

# **Longitudinal analysis in statistical offices**

**Discussion paper 05010**

*J.W.P.F. Kardaun, J.A. Loeve*

The views expressed in this paper are those of the authors  
and do not necessarily reflect the policies of Statistics Netherlands



Statistics Netherlands

Voorburg/Heerlen, December 2005

### Explanation of symbols

.	= data not available
*	= provisional figure
x	= publication prohibited (confidential figure)
—	= nil or less than half of unit concerned
—	= (between two figures) inclusive
0 (0,0)	= less than half of unit concerned
blank	= not applicable
2004–2005	= 2004 to 2005 inclusive
2004/2005	= average of 2004 up to and including 2005
2004/'05	= crop year, financial year, school year etc. beginning in 2004 and ending in 2005

Due to rounding, some totals may not correspond with the sum of the separate figures.

**Publisher**

Statistics Netherlands  
Prinses Beatrixlaan 428  
2273 XZ Voorburg  
The Netherlands

**Printed by**

Statistics Netherlands - Facility Services

**Cover design**

WAT ontwerpers, Utrecht

**Information**

E-mail: [infoservice@cbs.nl](mailto:infoservice@cbs.nl)

**Where to order**

E-mail: [verkoop@cbs.nl](mailto:verkoop@cbs.nl)

**Internet**

<http://www.cbs.nl>

© Statistics Netherlands, Voorburg/Heerlen  
2005.

Quotation of source is compulsory.  
Reproduction is permitted for own or  
internal use.

ISSN: 1572-0314

Key figure: X-10



Statistics Netherlands



## LONGITUDINAL ANALYSIS IN STATISTICAL OFFICES

*Summary: In order to serve as an orientation for future activities, an inventory was made of methods used in longitudinal analysis by National Statistical Institutes, together with a selection of longitudinal data sets maintained by or for NSI's. It turned out that in most cases the methodology as such was not an explicit, but an implicit part of longitudinal studies. A variety of more or less classical techniques was applied, depending on the situation.*

*Regarding the data sets: the UK, the USA, Canada and New Zealand are leading in the field, and have a long tradition of longitudinal surveys. Sweden and Denmark have possibilities for longitudinal analysis based on registers, but the use of it is mainly left to academia. Most longitudinal studies concern persons, as compared to enterprises.*

*In Statistics Netherlands there are 5 longitudinal surveys and several registers with longitudinal possibilities. Longitudinal analysis is limited and mostly published in Dutch, but the activity is increasing. There are a few methodological papers, already some years old.*

*Keywords: longitudinal analysis - longitudinal data sets - international overview*

## Contents

<b>1. Introduction.....</b>	<b>3</b>
<b>2. Inventory foreign offices.....</b>	<b>4</b>
Australia .....	4
Canada .....	6
Denmark .....	8
Germany .....	9
New Zealand .....	9
Sweden .....	10
Switzerland .....	12
United Kingdom .....	12
U.S.A. ....	15
Supranational .....	21
<b>3. Inventory Statistics Netherlands .....</b>	<b>21</b>
<b>4. Summary and conclusions.....</b>	<b>23</b>
<b>5. Appendices.....</b>	<b>25</b>
<b>5.1 References .....</b>	<b>25</b>
<b>5.2 List of acronyms .....</b>	<b>28</b>
<b>5.3 Not Included &amp; other sources of information .....</b>	<b>30</b>
<b>5.4 Index .....</b>	<b>32</b>

## 1. Introduction

At Statistics Netherlands, there is a growing availability of data sets with longitudinal information, such as the Social Statistical Database and the Health Care Database. Consequently, interest in methods for longitudinal analysis at Statistics Netherlands has increased. To acquire more knowledge about longitudinal analysis methods, a literature study is started within the Methods and Informatics Department (TMO). However, this is a very broad field of research, so a preliminary phase making an inventory is indicated. Therefore, this short study started out as an overview of the longitudinal methods used in foreign statistical offices. There is a scarcity of documents with methodological viewpoints on this subject, however, and the study turned out to become an inventory of major longitudinal data collections. The inventory concerns the statistical offices of Australia, Canada, Denmark, Germany, New Zealand, Sweden, Switzerland, United Kingdom, and the USA. To complete this overview, a short description of the situation at Statistics Netherlands is given.

### *Sources:*

As expected, we could not find much in the scientific journal literature about longitudinal methodology in foreign statistical offices. This is partly because of lack of material, partly because of difficulty in identifying such material. Instead of engaging the difficult task to survey the so called *grey literature*, we took the royal route, and sent out a couple of e-mails and telephone calls to get us informed by the “locals who know”. As soon as we included longitudinal data collections in our search, the internet became a useful source of information.

Here we thank all the persons who supplied us willfully information; after some deliberation, we decided not to mention them all by names, as in writing this report we will most likely have distorted their information in too large an extent.

Where suitable material was available, liberal use has been made of the “cut and paste” possibilities from websites and reports from other institutions. Especially for the USA, Canada and the UK, this report consists mainly of copied fragments from websites.

### *Inclusion criteria:*

- Only longitudinal studies in which NSI’s participate are considered. An exception has been made for the USA: as there is no single NSI in the USA, and US federal agencies quite often contract out work, selected federally sponsored longitudinal studies have been included, if they would fit statistical purposes.
- (Traditional) timeseries and timeseries analysis are not considered longitudinal analysis for the purposes of this survey. It is a requirement that measurements on the same individual units for at least two moments in time should be present.
- We aimed at including all methodological studies concerning longitudinal data in NSI’s.

— In countries with many longitudinal data sets, we selected the most important ones (no further refinement of this criterion); of the data sets / data collections a summary description is given.

— We did *not* include reports with emphasis on the results from the selected longitudinal studies.

#### *Coverage:*

With the type of search performed, it is likely that white spots remain uncovered in this survey. We did not aim at completeness, but stopped searching when the yield of continued searching decreased. The survey is sufficient to aid us in positioning the activities of Statistics Netherlands compared to sister institutes. *Yet, if readers are aware of omissions, or even better of a similar survey, not older than 10 years, the authors would appreciate very much a notification ([jkarn@cbs.nl](mailto:jkarn@cbs.nl); [jlve@cbs.nl](mailto:jlve@cbs.nl)).*

There is one methodological weakness in this type of search: as the web is predominantly in English, it is harder to find references to studies by NSI's in non-English speaking countries. This may be the reason that UK, Canada, USA, New Zealand are present more prominently in this study than is justified. We did look at the websites of the NSI's in some other languages, but this only overcomes part of the problem. Not every NSI has a website where one can see ongoing activities; some NSI's have a site that is oriented towards statistical results only. So, the surprising result that, e.g., France is not present in this inventory, might be an artifact rather than a finding.

## **2. Inventory foreign offices**

In this section, the results of the inventory are presented for the single countries. The countries are in alphabetical order.

To assist the readers' orientation, there is an indicator in the left margin, indicating the category of the following paragraph(s), where

- G = General,
- M = Methodological,
- D = Data collection,
- S = Summary.

### ***Australia***

- D The Australian Bureau of Statistics (ABS) ran a *Business Longitudinal Survey* (BLS). The BLS is a combination of four surveys of firm performance and factors affecting firm performance in the period 1994 to 1998. This produced a research database of firm level longitudinal data relating mainly to small and medium sized firms. A publication after the last year of the BLS [ANONYMOUS (1999)] presented only summary analysis, consisting chiefly of looking at annual changes (for example in the proportion of firms employing a certain number of employees) using simple



cross tabulations. No extensive results were published, as these were not considered reliable enough. A later *Business Longitudinal Database*, currently in development, will have learned from the previous experiences.

- D The *Survey of Employment and Unemployment (SEUP)* is a longitudinal survey where information was collected from the same individuals over three annual waves in the period 1994 to 1997. Interviews at each wave were used to obtain information on a person's current circumstances and on their labour market activities over the preceding 12 months (retrospective information). The information included characteristics such as labour market activity, job-seeking activities, industry, occupation and sector. Analysis of the SEUP was done by external researchers who had been contracted. They used transition probability models, including probit, logistic and multinomial logistic regressions. [CARINO-ABELLO (2000)].

As its preface states: The paper describes labour market dynamics across the Australian population over different transition periods, and identifies differences in the pattern of flows by age, sex and other characteristics. Modelling techniques are then used to tell us how important particular characteristics are [...] affecting people's transitions.

The sample consisted of three subgroups, concerning in total approximately 8.5 thousand persons. Three sets of weights for individual waves, and additional sets of weights for longitudinal regression were used.

- D The *Longitudinal Data Set (LDS)* is a one percent longitudinal sample of customer records of Centrelink, which is the agency responsible for administering the welfare and income support programs of the government. The data set tracks fortnightly payments to social security customers from 1995 to 2000. The records contain a number of customer characteristics and payment details. For the analysis of the LDS, use was made of event history analysis models including survival analysis (Kaplan-Meier and life table methods) to calculate survival and hazard functions. Furthermore, Cox regression was used to calculate hazard functions and to estimate the time-variant and time-invariant covariates that influence durations. A competing risk framework was used to model competing events. Also log-rank tests, logistic regressions and multinomial logit models were employed to study transitions. The results of the analyses will be published later on in 2004.
- D The *Longitudinal Study of Immigrants to Australia*<sup>1</sup> (LSIA) has two parts, LSIA1 and LSIA2, following 5 thousand and 3 thousand primary applicants in the periods 1993-95 and 1999-2000, respectively. Some information about the attached household was collected. Three collection moments for LSIA1 and two collection moments for LSIA2, covering at most 42 months of interval, were undertaken. Weighting to compensate for stratification to country of origin and for effects of sample loss was applied.

