

The Cross-National Equivalent File:
A product of cross-national research

Richard V. Burkhauser
Department of Policy Analysis and Management
Cornell University
Research Professor, DIW

Barbara A. Butrica
Social Security Administration

Mary C. Daly
Federal Reserve Bank of San Francisco

Dean R. Lillard
Department of Policy Analysis and Management
Cornell University
Research Affiliate, DIW

December 21, 2000

Funding for this research came from the National Institute for Aging, Program Project #1-PO1-AG09742-01, "The Well-Being of the Elderly in a Comparative Context" and from the Social Security Administration, Grant No. 10-P-98358-5-02, "The Role of Social Security Transfer Programs in Protecting Older People Transitioning into Retirement and Widowhood: A Cross-National Perspective.

Burkhauser, Richard V, Barbara A. Butrica, Mary C. Daly and Dean R. Lillard
"The Cross-National Equivalent File: A product of cross-national research" in Soziale Sicherung in einer dynamischen Gesellschaft (Social Insurance in a Dynamic Society) Irene Becker, Notburga Ott, and Gabriele Rolf (Eds.) Festschrift fuer Richard Hauser zum 65. Geburtstag Papers in Honor of the 65th Birthday of Richard Hauser.

The development of nationally representative cross-sectional survey data in most western industrialized countries enable researchers to measure the effects of alternative social policies and to compare the socio-economic characteristics of these countries' populations. While these data have enhanced cross-national research, they have also introduced new methodological issues related to the creation of comparably defined variables across countries.

Even the most sophisticated national surveys are unlikely to have cross-national comparability as a survey goal. Hence, while most national surveys use equivalent measures of age and gender, there is no international standard for measuring complex concepts like income, education, or employment. Thus, researchers interested in doing cross-national work must investigate the institutions, laws, and cultural patterns of a country in order to ensure that the variables they create for their analyses are equivalently defined across countries.

To reduce part of this burden, the Luxembourg Income Study (LIS) was developed. Over the last decade, LIS has brought together nationally representative micro-level household survey data from over 25 countries and attempted to make them comparable. LIS procedures for standardizing data are explained in Smeeding, O'Higgins, and Rainwater (1990) and in deTombeur, Milne, Warner, Gornick, and Randell (1994). This innovative standardization project greatly enhanced the ability of researchers to conduct cross-national comparative studies. However, it has two major limitations. First, LIS is based on national data sources that are confidential and not easy for the research community to access. Thus, researchers wishing to use the LIS data must access it through Luxembourg and must accept all LIS standardization rules without access to the original data sources. Second, the data are cross-sectional and hence, can not be used for dynamic analysis.

Here we describe a joint project that builds on the LIS model but does so using multiple waves of longitudinal data from Canada, Germany, Great Britain, and the United States. What distinguishes this project from other standardization projects is that the development of the data is driven by research questions. All equivalently defined variables flow from the research of experienced cross-national researchers who have developed cross-nationally comparable measures for their own analyses. Thus, the standardized data we maintain is an amalgam of the knowledge of many researchers answering a diverse set of questions.

In this paper, we first provide an overview of our equivalent data file project and the four data sets on which our standardized data are based. We then describe our methodological approach to the standardization of variables and provide a list of the comparable variables we have created. Finally, we provide an example of how these variables can be used in cross-national comparative analyses by showing how economic well being after labor force exit varies across countries.

Overview of the Cross-National Equivalent File (CNEF) 1980-1998

Researchers at Cornell University along with colleagues from the Institute for Social and Economic Research (ISER) at the University of Essex, the German Institute for Economic Research (Deutsches Institut für Wirtschaftsforschung - DIW) in Berlin, Statistics Canada in Ottawa, and the Survey Research Center at the University of Michigan have developed and tested algorithms that place information from four panel surveys into a framework of comparably defined variables for use in cross-national research. The panel surveys include the British Household Panel Survey (BHPS), the German Socio-Economic Panel (GSOEP), the Canadian Survey of Labour and Income Dynamics (SLID) and

the United States Panel Study of Income Dynamics (PSID). Using these panel surveys, researchers created a longitudinal micro-database known as the Cross-National Equivalent File 1980-1998 (CNEF). The CNEF is administered at Cornell University. It was funded from 1990 through 2000 by the National Institute on Aging and is now funded by the German Institute for Economic Research and Cornell University.

The CNEF was created to increase the accessibility and use of panel data among cross-national researchers and to assist current users of each panel survey in the creation of comparably defined cross-national variables. The CNEF unites comparably defined variables from these surveys in a single data file that can be used independently or in tandem with the original survey data. It is designed to allow cross-national researchers not experienced in panel data analysis to access simplified versions of these panels, while providing experienced panel data users with guidelines for formulating comparable variables across countries. Most importantly, the CNEF provides a set of constructed variables that are not immediately available on the original data sets. These variables include pre- and post-government household income, estimates of annual taxes paid by respondents, as well as household composition variables needed to construct household equivalence weights using commonly employed equivalence scales. Since the CNEF can be merged with the original surveys, users of the BHPS, GSOEP, SLID or PSID data can incorporate these constructed variables into existing analyses.

The CNEF currently contains data from 1980 to 1998. The data include standard demographic information, household income and its components, and individual data on employment and labor earnings. Also included are cross-sectional and longitudinal sample weights, and macroeconomic

indicators for each country. The CNEF is updated each year as additional waves of its four panels become available.

