report fewer extreme category answers, as do male respondents. Regarding age, we depict the effect of age categories in the case of a nonlinear effect, and the effect of the continuously measured variable age in case the effect is linear. Socially desired answers and not reporting the income increase linearly with age, while the use of extreme category answers is especially low for people aged 40-49, and increases thereafter.

More interesting is the general nonsignificance of the interviewer effects: only the interviewer specific progress has negative effects on income response, i.e., the higher the proportion of already conducted interviews in an interviewer's workload, the higher the

[7] The differences of the deviance statistics is approximately chi^2 distributed with the number of additional variables as degree of freedom. Note that the likelihood estimate is only approximative for binomial and poisson models.

income nonresponse. This may be a habitualisation or burden effect of the interviewer, in that she is increasingly less motivated to push the respondents to report their income. However the size of total interviewer workload has no effect. This points to a "late case" *respondent* effect, described later. The positive effect of the self reported impression to be able to convince new respondents (refreshment sample members) to participate at the SHP survey on the respondent use of extreme categories may be due to the fact that respondents on whom lots of efforts have to be spent to convince them to participate, give more extreme category answers ("satisfice" more). Similarly interesting, there is a time effect on socially desired answers and on income reporting: the later in the fieldwork period, the fewer socially desired answers and the higher the income nonresponse. The latter also co-varies with the relative progression of the interviews within interviewers. These effects are probably late cases effects: because the interviewer-respondent assignment is random, it takes longer to convince "difficult" cases who are consequently interviewed later. Presumably these respondents give fewer socially desired answers and refuse to report their income more often. Again, an interviewer habitualisation effect with respect to a decreasing motivation to push the respondents to give their income may also play a role.

Longitudinal models

We like to mention up front that the significance of the coefficients in the longitudinal models cannot be compared with those of the cross-sectional models due to the much higher sample size (16,398 person waves). Overall, the signs and the magnitudes of the coefficients are basically in line with those from the cross-sectional models, if applicable. Some specifics and variations must however be discussed.

We first note that the interviewer random effects are generally smaller than those in the 2004 cross sectional models. This is also a consequence of the cross-classified data structure: due to the inclusion of the wave, interviewer variance does not contain between-wave variation. Following this argumentation those who work during several waves have a reduced variation, and presumably a tendency to the mean. Moreover, given there is no between-wave variation included in the interviewer variance, there are fewer interviewers per wave (237/6~40) compared to the 2004 wave (81), probably causing a reduced variation.

The respondent specific random effect is comparatively small in the social desirability model, and high in the extreme category use model and particularly in the income nonresponse model. This shows that there is considerable between wave variation of the social desirability indicator, amounting to around 77% of the total variation. Similarly to the findings in Schräpler (2001) about stereotypical answers, this means that giving social desirable answers is a stable respondent trait only to a minor extent, but does rather depend on situational motivation and moods. In contrast the interviewer specific variation is negligible for the social desirability indicator. For extreme category answers and income nonresponse the interviewer variance proportion amounts to about 5%, the between wave variance proportion to 32% and 14%, respectively.

The latter shows that – unlike for the social desirability measure – income nonresponse can be considered a stable personality trait, because the lion's share of the total variation is due to the respondent. However, we wonder about the strong decreasing income nonresponse after wave 5, from 6.4% in 2003 to 4% in 2004. This may be due to the fact that in the first panel wave, it was communicated to the respondents that the panel is initially funded for five years. The expectation of a panel termination after five waves led to a high and selective attrition after wave five (Lipps 2006, 2007). However, the fact that also the longitudinal panel participants increase their income response shows that the longitudinal "loyal" subsample consists of persons who may have got a guilty conscience seeing that many of their cohabiters drop out in this year (Lipps 2006) and give better income reports.

With regard to the respondent coefficients, as in the cross-sectional models, they reflect correlates with respect to the indicator analysed. Only the higher prevalence of giving more extreme category answers by the middle age group is not in line with the findings from the cross-sectional models. The only significant interviewer variable is panel experience, which has negative effects on income nonresponse.