---

<sup>1</sup> Most information was found on the website of the Australian Department of Immigration and Multicultural and Indigenous Affairs, [www.immi.gov.au/research/lsia/lsia01.htm](http://www.immi.gov.au/research/lsia/lsia01.htm). (The role of ABS in this survey was not clear to us.)

- S In general, there seems to be no systematic employment of methods for longitudinal data analysis in the past at the Australian Bureau of Statistics. However, for recent analyses rather extensive survival analysis methods are used.

### *Canada*

- M At Statistics Canada there are two areas where longitudinal methods are of importance. On the one hand, many researches are conducted on social economic issues using longitudinal data. Topics in this field are for example, low income dynamics, labour market mobility, earning inequality, unemployment duration and school performance. The methods applied are the standard ones from literature. The method adopted in a research project depends on the questions to be asked and the nature of the dependent variable. Techniques employed include linear regression models, hazard models and panel logit/probit models.

On the other hand, analytical methods for longitudinal data are developed, particularly with respect to integrating the survey design aspects into data analyses. Often, fairly simple descriptive comparisons over time are made, such as in characteristics of specific subpopulations. Moreover, a lot of logistic and linear regressions and proportional hazards models are used. Most of the longitudinal surveys are now provided with sets of bootstrap weights so that people have a means of accounting for the survey design in their variance estimation. From the description of research projects on data analysis, Statistics Canada appears to be working in many different areas in the field of longitudinal analysis methods, with a focus on inferential problems and issues arising from the modelling of longitudinal data [KOVACEVIC, (2001)]. To this aim, for example, marginal models, generalized estimating equations, models for duration data and multi-level models are employed.

- D The *National Longitudinal Survey of Children and Youth* (NLSCY) consists of a sample of approximately 15 thousand Canadian children aged 0 – 11 years in 1994/5, interviewed every 2 years. Data collection is by Statistics Canada, which runs the study in partnership with Human Resources Development Canada; analysis appears to be done mostly by academic institutions [MICHAUD (1999)].
- D The *National Population Health Survey* (NPHS) is a panel of approximately 20 thousand persons, interviewed every two years during a 20 year period, starting 1994/95. The topics are health, health determinants, and health care use; linkage with administrative sources is provided. Every four years additional persons are added to the sample, in order to allow cross-sectional calibration. This dual view (longitudinal and cross-sectional) implies that separate meanings are given to some facts (e.g. out of country), according to the view, and that 7 different weights are needed. Special attention to variance estimation and sample maintenance is needed [YEO (1999)].
- D The *Survey of Income and Labour Dynamics* (SILD)<sup>2</sup> started in 1993, and follows approximately 15 thousand households for six years; every three year a “replenish-

---

<sup>2</sup> In other sources called Survey of Labour and Income Dynamics (SLID).

ing” sample is added. It is designed to serve both cross-sectional and longitudinal uses. Each use needs its own (set of) weights [GILES (1999), ROBERTS (2001), WEBBER (1998)].

- D The *National Graduate Surveys* is a survey where graduates from 1982, 1986, 1990 and 1995 were contacted 2 and 5 years after their graduation. Outcome variables were, e.g., employment, job satisfaction. The sample size is around 30 thousand.
- D The *Longitudinal Survey of Immigrants to Canada* (LSIC) interviews immigrants three times, 6 months, two years and four years after their arrival. A sample of 20 thousand persons from immigrants entering between Oct. 2000 and Sept. 2001 are questioned on voluntary basis about their jobs, education and housing, language problems, and coping behavior.
- D *School leavers survey*, 1991-1995, co-project of Human Resources Development Canada, Statistics Canada. In the late 1980s, Human Resources Development Canada (HRDC) commissioned Statistics Canada to conduct a survey to identify the characteristics and consequences of leaving school. Between April and June 1991, computer assisted telephone interviews (CATI) obtained demographic/background data, information on school experiences and post-school labour market experiences from 9,460 young people aged 18 to 20. The 1991 respondents were classified into one of three groups: continuers (still in high school), graduates (had received a high school certificate or diploma), or leavers (not in high school and did not have a high school certificate or diploma).

In early 1994, HRDC requested that Statistics Canada re-interview the 1991 survey respondents. *The School Leavers Follow-up Survey* (SLFS) was conducted between September and December 1995, again using the CATI instrument. The primary objective of the Follow-up was to study the school-work transitions of young people beyond high school, particularly: 1. from the end of high school to the first job and 2. after the first job (i.e. various school-work and work-work transitions).

The School Leavers Follow-up Survey gathers information on the respondents' labour market status and education/training beyond high school, for example: educational attainment; last week's (week preceding the interview) education and work; reference jobs (first and most recent reference job) [a reference job is one of sufficient duration, that is, a job involving at least 20 hours of work per week for a period of at least six consecutive months]; period between the last time in high school full-time and the beginning of the first reference job; period between the last time in high school full-time and the beginning of the job held in the week preceding the interview (or the beginning of the week preceding the interview); general work information.

- D Labour market activity survey (LMAS), 1986-1991. The data produced consists of: 1) measures of the pattern of employment and unemployment in Canada over a 12, 24 or 36 month period, 2) characteristics of paid jobs held during this period which are not available from other sources, and 3) socio/economic/demographic profiles for groups which are currently eligible for Employment and Immigration Canada (EIC) programs. It is a supplement to the Labour Force Survey. The survey collects

retrospective information for a 12 month period in an initial interview and combines this information with information for a second 12 month period collected in a subsequent interview conducted 12 months later.

- D Longitudinal administrative database This databank file, a longitudinal sample of tax-filers for the years 1982 to 2003, still ongoing, is made up of a wide variety of income and demographic variables. The statistics are derived from the annual tax file provided by the Canada Revenue Agency. It contains administrative data with demographic characteristics (families, lone parent families, marital status) and tax data.
- D Ontario child health survey (OCHS), 1983-[ongoing] is a long-term survey to evaluate the impact of early childhood experiences and development on later adult health, quality of life and functioning. This survey was conducted on a cost recovery basis for a client (the Canadian Centre for Studies of Children at Risk at McMaster University in Hamilton, Ontario).

The survey concerns all persons born between January 1, 1966 and January 1, 1979 whose usual place of residence was a household dwelling in Ontario (when the original sample was drawn), excluding those from Indian reserves, those living in a collective dwelling/institution and those living in dwellings constructed since June 1981.

- D Workplace and employee survey (WES), 1999 - [ongoing] is designed to explore a broad range of issues relating to employers and their employees. The survey aims to shed light on the relationships among competitiveness, innovation, technology use and human resource management on the employer side and technology use, training, job stability and earnings on the employee side. The survey is unique in that employers and employees are linked at the micro data level; employees are selected from within sampled workplaces. Thus, information from both the supply and demand sides of the labour market is available to enrich studies on either side of the market. There are over 100 research projects either completed or underway using the WES.

There are two reference periods used for the WES. Questions concerning employment breakdown use the last pay period of March for the reference year while other questions refer to the last 12-month period ending in March of the reference year.

- S At Statistics Canada a broad knowledge exists about different methods for longitudinal analysis. Most often, methods like linear and logistic regression and proportional hazard models seem to be employed, but research into many other methods is also carried out. There are a handful of longitudinal surveys, most of them in the *person* domain.

### ***Denmark***

- G Statistics Denmark has legal access to data from all public administrative registers. There even is a law requiring the authorities responsible for these registers to consult

with Statistics Denmark to ensure that the registers are appropriately organised for secondary use in statistics. These registers are the principle data source for Statistics Denmark. Other data collection methods are used when necessary, but they are only regarded as a supplement. Examples of registers used, are the Central Population Register, tax registers, wage registers, social benefits registers, the Central Register of Buildings and Dwellings, and so on. Thus, a lot of register data is available containing longitudinal information.

Longitudinal analyses on these data sets are mainly carried out by external researchers. Especially for research purposes, several integrated databases have been created by linking information from different statistical registers, such as the Demographic Database and the Fertility Database.

### *Germany*

- G At the Federal Statistical Office (Destatis) in Wiesbaden, there is a project to form longitudinal data sets from the cross-sectional microcensus. Another pilot project is concerned with the forming of an access panel. This is a panel of households who are recruited as they are leaving the microcensus. They will be available for regular participation in voluntary surveys of official statistics. Also, Destatis participated in a European research project about the harmonisation of European surveys. In this project different methods to collect data for the European Community Household Panel are compared. Furthermore, the project investigated important hypotheses about the data quality of panel surveys.
- S Very little data analyses are carried out at Destatis. This also holds for longitudinal analyses. Data analysis in Germany is generally considered to be the field of university or other private research institutes. However, there are some exceptions, such as a collaboration with DIW Berlin (Deutsches Institut für Wirtschaftsforschung). This pilot project using longitudinal microdata was meant to analyse the relation between the structure of costs and the profits in businesses.

