Longitudinal Studies Strategic Review

2017 Report to the Economic and Social Research Council

◆ **Authors**

Pamela Davis-Kean, University of Michigan, Review Chair

Raymond L Chambers, NIASRA, University of Wollongong, Panel Member

Leslie L Davidson, Columbia University, Panel Member

Corinna Kleinert, University of Bamberg, Panel Member

Qiang Ren, Peking University, Panel Member

Sandra Tang, University of Michigan, Research Assistant

Contents

Acknowledgements

The strategic review would not have been possible without the efforts of many individuals who provided their time, knowledge, and advice across the review period. First and foremost, I would like to thank the international review panel (Ray Chambers, Leslie Davidson, Corinna Kleinert, and Qiang Ren) for their dedication and commitment to the review and persistence/ integrity in making sure that all voices were heard and represented in the report. I was also fortunate to have the expert assistance of Sandra Tang to organize the review and be a liaison between the review panel and the ESRC office. Individual panel members are grateful to their research assistants Elisa Katharina Himbert, Natasha McMaster and Min Qin.

I would also like to thank the exceptional work of the ESRC office (particularly Charlotte Campbell, Maria Sigala, James Fenner, Jo Lake, and Rebecca Perring) for their assistance throughout the review process, specifically, the scheduling, organizing, editing, and fact checking, which was expertly handled by this team. I would also like to note the dedication and counsel during this process of Rebecca Fairbairn, who always clarified the scope and scale of this review and helped to ensure that the review panel received information and comments from the breadth of those who care about or are involved with ESRC's longitudinal studies. It is difficult to imagine that this review would have been successful without the level of expert assistance provided by Bridget Taylor. She provided outstanding support and help throughout the review. I am grateful and indebted to her for her assistance in all aspects of this review.

The Review Panel represented an international perspective on ESRC's longitudinal studies, and it was important to have an expert group that understood the issues of the UK to clarify any specific UK issues to the panel. I am grateful to have the expertise of the ESRC Review Steering Group which was chaired by Anna Vignoles with members: Mary De Silva, Jane Falkingham, Gordon Harold, Debbie Lawlor, Nyovani Madise, Joe McNamara, and Tom Wells. The Steering Group represented the UK's broad interest in longitudinal studies and from the beginning of the review process, provided their expert knowledge on longitudinal studies, users, and policy atmosphere.

I am also highly indebted to the CLS, Understanding Society, and CLOSER who provided, often with little notice, an array of information on the longitudinal studies. Their devotion to providing strong, scientifically-based social science was apparent in all the interactions with the Directors (Alissa Goodman/CLS; Michaela Benzeval/Understanding Society; Alison Park/CLOSER) of these organisations as well as researchers and scientists that work in these units. I appreciate all the time and effort they spent to assist the review panel.

Finally, the review panel members met individually (in person or by phone) with over 80 individuals who provided their opinions, thoughts, and research on both the current and future situation of the longitudinal studies (**Appendix 1**). These meetings and interactions provided a rich perspective to the review panel and I am very grateful for the time and energy they generously gave to make the review as comprehensive as possible.

Pam Davis-Kean
Chair of the International Panel

Executive summary

Context

The 2017 Longitudinal Strategic Review was commissioned to review the continuing scientific needs for longitudinal research resources, how these needs could be met by the Economic and Social Research Council (ESRC), and to offer recommendations on strategic and innovative ways to enhance this portfolio in the future.

The last **strategic review**[1] occurred in 2006, and since then there have been changes in both what the ESRC has invested in regarding their longitudinal portfolio, as well as normative changes in the UK population and the policies and context that impact this population. It is therefore timely to review the longitudinal data strategy of the ESRC.

Even during this review, the structure of UK government research funding is changing: the ESRC will now be part of the larger research funding body **UK Research and Innovation** (UKRI), and the locus of decision-making will evolve. Other funders of longitudinal investments (notably, the **Medical Research Council** (MRC) and **Wellcome**) have updated their strategies for funding longitudinal and cohort investments, and the ESRC has worked closely with these funders to develop potential avenues for greater synergies across funders. There have also been advancements and developments in legislation to make access to other sources of data (administrative and medical data, social media, smart meter data, consumer data, and harmonisation of multiple datasets) more feasible, and in technologies and methodologies to increase the value of an already rich source of data for the longitudinal data community. Other opportunities also continue to emerge such as the **Life Sciences Industrial Strategy**[2] and the regional digital innovation hubs initiative. These changes pose both opportunities and challenges for ESRC's continued investment in longitudinal studies and for social science research more broadly in the UK.

Project overview

It is within this changing context that the independent international review panel has gathered information and views through an open online consultation and in-person workshop, information requests and invitations, participation in events and use of published material, and meetings with key stakeholders and experts, exploring the continuing needs for evidence and priorities for investment in ESRC's longitudinal studies. These methods of gathering data from the community of users and stakeholders led to the initial review questions being honed down to three major priority questions, which then became the focus of the review panel's deliberations:

• What are the scientific and policy-relevant content needs for longitudinal data in the future?

• What are the advantages and challenges to using longitudinal data enhancements such as administrative data linkage and harmonisation?

• How can training, access, and promotion of longitudinal investments be continued and enhanced for both traditional and non-traditional research communities?

These questions and a summary of key findings from the review are provided below followed by specific recommendations that have been ordered by the panel's views on priority for implementation.

Key findings

1. What are the scientific and policy-relevant content needs for longitudinal data in the future?

• The UK is recognised globally as having significant strengths in longitudinal data, thanks to historic investments by the ESRC and other funders over many decades. The investment in these resources gives the UK social science community a competitive advantage in understanding critical population trajectories over the life course and across changing contexts.

---

[1] Martin, J. et al. (2006). **Strategic Review of Panel and Cohort Studies: Report to the Research Resources Board of the Economic and Social Research Council**. London: Longview.

[2] UK Office of Life Sciences. (2017). **Life Sciences: Industrial Strategy report**.

- ESRC's combination of longitudinal cohort and household panel studies covers the most urgent research questions of the scientific community, and provides an adequate data landscape to study a broad range of research and policy-related questions on different aspects of the life course. Government analysts consulted in the course of the review were overwhelmingly positive about the value of longitudinal data in informing policy.

- It is clear that only longitudinal studies can answer certain questions. Panel studies, for example, are particularly useful for estimating rapid changes and answering some policy issues, while cohort studies are particularly useful for providing rich information on different aspects of the life course.

- However, longitudinal surveys are expensive to maintain and the birth cohorts take a long time to generate useful insights about older age groups. There is a large gap in cohort measurements on young children. Thus, it is important to prioritise investments carefully.

- Cohorts established long ago are inevitably no longer fully representative of their age groups in the current UK population, and do not always provide the breadth of data required to look at devolved matters or particularly vulnerable groups. This can limit the usefulness of the existing studies in understanding some of the key issues of our age.

- Although numerous studies have been published from ESRC-funded longitudinal studies, the review panel has found it difficult to trace specific evidence of the 'instrumental' or 'direct' policy impacts from these investments within the time and resources allocated to this review. Impacts from these ESRC longitudinal data investments undoubtedly exist but are hard to pinpoint and quantify, in part because insights drawn from their use more typically contribute to 'conceptual' impact (or 'enlightenment') and thus act gradually to change the discourse, thinking, and common knowledge around an issue. If left unchecked, a lack of evidence is likely to undermine the case for future investments, given the general principle that research funding in the UK should be founded on an evidence-based scientific case.

- By building on the rapidly expanding access to greater quantities of administrative data, there may now be considerable opportunity to implement innovative approaches from the start in the design of a new study (or related studies) tracking key groups in a period of rapid change, such as younger cohorts where the UK lacks data to answer crucial policy and research questions.

2. What are the advantages and challenges to using longitudinal data enhancements such as administrative data linkage and harmonisation with longitudinal survey data?

- The ways in which longitudinal data is collected and analysed needs to continue to be state-of-the-art, particularly in terms of allowing for the impact of increasing data linkage and harmonisation, and the ESRC needs to take advantage of opportunities to help ensure that the UK maintains its leading position in these areas in the future.

- There is an opportunity to maximise the linkage of routine administrative and health data to our longitudinal studies and to realise benefits. These include improved representativeness in coverage, enhanced content and value of the data for research, and increased speed in creating insights.

- Use of an innovative sampling methodology – an administrative data population spine on which to base current and new longitudinal studies – would provide a key underpinning and transformative element of the UK's research data infrastructure; its creation would involve a number of challenges but the returns would be significant and reach beyond social science.

3. How can training, access and promotion of longitudinal investments be continued and enhanced for both traditional and non-traditional research communities?

- The level and range of training offered in the UK is world-leading. Nonetheless there are challenges to maintaining the right mix of skills and mentoring in the longitudinal studies community (particularly in the face of competition for quantitative methodological skills from other non-research sectors).

- There is widespread support for continuing and expanding a longitudinal data support centre approach, in combination with the **UK Data Service**, for making access and discoverability of longitudinal resources as straightforward as possible for researchers and policy makers.

- Better methods for tracking the use of ESRC's longitudinal data are also needed.

Recommendations (presented in order of priority)

1. Construct an administrative data spine that can be used for new and existing longitudinal studies supported by the ESRC and other funders

1.1 Policy-relevant research requires longitudinal data that is representative of a rapidly changing UK population. This in turn requires the future design of ESRC longitudinal data resources to be capable of dynamically representing this population. We therefore recommend that the ESRC develop and maintain a longitudinal administrative data spine with maximum population coverage that can be used as the basis for data linkage and as a sampling frame for its longitudinal surveys.

- o We recommend that ESRC negotiates with the appropriate government department(s) to obtain ongoing and guaranteed access to a register of unique personal identifiers with comprehensive UK population coverage, such as the NHS identifier, that can be used as an administrative spine for future longitudinal studies. This can also be a resource for data linkage for the ESRC's currently supported longitudinal studies and potentially those of other major funders.

- o We strongly advise that ESRC works collaboratively to ensure that it plays a leading role in developing UKRI's strategy for how the research councils (and other funders) can create a foundational longitudinal infrastructure that is open for research use.

- o We recommend that ESRC convenes a group consisting of representatives from government, legal, public, and academic spheres to advise on creation and management of this administrative data spine.

2. Commission a new birth cohort with an accelerated longitudinal design and additional funding for transition to adulthood for the Millennium Cohort Study

2.1 We recommend that the ESRC commissions a new UK population-representative birth cohort, and if possible, within a reasonable time frame, sampled from the administrative population spine so that data is available on those that participate (or not) as well as those that leave the study.

- o We recommend the use of an accelerated longitudinal design with sufficient power to examine key questions at the national and regional level or in vulnerable subgroups. This accelerated design entails that pivotal age periods (to be determined) can be ascertained across two or more ages simultaneously.

2.2 We recommend funding an additional sweep of the **Millennium Cohort Study** (MCS) to allow for an important life stage – transition to adulthood – be prioritised.

3. Continued funding for the household panel study and targeted innovative grants for existing cohorts

3.1 We recommend that ESRC continues investment in **Understanding Society**, subject to a strong scientific case and peer review, since it provides the backbone for observing changes in socioeconomic issues and household mobility in the UK, is based on a large national sample, enables international comparisons, and is highly valued for its flexibility in experimentation and importance to the policy community.

3.2 We also recommend that ESRC investigate the feasibility of basing future sweeps of Understanding Society on households identified via the longitudinal administrative data spine described in **Recommendation 1**.

3.3 We recommend that ESRC supports the **National Child Development Study** (NCDS), **British Cohort Study** (BCS70), and **Next Steps** through competitive and innovative grant proposals that can enhance the data in the existing cohorts, which may include targeted administrative data linkages and harmonisation as well as potential new data collection focused on important questions in line with ESRC's strategic priorities using a rigorous re-funding strategy.

3.4 We recommend that the ESRC commissions a mechanism for innovation in the cohort studies, similar to the innovation panel in Understanding Society, to continue supporting methodological development (piloting, testing, and experimentation) in their longitudinal investments.

4. Broaden consultations, extend sampling and introduce time-limited funding

4.1 We recommend that the ESRC ensures that study consultations are sufficiently broad and the processes are clearly articulated, to ensure an appropriate mix of academic and policy experts, and that the process of consultation is transparent.

4.2 We recommend that ESRC prioritises funding of studies (including the recommended new cohort) that consider ways to oversample devolved nations and communities of importance or policy interest. An important focus of this recommendation is to consider ways of establishing a representative population sample that allows for policy-relevant regional and community sub-samples that are large enough to be appropriately analysed. In particular, we recommend that this requirement is a priority when developing the longitudinal administrative data spine described in **Recommendation 1**, which would facilitate maintaining representativeness by allowing for refreshing of the sample over time.

4.3 We recommend that the data collection for any longitudinal investment new or ongoing be time-limited. Requests for data collection beyond the initial funding period should be submitted for review in a new proposal that outlines a strong scientific rationale accompanied with clear research questions related to why additional data should be collected at the time proposed. This will allow for innovations to be introduced to the studies and for ESRC to consider the representativeness and ongoing value of their investments in their portfolio of longitudinal studies.

5. Invest in data management and sharing

5.1 With additional funding and resources, we recommend that ESRC takes the lead in the UK across funders to expand the cross-collaborative functions of a longitudinal data resource centre. Such a centre would lead on carrying out and sharing research focusing on longitudinal analysis issues faced by all funders, would provide leadership in the use of data linkage and data harmonisation methods that could be used to facilitate their funded research, and promote international collaboration. We recommend that ESRC continues to support the functions of a longitudinal data resource centre to:

- o continue enhancement and innovation on data linkage, data harmonisation, impact, training, and discoverability

- o support the provision of topical data platforms (e.g. mental health, political science) for special interest data users

- o extend activities to cover the broader longitudinal investments across more UKRI longitudinal studies and beyond in collaboration with other funders

- o develop strong collaborations with other national and international groups.

5.2 We recommend that the ESRC strongly support continued funding of the UK Data Service, which has been central to the ease of access and increasing use of the data encompassed in the ESRC longitudinal investments.

5.3 Due to the broader availability of the ESRC's longitudinal studies and ubiquity of use, there is a need for better evidence of the use and impact of research findings and better practice at citing data. We recommend that ESRC supports longitudinal investments to develop innovative technology and tools, and better methods and measures to track and compile metrics on data use and sharing to enable demonstration of impact.

6. Data linkage and harmonisation

6.1 We recommend that ESRC collaborates within UKRI and with other funders to strongly promote, facilitate, and negotiate administrative data linkage (broadly defined) for researchers to achieve a step change in capacity to meet increasing demands for longitudinal information including at local authority level and for particular subgroups.

6.2 We recommend that all ESRC-funded longitudinal studies using data linkage will be required to provide appropriate paradata about the data linkage process, which can then inform analysis of the linked data.

6.3 We recommend that the ESRC continues to fund retrospective harmonisation of all studies of similar design, especially the cohort studies, where a scientific need is identified.

6.4 We recommend that ESRC funds or mandates that future longitudinal studies develop methods for prospective harmonisation to maximize use and further extend the value of the data, and provide leadership with other funders to harmonise data sources across the UK and internationally.

6.5 We recommend that all ESRC-funded longitudinal studies using data harmonisation are required to provide appropriate metadata about the harmonisation process and appropriate means of creating, storing, and sharing harmonisation-related metadata.

7. Review and expand training and develop a data dashboard for policymakers

7.1 Building on current efforts, ESRC should commission a review of the provision and organisation of all training applicable to longitudinal studies currently provided at multiple institutions, and ensure innovative and sufficient academic training at the doctoral and post-doctoral levels into the future to make best use of the unparalleled longitudinal data generated in the UK. This should include funding ongoing career development beyond the post-doctoral level to ensure well-placed senior mentorship in longitudinal research.

7.2 We recommend that ESRC, in collaboration with UKDS, funds the development of a centralised analysis platform aimed at policy users of its longitudinal data resources to access information related to these resources and data enhancements, and facilitate analysis. This dashboard would provide descriptive statistics and share data in a way that is accessible to users who are interested in longitudinal data, but with diverse interests and varying levels of statistical knowledge and methodological training. This may involve investment in new technologies to create such a dashboard.

Conclusion

It is clear that the ESRC has created a unique portfolio of longitudinal data resources for the research community that reaches far beyond the social sciences and the UK. It is hoped that the recommendations and priorities provided in this strategic review will continue to enhance the investments and the role of the ESRC as a leader in the collection sharing and use of longitudinal data for scientific and policy purposes.

Section 1: Background and method

1.1 Background

In 2016, the **Economic and Social Research Council** (ESRC) instigated a strategic review of the scientific and policy needs for investment in longitudinal research resources, including but not limited to their investments in longitudinal studies across the life course. The ESRC spends approximately 10% of its budget per year, every year, on longitudinal research and resources, so it is very important to ensure that these funds are used to best effect. Hence, the focus of the review is to ascertain the scientific relevance, sustainability, and contribution of these investments to the UK population, scientific community, policymakers, and to the broader international community, and whether the current portfolio is meeting the needs of these groups and if/how these needs could be met in the future. See **Figure 1** for an overview of ESRC's longitudinal resources.