Summary and Conclusion

In the article, we explore interviewer and respondent interview quality effects in the non-experimental central CATI Swiss Household Panel (SHP) survey which uses a random interviewer-respondent and interviewer-wave assignment. Using data from the 2nd through the 7th wave, cross-classified multilevel models are chosen in order to separate respondent, interviewer and wave effects in longitudinal variance components models. For the cross-sectional models, we use standard hierarchical multilevel models.

In the longitudinal models, we find an interviewer variance share of less than 1% for the social desirability measure, and around 5% for extreme category use and income nonresponse. While the respondent variance is less than a third (23%) of the wave variance (77%) for social desirability, it doubles (63%) those of the wave variance (32%) for extreme category use and is even six times as high (82%) as the wave variance (14%) for income nonresponse. The wave specific variations basically remain the same even after including the significant wave dummies. Socially desired answering thus appears to be very variable in individuals, with a much higher intrapersonal than interpersonal variation component. Interviewer effects are negligible. It can thus be concluded that giving social desired answers is not a fixed personality trait, but rather dependent on situational factors. A much smaller intrapersonal variation is calculated for giving extreme category answers and especially income nonresponse, compared with the interpersonal variation. Therefore, intrapersonal stability applies rather with respect to giving extreme value answers (“satisfice”) and particularly not reporting income. As to the interviewer specific share of the total variance, both amount to a substantial rate of around 5%.

With respect to covariates, we mainly find the expected respondent fixed effects like age, education, and culture, i.e. the difference between the Swiss German and the rest of Switzerland. Surprisingly few interviewer covariates are significant: Interviewer experience has only some positive effects on reporting income. Incidental interviewing the same respondent by an interviewer after one wave has no effect. Also interviewer attributes like trust with respect to data protection, and various satisfaction indicators do not play any role. Similarly, the interviewer-respondent matching variables sex, age, and education had no effect, once the (respondent) main effects are controlled.

In the cross-sectional two-level models we use data from the SHP 2004 wave and the 2004 interviewer survey. Here, we find slightly higher interviewer shares of the total effects, especially with respect to income nonresponse. We were especially curious about the differences between the original (1999) sample, then in their 6^{th} wave, and the refreshment sample, first asked in 2004, and equally representative of the Swiss population. The sample discriminating dummy is significant in none of the three models. This means that, although attrition in the SHP was considerable, the survey quality indicators considered are not significantly different for the mature and the refreshment sample, *if the respondent socio demographic characteristics are controlled*. In case of social desirability, the sample dummy is not even significant in the unconditional model.

We can identify a within wave seasonal effect on socially desired answers and income nonresponse. We encounter fewer socially desired answers and more income nonresponse the later the interview is conducted in the field. An explanation is that the more difficult cases must have first convinced to participate, before the interview starts. This time effect is also in line with known effects from the so called late cases (Stoop 2005, see also Kennickell 2000).

References

Andersen, A.-M. N. and Olsen, J. (2002). Do Interviewers' Health Beliefs and Habits Modify Responses to Sensitive Questions? A Study using Data Collected from Pregnant Women by Means of Computer-assisted Telephone Interviews. *American Journal of Epidemiology* **155** (1), 95-100

Billiet, J. and Loosveldt, G. (1988) Improvement of the Quality of Responses to Factual Survey Questions by Interviewer Training. *The Public Opinion Quarterly* **52** (2), pp. 190-211

Brick, J.M., McGuinness, R., Lapham, S.J., Cahalan, M., Owends, D. and Gray, L. (1995). *Interviewer Variance in two Telephone Surveys*. SESTAT Methodological Research Compendium, WESTAT (http://srsstats.sbe.nsf.gov/)

Campanelli, P. and O'Muircheartaigh C. (1999) Interviewers, Interviewer Continuity, and Panel Survey Nonresponse. *Quality & Quantity* **33**, pp. 59-76

Campanelli, P. and O'Muircheartaigh C. (2002) The importance of experimental Control in testing the Impact of Interviewer Continuity on Panel Survey Nonresponse. *Quality & Quantity* **36**, pp. 129-144