Cross-National Equivalent File Data Sources

The BHPS, GSOEP, SLID, and PSID collect similar information on family composition, income, employment, housing, and demographic characteristics. In addition, each survey collects information about a variety of personal attributes, opinions, and life choices. However, there are differences between the surveys. First, while each survey contains a core set of economic variables, each year some variables are added and deleted and some questions are reworded. The consequence is that some variables may no longer be comparable across surveys or within surveys over time. Second, the BHPS, GSOEP, SLID, and PSID differ in their data collection methods. The PSID differs in two distinct ways from the other panel studies. While all four panel studies collect information annually, since 1997 the PSID has collected data every other year. Further, the PSID interviews only the head of a household and gathers information about additional household members from this interview. In contrast, BHPS, GSOEP, and SLID interview all household members aged 16 and older. The SLID differs from the other panel studies because it has a rotating sample design. The SLID sample consists of two overlapping samples, each of which is followed for six years with the last three years of the older panel overlapping with the first three years of the newer panel. Where possible, CNEF variables are created to be equivalent across surveys and over time. Where not possible, the differences are noted in the documentation so that researchers can take them into account in their analyses.

The British Household Panel Survey (BHPS)

The BHPS began in 1991 with a sample of just over 5,500 households containing approximately 10,000 individuals. It was developed and carried out by the Institute for Social and Economic Research at the University of Essex with core funding from the University of Essex and the Economic and Social Research Council. Households were selected based on postal code of residence. The resulting sample represents the population of households with postal codes in England, Wales, and Scotland. All current BHPS families contain at least one member who was either part of the original 5,500 families or born to a member of one of these families. The same individuals are re-interviewed each successive year. Those who split-off from their original households are followed and the entire new household is included in subsequent samples. Children in original households are interviewed once they reach age 16 and continue to be interviewed when they leave their original households. For a more complete discussion of the BHPS data see Taylor, Brice, Buck and Prentice (1996).

The German Socio-Economic Panel (GSOEP)

The GSOEP is the public use version of the Socio-Economic Panel (SOEP), a longitudinal data set begun in 1984. It was developed in a former Special Research Unit '*Sonderforschungsbereich*' at the Universities of Frankfurt/Main and Mannheim in cooperation with the DIW, and initially financed by the German National Research Fund (DFG). In 1990, the DIW assumed control of the panel with funding from the Joint Federal-Land Commission for Promotion of Research Activities (Bund-Länder-Kommission für Forschungsförderung). The SOEP began with a sample of 6,000 households living in

the western states of the Federal Republic of Germany, including a disproportionate number of non-German migrant workers.

In November 1990, the eastern states of Germany were reunited with the western states of the Federal Republic of Germany. In June 1990, the DIW began a survey of families in the eastern states and merged these data with the existing SOEP population to provide a representative sample of reunited Germany. In 1995, a survey on families who immigrated into the western states of Germany after 1984 was added to the SOEP. All samples are included along with sample weights to allow for analyses that represent the population of reunited Germany.

The Department of Policy Analysis and Management at Cornell University provides an English Language version of the Public-Use file of the SOEP, the German Socio-Economic Panel (GSOEP), to researchers outside of Germany. For confidentiality reasons the GSOEP is a 95 percent sample of the full SOEP. It is this file, the GSOEP, that is used in the CNEF. For a more complete discussion of the GSOEP see Wagner, Burkhauser, and Behringer (1993). For a more complete discussion of the SOEP see Haisken-DeNew and Frick (1998).

The Panel Study of Income Dynamics (PSID)

The PSID began in 1968 with a sample of 5,000 households, representing a disproportionate number of low-income individuals. The PSID is administered by the Survey Research Center at the University of Michigan. The primary funder is the National Science Foundation but other federal government sources also contributed funds (e.g. NIA, Department of Health and Social Services, etc.) 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. Although the original sampling scheme disproportionately selected individuals from low-income families, a representative sample of the United States population can be obtained by excluding the original oversample from the data or by applying the sample weights provided with the data. Starting in 1997 the PSID began to administer its survey every other year and to no longer follow every member or related member of families in the low-income oversample population. For a more complete discussion of the PSID, data see Hill (1992).

The Survey of Labour and Income Dynamics (SLID)

The SLID began in 1993 with a sample of about 15,000 households, containing approximately 30,000 adults. It is 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. This three-year overlap was chosen to maintain continuity in the data. 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 1993 or 1996 household samples. For a more complete discussion of SLID data see the Statistics Canada website at:

www.statcan.ca/english/freepub/75f0011XIE/free.htm.

The Cross-National Equivalent File (CNEF)

The CNEF includes subsets of the original BHPS, GSOEP, SLID, and PSID data. The CNEF 1980-1998, the most recent release, includes 8 years of BHPS data from 1991 to 1998, 15 years of

GSOEP data from 1984 to 1998, 18 years of PSID data from 1980 to 1997, and 2 years of SLID data from 1993 and 1994. Additional years of SLID data, 1995 to 1998, have been put in CNEF form and are accessible via special permission from Statistic Canada. Access instructions are provided in the CNEF documentation.

Table 1 provides sample sizes for the four CNEF panels. As can be seen in Table 1, not all members of the CNEF sample provide an interview in every survey year. For instance, by 1998 the BHPS contained 21,616 individuals who had been surveyed at some time between 1991 and 1998. However, only 14,835 responded in that year. The other 6,781 did not. These 14,835 respondents lived in 6,010 households. The same basic pattern holds for the other three original data sets.

(Insert Table 1 here.)

