Gayle, Vernon, and Paul Lambert. "Quantitative longitudinal datasets." *What is Quantitative Longitudinal Data Analysis?*. London: Bloomsbury Academic, 2018. 13–30. *Bloomsbury Collections*. Web. 26 Apr. 2023. <<http://dx.doi.org/10.5040/9781472545268.0009>>.

Downloaded from Bloomsbury Collections, www.bloomsburycollections.com, 26 April 2023, 12:20 UTC.

Copyright © Vernon Gayle and Paul Lambert 2018. You may share this work for non-commercial purposes only, provided you give attribution to the copyright holder and the publisher, and provide a link to the Creative Commons licence.

2 Quantitative longitudinal datasets

In this chapter we will outline a range of quantitative longitudinal data resources that are suitable for social science research. The UK is home to the world's largest household panel study and longest-running large-scale birth cohort study. The Cohort and Longitudinal Studies Enhancement Resources (CLOSER) project aims to maximize the use of UK longitudinal data resources. Information on the UK's quantitative longitudinal data resources is regularly updated on the CLOSER website.¹ Rafferty et al. (2015) provide a useful introduction to UK data resources suitable for longitudinal data analyses.

Large-scale panel studies

The longest-running large-scale quantitative longitudinal dataset with a household panel design is the US Panel Study of Income Dynamics (PSID). PSID began in 1968 and is sometimes seen as the grandparent of all of the household panel surveys (Hill, 1992). The German Socio-Economic Panel (GSOEP) began in the mid-1980s (Burkhauser and Wagner, 1993). It follows a similar design to the PSID but benefitted from many of the methodological lessons that had been learnt. The GSOEP can be thought of as the elder sibling of the British Household Panel Survey (BHPS). Several nations now collect household panel data; these include the Household, Income and Labour Dynamics in Australia (HILDA) survey, the Swiss Household Panel (SHP) and the Canadian Survey of Labour and Income Dynamics (SLID) (see Wooden, Freidin and Watson, 2002; Budowski et al., 2001; Webber, 1994).

¹ See <http://www.closer.ac.uk/> (accessed 6 April 2016).

The British Household Panel Survey

The BHPS was a major panel study which began in 1991 (Taylor et al., 1996). It is available from the UK Data Archive (Study Number 5151). Members of the BHPS have now been incorporated into a larger study called Understanding Society (the UK Household Longitudinal Study), which will be discussed below. The BHPS was a multipurpose, or omnibus, survey that provided data resources suitable for investigating a wide range of research questions from across the social sciences. It followed a representative sample of individuals (the panel) for nearly two decades. The BHPS was based on yearly contacts with the households, usually in the autumn. It used an interview methodology where all members of sampled households aged 16 and above were interviewed for approximately 45 minutes each. There was an additional short household questionnaire which only one member of the household completed.

The BHPS sample was originally a stratified design drawn from records of postal addresses. All residents present at the selected addresses at the first wave of the survey were designated as panel members. These same individuals were re-interviewed each successive year. If they split away from the original household to form new households, they were still followed. All adult members of their new households were also interviewed. This was a very important feature of the BHPS design because it led to more wide-ranging data collection. In addition, the original survey was supplemented with new booster samples (which will be discussed below). The original BHPS survey contained 10,264 respondents (4,833 men and 5,431 women) in Wave A (1991) who were living in 5,511 households. By 2008 (Wave R) there were 14,419 respondents living in 8,144 households. By the time of the formal end of the BHPS, 43,272 adults had been interviewed at some time during the 18 waves of the survey.

The BHPS questionnaire covered a broad range of social science and policy interests. The topics included housing composition, housing conditions, residential mobility, education and training, health and the use of health services, labour market behaviour, socio-economic values, and income from employment, benefits and pensions. A core of questions were asked annually but some questions were asked less frequently. New questions were sometimes introduced and they often followed on from changing policies and changing research agendas. The variable

component of the BHPS has included questions on wealth and assets, additional health measures, ageing in retirement, quality of life, children and parenting, views on crime and neighbourhood, and social networks. Buck et al. (1994) showcase the wide range of topics that can be analysed using the BHPS data.