Diekmann, A. (2005) „Betrug und Täuschung in der Wissenschaft". Datenfälschung, Diagnoseverfahren, Konsequenzen. *Swiss Journal of Sociology* **31** (1), pp. 7-29

Esser, H. (193) *Response Set: Habit, Frame or Rational Choice?* In: Krebs, D. and Schmidt, P. (eds.): New Directions in Attitude Measurement, pp. 293-314. Berlin: de Gruyter

Frick, J.R. and Grabka, M.M. (2005). Item nonresponse on income questions in panel surveys: Incidence, imputation and the impact on inequality and mobility. *Allgemeines Statistisches Archiv* **89** (1), pp. 49-61

Gabadinho, A. and Budowski, M. (2002) Construction of Income Variables for Wave 2. *Swiss Household Panel Discussion Paper* **3-02**, Neuchâtel

Groves, R.M. and Magilavy, L.J. (1986) Measuring and explaining Interviewer Effects in centralised Telephone Surveys. *Public Opinion Quarterly* **50,** pp. 251-266

Groves, R.M., Fowler, F.J., Couper, M.P., Lepkowski, J.M., Singer, E. and Tourangeau, R. (2004). *Survey Methodology.* Wiley Series in Survey Methodology.

Heeb, J. and Gmel, G. (2001). Interviewers' and respondents' effects on self-reported alcohol consumption in a Swiss health survey. *Journal of Studies on Alcohol*, July, pp. 434-442

Hill, D. and Willis, R. (2001) Reducing Panel Attrition. A Search for effective Policy Instruments. *J. Human Resources* **36** (3), pp. 416-438

Hox, J., de Leeuw, E. and Kreft, I. (1991) *The Effect of Interviewer and Respondent Characteristics on the Quality of Survey Data: A Multilevel Model.* In: Biemer, Groves, Lyberg, Mathiowetz and Sudman (eds.): Measurement Errors in Surveys. New York: Wiley

Jäckle, A., Roberts, C. and Lynn, P. (2006) Telephone versus Face-to-Face Interviewing: Mode Effects on Data Quality and likely Causes. Report on Phase II of the ESS-Gallup Mixed Mode Methodology Project. *ISER Working Paper,* **2006-41**, Colchester, University of Essex

Japec, L. (2005) *Quality issues in interviewer surveys: some contributions.* Department of Statistics, Stockholm University. PhD Thesis

Japec, L. (2006) Quality issues in interviewer surveys: some contributions. *Bulletin de Méthodologie Sociologique* **90**, pp. 26-42

Kennickell, A. (2000) *What do the "late" Cases tell us? Evidence from the 1998 Survey of Consumer Finances.* Paper prepared for the 1999 International Conference on Survey Nonresponse, Portland, Oregon

Krosnick, J.A. (1991) Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys. *Applied Cognitive Psychology* **5**, pp. 213-236

Laurie, H., Smith, R. and Scott, L. (1999) Strategies for Reducing Nonresponse in a Longitudinal Panel Survey, *Journal of Official Statistics* **15**, pp. 269-282

Lipps, O. (2005). Interviewer and Respondent Effects in CATI Data – a multi-criterion, multilevel analysis. *Swiss Household Panel Discussion Paper* **2-05**, Neuchâtel

Lipps, O. (2006) Attrition of and within Households in the Swiss Household Panel – Roles of the Reference Person, Other Persons, and the Interviewer in the Course of Time. *Swiss Household Panel Working Paper* **6-06**, Neuchâtel

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

Loosveldt, G. and Carlton, A. (1997) Evaluation of Nonresponse in the Belgian Election Panel Study '91-'95. *Proceedings of the Survey Research Methods Section*, American Statistical Association

Loosveldt, G., Pickery, J. and Billiet, J. (2002) Item Nonresponse as a Predictor of Unit Nonresponse in a Panel Survey. *Journal of Official Statistics* **18** (4), pp. 545-557