Data development and structure

The CNEF provides cross-national researchers not experienced in panel data analysis with a simplified version of each country's panel. Thus, the CNEF is distributed as a stand alone data source, independent of the original surveys. The data are stored as rectangular data files - one for every year of survey data for each country. Each data file contains a comparable set of variables with 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), and macro-level variables (M) - and the last two 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.

A CNEF codebook identifies the components of each variable in sufficient detail to conduct rigorous analysis without referring to the original surveys. In addition, 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 completely comparable, whereas a code of “4” indicates that there is no comparable variable between the two surveys. These reliability codes are based on direct comparisons of the survey instruments as well as on knowledge of institutional differences across the countries. Table 2 provides a list of currently available CNEF variables, indication of availability in each data set, variable name, unit of analysis, and reliability code.

The CNEF also provides current users of the BHPS, GSOEP, SLID, or PSID panels with information designed to assist them in the creation of equivalently defined cross-national variables. The CNEF codebook documents the algorithms used to construct the comparable variables and includes the original survey variable names in the description. For example, the algorithm used to create the 1984 marital status variable for the United States is described in the PSID component of the code book as:

1984 algorithm: *if V30431 in (10, 20, 22) then D1110484 = V11065*

else if V30435 = 0 and V30432 Ge 18 then D1110484 = 6

else if V30435 = 0 and V30432 lt 18 then D 1110484 = 7

else if V30435 ne 0 then D1110484 = 1

else D1110484 =.M

The variables beginning with V are the variable names from the original PSID survey. Thus researchers who wish to verify that the CNEF recoding is suitable for a particular project can use these original variable names to refer to the PSID documentation distributed by the Survey Research Center at the University of Michigan. An analogous set of algorithms are provided for the other panels in the CNEF.

The CNEF includes the unique person and yearly household identifiers from the original surveys to allow users to more easily merge CNEF data with subfiles from the full BHPS, GSOEP, SLID, or PSID. Finally, in addition to providing comparably coded versions of existing survey variables, the CNEF contains 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 can not be computed without significant effort on the part of individual users.

As an example of the computational effort required to construct these types of variables consider total household income after taxes and transfers (post-government income). This variable is the sum of labor earnings, asset flows, private transfers, public transfers, imputed rental value of owner-occupied housing, and other income of all individuals in a given household minus federal income and payroll taxes. For the CNEF, this household-level variable is created and assigned to each individual in the household. Constructing post-government income for the CNEF using the SLID data and the pre-1993 PSID data is fairly straight-forward since all of the components are provided and are already in a yearly frame. Beginning in 1993, however, the PSID data no longer included tax burden estimates. To deal with this problem, CNEF staff estimate household tax burdens using the National Bureau of Economic Research tax simulation program written by Daniel Feenberg (see Feenberg and Coult,

1993). Butrica and Burkhauser (1997) provides a comparison of the new CNEF tax estimates for the PSID with the original PSID estimates for the years 1980 through 1992 together with a detailed discussion of each method of tax calculation. Once tax burdens are estimated in years from 1993 onward, post-government income can easily be created using the PSID data.

Similarly, although annual income measures are available in the BHPS, household tax burdens are not. Therefore, to construct post-government income using the BHPS data, household tax burdens must first be estimated. Since no known tax simulation program existed to estimate tax burdens under British tax law, Elena Bardasi, Stephen Jenkins, and John Rigg of the Institute for Social and Economic Research at the University of Essex wrote a tax estimation routine that assessed various income types and the appropriate tax rate (Bardasi, Jenkins, and Rigg, 1999). As with the PSID data, once tax burdens are estimated, post-government income can easily be created for the BHPS panel of the CNEF.

Constructing post-government income in the GSOEP is a much more complicated task. To begin, all income variables in the GSOEP are reported as average monthly amounts received. Thus, for cross-national comparability the first task is to annualize income by calculating the number of months in each year various types of income are received and multiplying this number by the reported average monthly amount. Next an estimated tax burden for households or individuals must be computed. Johannes Schwarze, a staff member of the DIW, developed and implemented a tax estimation routine similar in method to the one developed by the PSID staff. This tax package 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. (For a fuller discussion of this tax package see Schwarze, 1994.)

While post-government income and other created variables are not available in the original BHPS, GSOEP, SLID, and PSID surveys, they are available in the CNEF in a format that is comparable across the four data sources. Each of the created CNEF variables is the product of ongoing or completed research. Thus, variables included in the data file can be traced back to the specific papers in which they were developed. One of the benefits of this approach is that users can refer to the cited papers for examples of a variable's use and performance when considering it for analysis. The references to these papers are compiled and sent to user of the CNEF in documentation that accompanies the data.

Research with the Cross-National Equivalent File 1980-1998

To illustrate the usefulness of the CNEF to both new and experienced users we draw from Burkhauser, Lillard, and Valenti (2000) for an example of comparative research on the economic well-being of households after labor force exit in Great Britain, Germany, Canada, and the United States. Researchers studying labor force exits in different countries are confronted with a number of conceptual issues including, how to define a labor force exit, how to measure income replacement rates, and how to study relatively rare events while exploiting the longitudinal advantages of panel data. In this example we consider these issues for the four countries and describe how we take advantage of the longitudinal aspects of CNEF data.