The UK birth cohorts have set the standards by which other scientists across the world model and consider collecting information across the life course. As the primary funder of social science longitudinal studies, the ESRC has developed an ongoing investment portfolio that has been the flagship of social science data resources in the UK. More broadly, the ESRC has created a data infrastructure that promotes the training, collection, curation, sharing and analysis of data on the UK population over long periods of time and across the life course, providing valuable information on how people develop as contexts and policies are changing around them. Since the last **strategic review** was over 10 years ago, the ESRC wanted to assess the longitudinal data infrastructure to make sure that its investment continues to deliver and adapt to the changing landscape of scientific enquiry, survey research, and the data needs of all the communities that seek to use the data -- including research, government, charities, public policy, education, and for the participants of the various studies who have invested their time in providing these valuable data.

To this end, an independent **international review panel** comprising a multidisciplinary group of researchers with expertise in the use and methods of longitudinal data was commissioned to review the continuing scientific and policy needs for longitudinal survey data, how far the current investments meet these needs, and how these needs could best be met in the future. The panel was tasked with helping to create a vision for the next five to 10 years of the social science infrastructure that is needed to meet the needs of researchers with an interest in longitudinal and life course evidence (see **Appendix 2** for the review specification and remit). The panel was also tasked to consider the challenges and opportunities that may be important to address in their investment portfolio in the future. To accomplish this task the ESRC acquired information from multiple stakeholders (e.g., data users from science and policy, investment directors and staff, government departments and representatives, other research councils, charity funders) who could provide information and opinion on their needs and the longitudinal study investments.

1.2 Methods

Multiple methods were used for gathering initial data to help focus the review on challenges and opportunities in the future of longitudinal data investments (see **Annex A**). First, information was obtained through an open **web-based consultation**[3] that was advertised widely across social media and email campaigns to obtain information from as wide of a constituency as possible; over 600 responses were received. This information was used and elaborated on for a **workshop**[4] designed for researchers, policymakers, and other stakeholders in longitudinal research that was convened in January 2017 at Nuffield College in Oxford, UK. These two sources of information from the consultation and from in-depth discussions at the workshop were combined to refine the primary questions that the review panel were tasked to answer. These questions were:

- What are the scientific and policy-relevant content needs for longitudinal data in the future?

- What are the advantages and challenges to using longitudinal data enhancements such as administrative data linkage and harmonisation with longitudinal survey data?

- How can training, access, and promotion of longitudinal investments be continued and enhanced for both traditional and non-traditional research communities?

---

[3] Townsley, R. (2016). **Interim report: initial analysis of responses to the consultation**. Ruth Townsley Research

[4] ESRC. (2017). **ESRC Longitudinal Studies Review: Workshop summary report**.

Figure 1: Overview of ESRC's Longitudinal Resources

Core ESRC-funded Longitudinal Resources

Understanding Society
THE UK HOUSEHOLD LONGITUDINAL STUDY

(ISER, University of Essex)

Understanding Society is a large household panel study which began in 2009 and follows the lives of people living in all countries of the UK.

Incorporates the British Household Panel Survey (1990-2008) and consists of a main survey and an Innovation Panel.

CENTRE FOR LONGITUDINAL STUDIES

(IoE, University College London)

CLS is home to Britain's 4 major principally social science cohort studies:

• 1958 National Child Development Study

• 1970 British Cohort Study

• Next Steps

• Millennium Cohort Study

closer
Cohort & Longitudinal Studies Enhancement Resources

(IoE, University College London)

CLOSER brings together 8 UK longitudinal studies across social and biomedical science to maximise their use, value and impact.

CALLS-HUB
CENSUS & ADMINISTRATIVE DATA LONGITUDINAL STUDIES HUB

CeLSIUS
CENTRE FOR LONGITUDINAL STUDY INFORMATION & USER SUPPORT

ESRC funds 3 Research Support Units which support research using the national census-based longitudinal studies, as well as a hub to coordinate and promote the units.

SLS-DSU
SCOTTISH LONGITUDINAL STUDY DEVELOPMENT & SUPPORT UNIT

NILS-RSU
NORTHERN IRELAND LONGITUDINAL STUDY RESEARCH SUPPORT UNIT

Other ESRC co-funded Longitudinal Resources
(In receipt of ESRC support currently/recently)

Low and Middle Income Longitudinal Population Study Directory

CHILDREN OF THE 90s — Avon Longitudinal Study of Parents and Children — closer

bib We are family BORNINBRADFORD — Born in Bradford

ELSA — English Longitudinal Study of Ageing

NICOLA — Understanding Today for a Healthier Tomorrow — Northern Ireland's Cohort for the Longitudinal Study of Ageing

HAGIS Healthy Ageing In Scotland — Healthy Ageing In Scotland

Other studies in the CLOSER consortium (Not ESRC funded)

MRC National Survey of Health and Development (1946 cohort)

SOUTHAMPTON Women's Survey — Southampton Women's Study

HCS — Hertfordshire Cohort Study

closer

After these three areas were determined, the review panel in consultation with the ESRC met in person or via teleconference with over 80 key stakeholders (See **Appendix 1**), including members of both the social science and biomedical research communities, involved in these question topic areas to ascertain the current work being done in these domains, the achievements, the barriers that have been faced, and potential ideas to be addressed in the future for the longitudinal investment of the ESRC. Additionally, the review panel commissioned various data-seeking efforts that also informed this review, and these reports can be found in the Annex section online containing further resources. These meetings, conversations, and reports helped to shape the review, the recommendations, and the priorities of this report.

1.3 Context of the review

There have been a number of significant changes since the last review of ESRC's longitudinal investments. There are challenges to how large longitudinal surveys are undertaken in a number of respects; for example, shifting respondents' attitudes towards participation in such studies, survey costs, and technological advances in data collection. These challenges are offset by opportunities that make the universe of administrative and routine data potentially available for research, with changes in UK legislation around access to such data. In addition, changes in the policy environment in the UK, not least the increasing devolution (from UK level) of responsibilities for policy, are shifting policy needs.

During the review period, the research funding environment that ESRC operates in has changed with the formation of UK Research and Innovation (UKRI), which may provide a renewed emphasis on interdisciplinary approaches, but also budgetary pressures.

Wellcome and the **Medical Research Council** (MRC), who are other major funders of longitudinal data, also made significant changes in the way they approach their studies and investments. Both funders have recently introduced metrics to be used in making decisions about funding the continuation of existing cohorts. For potential new cohorts it needs to be clear that the research questions cannot be addressed using existing studies either in the UK or internationally. These metrics are part of a shift from a more traditional biomedical hypothesis-driven and study-team data ownership model towards a data as research infrastructure approach. Similarly, Wellcome published its new **Longitudinal Population Studies strategy**[5] that put an emphasis on integration of data, including linkage, harmonisation and close collaboration between funders including strategic co-funding of the most important studies.

The review panel was mindful of the continually evolving environment in the UK as well as the changes in the funding environment during the review period, and the contexts in which the review would be received. Thus, this review aims to offer strategic-level recommendations and identify priorities that provide flexibility for the ESRC in considering its strategic plans on longitudinal data investments in this dynamic environment.

1.4 Overview of current ESRC investments in longitudinal data

Based on the 2006 Strategic Review of Panel and Cohort Studies, additional new investments were recommended that included a new birth cohort and a data resource centre to aid in the use and value of longitudinal data. We briefly review the key investments to demonstrate the current breadth and strength of ESRC-funded social science longitudinal studies research infrastructure in the UK, and highlight important unique contributions of each study.

1.4.1 Birth and age cohorts
The ESRC has supported multiple population-representative studies to examine the life course of individuals born at various times in the UK. These studies comprise a set of what are referred to as the "birth cohorts" because they sample from individuals born in particular weeks or of particular ages at the same time. The primary birth and age cohort investments of the ESRC are:

• the 1958 National Child Development Study (NCDS)

• the 1970 British Cohort Study (BCS70)

• the Millennium Cohort Study (MCS)

• Next Steps (formerly the Longitudinal Study of Young People in England, LSYPE), an age cohort (not sampled at birth).

---

[5] Wellcome. (2017). **Longitudinal Population Studies Strategy: Wellcome's Longitudinal Population Studies Working Group**.

These cohorts are housed at the **Centre for Longitudinal Studies** (CLS) at the **UCL Institute of Education** which has allowed for a close connection with statistical and content specialisation in areas relevant to birth cohorts (education, health/wellbeing, economic circumstance, employment, family formation, parenting). More specific information on these studies is available in **Annex B**; Figure 2 provides a quick overview of the years these cohort studies were undertaken and the number of people who were initially in each.

Figure 2: Timeline of the CLS Cohort Studies
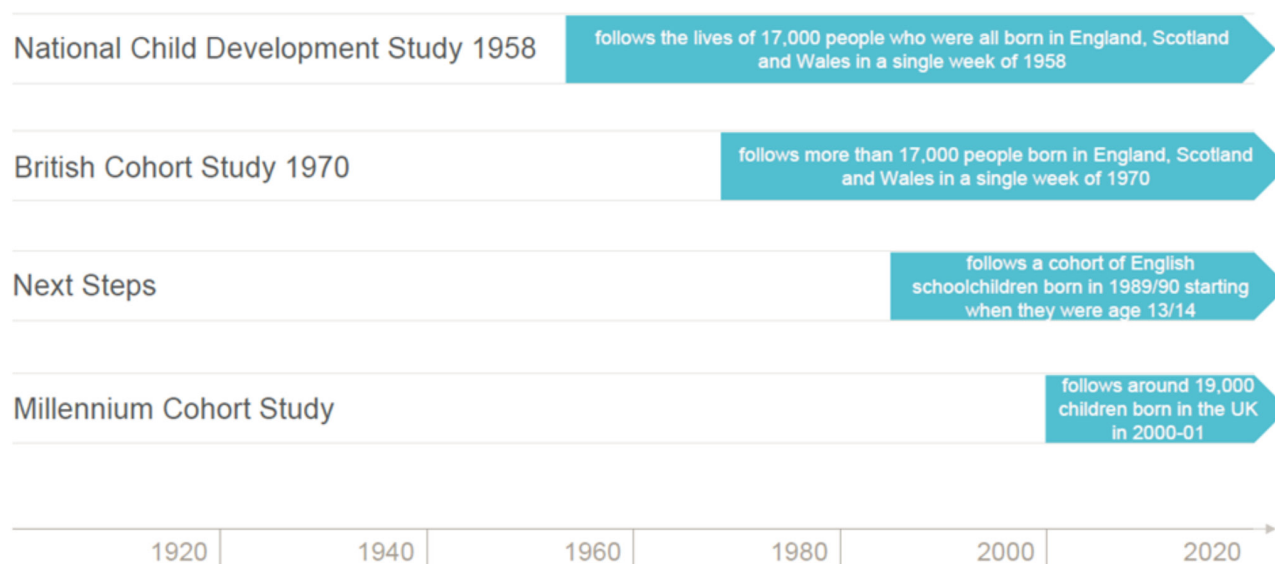
These studies have set standards for how individuals should be studied across time and have provided some of the most compelling evidence of the importance of various events or exposures to long-term outcomes of interest to the social science and government constituencies. Issues as diverse as breastfeeding infants and early indicators of dementia, and the changes in these topics over time, have been explored because these studies provided the important longitudinal data for the analyses (see **Annex C** for impact case studies). Thus, the birth cohorts have been a perennial positive investment for the ESRC. Even though each of the cohorts offer a unique view of the lives of individuals and families in the UK, we detail the most current of the birth cohorts, the Millennium Cohort Study (MCS), to give an example of the types of data and methods used in the birth cohorts as well as the types of users of this data source.

The MCS is the UK's most recent longitudinal birth cohort study which began collecting data on a sample of children born in the UK during the year 2000-2001. The goal of MCS is to provide a detailed portrait of children in the new century and to understand the influence of early family context on children's development and outcomes from childhood into adulthood. Although MCS is designed so that comparisons can be made with prior UK birth cohort studies, some changes to the study design were introduced in this new study to garner a sample that is reflective of the entire UK population of families with children born immediately after the Millennium and to incorporate new forms of data. At the study outset, over 19,000 children and their families were randomly selected from the Child Benefit (then a universal benefit) records within the UK to participate in MCS. The study sample contained oversampling among children from disadvantaged and ethnic minority families in England and children growing up in Scotland, Wales and Northern Ireland.

Data are collected directly from children (since age 3), their resident parents, and in more recent waves, the children's older siblings through a combination of questionnaires, cognitive assessments, and interviewer-administered physical measurements. The majority of the data is collected through computer-assisted personal interviewing (CAPI) and computer-assisted self-interviewing (CASI). All surveys and interviews are available in English, although this is not the first language of some of the participants. To date, a total of six waves of data have been collected when children were 9 months, 3-, 5-, 7-, 11-, and 14-years old. In the most recent wave, about 62% of the initial participants (11,726 families) completed interviews. A further wave of data collection is planned for 2018 when study participants are 17 years old.

The impact of adverse childhood experiences

Findings from the National Child Development Study, the British Cohort Study 1970 and the Millennium Cohort Study have shown how adverse childhood experiences (ACEs) can affect psychological, physical and academic outcomes throughout later life[6]; it is now understood that a combination of ACEs could affect individual mortality by as much as 20 years.[7] As a result of this knowledge, policymakers from across government are developing policy to equip frontline staff with the capacity to identify and support those most at risk.

The MCS, as well as the other birth cohorts, collects a range of information on household composition, family sociodemographic background, housing, schools, neighbourhoods, health, cognitive tests, and expectations. The topics for the surveys are based in large part on recommendations from survey consultations that include individuals involved with earlier birth cohort studies, funders, and other experts in the age group. In response to the changing technologies and need for more precision in data collection on health indicators, the most recent wave (age 14) collected DNA from saliva samples, physical and sedentary behaviour from wrist-worn accelerometers, and time use collected mainly from mobile phones and online. Additionally, to extend the survey data, for study participants who gave consent, data links between MCS and several administrative datasets across various domains (health, education, spatial) were made, and linkages to additional administrative datasets are in development as of 2017 (see **Section 3** on data linkages for additional information).

Based on user downloads and access of restricted data from the UK Data Service, where all ESRC-funded longitudinal studies data has been deposited over the last 10 years, the majority of MCS users are postgraduates (43%) and staff at an institute of higher education (38%), while 2% are central and local government staff, and 2% are NGO or registered charity staff (see **Annex L** for data on downloads and publications). MCS users come from primarily four disciplines: economics (25%), psychology (15%), sociology (12%) and medical (13%). The majority of users (83%) are from the UK with the next highest percent of users coming from the US (6%). The Republic of Ireland, Italy, and Germany each make up ~2% of the MCS users. For information on more metrics of study access, see **Section 4**.

The **Life Study** was intended to be the newest of the birth cohorts as recommended by the 2006 review, and data collection was started in 2014 but discontinued in 2016. It is discussed here because it represents a gap in the birth cohorts, and also highlights some of the challenges with longitudinal birth cohorts that will need to be addressed and considered in future investments. The Life Study was an innovative design intended to collect data on over 80,000 babies born between 2014 and 2018. Core topics that could have been addressed with this study were the impact of inequality, diversity and social mobility on children's life chances, education and school readiness, early life origins of illnesses, health and wellbeing in childhood, social, emotional and behavioural development as well as neighbourhoods and environment. The recruitment plan involved two components. In one, 60,000 women and their partners were to be recruited in selected NHS maternity units during their pregnancy. The initial proposal that was chosen during the grant competition did not include a random national population sample and so an additional 20,000 babies were to be recruited through the birth register, to allow for this sampling that is a required aspect of the ESRC birth cohorts. The sampling design for the larger cohort of women was based on geographical clustering for practical reasons, to have a range of social background and to oversample births to minority and ethnic groups. These women would be recruited through special Life Study Centres that would allow for assessments, including biomedical measures of the mother and child. The birth register sample was to be collected in the same manner as other birth cohorts, by sample members being approached and recruited at their homes.

Even though this study would have provided rich data on the pregnancy and early days of families with young babies, there were issues with the recruitment at the initial Life Study Centre and challenges in the pilot, through the use of the birth register, in getting families to agree to the study. ESRC and MRC ended the study in early 2016. For this reason, no

---

[6] Patalay, P. & Fitzsimons, E. (2017). **Mental ill-health among children of the new century: trends across childhood with a focus on age 14**. Centre for Longitudinal Studies: London.

[7] Brown, D., et al. (2009). Adverse childhood experiences and the risk of premature mortality. *American Journal of Preventative Medicine*, 37(5), 389-96.

fine-grained longitudinal data on early childhood is available UK-wide for any cohort born after 2000/01.[8] The reasons for ending the Life Study were complex and are outside the purview of this review; however, the issues related to this study being discontinued were an important part of the discussions and information received by panel members. These issues are incorporated throughout the review, and will be addressed again in the recommendation section regarding a potential new birth cohort.

One of the prominent issues that is relevant to the value of longitudinal data is why and for whom the data is collected. The ESRC birth cohorts have been described as the infrastructure studies of the social sciences. They are designed by principal investigators who are social scientists using a variety of ways to ascertain the questions (e.g. workshops, consultations). The data and samples for these infrastructure studies are collected on broad themes so that individual investigators can bring their own questions to these datasets and biomedical measures to understand trends and outcomes across the lifespan that can be generalised to the population. This approach differs from data that is collected and designed to answer specific questions or hypotheses of an individual researcher or investigator. Those datasets are often narrower in scope and concentrate on methods that will allow for the test of directional hypotheses, and are not necessarily intended to be generalized to a population but potentially to a clinical population. These infrastructure studies are also curated for general use (subject to appropriate conditions) by the data user community, which is in contrast to the medical model where data is often held by the research team that collected the data.