The BHPS was a prospective design with yearly contacts (i.e. an interview usually conducted each autumn). Time-constant information such as the person's ethnicity or their place of birth, which are not expected to change, were asked only at the first point of contact when a respondent entered the panel. The BHPS also contained retrospective data collection exercises. For example, respondents were asked about their marital histories, their fertility histories and their work-life histories. New entrants to the survey (e.g. someone entering a BHPS household in the third wave) may be asked retrospective questions out of the usual sequence, but considerable background work went into the harmonization of life history data that predates the survey.

A specific youth questionnaire was introduced into the BHPS in 1994, and this is known as the British Youth Panel. Young people entered this sample at age 11 and exited at age 15, when they entered the adult survey. The design is known as a 'rotating' panel design because the new 11-year-olds grew into the panel and 15-year-olds grew out of it. A more detailed account of the youth data is provided by Gayle (2005).

Towards the end of the 1990s following territorial devolution in the UK, extension samples were introduced into the BHPS. These are commonly known as 'booster samples'. Living in Britain was the original survey name of the BHPS and the booster samples are sometimes referred to as Living in Scotland, Living in Wales and Living in Northern Ireland. The extension samples allowed for territory (i.e. country)-specific research and comparative research within the UK to be undertaken. For example, using the Scottish booster sample it became possible to undertake both Scotland-specific analyses and research comparing Scotland with England. Laurie and Wright (2000) provide detailed information on the Scottish booster sample. In Wave I (1999) 1,458 new Scottish and 1,428 new Welsh households were added to the BHPS. In Wave K (2001) another booster sample in Northern Ireland was collected. Ermisch and Wright (2005) have compiled a special volume of research papers using the Scottish booster sample data.

UK Household Longitudinal Study – Understanding society

The BHPS was greatly extended and subsumed into the UK Household Longitudinal Study (UKHLS), which is also known as *Understanding Society* (Buck and McFall, 2012). The UKHLS shares common design features and data collection characteristics with the BHPS. It began in January 2009 and contains approximately 40,000 UK households, and is currently the largest household panel survey in the world. Hobcraft and Sacker (2011) provide an overview of the origins and development of the UKHLS. The UKHLS is available from the UK Data Archive (Study Number 6614).

The UKHLS is organized into a core of annual questions and topical modules which are rotated. The annual core makes up about half of the interview and collects information on important events that occurred between interviews. Many UKHLS questions were also asked in the BHPS and they facilitate continued longitudinal analyses of the BHPS data. An outline of the questionnaire content of the annual repeated measures, the rotating modules, the ethnic minority boost and the youth self-completion questionnaire is provided in Buck and McFall (2012 see Tables 1 and 2).

Buck and McFall (2012) highlight five key areas in which the UKHLS extends the research scope of the BHPS. First, the survey's very large sample size permits the study of research questions for which other longitudinal datasets are too small to effectively support analyses. For example, it better enables analyses at subregional levels and supports analyses of the geographical variations in policies (e.g. in the different territories of the UK). The large sample size also facilitates analyses of groups that usually have low coverage in nationally representative surveys, for example people living with disabilities or young mothers.

Second, the UKHLS contains an ethnic minority booster sample (which oversamples respondents from ethnic minorities) and collects additional measurements relevant to their social and economic experiences of living in Britain (see Berthoud et al., 2009). Third, data are collected on a wide range of research topics. While the main focus of the study is social and economic life, a much broader array of data are collected on health, well-being, psychological attributes, social attitudes and environmental behaviour.

Fourth, the UKHLS collects an innovative range of health indicators and biomeasures. For an introduction to the role and collection of these measures in social surveys, see Weinstein, Vaupel and Wachter (2007)

and Jaszcak, Lundeen and Smith (2009). Fifth, the UKHLS includes a special panel of approximately 1,500 households, which is known as the Innovation Panel (UK Data Archive Study Number 6949). The Innovation Panel is extremely valuable because it allows methodological developments to be tested within a realistic survey context, which mirrors the main UKHLS survey. The Innovation Panel informs decisions for the main survey but it also provides a platform for methodological research. Details on some of the early experimental and non-experimental studies conducted as part of the Innovation Panel are reported by Uhrig (2011).