### *New Zealand*

- G At Statistics New Zealand, there are a number of longitudinal surveys. Furthermore, there is an increasing access to a range of administrative data that is also longitudinal. To be able to analyse these data, Statistics New Zealand has been busy these last years with gaining and retaining skills for longitudinal data analysis, e.g. exploration of cluster analyses for describing longitudinal patterns, see HHLFS, below.

At the moment Statistics New Zealand is researching the construction of a population database for cohort modeling. The aim of this database is to integrate social and population statistics such that different kinds of cohorts can be studied.

Furthermore, a new longitudinal survey has started in October 2002, the *Survey of Family, Income and Employment*. Some of the key objectives are determining patterns for and relations between income levels, labour market activity, participation in education, and changes in family status. The first results are expected in 2004 and

will be primarily cross-sectional. The longitudinal character of the results is expected to build up from 2005.

- D The *Household Labour Force Survey* (HHLFS) is a household-panel survey, of approximately 28 thousand respondents; each household stays in the panel for 8 quarters. In Kuzmicich (2001), exploratory analyses with methodological aims are presented. These concern the application of CHAID<sup>3</sup> for imputation and the application of cluster analyses for describing longitudinal patterns.
- D The *Longitudinal Immigration Survey New Zealand* (LisNZ)<sup>4</sup> is a survey of approximately 8 thousand immigrants to New Zealand in 2004. There will be three waves of data collection, each with one year interval.
- D The *Longitudinal Survey of Income, Employment and Family Dynamics* (LSIEFD), later renamed to SoFIE, has a cohort design with a sample size of approximately 10 thousand households at wave one.

The LSIEFD will have a single fixed panel with a life of between six and eight years. Survey members will be interviewed annually about the year just passed on topics such as:

- income levels, sources and changes
- employment and education experiences
- family status and changes in family status experienced
- demographic information and health status.

Every two years, survey members will also be asked about their assets and liabilities.

Government granted full funding to run the survey in 2000 and the survey entered the field for its first wave of data collection in 2002.

The methodology developed and tested as part of this project is presented in detail in [ANONYMOUS(a), 2001].

### ***Sweden***

- G At the office of Statistics Sweden in *Stockholm* several data sets are available which can be used for longitudinal analysis. Most of them are surveys, for example the survey of living conditions (same respondents every eight years), the labour survey (every two years) and the income survey. In addition to surveys, some register data can be used. For example, tax records are used in studies on employment and income.
- M Most methods applied for longitudinal analysis are of the descriptive kind. Sophisticated methods are not often used. There are no regular publications on longitudinal

---

<sup>3</sup> CHAID, for *Chi-square Automatic Interaction Detection*, is an exploratory method used to study the relationship between a dependent variable and a series of predictor variables. CHAID modeling selects a set of predictors and their interactions that optimally predict the dependent measure. See: Hartigan, J.A. (1975), *Clustering algorithms*, New York, NY: John Wiley. For applications, see, e.g., the SPSS Answer Tree Program.

analyses, but ad hoc studies are occasionally carried out. An example is the study of response errors in categorical variables. The aim was to correct for the errors by examining individual changes. Also, some commissioned work was done on life table techniques in the field of survival analysis. Furthermore, an analysis of sick leave was carried out some years ago using structural equation models (LISREL). Possibly, studies in this field will be taken up again in the near future. Also used is a similar kind of technique, namely the method of partial least squares. This method was applied in some commissioned work for a study of customer satisfaction for the public sector.

- M At the office of Statistics Sweden in *Örebro* the most important longitudinal data sets are register based. Analyses on these sets are carried out by researchers, for example from universities, in medical, social and economical fields.

Methodology for the special characteristics of register based statistics is developed [WALLGREN (2002)], but the longitudinal use of the registers is mainly performed by academia.

- S On the whole it seems that there is no *systematic* application of longitudinal analysis methods at Statistics Sweden in Stockholm. However, now and again occasional studies are carried out where per study is decided which method would be the most suitable. Almost no longitudinal analyses are carried out by Statistics Sweden in *Örebro*, but quite a lot of interesting information is published by suitable presentations of different tables. Though staff members consider research on methods for longitudinal analyses to be important, management would have to provide the means to do this.

- D At the *Örebro* office of Statistics Sweden generally national registers are used. The four registers are: 1) Population; 2) Activity (i.e. corresponding roughly to Labour); 3) Businesses; 4) Real Estate [ANONYMOUS(b), 2001]. The development is to consider the registers as a *system* of registers.

The population register contains longitudinal registers on income, welfare (with some social indicators) and young persons entering the labour market. Furthermore, there is a multigenerational register (for example, to find the grandparents of an individual) which is mainly used for medical research.

- D The *Swedish Survey on Conditions of Living* <sup>5</sup>, since 1974/75, has included a panel since 1985 [ANONYMOUS, s.a.]. Topics were initially: health, economic resources, employment, working environment, education, and housing; later more topics were added. The number of respondents varies, say 11 thousand to 7 thousand per year; in the early years several methodological reports were published, mainly in Swedish

---

<sup>4</sup> [www.stats.govt.nz](http://www.stats.govt.nz) ; then search for "lisnz"

<sup>5</sup> Not to be confused with the Swedish Level-of-Living Survey (NLU), run by the University of Stockholm. This started in 1968, has had 5 additional observation cycles. The 1/1000 sample of the population at the start has been re-interviewed each cycle, with some linkage with administrative data.

Also not to be confused with LINDA, Longitudinal Individual Data for Sweden, run by Uppsala University, a register based panel over the period 1960 – 1998.

with some English reviews [THORSLUND (1985)]. The topics of these studies were mainly quality of data and non-sampling errors.

Since 1986 it serves a dual purpose of cross-sectional and longitudinal use. The sample is constructed by including half of the participants of 8 years before. Some methodological reports (on measurements of change and the methodology of panel studies) are available, but only in Swedish<sup>6</sup>.

### ***Switzerland***

- D The only longitudinal data set that we could find at the website of the Swiss Federal Statistical Office is the *Swiss Labour Force Survey* (LFS). The LFS, since 1991, is a CATI interview of a rotating panel of approximately 16 thousand households (after non-response). The sample frame is the PTT telephone directory, but persons must be living permanently in Switzerland to be included. Every year one-fifth of the panel is refreshed. For each new household a reference person is determined, who will be in the panel for five years. Besides the usual weights for the sampling fraction and for post-stratification, a factor for estimated attrition is used for every observation.<sup>7</sup>
- D The Swiss Federal Statistical Office participates in a joint-venture: the *Swiss Household Panel* (SHP). The SHP strives 1) to ensure a solid database for social reporting about stability and changes in living arrangements and well-being in Switzerland that complements data collected by the Swiss Federal Statistical Office, and 2) to promote the opportunities for quantitative social science research, by making high quality data available to the national and the international social science research community. In May 2002 the Swiss National Science Foundation and the Swiss Federal Statistical Office finalised a strategic agreement dealing with the continued operation of the Swiss Household Panel after 2004 in conjunction with the EU-SILC (European Statistics on Income and Living Conditions). Both institutions participate as equal partners in the collection of annual cross-sectional and panel data. Starting 2004, the SHP continues to be financed by Swiss National Science Foundation

The SHP data is compatible with various databases generated by the Swiss Federal Statistical Office (SFSO), in particular the two largest periodic population surveys, the Swiss Labour Force Survey and the Swiss Health Survey. In order to compensate for losses of panel membership due to different causes (moving abroad, untraceable addresses, refusals, and death), about 2,000 new households will be added in 2004 (wave 6) to the still existing 3,000 panelised households.

### ***United Kingdom***

- G Britain is a world leader in national longitudinal research studies and in their use for social science and medical purposes.

---

<sup>6</sup> As [ANONYMOUS (s.a.)] is not dated it cannot be said until when this last statement is accurate.



The Joint Centre for Longitudinal Research (JCLR) coordinates the three British birth cohort studies containing 35,000 life histories of people born in 1946 (NSHD), 1958 (NCDS) and 1970 (BCS70) - which claim to provide unique insights into the processes that shape human lives. The long-term effects of family circumstances and disruption, the causes of educational failure, the origins of chronic illnesses, inter-generational links in child health and development are examples of the topics that researchers have studied using these data. The JCLR and its participants are academic institutions, and these studies consequently do not meet the inclusion criteria. They are mentioned because of their influential and pioneering role. Besides the website of JCLR,<sup>8</sup> the site of the ESRC United Kingdom Longitudinal Studies Centre (ULSC)<sup>9</sup> at the university of Essex is a rich source information of activities on this type of research in the UK, in particular the Keeping Track project<sup>10</sup>, which aims to provide an up-to-date guide to major longitudinal sources of data. The project covers data sets collected by governmental, academic, private social research, medical and private industrial sources, including household panel surveys, studies following the health of individuals, birth cohort studies, studies following the quality of a product design, and administrative records.