The importance of numeracy

The Moser Report (1999)[9] used research based on the National Child Development Study to explore the relationship between mathematics qualifications and earnings[10], leading to the 2001 Skills for Life Strategy. The strategy made free literacy, numeracy and training available to all adults without Level 2 qualifications (GCSE).

Given that both of these approaches bring valuable data to the research community, the ESRC typically has funded birth cohorts that are nationally representative and create a foundational infrastructure in which multiple researchers and policymakers can benefit from the data. In the future, the ESRC will need to assess this approach to determine if it continues to be the best avenue for providing rich data on the human condition across the life course. Similarly, issues of recruitment and attrition will need to be addressed to make sure that state-of-the-art methods are being used and are successful. These concerns also need to be considered in other longitudinal investments like the panel study Understanding Society, which provides complementary value to the birth cohorts but uses the household as a sampling unit instead of the birth of an individual child within the family.

1.4.2 Panel studies
Understanding Society, the UK Household Longitudinal Study began in 2009 with the intent of providing a portrait of the full UK population across the age range as it changes over time. It was built upon and extended the **British Household Panel Survey** (BHPS) that was launched in 1991 and surveyed 10,000 persons living in 5,500 households. An additional 1,500 households each in Scotland and Wales were added to the sample in 1999 as well as 2,000 households in Northern Ireland in 2001. Since 1994, children aged 11-15 completed a short questionnaire. The BHPS collected data in yearly intervals until 2008 (18 waves). Data collection was devoted to socio-economic topics such as income and poverty, housing conditions, employment, and health. Occasionally information on family issues such as cohabitation, marriage, children and ageing as well as on neighbourhood was collected. The youth questionnaire contained questions on health behaviour, time use, and attitudes to jobs.

In response to the 2006 Strategic Review of Panel and Cohort Studies, BHPS was incorporated in 2009 into the newly launched Understanding Society, which covers a much larger nationally representative sample, initially of 40,000 households that were identified through a multi-stage sampling design. Ethnic-minority households were oversampled in wave 1 so that there are at least 1,000 adults from each of the five largest ethnic minority groups in England (Indian,

---

[8] However, regional cohorts, including the Avon Longitudinal Study of Parents and Children next generation (ALSPAC-G2; the offspring of the original ALSPAC participants), Born in Bradford (BiB) and Southampton Women's Survey, have collected data on births since this time.

[9] Moser, C. (1999), *Improving Literacy and Numeracy: A Fresh Start*. Department for Education and Employment.

[10] Dolton, P.J. and Vignoles, A. (2002). The return on post–compulsory school mathematics study. *Economica*, 69(273), pp.113-1.
DOI: **10.1111/1468-0335.00273**

Pakistani, Bangladeshi, Caribbean, and African) in the study. In the most recent wave of data collection for which data is available (wave 6), additional immigrant and ethnic-minority households were identified and added to account for individuals who entered the UK since the beginning of the study in 2009. Questionnaire instruments are translated into nine languages (Bengali, Punjabi in Urdu and Gurmukhi scripts, Welsh, Arabic, Somali, Cantonese, Urdu, and Gujarati). As a result of this oversampling strategy and subsequent sample booster, Understanding Society provides sample sizes large enough to allow for ethnic group-specific analyses. Original sample members who leave to start their own household continue to be followed by the study as long as they live in the UK. Babies born to sample members are included in the study as well. It is estimated that roughly 800 babies are added to the study each year. In future waves of the study (12+), Understanding Society plans to expand the sample to other family members and to collect data from sample members' partners who are outside the household. Currently, Understanding Society has collected seven annual waves of data, of which six are available for researchers; waves 8 and 9 are in the field at the time of writing. In November 2017 the first version of a joint harmonised dataset containing BHPS and Understanding Society data and yearly information from 1991 became available.

Household sampling makes BHPS and Understanding Society particularly suited for analysing household and family dynamics, mutual dependency of life courses within families and joint decision-making of partners. For example, in waves 1-8 about 3,800-5,800 children aged 0-4 and 4,200-5,300 children aged 5-9 were part of the sample. Information from sample members is collected via a joint household interview (conducted with one adult household member) as well as personal interviews with all household members 10 years or older. Youth (aged 10-15) and young adults (aged 16-21) receive specific questionnaires instead of the standard personal interviews for adults. Questionnaire modules targeted at respondent groups in certain life stages such as pregnant women or smaller children are increasingly being used (personal communication with Understanding Society team). Over a longer time span, this strategy – which has also been adopted by the **German Socio-Economic Panel** (GSOEP) – allows for gradually building up databases which enable answering life-course and child-related longitudinal questions.

A core set of measures on income, education, households, housing, employment, behaviours and health are collected annually with certain content rotating across years (see Table 1 in Buck & McFall (2011) for example of content across first four waves[11]). Due to this flexible study design, Understanding Society can collect data on time-specific events (London Olympics, Brexit), and has plans to incorporate person-event triggered data collection on job loss, retirement, bereavement and pregnancy in future waves.

In waves 2 and 3 of the study, biomedical data were collected from about 20,000 adult study participants in follow-up health assessments. DNA methylation profiling has been conducted on DNA samples from 1,200 participants from the BHPS subsample in the study. Additionally, administrative health and education records have been linked for study participants who gave consent. Data links to phones and sensors are in development as of 2017. Also, Understanding Society is part of an international network of panel studies and is part of the **Cross-National Equivalent File**, which harmonises data across studies on key variables (education, employment, demographics, health, satisfaction, income) and facilitates international comparisons.

Finally, a unique component of this study is the **Innovation Panel** (a separate sample), which allows for experimental and methodological assessment of data collection methods and instruments. Results from experiments with survey procedures, questionnaire design and non-experimental studies have advanced methods of survey data collection and informed data collection in international studies (e.g., **Health and Retirement Study** (HRS), **Panel Study of Income Dynamics** (PSID)), and in its own data collection strategy (i.e. use of web interviewing[12]). Overall, even though there are limitations to the Understanding Society data such as lack of information on the youngest children as they progress through development and sample size restrictions of various age groups, the panel study continues to be a strong asset in the longitudinal portfolio of the ESRC.

Within the last 10 years, Understanding Society had 12,821 data downloads and secure access users and MCS had 12,683, while all CLS studies combined had 50,255 data downloads and secure access users. (These and later data download figures (see **Annex L**) are from UK Data Service data; they may not be accurate in absolute terms

---

[11] Buck, N. & McFall, S. (2011). Understanding Society: design overview. *Longitudinal and Life Course Studies*, 3(1), 5-17.
DOI: **http://dx.doi.org/10.14301/llcs.v3i1.159**
[12] Understanding Society. (2017). Additional Information from Understanding Society produced for the Longitudinal Studies Review Panel.

(see **Section 4.2**), but provide a reasonable indication of relative use across datasets.) Similar to MCS, the majority of Understanding Society users are staff at an institute of higher education (37%) and postgraduates (35%) (see **Table 2c**, **Annex L**). Central government staff (5%) and local government staff (1%) comprise a higher percentage (and count) of users compared to MCS, which may be a result, at least in part, of Understanding Society's Policy Unit that responds to requests from government departments for research support and expertise as well. The percentage of NGO or registered charity staff using Understanding Society data compared to MCS is also slightly higher (3%).

With regards to academic discipline, users of Understanding Society are primarily from economics (33.5%) and sociology (21.1%). However, the other users are from a wider range of disciplines than MCS users (e.g., Statistics and Operational Research (8.4%), Social Policy and Administration (7.2%), Geography (6.5%)). This may, in part, be due to the study's digital and marketing communication strategy, which has over 3,500 individuals subscribed to the Understanding Society Newsletter and over 9,000 Twitter followers.[13] However, a much smaller percent of users of Understanding Society
are from the medical field as compared to the MCS (4% vs. 12.9%). Finally, similar to MCS, the majority of users (88%) are from the UK followed by the US, Germany and Italy, with each of these countries comprising less than 3% of the Understanding Society users.

1.4.3 Resource centre for longitudinal data: CLOSER

The 2006 review of the longitudinal studies recommended creating a centre that provides expertise in combining and enhancing longitudinal data. Based on this recommendation the **Cohort and Longitudinal Studies Enhancement Resource** (CLOSER) was created. CLOSER is a consortium of eight studies from the fields of medicine and social science, the UK Data Service and the **British Library**. Initially funded for five years with a recent extension of funding for two additional years until September 2019, the purpose of this consortium is to increase the use, value and impact of the UK's longitudinal studies.

To meet its strategic goals, CLOSER has created a search platform (**CLOSER Discovery**) that allows researchers with varying levels of data experience to find variables across all consortium studies. It is regularly updated and in future will also include metadata for additional studies beyond the eight CLOSER studies. CLOSER also promotes sharing of good practice between studies, and provides training workshops on a range of topics such as data management and life course research, in addition to workshops that use longitudinal datasets to train users on specific topic areas (health, educational attainment). The purpose of these workshops is to help develop a larger and more skilled user base for the longitudinal studies.

CLOSER also plays an important role in expanding and connecting the studies' data resources through data linkage and data harmonisation (see **Annexes D** and **E**). For data linkages specifically, CLOSER has been addressing challenges in gaining access to administrative data in collaboration with government personnel and key stakeholders. CLOSER also produced resources to promote the value of administrative data for research, and to help the studies that are part of the consortium overcome the barriers to data linkage that have plagued this important area of data science. In addition to administrative data linkages, CLOSER is working on retrospectively harmonising measures across all CLOSER studies in ten topic areas to date (e.g., body size and body composition, socioeconomic status, earning and income, etc.). A primary goal of this work is to augment the value of existing datasets and facilitate more cross-cohort research.

This consortium provides an important resource for the use of longitudinal studies in the UK. CLOSER, for example, is producing documentation on longitudinal methodology and examples of best practice for using longitudinal data that are shared with the research community. Additionally, in 2016 CLOSER appointed a Public Affairs Manager, who facilitates interactions between researchers and a wide range of policymakers, raises the profile of longitudinal studies and their evidence, which in turn helps to promote policies and thinking regarding data science and management.[14] CLOSER also produces a newsletter on longitudinal research and evidence geared toward a wide range in audience to further increase the impact of their work.

---

[13] **https://twitter.com/usociety** (as of January 2018)

[14] CLOSER. (2017). Additional information from CLOSER (Cohort & Longitudinal Studies Enhancement Resource) produced for the Longitudinal Studies Review Panel.
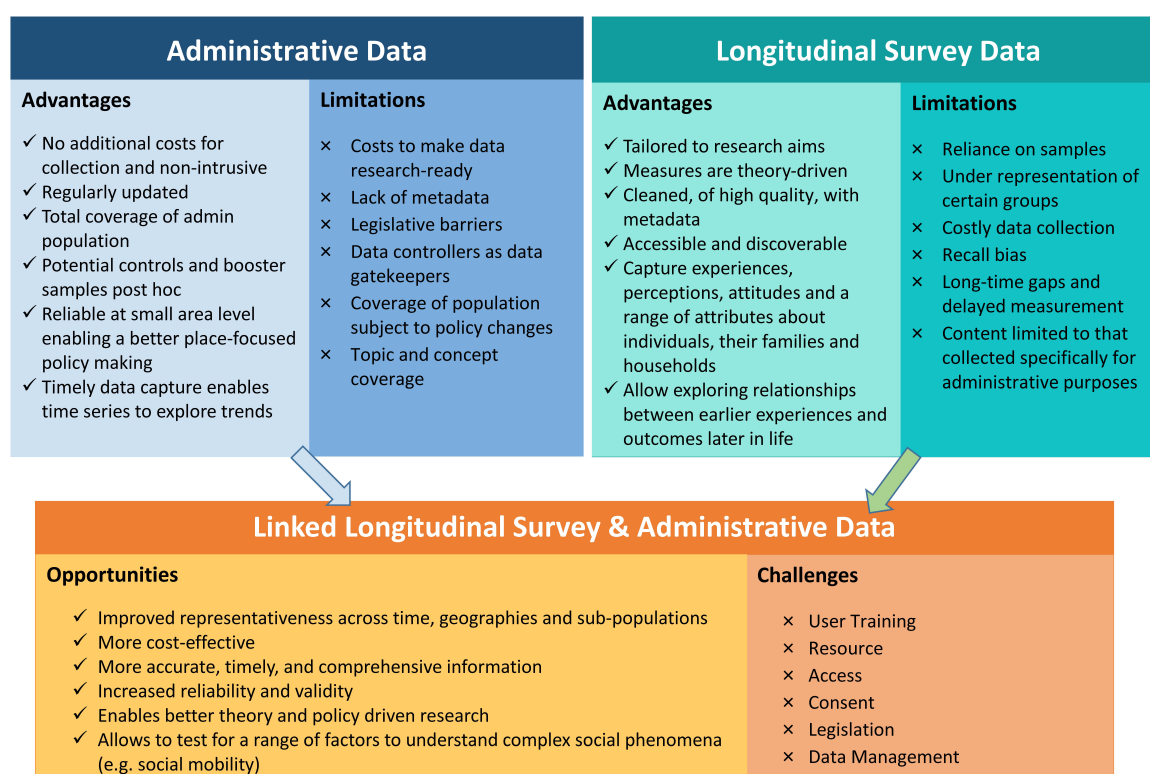
1.5 The UK social science longitudinal data landscape

The studies described above along with other investments such as the **Census Longitudinal Study**, **Born in Bradford**, CLOSER, the **UK Data Service** (UKDS) and **METADAC** (Managing Ethico-social, Technical, and Administrative issues in Data Access) constitute an impressive set of investments in longitudinal data and access. Central to the ESRC portfolio are Understanding Society (incorporating BHPS) and the Birth Cohort Studies. Each type of study represents different methodological designs and philosophies of longitudinal data. Thus, they allow for answering different research questions using different analysis methods and consequently address different groups of users. For example, economists are more interested in household survey data, due to detailed socio-economic information, but also because panel modelling and identifying causal effects in observational data usually calls for a long series of panel observations with consistent measures. In contrast, psychologists may prefer cohort study data due to their tailored developmental measures for different child ages.

In effect, the two hardly overlap with regard to contents and research potentials, but nicely complement each other. Regarding their core contents, studies in both areas cover sociodemographic and socioeconomic background information on the family. Even here, analysis potentials vary due to the large differences in survey design. For example, researchers who are interested in broadly analysing effects of economic inequalities and hardship on children might use Understanding Society to learn more about temporal changes in families' economic situation, ways in and out of poverty, and their effects on children's educational and employment histories in their teens and adult age. They might refer to the MCS to find out about impacts of economic deprivation on children's development and physical health, particular in the early life course.

In part, both have been enriched (or are planning to enrich) with biological and external data (e.g. geocoded data or administrative data), creating an even stronger set of data to understand the life course of the population. A graphical representation of how these various types of data can be used to enhance longitudinal data and research in the UK is provided in Figure 3. This figure indicates that one source or type of data is not adequate for understanding the important questions of interest to researchers and policymakers and that the biggest gains are to be found when these data are combined together, as is currently being done in the longitudinal studies and is recommended to continue in the future. It also indicates the continuing essential role of longitudinal surveys; see further on this below and in **Section 3**.

Figure 3. Advantages and Limitations of Linking Longitudinal Survey and Administrative Data

Administrative Data		Longitudinal Survey Data	
Advantages	**Limitations**	**Advantages**	**Limitations**
✓ No additional costs for collection and non-intrusive ✓ Regularly updated ✓ Total coverage of admin population ✓ Potential controls and booster samples post hoc ✓ Reliable at small area level enabling a better place-focused policy making ✓ Timely data capture enables time series to explore trends	✗ Costs to make data research-ready ✗ Lack of metadata ✗ Legislative barriers ✗ Data controllers as data gatekeepers ✗ Coverage of population subject to policy changes ✗ Topic and concept coverage	✓ Tailored to research aims ✓ Measures are theory-driven ✓ Cleaned, of high quality, with metadata ✓ Accessible and discoverable ✓ Capture experiences, perceptions, attitudes and a range of attributes about individuals, their families and households ✓ Allow exploring relationships between earlier experiences and outcomes later in life	✗ Reliance on samples ✗ Under representation of certain groups ✗ Costly data collection ✗ Recall bias ✗ Long-time gaps and delayed measurement ✗ Content limited to that collected specifically for administrative purposes

Linked Longitudinal Survey & Administrative Data	
Opportunities	**Challenges**
✓ Improved representativeness across time, geographies and sub-populations ✓ More cost-effective ✓ More accurate, timely, and comprehensive information ✓ Increased reliability and validity ✓ Enables better theory and policy driven research ✓ Allows to test for a range of factors to understand complex social phenomena (e.g. social mobility)	✗ User Training ✗ Resource ✗ Access ✗ Consent ✗ Legislation ✗ Data Management

Section 2: Current and future content needs of the longitudinal studies

In this section, we will assess the 'hot' topics, open problems and needs that social science researchers, policymakers and practitioners have today, which require data from longitudinal studies. Furthermore, we will examine how far the topics, areas of inquiry and research potentials of the ESRC longitudinal studies cover these demands, and discuss in which ways they might be enhanced to reach their full potential.