Mahalanobis, P.C. (1946) Recent Experiments in Statistical Sampling in the Indian Statistical Institute. *J Royal Statistical Society* **109**, pp. 325-378

Mayer, T.S. (2001). Interviewer Attitudes about Privacy and Confidentiality. *Research Report Series*, Statistical Research Division, U.S. Bureau of the Census

Mensch, B.S. and Kandel, D.B. (1988) Underreporting of Substance Use in a National Longitudinal Youth Cohort: Individual and Interviewer Effects. *Public Opinion Quarterly* **52** (1), pp. 100-124

Oksenberg, L. and Cannell, C. (1988) *Effects of Interviewer vocal Characteristics on Nonresponse.* In: Groves, R., Biemer, P., Lyberg, L., Massey, J., Nicholls II, W. and Waksberg, J.: Telephone Survey Methodology. New York: Wiley

O'Muircheartaigh C. and Campanelli, P. (1999) A Multilevel Exploration of the Role of Interviewers in Survey Non-Response. *J. Royal Statistical Society A* **162** (3), pp. 437-446

Pannekoek, J. (1988). Interviewer Variance in a Telephone Survey. *Journal of Official Statistics* **4** (4), pp. 375-384

Peneff, J. (1988) The Observers observed: French Survey Researchers at Work. *Social Problems* **35** (5), pp. 520-535.

Pickery, J. and Loosveldt, G. (1998) The impact of Respondent and Interviewer Characteristics on the Number of "No Opinion" answers. *Quality & Quantity* **32**, pp. 31-45

Pickery, J. and Loosveldt, G. (2001) An Exploration of Question Characteristics that Mediate Interviewer Effects on Item Nonresponse. *Journal of Official Statistics* **17** (3), pp. 337-350

Pickery, J. and Loosveldt, G. (2002). A Multilevel Multinomial Analysis of Interviewer Effects on Various Components of Unit Nonresponse. *Quality & Quantity* **36**, pp. 427-37

Pickery, J. and Loosveldt, G. (2004). A Simultaneous Analysis of Interviewer Effects on Various Data Quality Indicators with Identification on Exceptional Interviewers. *Journal of Official Statistics* **20** (1), pp. 77-89

Rasbash, J., F. Steele, W. Browne, and B. Prosser (2004). *A users guide to MLwiN version 2.0.* London, Institute of Education

Scherpenzeel, A. (2005). Interviewer Study 2004. Profiles of the MIS interviewers working for the SHP and effects on the data. *Swiss Household Panel Discussion Paper* **3-05**, Neuchâtel.

Schnell, R. (1991) Der Einfluss gefälschter Interviews auf Survey-Ergebnisse. *Zeitschrift für Soziologie* **20** (1) pp. 25-35

Schräpler, J.-P. (2001). Respondent Behaviour in Panel Studies. A case Study of the German Socio-Economic Panel (SOEP). *DIW Discussion Paper* **244**, pp. 118-156

Shoemaker, P.J., Eichholz, M. and Skewes, E.A. (2002). Item Nonresponse: Distinguishing between don't Know and Refuse. *Int. J. Public Opinion Research* **14**, pp. 193-201

Singer, E., Frankel, M.R. and Glassmann, M.B. (1983) The Effect of Interviewer Characteristics and Expectations on Response. *Public Opinion Quarterly* **47** (1), pp. 68-83

Smit, J.H. (1993) *Suggestive Questioning and Probing in Interviews - Occurrence, Causes and Effects*. Faculty of Social Cultural Sciences VU, Department of Social Research Methodology, University of Amsterdam. PhD Thesis.