The UKHLS comprises four main components. These are the general population survey, the ethnic minorities booster sample, the BHPS and the Innovation Panel. The ethnic minority booster sample was designed to provide samples of about 1,000 adults in five large minority ethnic groups in the UK (Indian, Pakistani, Bangladeshi, Caribbean and African) (Berthoud et al., 2009). Estimates from social surveys were used to identify geographical areas with an ethnic minority density of at least 5 per cent. Samples were drawn from these high-density sectors and households were initially screened for the presence of a member of a minority ethnic group.

Respondents from the BHPS entered the UKHLS in the second wave. The inclusion of the BHPS sample is scientifically important because it preserves the ongoing and already established panel of household data. The inclusion of the BHPS has lengthened the temporal coverage of the UKHLS data. As far as possible the Innovation Panel mirrors the main UKHLS sample. The Innovation Panel does not include households in Northern Ireland or households in the north of Scotland.

The overall target for the achieved initial sample was 40,000 households: approximately 26,000 from the general population sample, approximately 4,000 from the ethnic minority boost, approximately 8,400 from the BHPS and approximately 1,500 from the Innovation Panel. At the start of the study the total achieved number of households across the four samples was 39,802 containing 101,086 adults and children (Buck and McFall, 2012, 11). Further details are provided in Burton, Laurie and Lynn (2011).

The four main sample designs are similar and are multistage designs.² Lynn (2009b) details the unique features of each of the UKHLS samples,

² For an introduction to these methods of sampling and selection see (De Vaus, 2013) and for a more extended and technical review see Moser and Kalton (1971).

which we summarize below. The general population survey is a stratified, clustered equal probability sample of residential addresses drawn to a uniform design throughout all of the UK (including the north of Scotland). The Northern Irish sample is not clustered, however. Within Great Britain the primary sampling units are postal sectors. They are stratified by the nine official regions of England, plus Scotland and Wales, by population density and by ethnic minority density. This resulted in the systematic selection of 2,640 postal sectors with probability proportional to size (number of addresses). Within each sampled sector, 18 addresses were selected systematically, resulting in an equal probability sample of 47,520 addresses in Great Britain. In Northern Ireland 2,400 addresses were selected systematically from the Land and Property Services Agency list of domestic properties.

The overall response rate at the first wave for the general population survey was 57.2 per cent and for the ethnic minority boost was 56.9 per cent. The response rate for the general population was slightly lower than the 60 per cent target, but the ethnic minority boost was slightly higher than the 55 per cent target. These rates are common for omnibus surveys in the UK (Buck and McFall, 2012). Burton Laurie and Lynn (2011) provide in-depth information on the UKHLS response rates.

The constraints of undertaking such a large data collection exercise meant that the UKHLS fieldwork is spread over a 2-year period. The overall sample was partitioned into 24 monthly subsamples, each independently representative of the UK population. This means that differences over time within a wave can be compared using nationally representative subsamples, and it also means that quarterly or annual subsets of the data can be analysed independently.

There are three categories of respondents in the UKHLS. First, there are respondents who were in households at the start of the study, they are designated as original sample members (OSM). Every OSM will be followed throughout the life of the UKHLS as long as they continue to live in the UK. A child born to an OSM is automatically an OSM and is followed in the usual way. Second, there are individuals who join a household that includes an OSM. They are also interviewed in the study and they are designated as temporary sample members (TSM). For example, when an adult male OSM moves in with his female partner she will also be interviewed and will become a TSM. A TSM is not followed when they no longer live in a household with an OSM. The third type of respondent

is a permanent sample member (PSM). If a male TSM fathers a child with an OSM he becomes a PSM and remains potentially eligible for interview for the life of survey (Knies, 2014). This categorization of members and the 'following rules' are designed to mimic the demographic processes by which the population is reproduced. This includes births, deaths, partnership formations and dissolutions, and emigration. Therefore, the study's 'following rules' provide an organic selection method which over time represent the evolving pattern of families and households in the UK.

Computer-aided personal interviewing is the main method of data collection, although several data collection instruments are used. One household member completes the household enumeration grid, which collects general information and then provides the household interview, which takes about 15 minutes to complete. The adult household members (those aged 16+) undertake a personal interview which is usually completed in 32 minutes. The adult household members also undertake a self-completion questionnaire which usually takes less than 10 minutes to complete. There is a short proxy interview about adults who are unavailable for the interview. Information on children under the age of 10 is provided by a responsible adult (e.g. a parent) as part of the adult and household interviews. Respondents aged 10–15 are asked to fill in a self-completion questionnaire, which is a paper-and-pencil instrument.