2.1 The studies' coverage

Understanding Society follows a long-term content plan. The general priority topic areas are:

- income, wealth, consumption and expenditure

- health, wellbeing and related behaviours

- employment

- education

- family

Core questions on these issues are repeated in every wave. Rotating core questions appear in alternating waves or on a cyclical basis and cover, for example, family relationships, mental health, political and social engagement, psychological traits and time use. Stable characteristics such as family background, religion, chronic health conditions and retrospective questions on migration, employment status, partnership and fertility are asked only once. Most questions on time-changing characteristics are repeated frequently and regularly in unchanging form. Thus replication of measures is stressed in this study type, offering the advantage of describing long-term time trends and modelling individual changes over the life course by using advanced panel techniques. Household sampling makes BHPS and Understanding Society particularly suited for analysing household and family dynamics, mutual dependency of life courses within families and partners' joint decision-making. The large sample size allows relatively fine cohort comparisons and analysis of smaller subgroups in society. Furthermore, Understanding Society collects specific information for the largest UK ethnic minority groups and new immigrants. Youth (aged 10-15) and young adults (aged 16-21) receive specific questionnaires instead of the standard personal interviews for adults.

The cohort studies (i.e. NCDS, BCS70, Next Steps, MCS) collect follow-up information less frequently, particularly in later stages of the study. Instead of repeating the same set of questions in every wave they are strongly targeted to respondents' specific life stage. They focus on the intergenerational transmission of advantage and disadvantage; how early life circumstances and experiences influence outcomes later in life; individual change over time and the mapping of lifetime trajectories; and how health, wealth, family, parenting, education, employment and social attitudes are linked to each other, and how these aspects of life vary for different groups of people, and, by comparing between these studies, across cohorts. Accordingly, the main focus of the Millennium Cohort Study (MCS) lies in observing parenting (therefore it provides mother and father questionnaires), childcare and school choices, child behaviour and cognitive development, as well as child and parental health. The study also aims to survey the broader socio-economic context and collects data on parents' employment and education, income and poverty, housing, neighbourhood and residential mobility, social capital and ethnicity. Survey design and the strategy of targeting questions to life stages make the birth cohort surveys particularly useful for observing various dimensions of child development until adult age as well as transitions through childcare and educational institutions into adult age. The fact that MCS has collected direct information on both parents and older siblings makes this study particularly useful for examining linked lives in early childhood and the effects of the home learning environment on child development.

In sum, the ESRC longitudinal studies examined most closely in this review neatly complement each other. Regarding their core contents, studies in both areas cover sociodemographic and socioeconomic background information on the family. Nevertheless, due to the differences in survey design their analysis potentials vary. Both types of studies have been enriched (or are planning to enrich) with biological and external data (e.g., geocoded data or administrative data – see further in **Section 3**). Combined with the different survey information the two provide, they allow for quite different analyses, and therefore, this complementary approach is highly advisable.

Parents' wellbeing affecting childhood outcomes

Data from the Millennium Cohort Study and Understanding Society have shown how childhood outcomes are affected by parental well-being and that the quality of adult relationships is independent of the parents being married. This has shifted language in public policy away from focusing on the importance of parental marriage to the importance of parental relationships.[15]

This positive picture is also reflected in the studies' usage. Nearly all the ESRC longitudinal studies, in particular the core studies (BHPS and Understanding Society on the one hand, the birth cohort studies on the other hand), have been used extensively, which becomes apparent by their high and increasing download numbers, publications, and studies using their data (for a detailed discussion, see **Section 4**). Users come from a wide range of disciplines within the social sciences (see **Table 3e**, **Annex L** for the most common) and the usage and impact of the ESRC core studies also reaches beyond the social sciences, for example to biomedical and genetic science. Genetic data from the NCDS has achieved particularly high use in these fields (see further in **Section 4**).

2.2 Data users and the role of longitudinal studies

ESRC-funded longitudinal studies have a core group of users from different social science disciplines, including sociology, economics, political science, demography, educational science and psychology, as well as in biomedical science. The primary purpose of the longitudinal studies is to provide useful and relevant data for these disciplines and beyond, which allows users to answer novel research questions and use state-of-the-art methods. Since research nowadays is less isolated by field and discipline, the studies should foster interdisciplinary exchange and outreach to fields that work on questions that are related to traditional social sciences, for example geography, arts and humanities, or medical and biological science. Similarly, since research is less restricted by national boundaries, the studies should also be suited for international comparative research (see **Section 3** for technical aspects on this discussion). And finally, the studies should pose valuable and up-to-date data sources for questions posed by policymakers and practitioners in social and economic policy fields, which require longitudinal data. Therefore ESRC's longitudinal studies should be suited for reliably describing the prevalence of social phenomena, their change over time, and their consequences across the life course.

Use of longitudinal studies in genetics research

"The usage and impact of core ESRC-funded data and birth cohorts that contains biomedical and genetic information outside of the social sciences has been remarkable. Considering that the 1958 Birth Cohort has been one the most used datasets in GWAS discoveries to date, it has made a considerable impact on international biomedical, genetic and fundamental biological research. These studies have led to fundamentally new breakthroughs in cancer (notably breast, colorectal and prostate cancer), immunity, protein measurement and the nervous system." **See Annex C**.

ESRC longitudinal studies have to fulfil a number of different purposes and answer a wide range of research questions from different perspectives, user groups, and fields. In this sense, they are (and have to be) infrastructural tools. This broad and open nature of the studies also has a positive side effect: "Longitudinal studies by their nature are unique and future unspecified uses should not be underestimated."[16] However, according to our expert viewpoint, it is also necessary for each study to define a basic rationale that includes a set of thematic fields, overarching research questions, and analysis objectives that the study should be able to answer. This step is necessary to justify each study's funding and to point out to research and policy stakeholders the study's unique selling points in the landscape of UK longitudinal studies, and more broadly, data for research. Therefore, the ESRC longitudinal studies should reach a compromise between their function as a broad infrastructural tool on the one hand and targeted research tool on the other. In order to ensure that the ESRC studies fulfil the requirements stated in this section, it is necessary to test them against the interests, questions and needs of the scientific community, policymakers and practitioners from the fields described above.

---

[15] Harold, G. T., Acquah, D., Chowdry, H., & Sellers, R. (2016). What works to enhance inter-parental relationships and improve outcomes for children. Department for Work and Pensions (DWP), Ref: ISBN 978-1-78425-719-4, Ad hoc research report 32.

[16] Townsley, R. (2016). **Interim report: initial analysis of responses to the consultation**. Ruth Townsley Research

2.2.1 Representativeness of the longitudinal studies

It was clear throughout the data-gathering phase of the review the high esteem that many data users held for the longitudinal studies in the UK. However, one issue was consistently highlighted as a continued challenge: representativeness of data and the studies. Though not always explicitly stated, this representativeness is usually understood as applying to the UK population. In general it seems safe to say that representativeness is perceived by many of the social scientists that contributed to the review as a primary defining characteristic of an ESRC-supported longitudinal survey collecting data suitable for social research and for policy analysis, inasmuch as this property allows valid generalisation to the UK population of socio-economic attributes and behaviour observed on the survey sample.

We therefore continue to emphasise representativeness throughout this review and include data on the representativeness of the CLS studies and Understanding Society in **Annex F**. However, it should be noted that the representativeness concept is itself not a well-defined one, particularly when used without clarification. When claiming representativeness one needs to also be able to answer (implicitly or explicitly) the question: "Of what population is the sample representative?" That is, the "target population" for inference using the longitudinal sample data needs to be defined. In general this is the population of all socio-economic entities that could have been selected into the sample. Note that this does not require that each such entity have an equal chance of being selected, just that each has a non-zero chance. As a consequence there is no requirement that a representative sample be some sort of "mirror image" of the population.

The second important clarifier relates to what one can then infer about the population given the claimed representativeness of the sample. In effect, one needs to also have an answer to the question: "For what characteristics of the target population can the sample data provide information?" This implies that not all characteristics of a population can be inferred from a sample that is representative of this population. The question is usually answered by the creation of sample weights of one form or another, such that weighted sample statistics can provide statistically valid estimates of defined population characteristics. Here "statistical validity" means that comparability assumptions that are typically implicit in these weights are actually true.

In sum, representativeness is a key characteristic of any longitudinal dataset, but one that requires clarification when applied to a given sample, particularly regarding the definition of the population from which the sample is drawn ("representative of what?"), and also the characteristics of that population that can be inferred from the sample data ("representative for what?"). It can be seen that the concepts of representativeness and external validity are closely related, justifying the importance placed on representativeness in this review.

2.3 Researchers' needs: Priority areas and new demands

In the ESRC 2006 Strategic Review of Panel and Cohort Studies, eight priority areas for longitudinal research were highlighted:

- ageing population
- long-term effects of childhood experience
- the timing of transitions
- demographic shifts and mobility
- the biotechnology revolution
- immigration
- cultural diversity and inequality
- globalisation

In the autumn 2016 open consultation for the current review all these priority areas were still felt to be important, while two were rated as particularly relevant: long-term effects of childhood experience and ageing. However, since the topics were given in the survey and respondents were asked to rate each of these and rank the top three, this question format might have predetermined their answers, and we cannot be sure that the ranking really reflects their own content priorities.

Subsequently in the consultation, respondents could also mention two additional priority areas and name sub-fields in open-text format. We refer to these answers to assess their personal content priorities. When the additional answers are taken into account, differences in their topical interests turned up mostly in terms of terminology, emphasis, and conceptual breadth. The 2006 priority of cultural diversity and inequality was split into the areas of equality/inequality

on the one hand, and diversity/identity on the other hand. Health and wellbeing emerged as a separate scientific priority in its own right. Biosocial research and genomics updated the 2006 priority of the biotechnology revolution. The 2006 areas of globalisation, immigration and timing of transitions were each subsumed within one or more of the broader priority areas. In sum, the 2016 open consultation revealed seven main scientific priority areas that longitudinal studies need to address:

• Long-term effects of childhood and adult experiences

• Demographic shifts and mobility

• Health and wellbeing

• Equality and inequality

• Biosocial research and genomics

• Diversity and identity

• Ageing population

These areas differ in their conceptual breadth and how much they incorporate scientific sub-themes. Most of the frequently mentioned sub-themes again reflect core topics and content that are covered by the longitudinal studies. Most importantly, diverse aspects of social inequalities were mentioned, as **Figure 4** shows. Additionally, two domains of the life course seem to be particularly interesting to respondents: education, and mental health and wellbeing.

These fields were also mentioned quite often in the 2017 Oxford workshop and the panel's subsequent consultations with a wide range of stakeholders. Finally, it might be worth mentioning that despite the fact that ageing was rated particularly high in importance in the closed-format survey question on top priority areas, very few indications on sub-fields were mentioned in the open-ended follow-up question.

Figure 4. 'Top ten' sub-themes cited in the 2016 consultation, by respondent sector
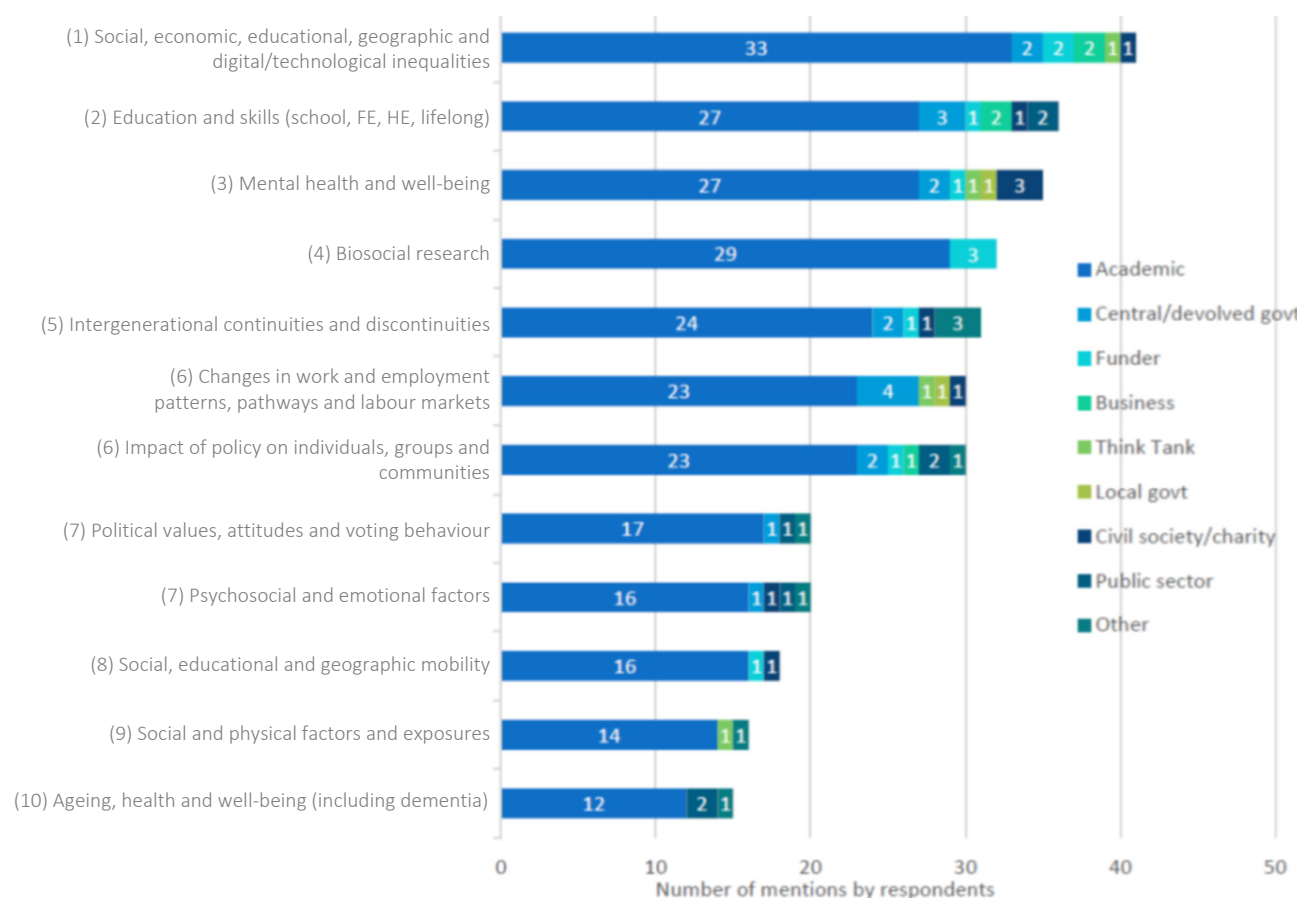
Closely connected to these indications, additional topics which cannot be answered sufficiently with the existing ESRC longitudinal studies turned up during our consultations.

A report[17] to the **Nuffield Foundation** remarks that currently there is limited data collected on negotiation processes and dynamics in separated families from a parental view or about the lives of non-resident parents and their households. Understanding Society currently provides the most comprehensive longitudinal data on this issue and is viewed as a strong dataset for studying family separation due to a reasonable sample size of currently separated families, a good range of data on family life, and the continued collection of data from parents and partners who leave the original household.

In the area of health and wellbeing the most pressing issues seem to be mental health, obesity and healthy ageing. Although a need for content on these issues was identified in the 2016 consultation and the 2017 workshop, experts in the field who met at a CLOSER workshop in summer 2017 on mental health issues state that the ESRC studies provide quite diverse information on these issues.[18] Rather, a more pressing problem in using these studies for studying mental health seems to be the variety of measures used. For future work, the workshop concluded that the existing measures should be validated and compared with a calibration study, with the potential goal of creating standardised variables that are better synchronized in different studies and can facilitate cross-study comparisons. Another problem is the fact that mental health patients often are not followed longitudinally in the studies and several high-risk groups are completely excluded from the samples, for example homeless persons or persons who are ill for long periods.[19] Since there is some research evidence that early recognition of mental health disorders could mitigate trajectories, Wellcome is currently considering investing in a mental health data platform with children/adolescents as a cornerstone.

The long-time impact of bullying

Data from the National Child Development Study has shown that bullying in childhood leads not only to poor childhood mental health, but also affects physical and mental health some 40 years after the bullying took place.[20, 21] Childhood bullying is associated with increased demands on mental health services up to mid-life[22], and with other indicators of societal economic costs.[23]

The most frequently mentioned sub-theme under the new priority area of diversity and identity was the need for data to understand changes to political values, attitudes, and voting behaviour, including public trust. Social and environmental values as well as civic engagement were also mentioned here. These answers reflect the fact that the longitudinal surveys have not covered central issues in political science very well, while political science surveys are typically not longitudinal. Relatedly, studying the consequences of policy changes, in particular Brexit, on individuals, groups, and communities was mentioned quite often by different actors during our consultations.

Finally, a new topic that was stressed frequently throughout our consultations was the broad issue of technological changes such as digitisation, robotics and artificial intelligence, and their impact on the labour market structure, working lives and inequalities, fertility, learning, and cognitive function. Children's interactions with the digital world are becoming increasingly important and seamless with the non-digital world. Children's interactions with technology need to be captured in cohort study data to aid understanding of their developmental effects.[24]

[17] Bryson, C., et al. (2017). **Understanding the lives of separating and separated families in the UK: what evidence do we need?** Summary report. Nuffield Foundation: London.

[18] ESRC. (2017). **ESRC Longitudinal Studies Review: Workshop summary report**.

[19] Personal communication (2017)

[20] Takizawa, R., Maughan, B., & Arseneault, L. (2014). Adult health outcomes of childhood bullying victimization: Evidence from a 5-decade longitudinal British cohort. *American Journal of Psychiatry*, 171, 777-784.