Smit, J.H., Dijkstra, W. and van der Zouwen, J. (1997) Suggestive Interviewer Behaviour in Surveys: An experimental Study. *J. Official Statistics* **13** (1), pp. 19-28

Snijders, T.A.B. and Bosker, R.J. (1999). Multilevel analysis. Newbury Park, California: Sage

Snijkers, G., Hox, J. and de Leeuw, E. (1999) Interviewer's tactics for Fighting Survey Nonresponse. *J. Official Statistics* **15** (2), pp. 185-198

Spiess, M. and Kroh, M. (2005) Documentation of Sample Sizes and Panel Attrition in the German Socio Economic Panel (SOEP) 1984-2004. *Data Documentation* **6**, DIW Berlin

Stokes, L. and Yeh, M.-Y. (1988) *Searching for Causes of Interviewer Effects in Telephone Surveys*. In: Groves, R., Biemer, P., Lyberg, L., Massey, J., Nicholls II, W. and Waksberg, J.: Telephone Survey Methodology. New York: Wiley

Stoop, I. (2005) The Hunt for the last Respondent: nonresponse in Sample Surveys. PhD Thesis, Utrecht University

van der Zouwen, J. (2006) Interviewer and Survey Researcher: Mutual Dependencies. *Bulletin de Méthodologie Sociologique* **89**, pp. 49-64 van der Zouwen, J. and Dijkstra, W. (1988) Types of inadequate Interviewer Behaviour in Survey Interviews: Their Causes and Effects. *Bulletin de Méthodologie Sociologique* **18**, pp. 5-20

van der Zouwen, J. and de Leeuw, E. (1989) The Relationship Between Mode of Administration and Quality of Data in Survey Research. *Bulletin de Méthodologie Sociologique* **19**, pp. 3-14

van Meter, K. (2005) Ongoing Research: Studying Survey Interviewers: A Call for Research and an Overview. *Bulletin de Méthodologie Sociologique* **87**, pp. 51-71

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

Winkler, N., Kroh, M. and Spiess, M. (2006) Entwicklung einer deutschen Kurzskala zur zweidimensionalen Messung von sozialer Erwünschtheit. *DIW Discussion Paper* **579**, DIW Berlin

Wave	**2004 cross sectional, 6,796 Resp., 81 matched Iwers**						**2000-2005 longitudinal, 2,733 Resp, 237 Iwers**					
Dep. Variable	**Social Desirability (Poisson Regr.)**		**Extreme Categories (Poisson Regr.)**		**Income NonResponse[176] (logit Regr.)**		**Social Desirability (Poisson Regr.)**		**Extreme Categories (Poisson Regr.)**		**Income NonResponse [177] (logit Regr.)**	
	V.C.	**Full**	**V.C.**	**Full**	**V.C.**	**Full**	**V.C.**	**Full**	**V.C.**	**Full**	**V.C.**	**Full**
Intercept	**.340**	**0**	**-1.421**	**-1.376**	**-2.822**	**-5.350**	**.428**	**0**	**-1.602**	**-1.550**	**-2.553**	**-3.687**
Wave 2		-		-		-				Base		
Wave 3		-		-		-						
Wave 4		-		-		-		.030				
Wave 5		-		-		-						
Wave 6		-		-		-		.069				-.514
Wave 7		-		-		-		.043				-.333
Interview Date		-.001				.006				.001		
RESPONDENTS												
Old Sample (6th wave)								-		-		-
Swiss German		-.219		-.163				-.190		.-188		-.335
Not in Labour Force												-.552
Full time employed						.308						
Higher education		**.303**		-.152				**.187**		-.112		
Age (continuous)		**.010**				.021		.009				.031
Age 40-49 years				-.073						.394		
Age 50-59 years				.062								
Age 60-69 years				.157						.306		
Age 70+ years				.229						.125		
Male				-.110				.044		-.111		