Some versions of the UKHLS are available via secure access, for example a version with linked data on detailed localities and other selected characteristics (UK Data Archive Study Number 6676) (see Rabe, 2011). These new resources are beneficial and, for example, Clark et al. (2014) use the spatial data linked to the survey data to explore life events and travel behaviour.

The depth and breadth of research questions that can be investigated with the UKHLS data are easy to illustrate. For example, Green, Poland and Willis (2016) studied companionship and community networks of older LGBT adults. Davillas, Benzeval and Kumari (2016) used the UKHLS to examine the association between body mass index and mental health. Clark and Coulter (2015) examined personal attitudes and attachments to neighbourhood. Chan and Ermisch (2015) used the data to study how close 'middle-aged' couples live to their parents. Zwysen (2015) studied the effects of a father's worklessness on young adults in the UK. McAloney et al. (2014) used the data to examine the smoking habits of young people and their mothers.

Cohort studies

Cohort studies are common in the social sciences and in health and medical research. A cohort study is a special case of a panel study. The panel is made up of a group of individuals who share a common characteristic or experience. The birth cohort study is a specific type of panel study where a group of individuals born around the same time are studied. Birth cohort studies can be found as geographically far afield as the Aberdeen Children of the 1950s Cohort Study (see Batty et al., 2004), the Stockholm Birth Cohort Study (see Stenberg and Vågerö, 2006), the US Early Childhood Longitudinal Study (see Flanagan and West, 2004), Growing Up in Australia: The Longitudinal Study of Australian Children (see Gray and Sanson, 2005) and Growing Up in New Zealand (see Morton et al., 2013).

The major British cohort studies

Britain is unique insofar as it has an unparalleled collection of large-scale nationally representative long-running birth cohort studies. A useful resource for getting preliminary information on British cohort datasets is the Cohort Directory,³ which is a searchable tool for some of the main UK cohort studies. A useful review is also provided by the Medical Research Council (see MRC, 2014).

The first of the British birth cohort studies was the 1946 National Birth Cohort, MRC National Survey of Health and Development (NSHD). It is currently the longest-running birth cohort study in the world (Pearson, 2016). The study aimed to address social and health policy questions in the years directly preceding the establishment of the UK National Health Service in 1948. The initial survey was a systematic sample of babies born in a single week in March 1946.

Data collection was frequent during infancy and the school years because of the pace of development and growth. Initial data were collected at birth and then the participants were re-contacted at ages 2 and 4. These two early childhood sweeps of data collection were primarily concerned with socio-economic differences in the infant's growth, development, and morbidity, and with the effect of care and the family's socio-economic

³ See <https://www.mrc.ac.uk/research/facilities/cohort-directory/> (accessed 6 April 2016).

circumstances upon health outcomes. During the school years the measurement of growth and health continued alongside a new stream of data collection focusing on education. In the youth phase measures relating to educational outcomes were collected, alongside data on topics such as delinquency. In the early adult phase collecting data on education, occupations and income came to the fore. In the middle phases of adulthood the study's health data collections were re-oriented. The focus became the measurement of physical and mental function, morbidity and mortality.

Wadsworth et al. (2006) provide a profile of the 1946 cohort, and Pearson (2016) has recently produced a popular science book on the cohort. The NSHD is accessible from the MRC Unit for Lifelong Health and Ageing. The data in the 1946 cohort has led to several landmark studies. Douglas and Blomfield (1958) showcased the first waves of data from the NSHD and charted the pre-school lives of the cohort. The study revealed that social and economic circumstances were generally improving in Britain but there was still a high degree of social inequality that affected the health of children in disadvantaged families. Douglas (1964) focused on primary school experiences and the transition to secondary school. This study was important because it provided empirical evidence that parents and pre-school circumstances had a substantial impact on outcomes in primary school, and this revealed high levels of social inequality. Douglas, Ross and Simpson (1968) focused on experiences and outcomes in secondary education and revealed enduring social inequality in the British education system. More recently, and reflecting the collection of data across the life course, studies such as Kuh et al. (2013) have used the data to study healthy ageing.