At the Office for National Statistics (ONS) there are several longitudinal data sets available. Examples are the ONS Longitudinal Study (LS), the Labour Force Survey (LFS, Great Britain), and the Claimant Unemployment Cohort (JUVOS Cohort).

*D Longitudinal Study (LS) by ONS, formerly OPCS.*<sup>11</sup>

The *Longitudinal Study* (LS) contains anonymised linked census and vital event data for one per cent of the population of England and Wales. Vital events include births, deaths, widowhood, cancer registrations, migration, enlistments and entries to long-stay hospitals. The sample originally included the 1971 Census of Population information, for people born on one of four selected dates of birth. These four dates were used to update the sample at the 1981, 1991 and 2001 Censuses and in routine event registrations.

A list of 79 working papers of LS/ONS and documentation can be found at [www.cls.ioe.ac.uk/Ls/workpap.htm](http://www.cls.ioe.ac.uk/Ls/workpap.htm).

---

<sup>7</sup> From [www.lisproject.org/les/ch/ch97gendesc.html](http://www.lisproject.org/les/ch/ch97gendesc.html)

<sup>8</sup> <http://www.cls.ioe.ac.uk/jclr.htm>

<sup>9</sup> <http://iserwww.essex.ac.uk/ulsc/>

<sup>10</sup> <http://iserwww.essex.ac.uk/ulsc/keeptrack/index.php>

<sup>11</sup> Most information was found on the site of Celsius (<http://www.cls.ioe.ac.uk/Ls/intro.htm>): “As of 1st December 2001, support for academic users of the LS is now provided by the Centre for Population Studies at the London School of Hygiene and Tropical Medicine. The School was awarded this contract as part of the ESRC/JISC Census of Population Programme, which continues to 31st July 2006. All enquiries for general information and proposals for new projects should be addressed to: CeLSIUS (Centre for Longitudinal Study Information and User Support)”

A comprehensive description [HATTERSLEY (1995)] is the LS Technical Volume. This 374-page volume covers the development of the LS from its inception up to the 1991 Census. Separate chapters are devoted to the history of the study, the scope for analysis (including international comparative work), its organisation and an overview of data sources used. It also examines the quality of the linkage of key socio-demographic variables from the 1971, 1981 and 1991 Censuses together with that of vital events (births, cancers, deaths) spanning this 20-year period. The volume is suitable for experienced LS researchers and those working in public health, academic, local or central government research with no previous knowledge of the study.

---

In an attempt to make this report slightly more palatable to the reader, the drone of the enumeration of all the studies, dry as dust, is interrupted below by a more detailed description of the LS, as an example of a long lasting and very fruitful study.

The ONS Longitudinal Study (LS) is a study of linked census and vital event data (births, cancers, deaths) held by the Office for National Statistics (ONS) relating to 1% of the population of England and Wales - about 500,000 individuals at any one census point.

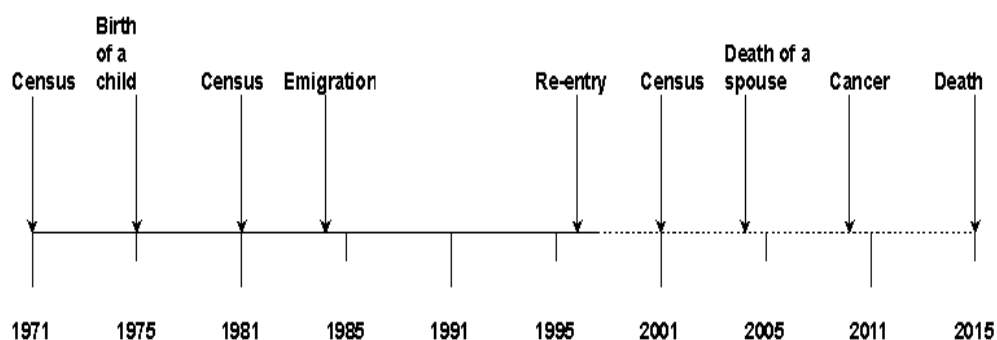
Initially, all people born on one of four dates each year and enumerated at the 1971 Census were selected for the study.

From 1971, the LS has been augmented with new births occurring on one of these four dates and as immigrants with the birth dates register with the National Health Service. Over the same period there have been "exits" from the study as sample members go to live abroad. As a result the LS differs from other longitudinal surveys where the sample is selected at one point in time.

Additional samples of individuals with one of the four LS birth dates were selected from the 1981 and 1991 Censuses. The dataset now includes longitudinal census data for three time points for approximately 400,000 individuals and information on others enumerated in their household. It is important to emphasise that the LS samples individuals and then attaches information about every individual living in the same household. It does not "follow up" household members in the same way as it does individual sample members, from census to census. Therefore, census data on other household members will be cross-sectional unless the LS member was enumerated in the same household at successive censuses.

Figure 1 shows the range of census and event data recorded in the LS for a hypothetical LS member, who first entered the study by giving an LS date of birth at the 1971 Census. This illustrates, for example, that census data will only be available for LS members enumerated in England and Wales. The availability of mortality data for the LS member's spouse offers scope for studies of post-bereavement mortality.

Figure 1: Census and event data recorded for a hypothetical LS member



The LS only includes data which are routinely collected by ONS. No data about individuals obtained from other government departments can be linked to the study.

- D The *Labour Force Survey* (LFS) is a quarterly sample household survey of households in Great Britain. The survey has a rotating sample design with each household included for five successive quarters, so data on the same individuals can be linked over time. The survey provides information on the labour market including data on (un)employment, occupation, hours worked, and earnings.
- D The *Claimant Unemployment Cohort* is a five per cent sample of all computerised claims for unemployment related benefits. Each time a person with a relevant national insurance number makes a claim, their details are added to the cohort file. The data collected include information about usual occupation, occupation sought, claim start date, and claim end date.
- S The ONS has a long history of organizing and fostering longitudinal studies, with extensive participation of academia.

### *U.S.A.*

- G As mentioned in the Introduction, the inclusion criteria for the U.S.A. are different from those for other countries. Not only longitudinal studies by the statistical offices of the various government departments are included, but also studies contracted out to other institutes. It is obviously a difficult and error-prone criterion, but the authors felt that by doing otherwise, too many relevant studies would have been missed.
- D The *National Longitudinal Surveys (NLS)*<sup>12</sup> are a set of surveys designed to gather information at multiple points in time on the labor market activities and other significant life events of several groups of men and women. For more than 3 decades, NLS data have served as an important tool for economists, sociologists, and other researchers. The set of surveys consists of:

*National Longitudinal Survey of Youth 1997 (NLSY97)* — Survey of young men and women born in the years 1980-84; respondents were aged 12-17 when first interviewed in 1997.

---

<sup>12</sup> <http://www.bls.gov/nls/home.htm>

*National Longitudinal Survey of Youth 1979 (NLSY79)* — Survey of men and women born in the years 1957-64; respondents were aged 14-22 when first interviewed in 1979.

*NLSY79 Children and Young Adults* — Survey of the biological children of women in the NLSY79.

*National Longitudinal Surveys of Young Women and Mature Women (NLSW)*— The Young Women's survey includes women who were aged 14-24 when first interviewed in 1968. The Mature Women's survey includes women who were aged 30-44 when first interviewed in 1967. These surveys are now conducted simultaneously in odd-numbered years.

*National Longitudinal Surveys of Young Men and Older Men*— The Young Men's survey, which was discontinued in 1981, includes men who were aged 14-24 when first interviewed in 1966. The Older Men's survey, which was discontinued in 1990, includes men who were aged 45-59 when first interviewed in 1966.

For the scientific impact of these national longitudinal surveys, see the National Longitudinal Surveys Annotated Bibliography (at [www.nlsbibliography.org/](http://www.nlsbibliography.org/)): *Descriptions of over 3,000 NLS-based journal articles, working papers, monographs, and dissertations stored in a searchable electronic database.*

- D The *Survey of Income and Program Participation (SIPP)*,<sup>13</sup> of the U.S. Bureau of the Census, collects data about source and amount of income, labor force information, program participation and eligibility data, and general demographic characteristics. It serves to measure the effectiveness of existing federal, state, and local programs; to estimate future costs and coverage for government programs, such as food stamps; and to provide improved statistics on the distribution of income in the country.

The survey design is a continuous series of national panels, with sample size ranging from approximately 14,000 to 36,700 interviewed households. The duration of each panel ranges from 2 1/2 years to 4 years. The SIPP sample is a multistage-stratified sample of the U.S. civilian noninstitutionalized population. For the 1984-1993 panels, a panel of households was introduced each year in February. A 4-year panel was introduced in April 1996. A 2000 panel was introduced in February 2000 for 2 waves. A 3-year 2001 panel was introduced in February 2001.

- D The *National Longitudinal Study of the High School Class of 1972 (NLS-72)* is the grandmother of the longitudinal studies designed and conducted by the National Center for Education Statistics (NCES) of the U.S. Department of Education. It may well be the richest archive ever assembled on a single generation of Americans.

Participants in the study were selected when they were seniors in high school in the spring of 1972, and in a supplementary sample drawn in 1973. The records include the "Base Year" survey; follow-up surveys in 1973, 1974, 1976, 1979, and 1986; high school records; and postsecondary transcripts (collected in 1984).