176 N (applicable) = 5,949
177 N (applicable)=14,516

INTERVIEWERS												
Experience[178]												-.135
Same interviewer last wave		-		-		-						
Number Interviews in wave												
Relative Progression Iw (0..1)						1.564		-				.384
Higher education								-		-		-
Male								-		-		-
Age								-		-		-
Would give Income								-		-		-
Satisfaction with the SHP								-		-		-
Satisfaction with Agency								-		-		-
New Resps easy to convince[179]				.050				-		-		-
Old Resps easy to convince								-		-		-
Questions easy to ask								-		-		-
Random Effect.: Iwer σ^2	.023	.009	.030	.022	.333	.296	.006	-	.031	-	.447	-
Random Effect: Resp σ^2	**2.678**[180]	**2.497**	**.523**	**.497**	**3.445**	**3.290**	**.293**	**.279**	**.408**	**.447**	**7.832**	**7.794**
Random Effect: Wave σ^2	**-**	**-**	-	-	-	-	**.997**	**1.075**	**.207**	**.234**	**1.316**	**1.724**
Interviewer variance share	.008		.054		.097		.005		.048		.047	
(Under) Dispersion Factor	**.814**	**.759**	**.159**	**.151**	1.047		**.303**	**.324**	**.063**	**.071**	**.400**	**.524**
-2*LL (IGLS Deviance)	20,299	19,548	-2,758	-3,254	-260	-2,129	40,436	40,518	-18,423	-17,430	-4,353	-5,831

Table 3: Fixed and Random Effects, Respondent: 2nd level, Wave: 1st level. Interviewer: (pseudo) 3rd level.

All listed Coefficients "significant": at least twice their standard error (s.e.). Bold: at least 10x their s.e., "-": not applicable/ not considered. Deviance statistics for poisson or binary models are approximative.

[178] For the 2004 model: years at MIS plus 1 if SHP is not the first project, according to interviewer questionnaire. For the 2000-2005 model: years working in the panel – 1.
[179] 0=very difficult, …, 10=very easy
[180] In binomial models the variance at the lowest level is constraint to the area under the logistic curve ($\pi^2/3 \sim 3.29$); see Snijders and Bosker (1999). Due to underdispersion, we have a variance of 2.678= .814 (dispersion factor) * 3.29 (constraint).

Acknowledgements

First of all, I would like to thank my co-authors Joachim Frick, Stephen Jenkins, Dean Lillard, and Mark Wooden for giving me the great opportunity to work with them on the exciting CNEF project. Also common work with Grant Benson, with whom I analyzed and discussed the SHARE call data, was both personally and scientifically enriching. I also thank Alexandre Pollien, with whom I elaborated and worked on the ESS contact form data and Kathrin Kissau, with whom I used individual register data from the canton of Lucerne to analyze nonresponse selection effects in a cantonal population survey.

For valuable comments on various papers, I would like to thank the colleagues from the former Swiss Household Panel, especially Ursina Kuhn. I am indebted to Brian Kleiner and Marieke Voorpostel from the Swiss Foundation for Research in Social Sciences (FORS) for critical proofreading and to Karsten Hank who gave me the idea to using multilevel models during our common time at the Mannheim Research Institute for the Economics of Aging (MEA) and also for amicable assistance after my departure from the MEA. I am grateful to the referees and editors of the journals, in which the papers that are collected here are published. Last but not least, I am indebted to Max Bergman for his continuous support, comments on an earlier version and encouragement of the habilitation, and the assessors for the time and effort invested to read this work and write a report.

Financial support for this research was mainly provided through the Swiss National Science Foundation, during my occupation at the Swiss Household Panel (SHP) and FORS. In the first years (1998 to 2007), the SHP was a joint project run by the Swiss National Science Foundation, the Swiss Federal Statistical Office and the University of Neuchâtel. Since beginning of 2008 the SHP is funded by the Swiss National Science Foundation, and is run at FORS. In addition to the SHP, I use call data from the 2004 and 2005 Swiss pilot of the EU-SILC and the early Release 1 of SHARE 2004. The Swiss pilot of the EU-SILC was funded by the Swiss Confederation and conducted by the Swiss Federal Statistical Office. SHARE has been primarily funded by the European Commission through the 5th framework programme. Release 1 of SHARE is preliminary and may contain errors that will be corrected in later releases. Also call data from the first three rounds of the European Social Survey (ESS) was used. I am indebted to the Norwegian Social Science Data Services (NSD) as the data archive and distributor of the ESS data. The CNEF data used have been made available by the Cornell University, the SOEP data by the German Institute for Economic Research (DIW Berlin), and the BHPS data by the UK data archive.