The second of the major British birth cohort studies was the 1958 National Child Development Study (NCDS). It mirrored the design of the 1946 study. The initial aims of the study were to address concerns regarding stillbirths and neonatal deaths. The cohort is a systematic sample of babies born in one week of March 1958. The original survey was not planned as a longitudinal study but funding became available to retrace the cohort. Further data collection exercises took place when the cohort members were 7, 11, 16, 23, 33, 42, 46, 50 and 55 and data continue to be collected.

The omnibus nature of the data collected on the 1958 cohort means that it has become an invaluable data source on such diverse topics as the effects of socio-economic circumstances on health, social mobility and changes in social attitudes. The NCDS is an exceptional resource

for studying ageing and changes over the lifecourse. Power and Elliott (2006) provide a profile of the 1958 cohort. Butler and Bonham (1963) provide a useful account of the origins of the study, and Davie, Butler and Goldstein (1972) provide useful results from the early phase of the study. The data can be accessed from the UK Data Archive (Study Number GN 33004).

The NCDS data have been used in a wide variety of applications across social science disciplines and in health research. For example, Furstenberg and Kiernan (2001) used the data to compare children who experienced parental divorce in childhood with those who were young adults when their parents divorced, to examine long-term effects of divorce. Bartley, Kelly and Sacker (2012) examined the relationship between financial adversity in childhood and lung function in midlife, and Gregg and Tominey (2005) examined the effects of youth unemployment on earnings in later life.

The genealogy of publications also reflect the lifecourse of the cohort. Bynner (1999) studied youth transitions, Berrington and Diamond (2000b) studied marriage and cohabitation patterns, Elliott (2002) studied women's employment behaviour after childbirth and Deary et al. (2013) investigated cognitive ability in middle age. The flexibility of the NCDS is illustrated by studies such as Elliott (2013), who undertook qualitative biographical interviews with a subsample of 170 members of the cohort.

The third of the major British birth cohort studies was the 1970 British Cohort Study (BCS70). Following the initial birth survey the cohort were followed up at ages 5, 10, 16, 26, 30, 38 and 42. The BCS70 mirrored the design of the two earlier birth cohorts, and the initial sample consists of babies born in a single week of 1970. The BCS70 is an omnibus study and collects information on a broad spectrum of topics relating to social and economic life and health. Elliott and Shepherd (2006) provide a profile of the 1970 cohort, and Butler, Despotidou and Shepherd (1997) provide information on the early phase of the study. The data can be accessed from the UK Data Archive (Study Number GN 33229).

The BCS70 data have been used in a wide variety of applications across social science disciplines and in health research. For example, Stewart-Brown and Haslum (1988) studied partial sight in childhood. Bynner and Parsons (2002) examined school-to-work transitions. Cooksey et al. (2009) used the BCS70 data to examine the relationship between mothers' employment and child development in a comparative framework with US data. Cable, Sacker and Bartley (2008) investigated the effects of

employment on psychological health in mid-adulthood. Viner and Taylor (2007) examined the adult outcomes of binge drinking in adolescence.

There was then a gap in the collection of large-scale nationally representative birth cohort studies in Britain. The next major birth cohort study was the Millennium Cohort Study (MCS), which was set up to follow the lives of children born at the turn of the new century. Data have been collected at birth, 9 months and at ages 3, 5, 7 and 11 years. Fieldwork has now been completed for the age 14 survey and there are plans to carry out another survey at age 17. The MCS covers the usual topics that would be expected in a birth cohort study but also collects information on topics such as parenting, family income and social capital.

The design of the MCS differs from that of the earlier birth cohorts. It is a sample of babies born throughout the year. The MCS covers the whole of the UK rather than just the three territories of Great Britain, and has boosted samples in Scotland, Wales and Northern Ireland to allow adequate sample sizes for analyses within each territory. The sample is geographically clustered, with an over-representation of deprived areas, and this increases the scope for including community-level analyses (Smith and Joshi, 2002). A major difference between the MCS and the earlier birth cohorts is that it has a complex sample design and selection procedure (see Plewis et al., 2007, Ketende and Jones, 2011b). Connelly and Platt (2014) provide a profile of the MCS, and Hansen (2014) provides detailed information on the first five surveys. The data can be accessed from the UK Data Archive (Study Number GN 33359).