---

<sup>13</sup> [www.sipp.census.gov/sipp/overview.html](http://www.sipp.census.gov/sipp/overview.html)

- D The objectives of the *National Longitudinal Mortality Study (NLMS)* are to investigate social, economic, demographic and occupational differentials in mortality (total and by cause) within a national sample of the U.S. population, as identified from the Current Population Survey (CPS) of the Bureau of the Census (n=637,162). The full study is larger, 1.3 million persons, but for confidentiality reasons, a subset of 5 of the original 12 Current Population Surveys is provided for the limited access data set.

Records from the Current Population Surveys (CPS) of March 1979, April 1980, August 1980, December 1980, and March 1981 are matched to the National Death Index to identify the occurrence and cause of death between 1979 and 1989. Extensive demographic, social, economic, and occupation information is collected in the CPS. The linkage of the individual social and economic data with the mortality outcomes provides the resource for extensive analysis.

- D *The NHANES I Epidemiologic Followup Study (NHEFS)* by the National Center of Health Statistics, is a longitudinal study designed to investigate the relationships between clinical, nutritional, and behavioral factors assessed in NHANES I<sup>14</sup> and subsequent morbidity, mortality, and hospital utilization, as well as changes in risk factors, functional limitation, and institutionalization.

The NHEFS cohort includes all persons 25-74 years of age who completed a medical examination at NHANES I in 1971-75 (n = 14,407).<sup>15</sup> It is comprised of a series of follow-up studies, four of which have been conducted to date. The first wave of data collection was conducted for all members of the NHEFS cohort from 1982 through 1984. It included tracing the cohort; conducting personal interviews with subjects or their proxies; measuring pulse rate, weight, and blood pressure of surviving participants; collecting hospital and nursing home records of overnight stays; and collecting death certificates of decedents.

Continued follow-up of the NHEFS population was conducted in 1986, 1987, and 1992 using the same design and data collection procedures developed in the 1982-84 NHEFS.

- D *National Epidemiologic Survey on Alcohol and Related Conditions (NESARC)*

---

<sup>14</sup> NEHFS is the only longitudinal part of a family of surveys: the National Health and Nutrition Examination Surveys (NHANES I, II, III, Hispanic HANES, and NHANES I Epidemiologic Followup Survey [NHEFS]) were designed to obtain information on the health and nutritional status of the United States population.

<sup>15</sup> The NHANES I (1971-1975) interviewed a sample of 31,973 persons aged 1-74 years. The sample was selected so that certain population groups thought to be at high risk of malnutrition (persons with low incomes, preschool children, women of childbearing age, and the elderly) were oversampled at preset rates. On completion of the survey, 23,808 of the interviewed sample were given a medical examination, and this information is also part of the NHANES I data collections.

The NESARC was conducted and sponsored by the National Institute on Alcohol Abuse and Alcoholism (NIAAA). The NESARC is the primary source for information and data on the U.S. population for:

- alcohol and drug use;
- alcohol and drug abuse and dependence; and
- associated psychiatric and other medical comorbidities.

NESARC was designed to be a longitudinal survey with its first wave of interviews fielded in 2001-2002. The second wave of interviews is planned for 2004-2005. The NESARC is a representative sample of the United States population and 43,093 Americans participated in the first wave of the survey. During wave 2, an attempt to re-interview all 43,093 of these respondents will be made. The target population of the NESARC is the non-institutionalized household population, 18 years and older, residing in the United States including the District of Columbia, Alaska, and Hawaii. Additionally the following non-institutional group quarters housing units were included as part of the NESARC sample: boarding houses, rooming houses, non-transient hotels and motels, shelters, facilities for housing workers, college quarters, and group homes. The sample provides estimates for the nation as a whole on topics related to alcohol and drug use, abuse and dependence and their associated disabilities.

- D *Add Health* and *Add Health 2000: A National Longitudinal Study of Adolescent Health* (University of North Carolina at Chapel Hill). The National Longitudinal Study of Adolescent Health (*Add Health*) is a nationally representative study that explores the causes of health-related behaviors of adolescents in grades 7 through 12 and their outcomes in young adulthood. *Add Health* seeks to examine how social contexts (families, friends, peers, schools, neighborhoods, and communities) influence adolescents' health and risk behaviors.

Initiated in 1994 under a grant from the National Institute of Child Health and Human Development (NICHD) with co-funding from 17 other federal agencies, *Add Health* is the largest, most comprehensive survey of adolescents ever undertaken. Data at the individual, family, school, and community levels were collected in two waves between 1994 and 1996. In 2001 and 2002, *Add Health* respondents, 18 to 26 years old, were re-interviewed in a third wave to investigate the influence that adolescence has on young adulthood.

Multiple datasets are available for study, and more than 1000 published reports and journal articles have used the data to analyze aspects of these complex issues. *Add Health* investigators hope this research will enable policy makers, researchers, health-care providers, and educators to better understand how to protect the health of young people in the US.

- D The *National Longitudinal Transition Study-2* (NLTS2), commissioned to begin in 2001 by the U.S. Department of Education, Office of Special Education Programs (OSEP), is a follow-up of the original National Longitudinal Transition Study. The original NLTS was designed and conducted by SRI for OSEP from 1985 through 1993 [SRI INTERNATIONAL, (1991)]. NLTS2 includes 12,000 youth nationwide who

were aged 13 through 16 at the start of the study (2000). Information will be collected over 10 years from parents, youth, and schools and will provide a national picture of the experiences and achievements of young people as they make the transit into early adulthood. The study will:

- Describe the characteristics of secondary school students in special education and their households.
- Describe the secondary school experiences of students in special education, including their schools, school programs, related services, and extracurricular activities.
- Describe the experiences of students once they leave secondary school, including adult programs and services, social activities, etc.
- Measure the secondary school and postschool outcomes of students in the education, employment, social, and residential domains.
- Identify factors in students' secondary school and postschool experiences that contribute to more positive outcomes.

D *National Longitudinal Study of Children and Families in the Child Welfare System* (The Research Triangle Institute (RTI))

In 1996, Congress directed the Secretary of Health and Human Services (HHS) to conduct a national study of children who are at risk of abuse or neglect or are in the child welfare system. The study was to include a longitudinal component; collect data on the types of abuse or neglect involved, agency contacts and services, and out-of-home placements; and yield reliable State-level data for as many States as feasible. In response, the Administration on Children, Youth, and Families (ACYF) awarded a contract for the National Longitudinal Study of Children and Families in the Child Welfare System. The purpose of the study is to describe the outcomes experienced by children and families who come to the attention of the child welfare system, and to gain an understanding of the factors, including system-level and service factors, as well as child and family characteristics, that contribute to those outcomes. This study will make available for the first time nationally representative longitudinal data drawn from first-hand reports from children and families or other caregivers, as well as service providers.

The study will provide information about 1) characteristics of children and families most likely to come into and stay in the child welfare system versus those who do not; 2) the service needs of children and families; and 3) the kinds of services actually provided and the factors that determine the kinds of services provided. In the longer term, the study will determine the outcomes for children and families who enter the system, examining the relationship between services provided and outcomes achieved for various types of children and families and how the children and families change over time.

The sample will represent the population of children and families who enter the child welfare system, and will include more than 6,000 children from 92 child welfare agencies nationwide. Participants will be children and families who entered the

child welfare system within a 15-month period, from late 1999 through the end of 2000. The study has collected child-and family-level data from children in the system, their caregivers, their caseworkers, other appropriate agency personnel, and administrative records. The first of three rounds of data collection began in fall 1999, with follow-up collections conducted 12 months and 18 months post-baseline. A fourth round, at 36 months post-baseline, is in the planning stage, and is contingent upon new funding. Both children who remain in the system and those who leave the system will be followed.

- D The *National Survey of Families and Households* (NSFH) consists of a national sample of 13,007 persons, including a main cross-section of 9,637 households plus an oversampling of blacks, Puerto Ricans, Mexican Americans, single-parent families, families with step-children, cohabiting couples and recently married persons. One adult per household was randomly selected as the primary respondent. Several portions of the main interview were self-administered to facilitate the collection of sensitive information as well as to ease the flow of the interview. The average interview lasted one hour and forty minutes. In addition, a shorter self-administered questionnaire was given to the spouse or cohabiting partner of the primary respondent.

A considerable amount of life-history information was collected, including: the respondent's family living arrangements in childhood, departures and returns to the parental home, and histories of marriage, cohabitation, education, fertility, and employment. The design permits the detailed description of past and current living arrangements and other characteristics and experiences, as well as the analysis of the consequences of earlier patterns on current states, marital and parenting relationships, kin contact, and economic and psychological well-being.