[21] Takizawa, R., Danese, A., Maughan, B., & Arseneault, L. (2015). Bullying victimization in childhood predicts mid-life risks for cardiovascular disease: A 5-decade birth cohort study. *Psychological Medicine*, 45, 2705-2715.

[22] Evans-Lacko, S., et al. (2016). Childhood bullying victimisation is associated with use of mental health services over 5 decades: A longitudinal nationally-representative cohort study. *Psychological Medicine, 47*, 127-135.

[23] Brimblecombe, N., et al. (under review). Long term economic impact associated with childhood bullying victimisation.

[24] Centre for Longitudinal Studies. (2017). **The value and future of birth cohort studies for social science and policy: Summary of Round Table meeting at the British Academy, July 10th 2017**.

Most importantly, the 2016 consultation, the 2017 workshop and many of the subsequent expert interviews echo a request for a new birth cohort study which could address the current data gap on UK children since 2000/01. The most important rationale for a new cohort study would be to maintain a continuous flow of data to track societal changes and their impacts on generations born in the 21st century, and to address policy questions related to these changes.

Regarding the design of a new birth cohort study, experts repeatedly stressed that a future study should be based on a UK nationally representative probability sample based on NHS unique personal identifiers (covering almost the entire UK population), which includes demographic information and allows linkage to medical records and potentially other administrative data (for a discussion of a new model of administrative data spine, see **Section 3**). A new birth cohort study might concentrate not only on children, but also could collect data on resident and non-resident fathers and mothers in order to adequately address the gene-environment interplay[25], to better understand the dynamics of separated families[26], and to enable more research on the intergenerational transmission of inequalities.[27] This would imply changes in the follow-up sampling design compared to current birth cohort studies, similar to Understanding Society.

The high interest in a new birth cohort study coincides with the current strong interest in early childhood's prediction of lifespan effects. Here respondents mentioned that the interplay not only of post-natal, but also pre-natal factors and exposures, as well as environmental factors (including pollution and climate change) might be of great research interest.[28] Taken together, these indications highlight the high and continuing relevance of long-term childhood effects, which accordingly turned out to be one of the two most important priority topics in the 2016 consultation.

2.4 Interdisciplinarity and non-user outreach

Since nowadays research is less isolated by field and discipline, and the UK funding landscape is converging with the formation of UK Research and Innovation (UKRI) in 2018, it is important that the longitudinal studies are increasingly useful for interdisciplinary exchange and cross-discipline fields that work on questions that intersect into the social sciences.

Traditionally, the ESRC longitudinal studies, particularly the birth cohort studies, have united biomedical, medical, epidemiological and social science research objectives, and have answered many relevant interdisciplinary questions. The most recent of these projects, Life Study, also suggested the limitations of interdisciplinary studies. In the social sciences an important prerequisite for establishing a high-quality longitudinal survey is to incorporate a relatively wide range of questionnaire content, whereas typical medical science surveys are much more heavily based on answering specific causal questions. Going along with these principles are different viewpoints and demands on representativeness, sample size, and study design. Combining both research models without a clear distribution of priorities and responsibilities has the potential for resulting in overburdening study content and sampling design. Accordingly, respondents of the 2016 consultation mentioned that a new birth cohort needs to be simpler and less ambitious than the Life Study.[29] However, from a multidisciplinary perspective we think that at the same time it is important to continue to incorporate interdisciplinary elements and innovation into the studies.

Consistent with this view, data users also stress that the collection of biomarkers and the study of genetic, epigenetic and other omics effects should continue to be included in the longitudinal studies, including to map children's early and later exposures and to continue their interdisciplinary tradition[30]. We agree with these claims, particularly because we see promising ground for novel research by using genetic research information in the longitudinal studies to examine correlations between certain genetic constellations and social attributes, such as educational attainment. An alternative idea for strengthening these research potentials which came up during our consultations was to link the longitudinal studies with **UK Biobank**, a large-scale medical study funded by the MRC, Wellcome and others. However, sample overlap between the studies may not be sufficient and this approach would cause severe problems in the realm of data harmonisation (for a more detailed discussion, see **Section 3**).

[25] Townsley, R. (2016). **Interim report: initial analysis of responses to the consultation**. Ruth Townsley Research

[26] Personal communication (2017)

[27] Personal communication (2017)

[28] Townsley, R. (2016). Interim report: **initial analysis of responses to the consultation**. Ruth Townsley Research

[29] Ibid., 27.

[30] Ibid., 27.

Today's non-users of the longitudinal studies who could be potential future users come not only from other disciplines, but also within the social sciences. Dependent on discipline, a significant proportion of researchers who could use the longitudinal studies' data for their work are not doing this. One main obstacle for this group is knowledge, know-how, and training (see **Section 4**); another challenge is differences in methodological approaches and empirical traditions. Particularly academics who use qualitative methods might feel that the longitudinal studies have nothing to offer them. However, some longitudinal studies have created opportunities for mixed methods research, for NCDS[30a], Understanding Society[30b] and the **Avon Longitudinal Study of Parents and Children** (ALSPAC) which show that triangulating quantitative and qualitative data and methods may produce very fruitful results.[31, 32, 33, 34]

2.5 Methodological requirements and innovations

The need to contribute to cutting-edge research requires data that not only depicts recent and ongoing changes in society, but also fulfils changed methodological needs and makes use of recent methodological advancements.

In the social sciences, particularly in fields such as economics and political science, testing causal effects has become increasingly important in recent years. Since it is hard to set up a causal frame based on observational field data, experiments and treatment/intervention studies have been increasingly established. Longitudinal studies are crucial tools in this strand of research, because they allow for controlling at least time-constant unobserved heterogeneity and for making use of natural experiments. This usually requires a series of consistent observations, at best from many different birth cohorts, in order to make use of exogenous shocks in a flexible manner. Within the UK, policy differences in the devolved administrations, particularly in Scotland and Wales, often provide natural experiments which may be exploited by the longitudinal surveys, provided that their populations are adequately represented in the surveys. Furthermore, controlled experiments may be integrated into the surveys. Here, the Understanding Society Innovation Panel is a valuable platform which could be used more frequently for short-term experiments, similar to the GSOEP Innovation Sample in Germany.

At the same time, recent and ongoing technological advances in data storage and processing have changed the data landscape tremendously and set the ground for using a larger amount of more complex data in empirical analyses. As a consequence, researchers increasingly have to take into account other data sources and integrate them in their survey-based analyses to handle innovative research questions, particularly on separating factors on the micro-level of individuals, families and households from environmental factors on the meso-level of neighbourhoods, schools and work organisations, and the macro-level of opportunity structures and institutions. As an example, studies of granular effects of spatial conditions and environmental hazards require longitudinal information based on geo-coded data as well as health records besides longitudinal survey data. Since few researchers have the know-how and the resources to build such complex databases on their own, this requires ready-to-use survey data enriched with spatial and administrative information wherever possible (for a discussion of data linkage, see **Section 3**, for a discussion of training needs, see **Section 4**). Similarly, technology has changed everyday communication. Today, social media as well as other kinds of internet-based communication produce vast amounts of big data (e.g., on social networks, new forms of social engagement, communication contents, consumption flows, etc.), which may pose interesting research areas for the social sciences, and possibly will be used increasingly in combination with survey data – at least if ethical issues around their usage are solved.

[30a] Elliott, J. (2013). Talkin' 'Bout My Generation': Perceptions of Generational Belonging Among the 1958 Cohort. *Sociological Research Online,* 18, 13-13. Elliott, J. (2010). Imagining a Gendered Future: Children's Essays from the National Child Development Study in 1969. *Sociology*, 44(6), 1073-1090.

[30b] Gush, Karon; Scott, James; Laurie, Heather (2015). Households' responses to spousal job loss: 'all change' or 'carry on as usual'? *Work, Employment and Society* (29:5), 703-719

[31] Stevens M, Perry B, Burston A, Golombok S, Golding J. (2003). Openness in lesbian-mother families regarding mother's sexual orientation and child's conception by donor insemination. *J Reprod Infant Psychol. 21* :347-62

[32] Nichol, J., Thompson, E. A., & Shaw, A. (2011). Beliefs, decision-making, and dialogue about complementary and alternative medicine (CAM) within families using CAM: a qualitative study. *The Journal of Alternative and Complementary Medicine, 17* (2):117-125.

[33] Sidebotham, P. (2001). Culture, stress and the parent–child relationship: a qualitative study of parents' perceptions of parenting. *Child: care, health and development. 27* (6):469-485.

[34] Gorell Barnes, G., Thompson, P., Daniel, G. & Burchardt, N. (1997). Growing Up in Stepfamilies. Oxford: Oxford University Press.

Nature or nurture?

Research using the ESRC supported longitudinal studies has shown that social experiences such as social isolation[35], parental separation in childhood[36] and unemployment[37] lead to increased biological stress markers and inflammatory functioning. The research has brought social and biological theory closer together and led to the emergence of the new field of biosocial research.

Technological advancements have not only changed the means of everyday communication, but they might also influence the modes of surveying. Participation in traditional face-to-face, telephone and postal modes has declined continuously. Both developments result in an increasing need for new, less obtrusive data collection methods, for example data collection based on existing or newly developed apps or wearable technology. Testing and application of both approaches have already started in the Understanding Society Innovation Panel (as discussed at two CLOSER workshops in 2017 on new technologies[38, 39]). While these new methods generate more precise data particularly on everyday activities such as movements in space, time use, or the use of social networks, they need thorough development and testing in order to assess the validity and reliability of the data.

In the future, surveys will constitute only a relatively small part of the social science data landscape. Their role is nevertheless extremely important. Only by means of longitudinal surveys is it possible to collect information on subjective perceptions, values and opinions, decisions and motives, evaluations, and feelings of individuals repeatedly over time. No other data source or collection method can generate this kind of data; therefore longitudinal surveys cannot be substituted.

2.6 International comparisons

The increased relevance of internationally comparative studies was mentioned quite a few times in the 2016 consultation as well as in the 2017 workshop. Data users explained that over the past 10 years increased opportunities for new comparative work have become available, due to advances in longitudinal studies in countries such as the USA, Germany, Japan and others. Various topics of cross-national interest were mentioned, for example:

• Comparing life-course trajectories to better understand how policy and circumstances affect wellbeing, health, development and other outcomes

• Comparing migrants' experiences in different countries to identify factors that foster successful integration

• Comparative historical analysis on areas such as income mobility

• Socio-economic issues such as savings practices or attitudes toward redistribution

• Political research topics such as attitudes regarding the law

• Specific health behaviour issues such as child obesity to assess the impact of policies

• Genetic issues would profit from larger sample sizes by pooling samples

The Royal Statistical Society mentioned, however, in their assessment of the longitudinal studies[40] that for now the proliferation of cohort studies around the world lacks a methodology to make sense of comparative analyses, and they recommend that ESRC and UK researchers should lead development work in this area. We agree with them on the need to increase efforts in retrospective and prospective harmonisation of longitudinal studies in different countries,

[35] Lacey RE, Kumari M, Bartley M. (2014) Social isolation in childhood and adult inflammation: evidence from the National Child Development Study. *Psychoneuroendocrinology*. 50:85-94. doi: 10.1016/j.psyneuen.2014.08.007.

[36] Lacey RE, Kumari M, McMunn, A. (2013) **Parental separation in childhood and adult inflammation: The importance of material and psychosocial pathways**. *Psychoneuroendocrinology*. 38(11) 2476-2484.

[37] Hughes A, McMunn A, Bartley M1, Kumari M. (2015) Elevated inflammatory biomarkers during unemployment: modification by age and country in the UK. *J Epidemiol Community Health*. 69(7):673-9. doi: 10.1136/jech-2014-204404. Epub 2015 Feb 19.

[38] CLOSER Knowledge Exchange Workshop: **New technologies to measure non-health topics in longitudinal studies** (4th May 2017)

[39] CLOSER Knowledge Exchange Workshop: **New technology and novel methods for capturing health-related data in longitudinal and cohort studies** (19th May 2017)

[40] Personal communication (2017)

in particular birth cohort studies, to enhance internationally comparative research and improve its methodological foundations. Since our recommendations in this regard are of technical nature, they are explained and justified in more detail in **Section 3** of this report.

2.7 Policymakers' needs

In general, the needs of policymakers and non-academics differ from the needs of the scientific community in three central aspects: first, they need results more quickly to be able to react in time and reach the affected population with policy measures; second, they have a large interest in observing changes over time to identify spheres of political action; and third, they often need data from intervention studies to evaluate measures in trial studies before implementing them. Prerequisites for data to be representative over time and to deliver reliable results are particularly high in the political arena. This explains why official statistics and large-scale administrative data are important information sources in this area, and why government often creates its own survey data (e.g. in the **Department for Digital, Culture, Media and Sport**). Even if time series are short, this is an important part of the data landscape that policymakers use. Despite having access to their own (administrative) data sources, some central government departments (e.g., **Department of Work and Pensions**, **HM Treasury**, **Department for Education**, **Department of Health and Social Care**) use the longitudinal studies in their work and in some cases link it with their department's administrative data.

The need for research is extensive and constant, whether basic, translational, clinical, epidemiological, public health, social, economic or policy research. Current government priorities in the Department of Health and Social Care, for example, identify two issues with a social science connection: lifestyle diseases, diabetes and obesity, on the one hand, and mental health and dementia on the other hand (for details, see their **Areas of Research Interest statement**). Co-funders of ESRC studies such as MRC commented on policymakers' needs in this field as well.[41] There is potential for a greater application of resources to address health challenges such as obesity, alcohol consumption, and healthy ageing. Therefore, they recommend greater use of the existing longitudinal studies to address public health challenges. More specifically, they recommend establishing effective models of two-way engagement between cohort study teams, policymakers, and practitioners to increase the impact of research outputs and potential for translation to inform evidence-based policies. We agree with this argument.

In addition to interest in more evidence on aspects of health, other government-identified areas of interest include more data on the gig economy; impact of leaving the EU on the economy, and individuals' skills, attitudes and behaviours; impact of events in early life on educational outcomes; and data on the crucial educational transitions today. For this latter area of interest, the large gap in birth cohort data is particular problematic. Given the need to make quick use of data, a new cohort starting with preschool or school children, not with new-borns, would be needed. Thus, apart from a stronger inclusion of policymakers in the phase of designing a new study, policy units in the study teams could provide a first port of call for such activities.

> ## Low-income households and child development
>
> The idea that socio-economic disparities in children's experiences in the pre-school period are a fundamental driver of later social inequality has had a major influence on government social mobility and early-years policy since 2010. The Millennium Cohort Study has been used to document significant differences in children's cognitive, socio-emotional and physical development at school entry across the full range of parental income groups – demonstrating to policymakers the many ways that less advantaged children's experiences hamper their development.

In some policy fields, there are important open questions regarding specific populations at risk, which are not covered adequately by the existing longitudinal surveys (i.e. high-priority local data, kids in foster care, jail inmates or homeless persons). Policymakers stress that evidence on these groups is needed to develop targeted measures for them.[42] Either these populations are covered in too small sample sizes in the current studies, as in the case of regional or local data, or they are absent because they are not included in the studies' sampling frames. Thus, some government departments have built their own longitudinal data resource by linking vast amounts of administrative data to evaluate programmes and to fulfil their research priorities.

[41] Personal communication (2017)
[42] Personal communication (2017)

A similar, but probably even more pressing issue is the lack of some studies, particularly the older birth cohort studies, to generate representative data on each of the devolved nations and tackle the related language and implementation issues.[43] Therefore a complex and mostly un-coordinated landscape of country and regional longitudinal studies has evolved, including for example **ALSPAC** in South-West England, **Next Steps** in England, and **Growing Up in Scotland**. Regional coverage of Understanding Society seems to be higher in comparison. While the samples in most individual local authorities are not large enough to support separate analysis, they are large enough to estimate between-local authority variance.

In sum, government analysts' usage of the longitudinal studies differs considerably between departments. Regarding more general challenges, differences in language and terminology were mentioned quite a few times. Therefore it seems crucial to invest further in capacity for bridging these differences (the Policy Unit in Understanding Society is a good example here). Another critical issue is the feeling of some policymakers that the data is not helpful for answering questions arising in their daily work, because they need a quick turnaround time or the longitudinal studies may not have the data needed to answer certain parliamentary questions. This problem could be solved on the one hand by programming a data dashboard which enables the complex data structures of longitudinal studies to be analysed much quicker and easier than before (see **Section 4**) to reduce turnaround time. On the other hand, we think it is advisable to give policymakers – and most importantly analysts from central departments and regional governments – an opportunity in consultations to feed their aims and needs into study design when designing new longitudinal studies or when revising the design and content of existing ones, so the longitudinal studies have the data needed to help answer the questions most analysts are most commonly asked. Furthermore, new organisations and platforms such as **Health Data Research UK** and working with **NHS Digital** will possibly increase the opportunities to integrate longitudinal survey data with health data (for details, see **Section 3**).

2.8 Summary

In general, the current portfolio of ESRC longitudinal studies covers the most important research objectives, content needs, and topics social science researchers are interested in today. Most importantly, this regards long-term effects of childhood development, the manifold dimensions of social equality and inequality, health and wellbeing, demographic changes, mobility and the ageing society. The two pillars of the ESRC longitudinal studies – household panels and birth cohort studies – have jointly provided a well-suited infrastructural base for carrying out state-of-the-art social science research. Understanding Society is increasingly becoming an established instrument for examining socio-economic household and family changes, and internationally the survey is highly acknowledged for its know-how and constant innovations in the field of survey methodology. The UK has a longstanding, leading role in the world in the development and implementation of large-scale birth cohort studies. Both achievements should be continued in the future.