Workers who exit the labor force risk experiencing a change in their economic well-being. In most countries, a mixture of private and public institutions exist to ameliorate the economic consequences of such exits, whether they are unforeseen labor force exits at younger ages or planned exits into retirement at older ages. On the public side, most social insurance systems provide income in the form of disability or workers' compensation benefits for younger workers who exit because of health conditions or retirement benefits for older workers who retire. Most countries also offer long-term unemployment benefits for workers of all ages as part of their social insurance system. In addition to these types of social insurance programs, which target long-term labor market workers, most countries also offer an array of means-tested welfare programs. Such programs typically provide a minimum social safety net universally or to nonworkers belonging to well-defined groups (e.g. older persons, disabled persons, lone parents, etc.). (See Aarts, Burkhauser and de Jong, 1998 for a fuller taxonomy of social welfare systems in a comparative context.)

While many studies focus on how government programs help maintain economic well-being after labor force exits, they tend to ignore the important role that private institutions play. Certainly in the United States, but also in Canada, Great Britain and increasingly in other industrialized countries, private employer fringe benefit packages protect the economic well-being of workers who exit the labor force due to redundancy, disability, or retirement. Furthermore, some households can use income from their accumulated wealth or from the increased market work of other household members to offset a given worker's lost earnings.

Studies, especially cross-national studies, of post-exit economic well-being often focus on how a given program replaces lost labor earnings. By focusing on benefits from a specific program such as

social security retirement, disability, or unemployment insurance, these studies attempt to gauge the potential post-exit income available to the households of workers who experience long-term labor market exits. The lack of comparable data, however, often restricts cross-national studies to either a comparison of a hypothetical average worker's earnings history and that worker's hypothetical subsequent social security benefits across various countries or the use of cross-sectional data from various countries to compare persons of a given age who are working relative to those who are not. (See Gruber and Wise, 1999 for an example of the former strategy and many studies using the comparable cross-sectional data from the Luxembourg Income Study for examples of the latter.)

Cross-national comparisons using hypothetical or cross-sectional data of this type are of limited value, however, in showing the actual economic risk to the household of a long-term labor market exit. These limitations arise both because the studies may fail to recognize variation across countries in the importance of social security insurance or any other government program in income replacement and because they are unable to trace the actual change in economic well-being across given households. Using data from CNEF, researchers can examine how labor earnings are replaced from both public and private sources after a worker leaves the labor force and do so by following actual and not hypothetical changes in household income.

In this example, taken from Burkhauser, Lillard, and Valenti (2000), we use the CNEF to summarize measures of economic well-being before and after a long-term labor force exit in Canada, Germany, Great Britain, and the United States. We define a long-term exit to have occurred if a man had at least three consecutive years of employment followed by at least two years of non-employment. A man is considered employed if he worked at least 52 hours for pay in a given year and is not

employed if he worked less than 52 hours for pay or received no labor earnings in the whole year.

Because labor force exits in a given year are relatively infrequent events, especially at younger ages, we use available CNEF data on age, annual work hours and labor earnings to realign our calendar year data into an event history framework to focus analysis on the workers who exit at a given age (or in a given age category) regardless of the year in which the exit occurred. This shift allows us to pool labor force exits across all years included in our data and thereby increase the sample of workers who exit in a given age category. (Very few men experienced more than one labor market exit over the period of our data.) The sample periods under study as a possible last year of long-term employment are income years 1990 through 1997 for the GSOEP, 1990 through 1996 for the PSID, 1990 through 1998 for the BHPS, and 1993 through 1998 for the SLID.

Because there are few differences across the countries before age 55, we present age specific exit risks for men age 55 and older. Figure 1 shows that the age-specific risk of an exit rises after age 55 in all four countries. With few exceptions, long-term age-specific exit rates are highest in Germany and lowest in the United States at all ages. German exit rates exceed 10 percent as early as age 58 and rise rapidly to nearly 30 percent by age 61. (Though data on the residents of the eastern states of Germany are available starting in 1990, we restrict our German sample to men with five years of continuous residence in the western states of Germany.) They approach 50 percent by age 64. Exit rates of men in the United States do not hit 10 percent until age 60 and do not hit 30 percent until age 65. Exit rates of men in Great Britain remain near 10 percent until age 62 at which point they begin to rise, peaking at age 65. Exit rates of Canadian men reach 10 percent by age 59 and remain between 10 and 20 percent until they rise sharply at ages 64 and 65.

To examine the economic consequences on the households of men who exit the labor force at different ages, we consider the amount and sources of household income before and after labor market exit for workers who exit at age 25 through 49, 50 through 61 and 62 and older. Here and throughout the rest of this example, we present income that is adjusted for household size. We divide all income figures by the square root of the number of people in the household. (There is no clear agreement on the appropriate equivalence scale to use in cross-national studies. The one proposed here is commonly used in literature. See Burkhauser, Smeeding and Merz, 1996 for a fuller discussion.)

We show that the sources and amounts of post-exit income vary both across age at exit and across country. Finally, we show that studies that focus on social security benefits to measure income replacement ratio (post-exit social security benefits divided by pre-exit labor earnings) will not only understate the actual change in household income following a long-term labor market exit but will do so disproportionately for countries like the United States and Canada where a greater portion of post-exit income comes from other sources.

This research not only provides an example of how the CNEF data can be used in cross-national research but also exemplifies the dynamic evolution of the CNEF data itself. Table 3 lists mean income from various sources in the households of men who exit the labor force at different ages. Many of these income sources, such as individual labor income (I11110\$\$), pre-government income (I11101\$\$), public transfer income (I11107\$\$), taxes (I11109\$\$), and post-government income (I11102\$\$), are directly available as CNEF variables. Others are sub-components of more aggregated CNEF income variables. For example, a focus of this study was on sources of income after labor force exits. Consequently, we wanted to consider employment-based pensions separately from the more

aggregate CNEF measure of household private transfer income (I11106\$\$) in which they are contained. To disaggregate employment-based pension income from the CNEF variable, we did two things. First, we consulted with researchers in Canada, Great Britain, Germany, and the United States about the measures of employer-based pension income available in each of the original data sources. Second, we merged those data with CNEF data and subtracted the sums from the aggregated CNEF variable measuring household private transfer income (I11106\$\$). By so doing, we preserved the additivity of our more detailed income sources to the aggregate measures included in the CNEF data. This measure of employer-based pension income will be further tested and added to the next wave of the CNEF data.