NSFH is funded by the National Institute of Child Health and Human Development (NICHD) and by the National Institute on Aging (NIA)

- D The *Wisconsin Longitudinal Study* (WLS) is a long-term study based on a random sample of 10,317 men and women who graduated from Wisconsin high schools in 1957. A companion sample contains comparable data for a randomly selected sibling of most respondents. Data collection was in 1957, 1964, 1975, 1977, 1992 and 1993. WLS data cover social background, youthful aspirations, schooling, military service, labor market experiences, family characteristics and events, social participation, psychological characteristics, health and well-being, and retirement. The Wisconsin Longitudinal Study is supported by the National Institute on Aging and run by the University of Wisconsin-Madison.
- D The *Baltimore Longitudinal Study of Aging* (BLSA), conducted by The NIA is America's longest-running scientific study of human aging, begun in 1958. BLSA scientists are learning what happens as people age and how to sort out changes due to aging from those due to disease or other causes. More than 1,400 men and women are study volunteers. They range in age from their 20s to their 90s.



### *Supranational*

- D The *European Community Household Panel* (ECHP) survey was launched by Eurostat in 1994. A sample of 60,000 households in 12 countries was identified that year. Another 13,000 households have been added since then in countries new to the European Union: Austria, Finland and Sweden; and Norway has also joined the group from 1998. All the adults in each household in the sample were initially interviewed in detail about their family structures, their employment and housing, and a wide range of other subjects. The same respondents have been re-interviewed each year since then, forming a representative 'panel' which can be used to follow the experiences of individuals, and whole families, over time. The annual interviews have continued through to 2001; the series has now ended. For Germany, Luxembourg and the UK, the original ECHP-samples have been discontinued after wave 3 (1996) - in these cases, data from existing national panels (SOEP, PSELL, BHPS) are used as a basis for "cloning" ECHP-data. When all the data are released, it will provide continuous coverage over an eight-year period. Eurostat proposes to replace it with a harmonised series of national data sets known collectively as the *European Union Survey of Income and Living Conditions* (EU-SILC).<sup>16</sup>

### **3. Inventory Statistics Netherlands**

At Statistics Netherlands, several kinds of longitudinal data sets are available. Not all of these sets were designed with the intent of gathering longitudinal information, but they can be used for this purpose. Some of the most important data sets containing longitudinal data are:

- a. the Regional Income Survey and the Income Panel Survey. For these surveys, data from fiscal registers are used. They contain information about several kinds of income in combination with some demographic variables.
- b. the Municipal Population Registers. These registers contain demographic variables for every resident of the Netherlands (approximately 17 million persons, since 1995).
- c. the Labour Force Survey. This survey uses a rotating panel to obtain information about the labour market situation.
- d. student cohort data about educational careers. For secondary education these data are collected by surveys, for higher education several registers are used. For secondary education, sampled students are followed annually for a maximum of 13 years. Both data from students and teachers are used. This type of longitudinal survey is carried out since 1965. For higher education, annual cohorts, since 1990, are followed administratively, as long as the students remain within the education system.

---

<sup>16</sup> It is not clear how much longitudinal possibilities these data sets offer.

- e. the registers on jobs and social benefits. These registers contain information for respectively, every job and every benefit in the Netherlands linked to persons by social security numbers.
- f. the Employment and Wages Survey. This is a business survey to provide information on the wages of employees (identified by social security numbers). Many businesses occur every year in the survey.
- g. the Social Statistical Database. In this database, data from administrative sources and surveys are linked. The database is used for the virtual census in 2001 instead of a door-to-door census and contains information about persons, jobs and social benefits. Longitudinal information can be obtained from the database, though it was not designed for this purpose.
- h. the Medical and Health Care Database. This database is still under development and contains information on individual level of hospital care and health care insurance, linked to the Municipal Population Registers.

The data sets mentioned above are the most obvious sources of longitudinal data at Statistics Netherlands, but they are not the only ones. For example, there is a strategic program on social dynamics in relation to the labour market. Within the framework of this program, several special longitudinal data sets are created by linking person data from for example the Labour Force Survey and the Regional Income Survey. Most of the (longitudinal) research is in progress, some publications have appeared, but are only available in Dutch. Many of these publications result from co-operation with a university. See BOUMAN(a,b) (2004), CORPELEIJN (2004), FLORQUIN (2004), FOUARGE (2004), MANTING (2004), NICOLAAS (2004), SPRANGERS (2004), TRAAG (2004).

Besides this strategic program, some longitudinal analyses are carried out in several fields. They are mostly concerned with unemployment, benefits and demographic topics. In general, the methods used are not very sophisticated. Sometimes, tables of transitions are used, but mostly, some kind of survival analysis is done, using for example (Cox) proportional hazard analyses or logistic regression. There are some publications about these applications, for example, Ekamper and Van Huis (2002), Gorter (1991), Janssen (2002), Klaus and Trimp (1996), and Opstal and Van de Pol (1991). Unfortunately, most of the publications are in Dutch. Besides these applications, there are some publications about methodology and methods related to longitudinal issues, such as Van de Pol (1989 and 1993), and Van Woudenberg (1996).

Another category are data sets that are not designed for and are not used longitudinally, but have inherent longitudinal information. E.g., at Statistics Netherlands, data for businesses are available which are partly longitudinal in character. For short-term statistics, a panel survey is used which is partially refreshed at the start of each year. However, the questions (often just one question) mainly concern turnover figures. To assess the quality of statistics at business level, information from earlier moments is used. No further longitudinal use of the data is made. The surveys used for long-term statistics (yearly figures) were not designed as panel surveys, but many businesses participate for several years.

In brief, methods for longitudinal data analysis are being used at Statistics Netherlands, but not much in the regular production of statistics.

#### 4. Summary and conclusions

The examples of longitudinal datasets and analyses described in this report can be distinguished primarily in two categories: *register based* and *survey/questionnaire based*. The first category can be found mainly in the Scandinavian countries, the latter one root in the Anglosaxon countries. This distinction is important for the view that can be applied to the data: survey based longitudinal studies must be planned and organized *prospectively*, by setting out and following a cohort or panel. Register based longitudinal studies have a “helicopter view” on the data over time, and cohorts can be followed *both prospectively and retrospectively*. Generally, registers have usually a smaller set of variables than questionnaires, and the variables are not always geared to the research question at hand. Of course, there are mixed varieties, where register data are supplemented with survey data.

Another major distinction is that between *person oriented* and *business oriented* longitudinal studies. The vast majority of studies that we found for inclusion in this inventory are person oriented studies. Their approach can reasonably be described as sociological, economical, educational, medical or multi-disciplinary. A special category is multi-generational. The range of topics covered consequently is very broad; due to the effort of following the cohort the number of participants usually is small (< 10 thousand persons), with some exceptions. The observation period varies between a few months to several decades.

The business oriented surveys face a challenge in following enterprises over time, as the identity of enterprises may not be constant over time (mostly due to mergers, split-ups, or change in activity.) Following enterprises over time is as complex as following households, rather than persons over time.

From a methodological point of view, often the additional complexities of longitudinal analysis as compared to cross-sectional ones are emphasized. Yet, in perspective, the additional complexity is not overwhelming. With a standard cross-sectional survey, the hardest issue is usually the non-response (compared to the sample design); for a panel based survey the initial non-response is followed by additional non-response (including attrition) in each wave. With register based longitudinal studies, the attrition depends heavily on the quality of the register. As a rule, the additional non-reponse is smaller than the initial non-reponse.

It is not easy to summarize the methodology of the studies in this report: as indicated in the introduction, the methodology is quite often not described explicitly, but implicitly. Or the methodology used depends on the question at hand, and is only described in the analytic report, or in the publications of the academic partners. We did not have enough time to read those.

It is our impression that in general, for the design of the data collections and the routine statistical publications, rather classical (longitudinal) methodology has been used (such as survival analysis methods, including life table techniques, proportional and Cox hazard models, and linear and logistic regression.) The more advanced methodology that likely will have been used for special analyses was out of sight for us. In many studies different weights for cross-sectional and longitudinal raising were used, but it is understood that optimising for cross-sectional purposes does not give optimal results for longitudinal purposes and vice-versa.

The purpose of this report is to describe a certain methodological approach in *Statistical Offices*; these come in a variety of constellations. There is variety in the scope of subjects covered by an NSI, and in the level of analytic activity permitted. Countries that have a central NSI, covering all topics, tend to be restrictive in the depth and detail of the single topics. Countries that have statistical services in many government departments, might allow for more depth, detail, or interest in course over time of the topics of their focus. As a third factor, the varying degree in which academic institutions in a country have access to micro-data of NSI's, influence the use of, and consequently the motivation for, longitudinal data collection.

What does all this mean for Statistics Netherlands? At Statistics Netherlands, currently most longitudinal data sets are based on registers. Only the Labour Force Survey (among persons) and the Employment and Wages Survey (among enterprises) are surveys with a panel composition. Most of the register-based longitudinal data sets are available for a couple of years, but rarely more than 10 years. These register data were meant primarily for cross-sectional purposes. Consequently, the longitudinal exploitation of these data sets is still in development. The register data can be combined with survey data, but the longitudinal aspect must come from the register part, as the same persons usually are not repeated in the surveys. Still, the (in itself non-longitudinal) survey part of the linked data can serve as one measurement in time, usually the starting or ending point of a (register defined) cohort. E.g., leaving school, becoming unfit for work (disabled), death. This type of questions does not fit in the traditional way of presentation of official statistics, and has to struggle to find a place in the work programme. Such a struggle is not specific for NSI's: also in academia there has been a lag of several decades between the moment that it became clear that "for this type of question we really need a longitudinal approach" and the moment that there are a sufficient number of running longitudinal studies [SCHAIE 2001].