In order to keep this role it is important for all studies to gradually adapt questionnaire content to changes in society, such as digitisation or changed patterns of ageing, and it seems reasonable to continue and increase investing in technological advancement to make use of the new technological possibilities for surveying individuals. However, what seems to be more pressing in this respect is to increase the analysis potential of the existing surveys and their questionnaire programmes by enriching them with other data sources, and by increasing harmonisation between the UK surveys and internationally.

If there are any 'structural holes' in the UK longitudinal social survey landscape today, they stem not so much from missing questionnaire content, but more from missing or unreliable information on particular subgroups in society, such as recent birth cohorts, the populations in the devolved nations, particular regions or localities, or high-risk populations. These gaps are either due to missing investments (such as a new birth cohort), missing coverage (high-risk populations), and/or limitations in representativeness or sample size issues (devolved nations, regions). Opportunities to address these issues would be to provide a more flexible study design, which could be reached easily with the availability of an administrative data spine. Existing studies would be linked with the spine and new study samples would be drawn from it (for a detailed description, see **Section 3**).

---

[43] Personal communication (2017)

Section 3: Enhancements of longitudinal studies: data linkage and harmonisation

Following the ESRC open consultation exercise in 2016 as well as the ESRC invitation-only workshop in January 2017, two related approaches to enhancing the existing datasets were singled out for special mention: linkages to administrative data and harmonisation with other datasets. In this section we detail the current state of these two aspects of data enhancement as far as the longitudinal studies are concerned, and also consider their use as tools for prospective studies.

3.1 Data linkage

The initial analysis of the responses to ESRC's open consultation survey[44] identified data linkage as the most frequently mentioned methodological and technological issue. It was the only methodological and technological issue mentioned by respondents from across all five primary research areas, showing the importance that this topic holds in any discussion of the uses and future role of administrative data linkage as far as the ESRC data investment portfolio is concerned.

Over the last 20 years, the UK has seen a digital transformation of public services[45] that has in turn led to more "linkable" administrative data, with key government departments now controlling very large and complex client databases. With devolved administrations in Northern Ireland, Wales and Scotland, there has been considerable interest (and flexibility) in linking the country-specific components of these databases to allow policy research and evaluation at local levels. The value of this linkage is now being recognised by central government, with the enactment of the **Digital Economy Act 2017** encouraging data sharing and an enhanced role for the **UK Statistics Authority** in this regard (see **Annex G** for a brief note on UK data legislation). In August 2017, the UK Office for Life Sciences released the Life Sciences: Industrial Strategy report[46], which recommends the creation of regional digital innovation hubs that use longitudinal-linked data from the UK Health System to provide "comprehensive and secure data in primary, secondary and tertiary care as well as social care and community data" for the development and testing of new health technologies.

At the same time, the sample survey paradigm underpinning data acquisition for scientific research has evolved, and linkage of survey data to external data sources is now an important research tool. In particular, analysts now routinely use data linked from multiple sources to improve inference. This is most notable in the health sector, where linkage is frequently employed to enhance data on clinical performance and patient health outcomes.[47] In order to unlock the research potential of routine data collected across the UK government, and following the recommendations of a report by the **Administrative Data Taskforce** (ADT)[48], the ESRC made a series of investments in 2013 to establish a UK **Administrative Data Research Network** (ADRN) that would provide safe access to linked de-identified administrative data. In arguing for this new research resource, the ADT report notes that the linked administrative data could be used in many areas of policy-relevant research, including social mobility (by linking data on education, training, employment, unemployment, incomes and benefits), causal pathways over the life course (by linking data on education, health, employment, incomes and wealth), poverty (by linking data on housing conditions, health incomes and benefits), and criminal offenders' mental health and tendency to reoffend (by linking data on (re)offending behaviour, incomes, benefits, and health). In particular, this report notes that "... value can also be gained from linking administrative data to other studies, including ongoing longitudinal and other surveys."[49] Linkages of this type play an important role in longitudinal resources, and are reflected in the activities of the Farr Institute and the strategic vision for **Health Data Research UK**. They have considerable potential for reducing the burden on respondents to such surveys and for improving the quality and extent of the information they provide.

---

[44] Townsley, R. (2016). **Interim report: initial analysis of responses to the consultation**. Ruth Townsley Research

[45] UK Cabinet Office, Government Digital Service & The Rt Hon Ben Gummer. (2017). **Government Transformation Strategy 2017 to 2020** [Policy Paper].

[46] UK Office of Life Sciences. (2017). **Life Sciences: Industrial Strategy report**.

[47] See references [4], [5], [6], [12], [13], [15], [17], [18] and [21] in Harron, K., Gilbert, R., Cromwell, D. & van der Meulen, J. (2016). Linking Data for Mothers and Babies in De-Identified Electronic Health Data. *PLoS ONE*, 11(10): e0164667. **https://doi.org/10.1371/journal.pone.0164667**

[48] Administrative Data Taskforce (2012). **The UK Administrative Data Research Network: Improving Access for Research and Policy**.

[49] Ibid., 49; Page 1

Having said this, it should also be recognised that administrative data linkage is no panacea for UK longitudinal research. Administrative data, by definition, are collected for administrative, and not scientific, purposes. And though these data often have longitudinal structure, they typically do not address the scientific issues that are the focus of the data collected in the longitudinal studies currently supported by the ESRC. Furthermore, data ownership is complex (see **Annex H**). As a consequence, the main advantage of administrative data linkage is the capacity to extend and enhance, rather than replace, a properly targeted longitudinal survey.

Longitudinal data linkage is the linkage of longitudinal survey data to a range of other (often also longitudinal) administrative data, such as health, tax, welfare and educational records, open or free data, as well as to "big data" such as digital footprints. This linkage can range from a simple collecting together of records belonging to the same person across the different systems, to full integration of these records including the creation of a cross-validated and harmonised time structure. The UK has extensive administrative data resources as well as what is internationally recognised as the best set of longitudinal survey data resources in the world. Linking across these two data domains builds on this unique combination and can be expected to provide considerable benefits, some of which are[50]:

• More efficient data collection and lower participant burden, particularly when information directly obtained from a respondent can be replaced with better quality data available from an external source. This would allow surveys to focus on collecting valuable information that can be obtained only in this way.

• Increased completeness of data as well as increased information for correction of participant bias (e.g., due to missing data).

• Collection of information that cannot be obtained from survey participants, or is subject to reliability concerns, thus expanding the potential of the dataset obtained directly from these participants. For example, collecting more reliable data on retrospective or hard to remember events may be possible via data linkage.

• A potential increase in coverage for the survey by allowing the study of subpopulations who are inadequately covered by the traditional data collection process, but have substantial contact with service providers.

• A potential increase in the representative use of the survey data by allowing subpopulations present in the survey sample to be nested within a population for which linked administrative data are available and hence facilitate population-level analysis for these subpopulations.

Computerised record linkage is now ubiquitous in many areas of health and social research, which in many cases is a consequence of the move to more digitised public services and thus more electronically linkable records. This is particularly the case when there are common variables (identifiers) stored on these datasets. Recognising the growing importance of data linkage in research, dedicated data linkage centres were set up in Australia and Canada in the mid-1990s. Facilities for carrying out data linkage are now relatively widespread, and include dedicated data linkage centres in Brazil, the US, New Zealand, England, Wales, Scotland and Northern Ireland. From a more UK perspective, Wellcome, a major funder of UK and international longitudinal studies, has recently released its longitudinal population studies strategy[51]. This strongly supports further data linkage in longitudinal studies, and in particular the development of methods for improved data linkage.

The creation of longitudinal survey databases with links to various administrative databases is half the story. The other half is what is done with these linked datasets. Here progress has been much slower and additional methodological work is needed. Linked data are subject to a variety of potential biases, including consent bias, missed links bias, and incorrect links bias[52, 53]. Creation of linked datasets is currently running far ahead of our capacity to analyse the linked data while accounting for these biases in a statistically rigorous fashion[54]. Current methods for carrying out appropriate

[50] Correspondence with Andy Boyd (2017) summarising in progress work on record linkage in cohort and longitudinal studies conducted as part of the ESRC CLOSER cohort consortium and based on insights from the Wellcome Trust funded Project to Enhance ALSPAC through Record Linkage (PEARL).

[51] Wellcome. (2017). **Longitudinal Population Studies Strategy: Wellcome's Longitudinal Population Studies Working Group**.

[52] Harron, K., Gilbert, R., Cromwell, D. & van der Meulen, J. (2016). Linking Data for Mothers and Babies in De-Identified Electronic Health Data. *PLoS ONE*, 11(10): e0164667. **https://doi.org/10.1371/journal.pone.0164667**

[53] Sakshaug, J. W. & Antoni, M. (2017) Errors in Linking Survey and Administrative Data. In Biemer, P., Eckman, S., Edwards, B., de Leeuw, E., Kreuter, F., Lyberg, L., Tucker, C. & West, B. (eds.). *Total Survey Error in Practice*. John Wiley & Sons Ltd. DOI: **10.1002/9781119041702.ch25**

[54] See chapters 4 - 6 in Harron, K., Gilbert, R., Cromwell, D. & van der Meulen, J. (2016). Linking Data for Mothers and Babies in De-Identified Electronic Health Data. *PLoS ONE*, 11(10): e0164667. **https://doi.org/10.1371/journal.pone.0164667**

analyses of linked data depend on knowledge of how effective the linkage is, and there do not appear to be standards for end-users to access the relevant paradata (i.e. data about how the data are obtained) relating to the linkage process. Further contributing factors to this low level of statistical rigour are a lack of high level methodological training in the statistical issues that arise when advanced statistical and data-mining techniques are used to link datasets for analysis, and the lack of support for such analyses in the statistical software commonly used by social scientists. This is in sharp contrast to the sophisticated methodologies, and associated software implementation, that have been developed to account for sampling and attrition biases when analysing data collected from respondents to longitudinal surveys.

In summary, data linkage is now a well-established and mature research methodology, with a long history of international implementation, particularly in countries with centralised registers and unique personal identifiers (e.g. the Nordic countries). Despite these advances, challenges also remain related to performing data linkage and using linked data. Looking further afield, **Annex I** details some current initiatives involving data linkage in Brazil, Canada, Australia and New Zealand, while **Annex J** provides an introduction to the more widely used linkage methods.

3.1.1 ESRC-supported data linkage in the UK: Opportunities and challenges

From discussions at the ESRC workshop in January 2017, and subsequent discussions with data administrators, research funders, and researchers involved in data linkage in the UK, there appear to be three generic linkage scenarios of most interest to ESRC for enhancing existing longitudinal data. These are:

- **Linking an established longitudinal survey dataset to one or more administrative registers or other administrative datasets in order to enhance the research potential of the longitudinal dataset.** All the ESRC's longitudinal studies are implementing this approach. A further example of this is ALSPAC, which is linked to Department for Education registers (National Pupil Database, Annual School Census), National Health System registers (Hospital Episode Statistics, Clinical Practice Research Datalink), and Office for National Statistics registers (Cancer Registry, Death Registry)[55].

- **Linking two or more administrative databases in order to create a population-level longitudinal dataset.** An example is the **Brazilian 100 million cohort study**, which is constructed by linking the longitudinal Cadastro Único database, defined by all individuals receiving Bolsa Familia cash payments, to the databases making up the Brazilian Unified Health System.

- **The use of contextual linkage to enhance the research potential of a longitudinal dataset.** Contextual linkage is where group-level information is associated with individuals in the group in order to provide contextual information relevant to individual-level analysis. An example is the Netherlands Cohort Study on Diet and Cancer, where information on individual location has been linked to geospatial coordinates for recording exposure to particulate air pollution[56]. In the UK the postcodes of households participating in the Understanding Society study have been matched to the National Statistics Postcodes Directory, allowing the provision of a range of geographical identifiers and linked geographical data[57].

The ESRC 2006 Strategic Review of Panel and Cohort Studies made a number of recommendations associated with data linkage. These included "ESRC should ... secure greater access to relevant administrative data for longitudinal studies"[58]; "there should be support for data enhancement and potentially data collection ... through addition of relevant administrative data"[59]; and "the potential for adding or linking additional data to existing and new studies should be fully explored, including administrative, biomedical (including DNA), environmental and other data".[60] As part of its response to this review the ESRC commissioned the **Cohort and Longitudinal Studies Enhancement Resource** project (CLOSER, commenced 2012), which aims to maximise the use, value and impact of the data collected in the UK's longitudinal studies, and as part of this mission includes a work package to help cohort and longitudinal studies overcome barriers to data linkage, by facilitating their linkage to health, educational, and geographic data.[61]

---

[55] http://www.bristol.ac.uk/alspac/researchers/our-data/linkage/

[56] Hoek, G., et al. (2002). Association between mortality and indicators of traffic-related air pollution in the Netherlands: a cohort study. *Lancet*, 360(9341), 1203-1209. DOI: http://dx.doi.org/10.1016/S0140-6736(02)11280-3

[57] https://www.understandingsociety.ac.uk/about/data-linkage#part2

[58] Recommendation 4, page 6 in Martin, J. et al. (2006) **Strategic Review of Panel and Cohort Studies: Report to the Research Resources Board of the Economic and Social Research Council**. London: Longview.

[59] Recommendation 13, page 9 in ibid., 59

[60] Recommendation 16, page 10 in ibid., 59

[61] https://www.closer.ac.uk/about/areas-work/

Data linkage is now part of all core ESRC-supported longitudinal datasets. The extent of this activity is demonstrated by, but not limited to, the linkage outlined in **Annex D**. This details ongoing and planned data-linking activity for each of the CLOSER studies. Figure 5 below summarises the achieved, ongoing, and planned data linkage in the longitudinal studies covered by CLOSER. There is clearly a considerable amount of linkage activity, particularly in the health domain, and particularly as far as ALSPAC, MCS, and Understanding Society are concerned. There also appears to be a lot of potential linkage of employment, benefits and earnings data which is "in development", perhaps reflecting the sensitivity of the corresponding administrative data sources and difficulty in gaining department permission to access this data.

Figure 5: CLOSER studies data linkage activities

Domain	Dataset	Hertfordshire Cohort Study	1946 NSHD	1958 NCDS	1970 BCS	ALSPAC	Southampton Women's Study	Millennium Cohort	Understanding Society	Next Steps
Health	Civic Registry	Established	Established	Established	Established	Established	No Current Plans	Established	In Development	Established
	Maternity Records	No Coverage	No Coverage	No Coverage	In Development	Established	Planned for the Future	Established	In Development	No Current Plans
	Secondary and Community Care	Established	Established	In Development	In Development	Established	Planned for the Future	In Development	In Development	In Development
	Primary Care	No Current Plans	Planned for the Future	No Current Plans	No Current Plans	Established	Planned for the Future	Planned for the Future	No Current Plans	No Current Plans
	Social Care	No Current Plans	No Current Plans	No Current Plans	No Current Plans	In Development	No Current Plans	No Current Plans	No Current Plans	No Current Plans
Education	Early Years	No Coverage	No Coverage	No Coverage	No Coverage	No Coverage	No Coverage	No Coverage	No Current Plans	No Current Plans
	School	No Coverage	No Coverage	Established	No Coverage	In Development	Planned for the Future	Established	Established	Established
	FE/HE	No Coverage	No Coverage	No Coverage	No Coverage	In Development	No Current Plans	Planned for the Future	Established	Established
	Children in Care	No Coverage	No Coverage	No Coverage	No Coverage	In Development	No Current Plans	No Current Plans	No Current Plans	No Current Plans
Employment	Employment	No Current Plans	No Current Plans	In Development	In Development	Planned for the Future	No Current Plans	No Current Plans	In Development	No Current Plans
	Earnings	No Current Plans	No Current Plans	No Current Plans	In Development	Planned for the Future	No Current Plans	No Current Plans	In Development	No Current Plans
	Benefits	No Current Plans	No Current Plans	No Current Plans	In Development	Planned for the Future	No Current Plans	No Current Plans	In Development	No Current Plans
Criminality	Conviction/Cautions	No Current Plans	No Current Plans	No Current Plans	No Current Plans	In Development	No Current Plans	Planned for the Future	No Current Plans	In Development
Spatial	Geo-Spatial	Established	Established	Established	Established	Established	Established	Established	In Development	Established
	Environment Exposures	No Current Plans	Established	No Current Plans	No Current Plans	Established	No Current Plans	Established	No Current Plans	No Current Plans
	Neighbourhood, Services, Green Space	No Current Plans	Planned for the Future	No Current Plans	No Current Plans	Established	No Current Plans	Established	No Current Plans	No Current Plans
Consumer	Credit Histories	No Current Plans	No Current Plans	No Current Plans	No Current Plans	No Current Plans	No Current Plans	No Current Plans	In Development	No Current Plans
	Mobile Phones & Personal Sensors	No Current Plans	Planned for the Future	No Current Plans	No Current Plans	In Development	No Current Plans	No Current Plans	In Development	No Current Plans
	Social Media	No Current Plans	No Current Plans	No Current Plans	No Current Plans	In Development	No Current Plans	No Current Plans	In Development	No Current Plans
Other	Participant Tracing	No Current Plans	Established	Established	Established	Established	No Current Plans	Established	Planned for the Future	No Current Plans
	Home Energy	No Current Plans	No Current Plans	No Current Plans	No Current Plans	No Current Plans	No Current Plans	No Current Plans	No Current Plans	No Current Plans
	Vehicle Registration	No Current Plans	No Current Plans	No Current Plans	No Current Plans	No Current Plans	No Current Plans	No Current Plans	No Current Plans	No Current Plans
	Voter Registration	No Current Plans	No Current Plans	No Current Plans	No Current Plans	No Current Plans	No Current Plans	No Current Plans	No Current Plans	No Current Plans

Linkage Key: Established · In Development · Planned for the Future · No Current Plans · No Coverage

This table provides a general overview of most of the linkage activity among studies. It aggregates smaller distinct datasets focused on the health domain, and generalises linkage status across the four home nations of the UK and Northern Ireland (Produced by Andy Boyd, Autumn 2017).

