Schmollers Jahrbuch 127 (2007), 627 – 654
Duncker & Humblot, Berlin

The Cross-National Equivalent File (CNEF) 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 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]

[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 go to Richard V. Burkhauser, Gaétan Garneau, Robert Schoeni and Gert G. Wagner for their comments on previous drafts of this paper.

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 that 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 / 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, they 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, 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

[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.

[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

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-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 the 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 University of Michigan, the Melbourne Institute of Applied Economic and Social Research at the University of Melbourne, and the University of Neuchâtel.

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

5 Principal investigators were Richard V. Burkhauser, then Maxwell School at Syracuse University, NY (USA) 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.

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 were income and demographic characteristics of respondents to the PSID and the 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 the SLID. This extension was natural because much of the income focus of the PSID and the SOEP was also present in the BHPS and the SLID – surveys with designs that were informed by the experiences of the PSID and the 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 / Schluter, 2003; Jenkins / Schluter / Wagner, 2003; Burkhauser / Giles / Lillard / 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 / 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 vary across time and countries (see Burkhauser / Schmeiser / 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 – 2005 for the SLID, 1999 – 2005 for the SHP, and 2001 – 2005 for the HILDA Survey. Sample sizes of individual respondents (adults and children) by year up to 2005 are listed in Appendix 2 – pooled across all six surveys, the total number well exceeds 2.57 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 col-

[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 / Schwarze, 1997; Burkhauser / Crews-Cutts / Lillard, 1999).

lects 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 if 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, judgement, 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 the LIS, reduces the burden each individual cross-national researcher faces.

The effort required to compute after-tax income varies across the different country panel surveys. In the SLID and the 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 to 1993 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 / Coutts, 1993).[9] TAXSIM has thus been used to estimate the PSID household taxes for all years in the 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. These annual tax values are combined with the annualized components of income to create a measure of household post-government income.

---

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.

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 the CNEF

All six panel surveys in the 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 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 the 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.

[10] In general, the definition of the 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.

[11] The 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.

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 Government 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 mode	Waves 1 – 5 (1968 – 1972) PAPI. Since wave 6 (1973) Mainly telephone. Since wave 26 (1993) CATI.	Waves 1 – 14 (1984 – 1997) PAPI. Since wave 2 (1985) mixed mode (face-to-face and self-completion).	Waves 1 – 9 (1991 – 1999) PAPI plus short self-completion questionnaire. Since wave 10 (2000) CAPI.	Since wave 1 (1993) CATI.	Since wave 1 (2001) PAPI plus self-completion questionnaire. Telephone used as mode of last resort.	Since wave 1 (1999) CATI.

		Since wave 15 (1998) began migrating to CAPI.	Since wave 3 (1993) use short telephone interview as last resort.			
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.	All originally sampled household 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 sample size	4,802 families.	5,921 households.	5,538 households.	15,006 households.	7,682 households.	5,074 households.
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).

To be continued next page

Table 1 (continuation)

Feature	*PSID*	*SOEP*	*BHPS*	*SLID*	*HILDA Survey*	*SHP*
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 = 68 %. East German sample, fully interviewed households = 70 %. 1998 refresher sample, includes partially interviewed households = 54 %.	Partial households = 74 %. Full households = 69 % (includes proxy interviews). 1999 Scottish / Welsh sample, partial households = 63 %. 2001 Northern Ireland sample, partial households = 69 %.	93 %	Partial households = 66 %. Full households = 59 %.	Partial Households = 49 %.

		2000 new sample, partial households = 52 %. 2006 new sample, partial households = 41 %.				
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 purpose research.	Deposited in UK Data Archive.	Currently only available via remote access or on-site access at Statistics Canada.	CD-Rom. Access restricted to bona fide researchers for specific purpose research.	CD-Rom. Access restricted to bona fide researchers for specific purpose research.

[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, 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.

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 the SLID allows proxies to be interviewed. The 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. The BHPS and SOEP increasingly interview with the assistance of a laptop computer (computer-assisted personal interviewing, or CAPI). Mixed-mode surveying takes place in the SOEP and the HILDA, with self-completion becoming more prevalent in the 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 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, the 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 the SOEP, interviews were completed with all household members at 69 and 65 percent of cases respectively.[14] Wave 1 re-

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.

sponse rates for both the BHPS and the 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.

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 the 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 the 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

[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".

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

(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 the 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.

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 the 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, 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 were surveyed again in 2002 and 2007. The PSID has also been a leader in tracking information about sample members who have died. The 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 / soep)

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 success-

fully coping with the unique event of the extension of its survey territory. In 2008, the SOEP will collect its 25th wave of data. Over the period 1994 to 2001 (i.e., in SOEP waves 11 to 18) the 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 themselves 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, the 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 / Wagner, 2006).

In 2002, and again in 2007, wealth data was collected at the individual level which – unlike many other studies, including the SOEP in 1988, 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, the SOEP micro data has been complemented by a survey of the interviewer staff in 2007, thus greatly improving the potential for analyses of interviewer-respondent effects.

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

The BHPS began its fieldwork in the autumn of 1991 and has been following and re-interviewing respondents ever since. The wave 1 sample consists of

[16] See Wagner / Frick / Schupp, 2007 for a 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 / Frick, 2005.

some 5,500 households and 10,300 individuals drawn from 250 areas of Great Britain. The BHPS was supplemented in wave 4 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 included 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 UK ECHP, 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 Household Longitudinal Study (UKHLS), 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

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

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 the 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 the 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 the SLID means that it has very high quality data on post-government income for the SLID respondents who have consented to have their tax information appended (currently about 80 percent of the SLID respondents give their consent). While the 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 the 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 the SLID. Like the 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 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 the BHPS and the 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 the 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]

---

[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.

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/Tillmann, 2004). Thus income related variable requirements from the CNEF are only partly met in the first few SHP panel waves, but some 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 the 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 Sozialwissenschaf-*

ten (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.

7. How to get Access to the CNEF Data

Data availability is influenced by national data privacy regulations. Because the original PSID data are publicly available, we are able to post the PSID-

[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).

CNEF files via the CNEF-website for public use. To access the BHPS-CNEF, SOEP-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 the SLID-CNEF files you must first be a registered CNEF user. The 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.[21]

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>*.

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 54B (5), S279 – S290.

Burkhauser, R. V. / *Frick,* J. R. / *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 60B (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.

[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.

[21] A recent change in the SLID access policy allows researchers to access the CNEF-SLID data files through the network of Research Data Centres (RDC) that are located throughout Canada. Researchers will need to go physically there while following regular rules of the Research Data Centres. The RDC network consists of 14 Research Data Centres, 6 branch RDCs and the Federal Research Centre 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, 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), Berne.

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	B, G, H, CH[22]	L11102_xxxx

To be continued next page

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

Table A1 (continuation)

Label	Data	Variable name
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[23]	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

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

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
Population Factor for W11108_xxxx	B	W11111_xxxx
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		
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
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 SOEP, 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) SOEP: 1984 – 2005 Survey Years
(H) HILDA: 2001 – 2004 Survey Years
(P) PSID: 1980 – 2003 Survey Years
(S) SLID: 1992 – 2005 Reference Years
(CH) SHP: 1999 – 2005 Survey Years

[a] Race in the BHPS and the 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 SOEP, 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	139246	17209	10666
2005	22918	25544	15627	91569	17469	8550
Total observations (Person * Years)	449.767	416.970	261.486	1.292.722	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.