## 5. Appendices

### 5.1 References

- ANONYMOUS (1999). *Business Growth and Performance Survey; small and medium enterprises*. Canberra, doc. nr 8141.0.1997-98.
- ANONYMOUS (s.a.) Appendix 16 : *The Swedish Survey of Living Conditions, design and methods*. [English version of a Swedish original]. Stockholm: Statistics Sweden.
- ANONYMOUS(A) (2001). *A Longitudinal Survey of Income, Employment and Family Dynamics*; Feasibility Project; Final Report. Wellington : Statistics New Zealand. ISBN 0-478-20779-4.
- ANONYMOUS(B) (2001). The future development of the Swedish Register system; Final report of the Register Project and decisions of the Director-General. Statistics Sweden, Stockholm, ISSN 0283-8680, 2001.
- BOUMAN A.M. (2004). Financiële gevolgen van de beëindiging van ongehuwd samenwonen voor man en en vrouw. *Bevolkingstrends*, 2004-III: 67-70. CBS, Voorburg/Heerlen.
- BOUMAN A.M. (2004). Financiële gevolgen van echtscheiding voor man en vrouw. *Bevolkingstrends*, 2004-II: 19-23. CBS, Voorburg/Heerlen.
- CARINO-ABELLO A., PEDERSON D., KING A. (2000). *Labour Market Dynamics in Australia; an application using the 1994-97 Survey of Employment and Unemployment*; Occasional Paper. Canberra, doc. nr 6293.0.00.006.
- CORPELEIJN A., KERKHOFS M. (2004). Arbeidsmarkttransities van oudere werknemers 2000-2001. Overgangen tussen werken en niet-werken, wisseling van baan, bedrijf en bedrijfstak. *Sociaal-economische trends*, 2004-IV: 19-29. CBS, Voorburg/Heerlen.
- DEEG D.J.H.. *Experiences from longitudinal studies of aging; conceptualization, organization, and output*. Nijmegen: Netherlands Institute of Gerontology, 1989.
- EKAMPER P., HUIS M. van (2002). Verhuizingen vanwege veranderingen van de huishoudenssamenstelling. *Maandstatistiek van de bevolking*, no 8: 14–17. CBS, Voorburg/Heerlen.
- FLORQUIN F. (2004). Economie en techniek meest lonende studies. *Sociaal-economische trends*, 2004-I: 38-46. CBS, Voorburg/Heerlen.
- FOUARGE D., SCHILS T., HUYNEN B. (2004). To retire or continue working? An analysis of the early retirement behaviour of Dutch workers. OSA publicatie A207, ISBN 906566 2723.

- GILES Philip (1999). *Overview of the Survey of Labour and Income Dynamics (SLID)*. Workshop on Longitudinal Research in Social Science: A Canadian Focus, Windermere, Manor, London, Ontario, Canada, October 25-27, 1999.
- GORTER D. (1991). *The relation between unemployment benefit, reservation wage and search duration*. CBS-report, BPA-no 2663-91-M1. CBS, Voorburg/ Heerlen.
- HATTERSLEY L., Creeser R. (1995) *The Longitudinal Study 1971 to 1991: History, organisation and quality of data, OPCS series LS no. 7*, London: The Stationery Office (formerly HMSO), ISBN: 0 11 691637 0).
- JANSSEN J.P.G. (2002). Gemengd huwelijk en echtscheiding: een dynamische analyse op registergegevens. *Maandstatistiek van de bevolking*, no 9: 4-10.
- KLAUS J., Trimp L. (1996). *Sociale uitkeringen en werkhervatting bij paren*. CBS-nota, BPA-no H01518-96-SIV. CBS, Voorburg/Heerlen.
- KOPPES L.L.J., KEMPER H.C.G. Review of AGAHLs and Other Observational Longitudinal Studies on Lifestyle and Health from Adolescence into Adulthood. *Med. Sport. Sci. (Basel)*, 2004;47, pp. 21-29.
- KOVACEVIC Milorad, ROBERTS Georgia (2001). Analysis of Duration Data from Statistics Canada's Longitudinal Surveys: What Methods are Recommended? *Discussion paper presented at the Advisory Committee on Statistical Methods, April 11, 2001*.
- KUZMICICH Geoff, WIGBOUT Max (2001). *A longitudinal look at some data of the Household Labour Force Survey*. Wellington: Statistics New Zealand. ISBN 0-478-20778-6
- MANTING D., LOEVE J.A. (2004). Economic circumstances and union dissolution of couples in the 1990s in the Netherlands. *Discussion paper 04006*. CBS, Voorburg/Heerlen.
- MICHAUD Sylvie (1999). *The National Longitudinal Survey of Children and Youth - Overview and Changes after Three Cycles*. Workshop on Longitudinal Research in Social Science: A Canadian Focus, Windermere, Manor, London, Ontario, Canada, October 25-27, 1999.
- NICOLAAS H., SPRANGERS A., ZORLU A., HARTOG J. (2004). Migranten: wie komen, wie gaan terug en wie laten hun gezin overkomen? *Bevolkingstrends*, 2004-II: 36-42. CBS, Voorburg/Heerlen.
- OPSTAL R. van, POL F. van de (1991). *The transition from school to work in 1979-1987 in the Netherlands*. CBS-report, BPA-no 3256-90-M1-1. CBS, Voorburg/Heerlen.
- POL, F. van de (1989). *Issues of design and analysis of panels*. Sociometric research foundation, Amsterdam.
- POL, F. van de (1993). *Weighting panel survey data*. CBS-report, BPA-no 6040-93-M1. CBS, Voorburg/Heerlen.

- ROBERTS Georgia, KOVACEVIC Milorad, MANTEL Harold, PHILLIPS Owen (2001). Cross-sectional Inference Based on Longitudinal Surveys: Some Experiences with Statistics Canada Surveys. *Federal Committee for Statistical Methods Conference, Washington, October 2001*, Working Paper 34, Part 4, p. 25-34.
- SCHAE K.W., HOFER, S.M. *Longitudinal Studies in Aging Research*, p 53 77 in: BIRREN JE, SCHAE KW. *Handbook of the Psychology of Aging*. San Diego, etc: Academic Press, 2001.
- SPRANGERS A., ZORLU A., HARTOG J., NICOLAAS H. (2004). Immigranten op de arbeidsmarkt. *Sociaal-economische trends*, 2004-II: 27-37. CBS, Voorburg/Heerlen.
- SRI INTERNATIONAL (1991). *Youth With Disabilities: How Are They Doing? The First Comprehensive Report from the National Longitudinal Study of Special Educational Students*. Washington, DC: SRI International.
- THORSLUND Mats, WÄRNERYD Bo (1985). *Methodological research in the Swedish Survey of Living Conditions. Problems in Measurement and Data Collection*. Social Indicators Research, 16 (1985) 77-95.
- TRAAG T., VALK J. van der, VELDEN R. van der, VRIES R. de, WOLBERS M. (2004). Een verklaring voor het effect van opleiding op de arbeidsmarktpositie van schoolverlaters. *Sociaal-economische trends*, 2004-III: 26-33.
- WALLGREN Anders, WALLGREN Brit (2002). *Estimation in Register Statistics*. The second Seminar on Strategies for Social and Spatial Statistics, Statistic Denmark, Copenhagen, 5 September 2002.
- WEBBER Maryanne, SHERIDAN Mike (1998). *Labour market dynamics: information from new Canadian longitudinal sources*. Statistics Canada, cat. Nr 98-15, prod. Nr. 75F0002M.
- WOUDENBERG, T.A. van (1996). *Record linkage of files of employment offices*. CBS-nota, BPA-no 8093-96-RSM. CBS, Voorburg/Heerlen.
- YEO Douglas (1999). *After the First Steps: The Evolution of the National Population Health Survey*. Workshop on Longitudinal Research in Social Science: A Canadian Focus, Windermere, Manor, London, Ontario, Canada, October 25-27, 1999.

## 5.2 List of acronyms

Studies in *italics* are not described, only mentioned, in this report.