Dex and Joshi (2006) edited an extensive volume containing work on pregnancy, childbirth, social origins, childhood development, childcare, and employment and parenting. This was followed by Hansen, Joshi and Dex (2010), which is a more extensive volume of work covering the first five years of the lives of the MCS members. Platt et al. (2014) showcase findings from the age 11 survey.

The multidisciplinary nature of the MCS data supports a wide range of analyses. For example, Quigley, Kelly and Sacker (2007) investigated breastfeeding and hospitalization for diarrheal and respiratory infections. Kelly et al. (2009) examined variations in birth weight among ethnic groups. Sabates and Dex (2015) examined the impact of multiple risk factors on young children's cognitive and behavioural development. Connelly (2011a) examined the drivers of unhealthy weight in childhood, and Chanfreau et al. (2015) investigated the disparities in the take-up of school sports activities.

The similar structure and scope of the British Birth Cohorts has facilitated cross-cohort comparisons. Brown and Elliott (2010) offer an outline of the 1958 and 1970 cohorts, and demonstrate the potential for longitudinal analyses and cross-cohort comparisons. Bynner and Joshi (2002) studied inequality and opportunity in education using the NCDS and the BCS70 data. Breen and Goldthorpe (2001) studied social class, mobility and merit through a cross-cohort comparison of 1958 and 1970 data. Connelly (2011b) studied the influence of childhood test scores, and family background and occupational positions in adulthood using the NCDS and the BCS70.

Data from the 1946, 1958 and 1970 cohorts have been used by Bukodi and Goldthorpe (2011) to study returns to higher education and by Bukodi et al. (2015) to study social mobility. Chan and Boliver (2013) examined grand-parental effects on social mobility using these datasets. Fertig (2010) used data from the NCDS, BCS and MCS to examine smoking behaviour, and Connelly and Gayle (2015) used these three cohorts to investigate changing socio-economic inequalities in childhood cognitive test performances.

Other British birth cohort studies

In addition to the major birth cohort studies, the UK data portfolio also includes some other large-scale birth cohorts. Growing Up in Scotland (GUS) is a birth cohort study following the lives of a large sample of children born in Scotland. It began in 2005 and collects a wide spectrum of information, for example on family circumstances, child development, health, pre-school experiences, education, physical activity and parenting styles. A description of the study is provided by Anderson et al. (2007). The data can be accessed from the UK Data Archive (Study Number SN 5760).

The GUS data have been used in a variety of social science and health projects. For example, Parkes, Sweeting and Wight (2015) used the GUS data to investigate parenting stress and parent support among mothers with high and low education. Parkes, Sweeting and Wight (2016) used the data to examine early childhood precursors and school age correlates on internalizing problem trajectories in young children. Parkes, Sweeting and Wight (2014) examined family and school influences on children's social and emotional well-being. Skafida and Treanor (2014) examined how family income predicts change in children's diets.

The Avon Longitudinal Study of Parents and Children (ALSPAC) is a birth cohort study with a regional focus. It considers multiple genetic,

epigenetic, biological, psychological, social and other environmental exposures in relation to a similarly diverse range of health, social and developmental outcomes. The samples were recruited from the Avon area, which is in the Southwest of the UK. ALSPAC has a data-sharing and access policy and proposals to access data must be made through the ALSPAC Executive Committee.⁴ A profile of the cohort is provided by Boyd et al. (2012).

ALSPAC has generally supported medical and health research but has also been used effectively in social science research. For example, Melotti et al. (2011) studied adolescent tobacco and alcohol use. Jerrim et al. (2013) undertook an innovative study of the socio-economic gradient in children's reading skills by examining the role of genetics. Mokrysz et al. (2016) studied the effect of cannabis use on education and intelligence.

Other longitudinal studies

There are a number of other longitudinal studies within the UK social science research data portfolio. In this section we review a small number of these data resources to illustrate the diversity of the existing quantitative longitudinal datasets.