Following the publication of the 2012 Administrative Data Taskforce (ADT) report Improving Access for Research and Policy in October 2013, the ESRC commissioned the ADRN consisting of four Administrative Data Research Centres (ADRCs), one in each country of the UK, and an Administrative Data Service (ADS). The aim of ADRN was to give accredited researchers safe access to de-identified data from linked administrative databases in a secure environment.

In November 2016, the ESRC published the ADRN mid-term review report[62], which found that although the ADRN had established a safe infrastructure for linking administrative data, the network had a major issue in that it did not have the UK-wide data that was required for this infrastructure to have its intended impact. The evaluation found that after two

---

[62] Research Project Management. (2016). **Administrative Data Research Network (ADRN) Mid-Term Review Report: November 8, 2016**. Toronto: Research Project Management.

years of the ADRN operation, progress on accessing government administrative data had not been good, though it was better for Devolved Administrations (public bodies in Wales, Scotland and Northern Ireland) than large Whitehall departments. For example, the Administrative Data Research Centre in Wales (ADRC-W) is one of four UK centres that were set up as part of the ADRN. This centre has the capacity to link administrative data from a number of Welsh sources, including the **Welsh National Pupil Database**, which contains linked individual pupil records for all children in Wales in the state school system, and the **Secure Anonymised Information Linkage** (SAIL) Databank, which contains administrative datasets relating to general practitioner, vital, and hospital admission records for Wales. In Scotland, the Scottish Longitudinal Study is a large database created by linking Scottish administrative data, including census data, vital events, NHS data, and education data. It is part of the ADRC in Scotland, and researchers access its linked data via the ESRC-funded **Scottish Longitudinal Study Development and Support Unit**, which is part of the Census Longitudinal Studies. Similarly, the ADRC in Northern Ireland is actively involved in creating linked administrative datasets for research, focusing on census, health, and education data for Northern Ireland.

All three of the ADRCs in the devolved UK countries benefit from historically good and trusting partnerships with their respective devolved administrations and statistical agencies, and operate with the full cooperation of the governments of their respective countries in terms of access to country-specific administrative data. This has undoubtedly been one of the most important reasons why they currently provide an effective data linkage resource for social science research. Unfortunately, the size and complexity of UK-wide administrative databases, as well as the legal barriers to data-sharing with non-government entities (i.e. universities and other research institutions), has meant that progress on data linkage requests related to England has been slow. Major factors that are perceived as being responsible for this are:

• lack of transparency about the structures of the relevant administrative databases and their quality

• inconsistent access conditions

• time delays and and lack of transparency around decision making regarding access

• inadequate resources within government departments to meet these requests

• complex and unclear legal framework for access

There is also a perception of an administrative culture where linkage requests that are tied to policy questions of interest are being viewed more favourably than "curiosity-driven" linkage requests.

3.1.2 Privacy and security issues for linked longitudinal data

Although a growing number of ESRC-supported longitudinal data resources have been successfully linked to administrative databases, it is clear that this linkage has rarely been straightforward or cost-free, even when unique personal identifiers are available for all individuals in the datasets to be linked. There have been and still are many barriers to successful linkages. In particular, there are issues related to privacy and consent that make it difficult to link survey with administrative data, as well as issues around the trust that various agencies have in the security and privacy of the linked data. Linkage of survey data to an administrative database almost always requires that some form of consent to the linkage operation needs to be obtained from the survey respondent. This will depend on the degree of shared trust – by the individuals concerned, as well as by the organisations responsible for their data – that the linkage will not lead to their identification.[63]

In a statement to this review on current developments in obtaining consent for linkage, Andy Boyd and Alison Park of the CLOSER project note that the legal necessity for consent is possibly misunderstood, as consent is typically not the appropriate legal basis for studies to meet Data Protection legislation requirements. In January 2018, the Information Commissioner's Office (ICO) published a **Guide to the General Data Protection Regulation** stating that the biggest change would be for public authorities, which include UK universities, who now need to consider the new 'public task' basis first for most of their legal processing of personal data, and have more limited scope to rely on consent or legitimate interests.[64] It is therefore important that everyone, including data controllers and researchers, understands

---

[63] See findings of ESRC's public public dialogues: **http://www.esrc.ac.uk/public-engagement/public-dialogues/public-dialogues-on-using-administrative-data/**

[64] **https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/#ib1**
(accessed 15th January 2018)

why and when consent should be used. They also observe that "Changing expectations regarding the 'validity' of consent is proving problematic for longitudinal studies; with the consent wording being perceived as 'outdated' or the intervals between 'refreshing' consent being considered overly long." The full text of their comments is set out in **Annex K**.

There is also growing concern that linking data relating to individuals provides information that can be used for identification of these individuals, and consequent potential breaches of confidentiality. This is in addition to the existing identification risks associated with release of data from the individual databases being linked. In response to these concerns, the Organisation for Economic Co-operation and Development (OECD) in 2016[65] set out recommended ethical principles, as well as corresponding governance procedures, that should be followed when linked data containing potentially sensitive information are released for research purposes. In essence, these procedures require informed consent that is "future-proof", allowing the linked data to be used in ways that may have been unanticipated at the time of linkage. However, such consent also requires a believable set of data security provisions to be in place to ensure that confidentiality continues to be preserved irrespective of how the linked data are used in the future. This sets an extremely high bar that needs to be cleared if the de-identified linked data are to be made publicly available, and inevitably leads to restricted access of one form or another.

It should also be noted that there is considerable ongoing research that tackles the confidentiality vs. analytic convenience issue from the other end, seeking to develop privacy-protecting data processing procedures that result in anonymised linked data.[66] These procedures may assist with confidentiality issues, though they typically also make the data more difficult to use because of more complex data analysis requirements.

The privacy and consent issues discussed in the preceding paragraphs indicate that there are concerns that linkage can increase the risk of respondent identification. However, the benefits of linkage, as already noted, are large. In addition, it is important to ensure that the public at large understands the benefits of research with linked longitudinal data, and a more proportionate stance to risks can be adopted by data controllers and the public.[67] The ESRC provides co-funding to **Understanding Patient Data**, a Wellcome-led initiative to support discussions with the public, patients and health care professionals about health and social care data.

3.1.3 Integrating data linkage and longitudinal survey design

Many of the discussions carried out with stakeholders during meetings with panel members touched on the potential for using data linkage in more innovative ways, and not primarily as a method for enhancing the scope of information collected in the different longitudinal surveys. In this context a recurring theme was that of the ESRC moving towards an integrated longitudinal data resource based on maximising its use of data linkage. This can be achieved by standardising the designs of its various longitudinal surveys so that they all use the same longitudinal population register (i.e. a population spine) as their sampling frame, and with all ESRC research-related linkage of different administrative and survey data sources harmonised to this spine.

Data linkage has the potential to drive the design of longitudinal surveys, by using a suitable population register as a "spine" and linking all (or as many as possible) ESRC longitudinal data collections to it. This will enhance the use of these studies and will also enable future ESRC longitudinal studies with sample designs based on this spine to be representative of their target UK populations. It also has the potential to ensure consistency between ESRC longitudinal surveys and future official statistics collections, including the ONS plans for an administrative data-based census beyond 2021.

Obviously, this approach will have sampling implications for Understanding Society, which presently uses area-based sampling for household selection. However, the conduct of Understanding Society should remain essentially unchanged once the current sample is linked to the spine, and many of the sampling problems arising from the need to maintain the UK representativeness of the Understanding Society sample will be easier to deal with given the longitudinal population data that will be available on the spine.

---

[65] OECD. (2016). Research Ethics and New Forms of Data for Social and Economic Research. OECD Science, Technology and Industry Policy Papers, No. 34. OECD Publishing: Paris. **http://dx.doi.org/10.1787/5jln7vnpxs32-en**

[66] Dwork, C. & Yekhanin, S. (2003). Database Privacy. Retrieved from: **https://www.microsoft.com/en-us/research/project/database-privacy/?from=http%3A%2F%2Fresearch.microsoft.com%2Fresearch%2Fsv%2Fdatabaseprivacy%2F**

[67] See findings of ESRC's public dialogues: **http://www.esrc.ac.uk/public-engagement/public-dialogues/public-dialogues-on-using-administrative-data/**

Linking the current ESRC-funded birth cohorts to the spine will likely generate the largest benefit, since it will mean that information will be available on the spine regarding those areas where the cohorts lack sample size or representation as far as the UK population of the relevant age is concerned. There will also be advantages for the more regionally focused studies (e.g., ALSPAC), since linkage to the spine will open up access to more sources of administrative data.

The most important innovative use of an administrative data spine will be as the primary sampling frame for any future cohort and longitudinal studies. The advantage of knowing the characteristics of the population that are recruited into the study and those that are not has important analytic advantages, including the capacity to assess the extent of cross-sectional non-response bias and to define sample weights for cross-sectional inference. The information held on the spine should also allow the development of more powerful ways of dealing with panel attrition and help reduce the burden on respondents by reducing the amount of information that a survey needs to collect from (and thus time with) the respondent. For areas such as education and health outcomes, the information from administrative data may be of higher quality than that obtained from surveys that rely on self-reporting on various outcomes.

Furthermore, the spine will make it relatively straightforward to carry out cohort comparison and sample boosts to improve representation of migrants and other new entrants into the population, as well as provide a vehicle for selection of new "accelerated" cohorts from age groups not covered by the current cohorts in order to address emerging research issues. Obviously, the ability to link other administrative data to the spine will also mean that panel and cohort studies will then have access to the same external data, ensuring consistency in analyses carried out using these data.

The main challenge will be to create an administrative data-based population spine with comprehensive UK coverage which is acceptable to data controllers and the UK public. Recent changes to the way the UK government plans to share its administrative data for research purposes (the Digital Economy Act 2017) coupled with the establishment of NHS Digital mean that it may be possible to negotiate the creation and upkeep of such a research resource. Since the spine will be useful for sample design and selection in other research areas (e.g., those funded or co-funded by the MRC), one would hope that its creation would also receive the strong support of UKRI.

3.1.4 Summary

Data linkage for longitudinal surveys is a very active and rapidly growing area with many advantages to social science researchers. Linking these surveys to administrative data has considerable potential for adding significant value to the survey data. However, such linkages are still a challenge in the UK. This is mainly because the owners of the administrative data sources have tended to view linkage to research-oriented databases as potentially risky. These attitudes are changing, but there are still barriers to data linkages, and it is unclear when these issues will be resolved.

Discussions held during the January 2017 ESRC workshop considered the possibility of basing longitudinal research on linked population registers as a way of substantially reducing the costs of UK investment in longitudinal surveys. However, these quickly concluded that in the UK context there is no real possibility that data linkage can be used to effectively replace longitudinal surveys with linked data obtained from longitudinal administrative databases. This is because UK administrative databases do not contain the rich social data that may be available from population registers in other (e.g., the Nordic) countries. Consequently, linking UK administrative data sources on their own will not provide the breadth and depth of data currently collected in ESRC longitudinal surveys (see Section 2). These longitudinal surveys, however, can be enhanced by linked administrative data, providing a rich set of data on the condition of the UK population.

Data linkage has considerable attractions for research funding agencies, since it offers what should be an economical way of extending the types of research that can be carried out using the data from their funded longitudinal studies. This particular longitudinal survey model is currently in use in New Zealand, and regional versions of linkage to data spines are operating with some success in Wales and Scotland. However, we were not able to see any overarching strategy for what is linked and how it is linked as far as the current set of ESRC longitudinal data collections is concerned. This will be a challenge to address in the future, with the opportunity to harmonise this data linkage across different longitudinal studies.

As noted earlier, there are risks associated with adopting this strategy for future ESRC longitudinal data, primarily from public acceptance of researcher access to the longitudinal data on the spine and also from ensuring guaranteed access to it, since it will by default be based on administrative data sources. However, the gains from adopting this strategy, particularly in terms of integrating and harmonising the data collected in the various age cohorts as well as by Understanding Society, and in the increased representativeness and flexibility of these different studies, will be considerable.

3.2 Data harmonisation

Data harmonisation is another method that can be used to enhance existing data. The purpose of data harmonisation is to achieve or improve the comparability of different surveys and measures collected in the past (retrospective harmonisation) or to be collected in the future (prospective harmonisation). Importantly, data harmonisation can be applied to any step of the survey or data collection process, including study design, choice of variables, question wording, questionnaire design, sampling, data collection, coding, editing or documentation.

Retrospective harmonisation is a method that can be used to facilitate comparative research of existing survey data or for instances where intensive early planning was not possible because of financial or policy constraints. The items to be harmonised sometimes are not comparable, but are assessed and edited to achieve commonality. For example, the data collected from a sampled individual in one sweep of a cohort will (hopefully) be internally consistent, but the joint data collected from the same individual in separate sweeps may not be internally consistent because of changes in item definitions or collection modes (face to face, telephone, internet, other responsible person, etc.) between the two sweeps. Data harmonisation involves "massaging" these joint data to remove, as much as possible, the impact of these changes without losing too much value or meaning.

In contrast, prospective harmonisation is a method that can be used during the initiation of new studies or new sweeps of existing studies. Also known as input harmonisation, prospective harmonisation is most often applied to surveys administered in different countries and used to enforce strict standards and protocols from the beginning of the survey process. The standards and protocols are applied similarly across survey procedures, and common questionnaire items are asked across studies. Although some adaptations may be allowed for individual data collection sites, this approach aims to ensure a high degree of comparability.

Data harmonisation has several advantages. First, it is a flexible method that can be applied to a wide range of study designs (e.g., linkage studies[68], surveys, and biomedical studies), including within and across longitudinal studies. For example, to understand social phenomena (e.g., social inequality, population dynamics, and public opinion) over time and in different populations, studies need to harmonise measures or data across different sweeps within the same study and/or across different studies.

Second, data harmonisation can be implemented with national and international studies. Harmonisation of cohort studies can facilitate cross-cohort comparisons at country level in the UK (e.g., England, North Ireland, Scotland and Wales), while harmonisation of common variables in panel surveys carried out in different countries facilitates international comparisons (e.g., **Gateway to Global Ageing Data**; **Gender and Generations survey** (GGS)). Indeed, harmonisation is a prerequisite for international comparative research, which is based on microdata, and may increase the understanding for social phenomena across different cultures.

Third, harmonised metadata, which is documentation of an agreed set of data characteristics for key analysis variables associated with a specific study, is an important output of data harmonisation. Not only does harmonised metadata have the ability to provide conceptual and methodological clarification within and across studies, it can also provide additional information on the quality of shared data by highlighting information on variable definitions, data sources, methods of estimation, and other information that allow users to gain a better understanding of the variables of interest.

3.2.1 Data harmonisation in UK
Given the potential advantages of data harmonisation for increasing the value and (re)use of longitudinal studies and improving data quality, the ESRC has funded a number of harmonisation activities with their longitudinal investments, some described below.

---

[68] Linkage studies are used to map genetic loci by use of observations of related individuals. See Dawn Teare, M. & Barrett, JH. (2005). Genetic linkage studies. *Lancet,* 366(9490):1036-44. DOI:**10.1016/S0140-6736(05)67382-5**

3.2.1.1 Retrospective harmonisation

To date, the majority of ESRC-funded longitudinal data harmonisation activities have been retrospective. For example, CLS has harmonised data across four cohort studies (BCS70, NCDS, Next Steps, and MCS) to answer questions related to weight, cognitive ability, and socioeconomic inequality. Three published papers[69, 70, 71] provide detailed information on what has been done to harmonise this data in the longitudinal cohort studies.

With BHPS and its successor Understanding Society, the responsible research team has tried where possible to retrospectively harmonise measures of interest with UK government surveys and other international household panel studies. Thus far, most of the retrospective harmonisation work with the panel studies has been with data from other household panel studies available in the **Cross National Equivalent File** (CNEF) platform. In recent years, for example, some topics such as social inequality, social wellbeing, educational return and ageing studies have been harmonised across different societies (US, UK, China, Germany and Netherlands)[72] A result of this work is a harmonised dataset containing annual data since 1991 from the BHPS, Understanding Society and CNEF panel studies.

It is important to note that CLOSER has facilitated and supported much of the retrospective harmonisation work that has occurred or is currently in progress. In conjunction with the cohort studies, currently CLOSER is working on ten harmonisation projects focused on obesity, socioeconomic status, biological samples, income, wellbeing, housing and biomarkers (see **Annex E**). The decision of which areas to focus on were based on researchers' experience and expertise in particular areas, rather than on a push for harmonisation across the board in all the studies. CLOSER's harmonisation activities primarily with cross-cohort studies have produced useful scientific outputs for researchers that demonstrate the value of harmonisation activity (harmonised metadata) including datasets, guidance documentation, workshops and training courses.