Table 3 disaggregates mean household income into its component parts. Pre-government income (i.e. household income from all non-government sources) consists of: own labor earnings (row 1), the labor earnings of all other household members (row 2), income from employer-based pensions (row 3), and all other private sources of income (row 4). Because the level of disaggregation varies across the four data sets, the discrete components of this final category include such things as interest and dividends, imputed rent from owner-occupied housing, etc.

The mean values contained in the first four rows of each column sum to mean household pre-government income (row 5). The next three rows contain sources of government income—social insurance pensions (e.g. social security or national insurance pension, etc.), other public transfers (e.g. welfare payments, payments to lone mothers, etc.), and taxes (federal income and payroll taxes). Post-government income (row 9) comprises total household income net of taxes (rows 5, 6, and 7 minus 8). A definition of each source of income used in CNEF is contained in the CNEF documentation.

Table 3 provides information on mean average post-tax, post-transfer household income (i.e. total gross household income minus all taxes) as well as by key sources of that income for two years before and two years after a labor market exit of men in the four countries in the 1990s. By definition, own labor earnings falls to near zero in the two years following labor market exit in all countries.

Table 3 shows that the sources of household income that replace lost labor earnings in the years immediately following a long-term exit from the labor market vary both within a country, by age of exit, and across our four countries. Social security plays an important role in replacing the lost earnings of men who exit the labor market after age 61 in all countries, but it is far more important in Germany than in the other countries. Social security income plays much less of a role for men who exit the labor force at middle age. Only in Germany does social security continue to play a dominant role. But other public transfer programs are important for the economic well-being of households whose men exit in middle age, except in the United States. At younger ages, other public transfers continue to be important in all countries, except in the United States. This variation in the relative importance of sources of post-exit income has important implications for interpreting various measures of income “replacement rates” across countries.

Table 4 shows how completely labor earnings are replaced, after labor force exit, by social security benefits and private pension benefits. Here the relative success of social security benefits is measured by the median ratio of total household post-exit social security benefits to pre-exit own labor earnings. The median replacement rate of private pension benefits is similarly defined as the ratio of household post-exit private pension benefits to pre-exit own labor market earnings. Finally, the replacement rate of total income is measured as the median ratio of post-government household income

in the years just after labor market exit to post-government household income in the years just before labor market exit.

In the context of cross-national comparative research it is particularly apt to focus on the social security earnings replacement because that measure is often used not only to show how much social security income replaces a typical worker's earnings in a country but is also used to infer how much a household's income is likely to fall after a long-term labor market exit. Table 4 shows that simple social security replacement rates substantially understate how much total post-government household income is available following such an exit and does so disproportionately for the United States and Canada.

The median social security replacement rate for men exiting the labor force at age 62 or older in Germany is 56.8 percent, substantially more than the 33.0 percent median social security replacement rate for men exiting at those ages in the United States. A much different story emerges, however, if one compares these results to the total income replacement rate. Using this measure, which includes all sources of income, the median total replacement rate for those same men in Germany is 78.0 percent and it is 55.9 percent for those men in the United States. The change is even greater in Canada where the median rates are 28.3 percent and 84.2 percent respectively. Higher median private pension replacement rates explain part of this difference across countries. While the median total replacement rate in the United States continues to be lower for men who exit at older ages than in the other three countries, it is less so than the median social security replacement rate.

Summary

The CNEF allows experienced and novice users to perform cross-sectional and longitudinal comparative analyses of Canada, Germany, Great Britain, and the United States. In contrast to other

cross-sectional data files, the CNEF allows researchers substantial freedom to modify the data by not only providing a number of created variables when appropriate but also providing detailed descriptions of how these variables were created. Since the creation of equivalent variables across countries is research-based, the data file is accompanied by numerous examples of each variable's use in a research application. While the CNEF still contains a relatively small subset of the variables included in the original BHPS, GSOEP, SLID, and PSID data, it is growing each year as cross-national researchers explore new areas and contribute carefully considered equivalently defined variables.

References

- Aarts, Leo J. M., Burkhauser, R.V., de Jong, P.R. “Convergence: A Comparison of European and United States Disability Policy.” In Terry Thomason, John Burton, and Douglas Hyatt (eds.). *New Approaches to Disability in the Work Place*. IRRA Research Volume, (1998), pp. 299-338.
- Bardasi, E., Jenkins, S. P. and Rigg, J.A. (1999) “Documentation for Derived Current and Annual Net Household Income Variables, BHPS Waves 1-7.” Institute for Social and Economic Research Working paper 99-25, University of Essex: Colchester.
- Burkhauser, R.V., Lillard, D.R. and Valenti, P.M. (2000), *Long-Term Labor Force Exit and Economic Well-Being: A Cross-National Comparison of Public and Private Income Support*. Paper presented at the Fourth International Conference of GSOEP Users, July 5-7, Berlin.
- Burkhauser, R.V. ,Smeeding, T.M. and Merz, J. (1996), “Relative Inequality and Poverty in Germany and the United States Using Alternative Equivalence Scales.” *The Review of Income and Wealth*. 42 (4): 381-400.