The Youth Cohort Study of England and Wales (YCS) was a major longitudinal study that began in the mid-1980s. It was a large-scale nationally representative survey funded by the government and was designed to monitor the behaviour of young people as they reached the minimum school-leaving age and either remained in education or entered the labour market. The YCS survey collected detailed information on the young person's qualifications and experiences of education, as well as information on employment and training. A limited amount of information was collected on the young person's personal characteristics, their family and circumstances at home. The YCS sample was nationally representative of Year 11 pupils in England and Wales. A large sample from an academic year group (a cohort) was contacted in the spring following school Year 11. The young people were usually aged 16–17 when they were first contacted. The main data collection instrument was a postal questionnaire. The cohort

⁴ See <http://www.bristol.ac.uk/alspac/researchers/access/> (accessed 14 June 2016).

members were usually re-contacted and surveyed on at least two further occasions (e.g. at ages 17–18 and 18–19).

Currently, 13 cohorts of YCS data can be accessed from the UK Data Archive (Study Number GN 33233). A special combined cohort dataset, which includes data on Scottish pupils, is also available from the UK Data Archive (Study Number 5765). The YCS was primarily a monitoring tool and was not specifically designed for social science research. A number of challenges are associated with analysing YCS data, most notably inadequate documentation (see Croxford, 2006). Despite these challenges a range of analyses usually orientated towards education and employment have been undertaken using the YCS. Drew, Gray and Sime (1992) is an example of analyses of the older cohorts, and Gayle, Murray and Connelly (2016) is an example of analyses that span a number of YCS cohorts.

The English Longitudinal Study of Ageing (ELSA) is a longitudinal study of a representative cohort of men and women living in England who were initially aged 50 years and above. ELSA was designed as a sister study to the US Health and Retirement Study. ELSA is multidisciplinary in its orientation and involves the collection of economic, social, psychological, cognitive, health, biological and genetic data. The study began in 2002 and data are usually collected at two-year intervals. Data are collected using computer-assisted personal interviews and self-completion questionnaires, with additional nurse visits for the collection of health and physical measures. ELSA is harmonized with ageing studies in other countries to facilitate international comparisons, and administrative data have been linked to the sample. Steptoe et al. (2013) provide a profile of the ELSA cohort. The data can be accessed from the UK Data Archive (Study Number GN 33368).

The Northern Ireland Cohort Longitudinal Study of Ageing (NICOLA) has recently been established. NICOLA will closely follow the comprehensive approaches undertaken by ELSA and The Irish Longitudinal Study of Ageing (TILDA). Anderson, Boyle and Sharp (2008) proposed a Scottish longitudinal study of ageing, and preparatory and pilot work for the Healthy Ageing in Scotland study is currently in progress. Kaiser (2013) provides a useful overview of a series of international studies and compares their methodologies and key policy themes and also outlines some key findings.

The UK is also uniquely served by three quantitative longitudinal studies based on census and administrative data. The longest running of the three is the Office for National Statistics Longitudinal Study (ONS-LS), which covers England and Wales. The ONS-LS is a longitudinal sample of around

1 per cent of the total population of England and Wales. It is a systematic sample of people born on four birthdates spaced across the calendar year. The ONS-LS contains data from the decennial UK Censuses and other sources including routine event registrations (e.g. births, deaths, immigration and embarkation). It has linked records at each UK Census since 1971 (for the people born on one of four selected dates in the calendar year). Therefore, the ONS-LS is usually considered as being a systematic sample. These four dates were used to update the sample at the 1981, 1991, 2001 and 2011 Censuses, and new ONS-LS members enter the study through birth and immigration if they are born on one of the four selected birth dates.

The ONS-LS began in 1971 and the dataset were created and are maintained by the ONS. Researchers can access the ONS-LS data either in approved safe environments or by remote access. Goldring and Newman (2010) provide a useful outline of the ONS-LS and a commentary on its research value. The data have supported a wide spectrum of social science and health analyses. For example, Platt, Simpson and Akinwale (2005) used the data to examine social change in minority ethnic populations. Sturgis and Buscha (2015) examined changes to social mobility. Boyle, Norman and Popham (2009c) used the data to examine social mobility and health, and Bartley and Plewis (1997) used the data to look at socio-economic differences in health.