ABS	Australian Bureau of Statistics
ACYF	Administration on Children, Youth, and Families
BCS70	<i>Birth Cohort Study 1970</i>
BHPS	<i>British Household Panel Survey</i>
BLS	Business Longitudinal Survey
BLSA	Baltimore Longitudinal Study of Aging
CATI	Computer assisted telephone interview
CATI	computer assisted telephone interviews
CBS	Centraal Bureau voor de Statistiek, a.k.a. Statistics Netherlands
CeLSIUS	Centre for Longitudinal Study Information and User Support
CHAID	Chi-square Automatic Interaction Detection
CPS	Current Population Survey
DC	District of Columbia
DIW	Deutsches Institut für Wirtschaftsforschung
ECHP	European Community Household Panel
EIC	Employment and Immigration Canada
ESRC	Economic and Social Research Council
EU-SILC	European Union Survey of Income and Living Conditions
GSOEP	<i>German SOEP, see SOEP</i>
HANES	See NHANES
HHLFS	Household Labour Force Survey
HHS	Department of Health and Human Services
HMSO	Her Majesty's Stationary Office
HRDC	Human Resources Development Canada
ISBN	International Standard Book Number
ISSN	International Standard Serial Number
JCLR	Joint Centre for Longitudinal Research
JISC	Joint Information Systems Committee
JUVOS	Claimant Unemployment Cohort
LDS	Longitudinal Data Set
LFS	(Swiss) Labour Force Survey
LINDA	Longitudinal Individual Data set (for Sweden)
LisNZ	Longitudinal Immigration Survey New Zealand
LISREL	software package for structural equation modeling
LMAS	Labour market activity survey
LS	Longitudinal Study
LSIA, -1, -2	Longitudinal Study of Immigrants to Australia (study 1 and 2)
LSIC	Longitudinal Survey of Immigrants to Canada
LSIEFD	Longitudinal Survey of Income, Employment and Family Dynamics
NCDS	<i>National Child Development Survey</i>
NCES	National Center for Education Statistics (USA)
NESARC	National Epidemiologic Survey on Alcohol and Related Conditions
NHANES	National Health and Nutrition Examination Survey
NHEFS	NHANES I Epidemiologic Follow-Up Study
NIA	National Institute on Aging
NIAAA	National Institute on Alcohol Abuse and Alcoholism
NICHD	National Institute of Child Health and Human Development



NICHD	National Institute of Child Health and Human Development
NLMS	National Longitudinal Mortality Study
NLS	National Longitudinal Surveys, <i>see</i> NLSY79, NLSY97, NLSW
NLSCY	National Longitudinal Survey of Children and Youth
NLSW	National Longitudinal Surveys of Young Women and Mature Women
NLSY, -79, -97	National Longitudinal Survey of Youth 1979 and 1997
NLTS, -1, -2	National Longitudinal Transition Study
NLU	(Swedish) Level-of-Living Survey
NPHS	National Population Health Survey
NSFH	National Survey of Families and Households
NSHD	<i>National Study of Health and Development (=British birth cohort study 1946)</i>
NSI	National Statistical Institute (generic term)
OCHS	Ontario child health survey
ONS	Office for National Statistics (UK)
OPCS	Office of Population Censuses and Surveys
OSEP	Office of Special Education Programs (USA)
PSELL	<i>Panel Socio-économique / Liewen zu Lëtzebuerg</i>
PSID	<i>Panel Study of Income Dynamics</i>
PTT	Post, Telegraph, Telephone
RTI	Research Triangle Institute (USA)
SEUP	Survey of Employment and Unemployment
SILD	Survey of Income and Labour Dynamics
SIPP	Survey of Income & Program Participation
SLFS	School Leavers Follow-up Survey
SLID	Survey of Income and Labour Dynamics
SOEP	Sozial Oekonomisch Panel
SoFIE	Later name of LSIEFD
SPSS	Statistical Package for the Social Sciences, trademark of SPSS, Inc.
SRI	SRI International, Stanford Research Institute
ULSC	United Kingdom Longitudinal Studies Centre
WES	Workplace and employee survey
WLS	Wisconsin Longitudinal Study

### 5.3 Not Included & other sources of information

#### *Studies not included:*

The following studies are not included in this survey for the stated reasons. This list could be extended by great lengths. These are the studies that we encountered, and the list serves the purpose of helping the readers (and us!) who have only heard about a certain study and wonder why it is lacking in this survey.

AUS — The *Longitudinal Survey of Australian Youth* (LSAY) (and predecessors) is managed jointly by the Australian Council for Education Research and the Commonwealth Department of Education, Science and Training. ABS seems not involved.

CAN – *Aging in Manitoba longitudinal study (AIM)*, run by University of Manitoba.

CAN – *Longitudinal immigration database*, run by Citizenship and Immigration Canada.

CAN – *Longitudinal study of immigrants, 1969-1971, 1976*, run by Dept. Of Manpower and Immigration, later by Employment and Immigration Canada.

CAN – *Social change in Canada, 1977-1981*, run by York University.

CAN – *Canadian study of health and aging, 1991-1996-2001*, run by University of Ottawa, sponsored by Health Canada.

CAN – *Canadian out-of-employment survey, 1995*, run by Human Resources Development Canada (and distributed by Statistics Canada)

F – The *French Household Panel*, or *Enquête Socio-Economique auprès des Ménages Lorrains* (ESEML), also known as *Panel des Ménages Lorrains*, because we could not find sufficient information about this study. INSEE is co-sponsor.

UK – *National Child Development Survey (NCDS)* is run by the Institute of Education, not by ONS.

UK – *British Household Panel Survey (BHPS)* run by University of Essex.

US – *Panel Study of Income Dynamics (PSID)* by the Institute of Social Research, Ann Arbor, sponsored by a mix of private and federal government institutions.

US – *National Survey of America's Families*, by the Urban Institute, Washington, DC.

D – *German Socio-economic Panel (GSOEP)* is run by the Deutsches Institut für Wirtschaftsforschung.

*Reviews of longitudinal studies in particular area's* (suggestions for enlargement of this list welcome):

Aging: see DEEG (1989), SCHAIE (2001)

Growth and development: see KOPPES (2004)

*Websites with lists of longitudinal studies (URLs lastly verified November 2004):*

<http://iserwww.essex.ac.uk/ulsc/keeptrack/index.php> Keeping Track project of the ULSC (as mentioned before).

[http://www.chass.utoronto.ca/datalib/other/longitudinal\\_DATASETS.doc](http://www.chass.utoronto.ca/datalib/other/longitudinal_DATASETS.doc) : List of longitudinal, panel, and time-sensitive datasets.

<http://psidonline.isr.umich.edu/Guide/RelatedSites.html> This page provides link to a list of long-term panel data sets in different countries that could be used for cross-national comparison with the PSID.

[http://www.csha.ca/r\\_relevant\\_site\\_links.asp](http://www.csha.ca/r_relevant_site_links.asp). This site provides a list of long-term panel data sets in the field of aging.

## 5.4 Index

- Aging in Manitoba longitudinal study, 32
- Baltimore Longitudinal Study of Aging, 22
- Birth Cohort Study 1970, 13
- British Household Panel Survey, 22, 32
- Business Longitudinal Survey, 4
- Canadian out-of-employment survey, 32
- Canadian study of health and aging, 32
- Claimant Unemployment Cohort, 14, 16
- Current Population Survey, 17
- Employment and Wages Survey, 23
- Enquête Socio-Economique auprès des Ménages Lorrains, 32
- European Community Household Panel, 22
- European Union Survey of Income and Living Conditions, 22
- French Household Panel, 32
- Household Labour Force Survey, 10
- Income Panel Survey, 22
- Labour Force Survey
- Canada, 8
  - Dutch, 23
  - Swiss, 12
  - UK, 14, 15
- Labour market activity survey, 8
- Level-of-Living Survey (Swedish), 12
- Longitudinal administrative database, 8
- Longitudinal Data Set, 5
- Longitudinal immigration database, 32
- Longitudinal Immigration Survey New Zealand, 10
- Longitudinal Individual Data set (for Sweden), 12
- Longitudinal Study (of OPCS), 14, 15
- Longitudinal study of immigrants, 32
- Longitudinal Study of Immigrants to Australia, 5
- Longitudinal Survey of Australian Youth, 32
- Longitudinal Survey of Immigrants to Canada, 7
- Longitudinal Survey of Income, Employment and Family Dynamics, 10
- Medical and Health Care Database, 23
- Municipal Population Registers, 23
- National Child Development Survey, 13, 32
- National Epidemiologic Survey on Alcohol and Related Conditions, 18, 19
- National Graduate Surveys, 7
- National Health and Nutrition Examination Survey, 18
- National Institute of Child Health and Human Development, 19
- National Institute on Alcohol Abuse and Alcoholism, 18
- National Longitudinal Mortality Study, 17
- National Longitudinal Study of Adolescent Health, 19
- National Longitudinal Study of Children and Families in the Child Welfare System, 20
- National Longitudinal Study of the High School Class of 1972, 17
- National Longitudinal Survey of Children and Youth, 6
- National Longitudinal Survey of Youth 1979 and 1997, 16
- National Longitudinal Surveys, 16
- National Longitudinal Surveys of Young Men and Older Men, 16
- National Longitudinal Surveys of Young Women and Mature Women, 16
- National Longitudinal Transition Study, 20
- National Population Health Survey, 6
- National Study of Health and Development, 13
- National Survey of America's Families, 32
- National Survey of Families and Households, 21
- NHANES I Epidemiologic Follow-Up Study, 18

Ontario child health survey, 8	Social Statistical Database, 3, 23
Panel des Ménages Lorrains, 32	Socio-economic Panel (Germany), 22, 32
Panel Socio-économique / Liewen zu Lëtzebuerg, 22	Survey of Employment and Unemployment, 5
Panel Study of Income Dynamics, 32	Survey of Income and Labour Dynamics, 7, 8
Regional Income Survey, 22, 23	Survey of Income and Program Participation, 17
School Leavers Follow-up Survey, 7	Swiss Household Panel, 12
School leavers survey, 7	Wisconsin Longitudinal Study, 21
Social change in Canada, 32	Workplace and employee survey, 8