- Butrica, B.A. and 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, NY: Syracuse University.
- deTombreur, C., Milne, D., Warner, U., Gornick, J. and Randell, R. (1994). "LIS Information Guide." Luxembourg Income Study Working Paper No. 7. Walferdange, Luxembourg.
- Feenberg, D. and Coutts, E. (1993). "An Introduction to the TAXSIM Model." *The Journal of Policy Analysis and Management*. 12 (1): 189-194.
- Gruber, J. and Wise, D. A. (Eds.) (1999) *Social Security and Retirement Around the World*. Chicago: The University of Chicago Press.
- Haisken-DeNew, J. and Frick, J. (Eds.) (1998) DTC: Desktop Companion to the German Socio-Economic Panel Study (GSOEP), Version 2.2, German Institute for Economic Research, Berlin.
- Hill, M. (1992), *The Panel Study of Income Dynamics: A User's Guide*. Beverly Hills, CA.: Sage Publications.

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. NY: Syracuse University.

Smeeding, T.M., M.O'Higgins, and L. Rainwater. (1990). *Poverty, Inequality, and the Distribution of Income in an International Context: Initial Research from the Luxembourg Income Study (LIS) Project*. London: Wheatsheaf Book and Washington, D.C.: The Urban Institute.

Taylor, M.F., Brice, J., Buck, N., and Prentice, E. (Eds.)(1996) *British Household Panel Survey User Manual Volume A: Introduction, Technical Report and Appendices*. Colchester: University of Essex.

Wagner, G.G., Burkhauser, R.V. and Behringer, F. (1993), "The English Language Public Use File of the German Socio-Economic Panel". *Journal of Human Resources*, 28 (2): 429-433.

Table 1. Cross-sectional and longitudinal sample size for the BHPS, GSOEP, SLID, and PSID

Year	BHPS				GSOEP			
	Individuals			Households	Individuals			Households
	Respondents	Non-Respondents	Total		Respondents	Non-Respondents	Total	
1980	---	---	---	---	---	---	---	---
1981	---	---	---	---	---	---	---	---
1982	---	---	---	---	---	---	---	---
1983	---	---	---	---	---	---	---	---
1984	---	---	---	---	15,321	15,918	31,239	5,624
1985	---	---	---	---	13,731	17,508	31,239	5,053
1986	---	---	---	---	13,079	18,160	31,239	4,831
1987	---	---	---	---	12,856	18,383	31,239	4,771
1988	---	---	---	---	12,212	19,027	31,239	4,571
1989	---	---	---	---	11,783	19,456	31,239	4,445
1990	---	---	---	---	17,304	13,935	31,239	6,472
1991	13,780	7,836	21,616	5,509	16,909	14,330	31,239	6,358
1992	13,151	8,465	21,616	5,227	16,504	14,725	31,239	6,326
1993	13,104	8,512	21,616	5,232	16,176	15,063	31,239	6,298
1994	12,851	8,765	21,616	5,127	16,503	14,736	31,239	6,442
1995	12,549	9,067	21,616	5,034	16,973	14,266	31,239	6,605
1996	12,720	8,896	21,616	5,066	16,558	14,681	31,239	6,525
1997	14,385	7,231	21,616	6,092	16,186	15,053	31,239	6,442
1998	14,835	6,781	21,616	6,010	17,758	13,481	31,239	7,274
Year	PSID				SLID			
	Individuals			Households	Individuals			Households
	Respondents	Non-Respondents	Total		Respondents	Non-Respondents	Total	
1980	18,921	31,900	50,821	6,533	---	---	---	---
1981	18,949	31,872	50,821	6,620	---	---	---	---
1982	19,173	31,648	50,821	6,742	---	---	---	---
1983	19,425	31,396	50,821	6,852	---	---	---	---
1984	19,511	31,310	50,821	6,918	---	---	---	---
1985	19,718	31,103	50,821	7,032	---	---	---	---
1986	19,553	31,268	50,821	7,018	---	---	---	---
1987	19,579	31,242	50,821	7,061	---	---	---	---
1988	19,617	31,204	50,821	7,114	---	---	---	---
1989	19,600	31,221	50,821	7,114	---	---	---	---
1990	19,871	30,950	50,821	7,328	---	---	---	---
1991	19,880	30,941	50,821	7,375	---	---	---	---
1992	20,264	30,557	50,821	7,561	--	--	--	--
1993	21,359	29,462	50,821	7,873	41,902	292	42,194	15,891
1994	23,488	27,333	50,821	8,659	40,095	3,625	43,720	15,601
1995	23,130	27,691	50,821	8,570	82,669	5,568	88,237	31,872
1996	22,986	27,835	50,821	8,517	83,172	8,556	91,728	32,392
1997	19,100	31,721	50,821	6,748	80,905	13,322	94,227	31,726
1998	---	---	---	---	80,964	15,032	85,996	32,023

Source: Cross-National Equivalent File 1980-1998, SLID Equivalent Files 1995-1998