More recently two longitudinal studies based on census and administrative data have been developed, one covering Scotland and another covering Northern Ireland. These two studies are similar in design to the English and Welsh data. Each of the three studies has a slightly different set of data access arrangements but follows a similar procedure and usually requires users to be trained, data analysis work to be carried out in a safe environment and research outputs to be given official clearance. Boyle et al. (2009b) provide a profile of the Scottish Longitudinal Study (SLS), and Ralston, Gayle and Lambert (2016) provide an example of a recent analysis using the SLS. O'Reilly et al. (2012) provide a profile of the Northern Ireland Longitudinal Study (NILS), and O'Reilly et al. (2015) provide an example of analyses using the NILS data.

Cross-national and comparative longitudinal studies

The collection of nationally representative social survey data in many different countries has sometimes enabled researchers to undertake

cross-national and comparative longitudinal analysis. Some social surveys are deliberately designed to facilitate cross-national research. Others originate as national surveys, and only at a later stage is an attempt made to bring their data into a cross-national framework (see Harkness, Van De Vijver and Mohler, 2003, Harkness et al., 2008). Not all cross-national surveys have a longitudinal element; however, reflecting the sustained and substantial work that is involved in developing cross-nationally comparative survey instruments, many of the designed cross-national social survey resources have been maintained over a long period of time. This makes them suitable for longitudinal analysis (i.e. they feature repeated cross-sectional, or repeated contacts, designs). In particular, there have been a great many endeavours in generating cross-nationally comparative repeated cross-sectional data that are suited to comparing trends over time across different countries. One notable example is the Luxembourg Income Study (see Smeeding, Jesuit and Alkemade, 2002). There are many other similar projects, some of which we will discuss further in Chapter 3.

Any cross-national survey design faces considerable challenges in establishing the comparability of measures across countries. Some survey measures (e.g. age and gender) may have a natural equivalence across societies, but many other measurements (e.g. income, welfare benefits and education) will not be equivalent, and in some cases there may not be any agreed-upon international standards for collecting or comparing these measures (Connolly, Gayle and Lambert, 2016a). Therefore, researchers using datasets from multiple countries are forced to put a great deal of work into assessing and ensuring the comparability and equivalence of survey measures (see Hoffmeyer-Zlotnik and Warner, 2014; Jowell et al., 2007). Johnson (1998) provides a useful overview of the methodological issues associated with undertaking research with cross-national datasets. Burkhauser and Lillard (2005) provide a valuable account of the scope and limitations of a number of projects that have endeavoured to bring together cross-national social survey datasets.

While most research using panel surveys focus on one particular country, the existence of similar panel studies in several nations presents opportunities for internationally comparative research. Research designs sometimes involve bespoke comparisons between a small number of panel studies from different countries, based upon access to and analysis of the original data (e.g. Hadjar and Samuel, 2015). There have also been some explicit attempts to construct cross-national and

comparative longitudinal panel survey datasets. Two notable examples are the European Community Household Panel (ECHP) and the European Union Survey of Income and Living Conditions (EU-SILC). The ECHP has now been discontinued (Frick et al., 2007). Lillard and Burkhauser (2006) provide an account of the Survey of Health, Ageing and Retirement in Europe, which is a multidisciplinary and cross-national dataset.

The Cross-National Equivalent File (CNEF) project is an attempt to bring together data, *ex post*, from ongoing panel studies. The origins of the CNEF are described in Burkhauser et al. (2000). The CNEF 1970–2013 contains equivalently defined variables from the BHPS, the HILDA Survey, the Korea Labor and Income Panel Study (KLIPS), PSID, the Russia Longitudinal Monitoring Survey (RLMS-HSE) (new in 2016), the SHP, SLID and the GSOEP⁵. It is likely that other studies will be included at a later date.

Conclusions

The steadily expanding wealth and volume of large-scale longitudinal social survey datasets means that many social research questions can be addressed by using these resources. Sometimes more than one data resource might be used to answer the same research question. In such cases it is not always clear which data resource is optimal, but considerations of the size of the dataset, and the resolution of the variables available in the dataset, are important. In general when the main interest is describing longer term social trends, repeated cross-sectional surveys are often the most feasible source of longitudinal data. For some studies, data collected from repeated contacts from the same individuals will be required. Cohort designs tend to favour the study of growth or development in specific contexts. Due to their design, household panels tend to be particularly flexible for investigating micro-level change in short time periods in a wider range of contexts.

⁵ See <https://cnef.ehe.osu.edu/> (accessed 4 April 2016).