Table 2. Variables included in the Cross-National Equivalent File 1980-1998

Variable Description	Data set	Variable Name	Unit of Analysis	Reliability Code
Demographics:				
Age of Individual	B, G, P, S	D1110191 - D1110198	I	1
Sex of Individual	B, G, P, S	D11102LL	I	1
Race of Household Head ^a	B, P, S	D1110391 - D1110398	H, I	4
Marital Status of Individual	B, G, P, S	D1110491 - D1110498	I	1
Relationship to Household Head	B, G, P, S	D1110591 - D1110598	I	1
Number of Persons in Household	B, G, P, S	D1110691 - D1110698	I	1
Number of Children in Household	B, G, P, S	D1110791 - D1110798	I	1
Education With Respect to High School	G, P, S	D1110891 - D1110898	I	2
Number of Years of Education	G, P, S	D1110991 - D1110998	I	2
Disability Status of Individual	B, P, S	D1111091 - D1111098	I	1
Satisfaction With Health	B, G, P, S	D1111191 - D1111198	I	1
Employment:				
Annual Work Hours of Individual	B, G, P, S	E1110191 - E1110198	I	1
Impute Annual Work Hours of Individual	B	E1120191 - E1120198	I	1
Employment Status of Individual	B, G, P, S	E1110291 - E1110298	I	1
Employment Level of Individual	B, G, P, S	E1110391 - E1110398	I	1
Primary Activity of Individual	B, G, P, S	E1110491 - E1110498	I	1
Occupation of Individual	B, G, P, S	E1110591 - E1110598	I	1
1 Digit Industry Code of Individual	B, G, P, S	E1110691 - E1110698	I	1
2 Digit Industry Code of Individual	B, G, P, S	E1110791 - E1110798	I	1
Household Composition:				
Number HH members age 0-14	B, G, P, S	H1110191 - H1110198	H	1
Number HH members age 15-18	B, G, P, S	H1110291 - H1110298	H	1
Number HH members age 0-1	B, G, P, S	H1110391 - H1110398	H	1
Number HH members age 2-4	B, G, P, S	H1110491 - H1110498	H	1
Number HH members age 5-7	B, G, P, S	H1110591 - H1110598	H	1
Number HH members age 8-10	B, G, P, S	H1110691 - H1110698	H	1
Number HH members age 11-12	B, G, P, S	H1110791 - H1110798	H	1
Number HH members age 13-15	B, G, P, S	H1110891 - H1110898	H	1
Number HH members age 16-18	B, G, P, S	H1110991 - H1110998	H	1
Number HH members age 19+ or 16-18 and indep.	B, G, P, S	H1111091 - H1111098	H	1
Indicator - Head in HH	B, G, P, S	H1111191 - H1111198	H	1
Indicator - Wife/spouse in HH	B, G, P, S	H1111291 - H1111298	H	1
Yearly Income:				
Household Pre-Government Income	B, G, P, S	I1110191 - I1110198	H	1
Household Post-Government Income	B, G, P, S	I1110291 - I1110298	H	1
Household Labor Income	B, G, P, S	I1110391 - I1110398	H	1
Household Asset Income	B, G, P, S	I1110491 - I1110498	H	2
Household Imputed Rental Value	B, G, P, S	I1110591 - I1110598	H	2
Household Private Transfers	B, G, P, S	I1110691 - I1110698	H	1
Household Public Transfers	B, G, P, S	I1110791 - I1110798	H	1
Household Social Security Pensions	B, G, P, S	I1110891 - I1110898	H	1
Total Household Taxes	B, G, P, S	I1110991 - I1110998	H	2
Individual Labor Earnings	B, G, P, S	I1111091 - I1111098	I	1
Household Federal Taxes	G, P	I1111191 - I1111198	H	1
Household Social Security Taxes	G, P	I1111291 - I1111298	H	1
Household Post-Government Income (TAXSIM)	P	I1111391 - I1111398	H	1
Total Household Taxes (TAXSIM)	P	I1111491 - I1111498	H	1
Household State Taxes (TAXSIM)	P	I1111591 - I1111598	H	1
Household Federal Taxes (TAXSIM)	P	I1111691 - I1111698	H	1

Impute Household Post-Government Income	B, G, S	I1120291 - I1120298	H	1
Impute Household Labor Income	B, G, S	I1120391 - I1120398	H	1
Impute Household Asset Income	B, G, S	I1120491 - I1120498	H	1
Impute Household Imputed Rental Value	B, G, S	I1120591 - I1120598	H	1
Impute Household Private Transfers	B, G, S	I1120691 - I1120698	H	1
Impute Household Public Transfers	B, G, S	I1120791 - I1120798	H	1
Impute Household Social Security Pensions	B, G, S	I1120891 - I1120898	H	1
Impute Total Household Taxes	G, S	I1120991 - I1120998	H	1
Impute Individual Labor Earnings	B, G, S	I1121091 - I1121098	I	1

Location:

Area of Residence ^b	B, G, P, S	L1110191 - L1110198	H	1
Region of Residence	B, G	L1110291 - L1120198	H	1

Weights:

Individual Weight	B, G, P, S	W1110191 - W111019	I	1
Household Weight	B, G, P, S	W1110291 - W111029	H	1
Longitudinal Weight	B, G, P, S	W1110391 - W111039	I	1
Population Factor	B, G, P, S	W1110491 - W111049	Y	1
Individual Weight - Immigrant Sample	G	W1110591 - W111059	I	1
Household Weight - Immigrant Sample	G	W1110691 - W111069	H	1

Identifiers:

Unique Person Number	B, G, P, S	X11101LL	I	1
Household Identification Number	B, G, P, S	X1110291 - X1110298	H	1
Individual in Household at Survey	B, G, P, S	X1110391 - X1110398	I	1
Oversample Identifier	G, P, S	X11104LL	I	1

Macro-level Variables:^c

Consumer Price Index	B, G, P, S	Y1110191 - Y1110198	Y	1
Median Pre-government Household Income	B, G, P, S	Y1110291 - Y1110298	Y	1
Median Post-government Household Income	B, G, P, S	Y1110391 - Y1110398	Y	1
Median Post-government Household Income (TAXSIM)	P	Y1110391 - Y1110398	Y	1

^aRace in the BHPS and SLID is reported for all sample members. In the PSID, race of the household head and wife is reported.

^bArea of residence is the Local Authority District of Residence in the BHPS, the Bundesland in the GSOEP, and the US state in the PSID. 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.

^cBecause macro-level variables do not vary across individuals or households, they are only listed in the codebooks for reference purposes.

Table 3. Mean Household Income of Men Before and After Exit, by Source, Country, and Age at Exit

	United States (Dollars)		Germany (Marks)		Britain (Pounds)		Canada (Dollars)	
	Before	After	Before	After	Before	After	Before	After
Income Source								
Aged 25 through 49								
Own market work	15,825	339	28,442	148	9,488	1,959	22,781	1,495
Others market work	9,588	7,403	7,812	9,549	2,548	3,166	11,777	11,729
Employer-based pensions	859	1,262	91	393	78	361	492	1,251
All other private	3,998	3,727	3,374	3,219	1,570	2,259	1,318	6,371
Pre-government	30,271	12,731	39,719	13,309	13,685	7,746	36,368	20,846
Social insurance pensions	1,571	869	645	2,307	144	491	344	778
Other public transfers	665	938	1,381	8,732	843	2,563	816	7,097
Taxes ^a	6,266	1,811	11,731	3,440	2,963	1,395	1,627	0
Post-government	26,241	12,727	30,015	20,907	11,708	9,404	35,901	28,721
Aged 50 through 61								
Own market work	29,984	730	42,071	2,740	10,798	149	40,212	3,372
Others market work	12,878	7,572	13,813	12,048	5,596	3,811	21,739	12,710
Employer-based pensions	1,900	8,874	47	2,326	448	3,375	1,481	13,898
All other private	10,083	12,349	5,975	6,655	3,718	3,678	2,599	18,980
Pre-government	54,845	29,525	61,906	23,769	20,561	11,013	66,031	48,960
Social insurance pensions	667	2,151	792	12,387	412	889	203	681
Other public transfers	497	443	505	5,282	486	1,761	1,189	5,224
Taxes ^a	13,691	4,744	18,436	5,752	4,116	895	12,330	14,005
Post-government	42,318	27,375	44,768	35,687	17,342	12,767	55,093	40,860
Aged 62 and over								
Own market work	23,277	355	40,433	143	10,213	71	27,656	1,941
Others market work	7,064	4,203	11,421	5,330	3,112	1,828	9,538	6,114
Employer-based pensions	3,724	7,056	38	2,206	1,931	3,386	2,199	8,281
All other private	9,398	6,944	7,067	8,946	4,036	4,236	2,405	9,278
Pre-government	43,464	18,557	58,958	16,625	19,292	9,520	41,798	25,614
Social insurance pensions	3,048	7,023	4,470	22,213	1,202	3,417	804	6,568
Other public transfers	129	162	145	738	267	511	397	1,034
Taxes ^a	9,519	2,831	15,497	3,403	3,015	404	2,289	448
Post-government	37,122	22,910	48,076	36,173	17,746	13,044	40,710	33,664

Source: PSID income years 1988-1996, GSOEP income years 1988-1997, BHPS income years 1990-1997, SLID income years 1993-1998. Sample sizes for men aged 25-49, 50-61 and 62 and over respectively are: PSID (156,73,82), GSOEP (74,241,72), BHPS (47,71,60), and SLID (134, 179, 83).

Notes: Household income is averaged over years t-2 and t-1 (before) and over years t+1 and t+2 (after), where (t) is the last year of long-term employment. Income values are in constant 1996 United States dollars, 1995 German marks, 1996 British pounds, and 1997 Canadian dollars.

^aTaxes in the BHPS include both taxes, occupational pension and national insurance contributions.

Table 4. Median Private Pension, Social Security and Total Income Replacement Rates of Men, by Country and Age of Exit (Percentages)

Country	Age 25 through 49			Age 50 through 61			Age 62 and over		
	Social Security	Private Pensions	Total Income	Social Security	Private Pensions	Total Income	Social Security	Private Pensions	Total Income
United States	0.0	0.0	44.3	0.0	28.4	61.2	33.0	25.3	55.9
Germany	0.0	0.0	58.3	28.9	0.0	77.1	56.8	0.0	78.0
Great Britain	0.0	0.0	87.9	0.0	17.3	74.7	49.6	25.0	74.0
Canada	0.0	0.0	75.9	0.0	20.5	71.3	28.3	19.6	84.2

Source: PSID income years 1988-1996, GSOEP income years 1988-1997, BHPS income years 1990-1997, SLID income years 1993-1998. Sample sizes for men aged 25-49, 50-61 and 62 and over respectively are: PSID (156,73,82), GSOEP (74,241,72), BHPS (47,71,60), and SLID (134, 179, 83).

Notes: The social security earnings replacement rate is the ratio of household social security income to own labor income. The private pension earnings replacement rate is the ratio of household private pension income to own labor earnings. The total income replacement rate is the ratio of post-government household income after and before labor force exit. In all cases, income before labor force exit is averaged over years t-1 and t-2. Income after labor force exit is averaged over years t+1 and t+2. The median value of each is reported in the cells of this table. The actual median person is different in each cell.

Figure 1. Age Specific Labor Force Exit Rates of men in the 1990s, by Country