Despite the progress made in retrospective data harmonisation, at present, retrospective harmonisation is considered very challenging due to issues of how variables have changed over time and varied between cohort/panel studies. There continue to be challenges such as reviewing and updating the questionnaire content regularly; involving users earlier in the process; and systematically investigating how other studies collect data on new topics, to either harmonise with them or ask study-specific questions. In some instances it simply is not possible to harmonise variables, either because the questions asked are too different or because they were not asked at comparable life stages. Even when harmonisation has taken place, it can be difficult to establish clearly whether a difference between studies reflects a 'real' difference or measurement error. Sometimes study differences can be caused not just by the wording of the question and/or the response categories, but also the context of the survey and question mode.

In conversations with researchers engaging in retrospective data harmonisation, it was noted that some (often older) datasets lack the level of documentation required to facilitate harmonisation. As an example, marital status was measured differently not only across different studies in the UK (e.g., NCDS, Understanding Society, and ELSA), but also for the same study (NCDS) at different time points. To further complicate matters, in the UK there is a wider range of partner arrangements other than marriage at present compared to in the past. Thus, retrospective data harmonisation requires that researchers take into account how the UK has changed over the lifetime of the studies.

Similarly, there are many challenges in harmonising variables such as family income, education attainment, etc. across international studies which vary widely, and whose 'purchasing power' can differ based on each country's context. Thus, although harmonisation has advantages for comparative research, many social phenomena are culturally specific, and using a standard set of definitions may lead to loss of cultural specificity. Moreover, keeping the definition of variables constant across time (as in a panel study) may mean that they fail to reflect changes in important phenomena such as classification of occupations, family/household income, education and marital status, and fail to capture phenomena

[69] Bann, D., Johnson, W., Li, L., Kuh, D. and Hardy, R. (2017). Socioeconomic Inequalities in Body Mass Index across Adulthood: Coordinated Analyses of Individual Participant Data from Three British Birth Cohort Studies Initiated in 1946, 1958 and 1970. *PLoS Med* , 14(1): e1002214. **https://doi.org/10.1371/journal.pmed.1002214**

[70] Goisis, A., Özcan, B. & Myrskyla, M. (2017). Decline in the negative association between low birth weight and cognitive ability. *Proc Natl Acad Sci U S A* , 114(1): 84-88. DOI: **10.1073/pnas.1605544114**

[71] Ploubidis, GB., Sullivan, A., Brown, M. & Goodman, A. (2017). Psychological distress in mid-life: evidence from the 1958 and 1970 British birth cohorts. *Psychol Med* , 47(2):291-303. DOI: **10.1017/S0033291716002464**

[72] See joint research project: **http://csr.pku.edu.cn/EN/AcademicResearch/station/data/**

that are novel to the social context (e.g., growth of social media and the digital economy). Importantly, since harmonisation depends both on what was asked and how it was asked, it can only provide an approximate indicator of these variables when they differ. Thus, the best way to harmonise international data is to use survey designs and questionnaires that integrate well with each other and are comparable, factors that may be managed more easily through prospective harmonisation.

3.2.1.2 Prospective harmonisation

In contrast to retrospective harmonisation, in general there has been limited prospective harmonisation with longitudinal data to date (for an exception see ELSA's work in cross-national comparisons), and particularly with ESRC's longitudinal investments. An ESRC-funded project of prospective harmonisation as part of **A Comparative Study of Life Course and Family Dynamics**, a joint research project that has harmonised more than 100 variables for NKPS in Netherlands and CFPS in China and combined them with already harmonised data for Understanding Society, PSID in the US, and GSOEP in Germany across nine topics such as education attainment, family income, marriage status, etc. Due to the focused content area of the research project, however, the harmonised variables are limited.

CLS routinely harmonises questions across all of the CLS birth cohort studies, and is also making progress in prospective data harmonisation in Next Steps. For instance, with the recent addition of Next Steps to the CLS portfolio, CLS is strategically harmonising questions asked of Next Steps study members during their mid-20s with comparable questions asked of study members from BCS70 and NCDS.

Thus, in the future, Understanding Society may want to explore expanding prospective harmonisation with studies from other countries that use similar questionnaires. In addition, ESRC may want to promote additional prospective harmonisation activities across a wider range of content areas and longitudinal investments (i.e. panel and cohort studies) using ELSA's successful model of prospective data harmonisation with cross-national and international studies as a model.

3.2.2 Summary

Data harmonisation is a necessary process for understanding social phenomena and is an important tool in enhancing existing cross-sectional and longitudinal data. It is also the case that harmonisation could save the cost of directly collecting data in a survey and enrich our understanding of a phenomenon from many angles. It is not surprising that data harmonisation was identified as one of the "top ten" methodological and technological issues in the 2016 online consultation exercise, hosted on the ESRC website, on the future need for longitudinal studies in scientific and policy-relevant research.

The initiatives of data harmonisation currently underway in the UK are a good start towards a comprehensive programme aimed at enhancing ESRC longitudinal surveys via data linkage. Retrospective (output) harmonisation supports comparability and long-term trends in variables both within the UK and internationally. Prospective (input) harmonisation activities such as integration of survey design, questions and data collection have the potential to make longitudinal studies better and more cost-effective, with fewer burdens on participants. These efforts also make the data more discoverable and accessible, which in turn enhance the use of longitudinal studies.

In order to increase the value and use of rich longitudinal studies in the UK, some research groups have already undertaken harmonisation for international comparisons or understanding social changes over time in the UK. CLOSER and CLS plan to do retrospective harmonisation for UK's birth cohort studies, extending the value of longitudinal studies.

Even with all the data harmonisation activities currently underway, there are still many challenges to address. Commonly faced issues include:

- difficulty in harmonising variables either due to the questions being asked in different ways or asked at different ages

- restrictions in using personal information in surveys

- lack of understanding and agreement across disciplines, which may make harmonising representative longitudinal studies problematic

Section 4: Access, discoverability and training

4.1 Background

The 2006 Strategic Review of Panel and Cohort Studies endorsed increasing access to longitudinal data. It called for increased capacity development in the UK to design, implement, and analyse longitudinal studies.[73] At this time the ESRC had recently established the National Centre for Research Methods (NCRM) at the University of Southampton (2004) to increase the training and capacity-building as well as produce methodological innovation in the use of complex longitudinal data. The review cited the need to augment the work performed by this centre with a coordinated national programme working at several levels to address the issues related to increasing access to, discoverability of, and training in ESRC-funded longitudinal data.[74]

Since the 2006 strategic review, the ESRC has taken numerous strategic steps to address and improve capacity regarding these areas, including the formation of CLOSER with a core remit including both training and discoverability, as well as maintaining support for the UKDS. Even with this increased investment by the ESRC, access, discoverability and training were still identified as issues for concern by respondents in the 2016 consultation and 2017 workshop. For example, subthemes identified in the report of the 2016 consultation included access to longitudinal studies data, training data analysts and users, and need for support for data users in conducting complex analyses. To explore these subthemes further, the January 2017 workshop at Oxford included one breakout session on access and discoverability among other issues and one on training and capacity-building.[75, 76]

Since the 2006 strategic review, the ESRC has taken the following steps to further facilitate access to and discoverability of longitudinal studies:

- The UK Data Service (UKDS) at the University of Essex became the depository for all ESRC-funded longitudinal studies in 2007 and a centralised access point to many of the longitudinal studies within the UK.[77]

- The Cohort and Longitudinal Studies Enhancement Resource (CLOSER) was established in 2012 and funded jointly by the ESRC and the MRC to 2017, with an extension to 2019. Its remit is broad and designed to enhance the discoverability of data across eight national and regional longitudinal studies, to promote data harmonisation and data linkage, and to facilitate and expand training efforts.

- The **METADAC** (Managing Ethico-social, Technical, and Administrative issues in Data Access) was formed in 2015 and complements the work of the UKDS, to provide a multi-agency, multi-study data access structure of complex data and biosamples that serves several of the UK's major longitudinal studies and provide a scalable mechanism to incorporate additional studies in the future.

- The Expert Advisory Group on Data Access (EAGDA) was established in 2012 and funded through 2017 by ESRC, Wellcome, the MRC, and Cancer Research UK to provide strategic advice on data access for cohort studies and genetic research.[78] Besides its work in ethical and legal issues, EAGDA is supportive of data access and good practice in governance.

---

[73] Recommendations 25-28, pages 11-12 in Martin, J. et al., (2006) **Strategic Review of Panel and Cohort Studies: Report to the Research Resources Board of the Economic and Social Research Council**. London: Longview.

[74] Sections 8.7 & 8.9 in ibid., 74

[75] Townsley, R. (2017a). **Further analysis of responses to the consultation. Paper 3: The capacity and infrastructure for longitudinal research**. Ruth Townsley Research.

[76] Townsley, R. (2017b). **Further analysis of responses to the consultation. Paper 4: Use and accessibility of longitudinal resources**. Ruth Townsley Research.

[77] Van den Eynden, V., Vlad, A. & UK Data Service. (2017). **Towards Open Research: what ESRC-funded researchers tell us**.

[78] Expert Advisory Group on Data Access. (2017). Additional Information from EAGDA produced for the Longitudinal Studies Review Panel.

- The Administrative Data Research Network, not specific to longitudinal studies, was founded by the ESRC in 2011 to improve the access to government administrative data to enhance research and policy.[79]

- Recently, the **Low and Middle Income Longitudinal Population Study Directory** (LMIC LPS Directory) was developed by the **Institute for Fiscal Studies** with funding from the ESRC, the MRC and Wellcome to provide data on studies that may be of interest to researchers examining issues of health and economic change.

4.2 Accessing and use of longitudinal data

Despite the evidence for the substantial utilisation of these resources, users and potential users continue to note barriers to, access to, and use of longitudinal data. It is unclear why this continues to be an issue given the investments, innovations and new technology that have been used to make the data more accessible. To understand these issues we reviewed available information for any evidence of barriers to access or use. We also reviewed statistics on downloads of data curated at UKDS to evaluate if there was clear evidence of gaps in user types, disciplines or country of origin. These served as a rough proxy measure of access to ESRC-funded longitudinal data. It should be noted, however, that the download of data does not necessarily mean productive use (CLS estimates a ratio of about 50% of downloads or projects result in publications[80]), and that a single download may be used by multiple researchers. Thus, another measure of productive use, publications, was also examined in order to review access. This includes peer-reviewed academic publications, government and/or charitable reports.

Statistics on data downloads of the ESRC longitudinal studies over the decade between 2007 and 2016 suggest that there have been over 90,000 downloads of longitudinal study data by users in that timeframe. Importantly, there has been a sustained year-on-year increase in the number of downloads of the three national birth cohorts (NCDS, BCS70 and MCS) and the panel studies (Understanding Society, BHPS) over this same timeframe (see **Section 2**, **Annex L** for trends in downloads by study).

4.2.1 Access by type of user and by discipline
Though there is variability in the number of data downloads of individual studies, there is remarkable consistency in characteristics of those downloading data across longitudinal studies. UKDS provides 17 codes for type of user, but three types account for 91% of all downloads: postgraduate students, staff at an institute of higher education, and undergraduate students (**Table 2c**, **Annex L**).

The coding for discipline by UKDS is sometimes difficult to interpret so it is not possible to get estimates of use of the data by discipline at a high level; discipline of user is single-coded and uses a pre-determined list of disciplines, which may provide a simplified picture. Bearing in mind these limitations, the statistics indicate use by a wide range of disciplines. It is evident that the primary disciplines downloading both CLS and Understanding Society/BHPS data are Economics and Econometrics (38% and 34%) followed by Sociology (13% and 21%) (**Table 2d**, **Annex L**). Given the variability in coding for some disciplines, it is not possible, for instance, to get a robust estimate of health-related use, where there are separate codes for "Other studies and professions allied to medicine", "nursing", "community", and separately "hospital-based" clinical subjects, "biology", and "pre-clinical studies". As a result, we have not included tables on download use by discipline in **Annex L**. Improving and consolidating the coding options would facilitate future investigation into gaps and potential barriers to specific disciplines.

4.2.2 International access
Substantial consistency is also found in patterns of use of ESRC-funded longitudinal studies by international researchers. As expected UK researchers represent the vast majority of use, but international users represent 20% of CLS longitudinal study birth cohort downloads and 12% of Understanding Society data downloads (**Table 2e**, **Annex L**). The majority of international downloads go to researchers in a very small number of countries. These include the US, Germany and Italy, with other countries accounting for 1% or less of downloads. There are important barriers related to the international use of these datasets. Challenges in access to sensitive data due to the data redistribution license agreements with data providers, for example, make it difficult for those outside the UK to access data.

---

[79] Administrative Data Taskforce. (2012). **The UK Administrative Data Research Network: Improving Access for Research and Policy**.

[80] Personal communication (2017)

4.2.3 Access to sensitive data

Currently, access to sensitive data, such as biological samples and measures, from NCDS and Understanding Society (and also ELSA) is managed by METADAC, and new data and samples from BCS70 and MCS will be added in due course. METADAC reviews research proposals and approves, defers or denies them. Between 2007 and 2016, 88% of the 210 completed applications were approved (**Table 3a**, **Annex L**), with 72% of these from the UK (**Table 3d**, **Annex L**).

METADAC receives a slightly larger proportion of international applications (25.5%) compared to downloads of CLS and Understanding Society data through UKDS, which achieve around 12-22% international use (**Tables 3d** & **2e**, **Annex L**). The disciplines of applicants to METADAC is also, unsurprisingly, different to UKDS users, with epidemiology, chemistry, biosocial studies and genetics all with a greater than 10% share of the total METADAC applicants (**Table 3e**, **Annex L**). This is compared to economics and econometrics, sociology and psychology which were the UKDS user disciplines with an equivalent share of the total UKDS downloads (**Table 2d**, **Annex L**).

A large majority (88%) of the applications to METADAC between 2007 and 2016 were for NCDS data (**Table 3b**, **Annex L**), which includes DNA, biomedical samples and genome-wide association and exome sequencing data (**Table 3c**, **Annex L**). A recent paper has highlighted the particularly high use of NCDS genetic data, finding that NCDS is in the top four of all datasets ever used in GWAS discoveries[81] and used in double the number of papers than the Wellcome Trust Case Control Consortium, for instance, which was specifically designed for GWAS research. The other top datasets used (e.g. Framingham Heart Study) were generally longitudinal cohort data that asked a wide variety of questions and were not specifically hypothesis driven, suggesting that a broader inclusion of questions incites wider usage by multiple disciplines over time.[81a]

4.2.4 Publications

Another metric of access and use of ESRC-funded longitudinal studies is publications. Both Understanding Society and the individual birth cohorts keep records, as far as they are able, of all publications arising from the longitudinal study data they hold by year and by country of the first author (see **Section 1**, **Annex L**).[81b] These include peer-reviewed research papers, government reports and other academic papers. Publications do not always reliably cite the data they use and data users do not always report their publications to UKDS or the studies.

Similar to the UKDS data download statistics, publication trends show a remarkable year on year increase over the last decade (2007-2016) of publications arising from the use of Understanding Society, BHPS and the three birth cohorts (**Table 1a**, **Annex L**). Note that Understanding Society publications data were not collected until 2008, while the others are from the full decade. In addition, Understanding Society, including the Innovation Panel which began in 2010, has recorded 594 publications over nine years and BHPS recorded 1,549 over the whole decade. The three birth cohorts held by CLS report 2,752 publications during the same time period. It was noted that there were also publications that did not clearly indicate the use of the longitudinal studies and so the numbers provided in **Annex L** are likely lower than the true number of publications using the studies. This may be due to the easy availability of the data, which is a strength of the ESRC longitudinal studies but makes it difficult to track all the potential uses of the data.

4.3 Discoverability of longitudinal study data

The review panel did not find evidence of a significant problem with discoverability of the ESRC-funded longitudinal studies. Both CLOSER and UKDS have separate discovery services (each offering different and complementary functionality) designed to enable interested consumers to find and access detailed information on the structure, measures and variables included in each of the longitudinal studies curated at UKDS and/or included within CLOSER. When searching the websites of CLS, Understanding Society and other ESRC-funded studies for access to data, users are immediately directed to the **UKDS discovery site** where they can find more information and instructions on how to gain access.

CLOSER Discovery (launched in 2017) is a major achievement of CLOSER. CLOSER Discovery has already served 2083 unique users (1 April 2016-15 December 2017) and considering an earlier six months period (May-November) in 2016 there were 984 sessions compared to 2,135 in the same months in 2017.[82] It is a user-friendly search engine which allows users to browse the content of eight UK longitudinal studies, providing metadata about what, when, and how data have been collected over 70 years of data collection, at variable level, allowing both contemporaneous cross-study and intra-study comparison over time.

[81] Mills, M. C. & Rahal, C. (under review). The Anatomy of GWAS.

[81a] The NCDS resource became available earlier than those for other studies, see METADAC.

[81b] Due to differences in how CLS and Understanding Society handle combined studies publications, the figures may not be directly comparable.

[82] Personal communication (2017)

UKDS Discover is the UK Data Service's search and browse application providing access to data and related resources. The UK Data Service provides access to over 6,000 digital data collections, including the ESRC-funded longitudinal studies, for research and teaching purposes. UKDS Discover allows bona fide researchers to access data collections directly, subject to the deposit and licensing conditions of the data.

With the advent of the ESRC's **Secondary Data Analysis Initiative** (SDAI), which is a funding scheme specifically designed to encourage greater secondary use of ESRC-funded data resources including the longitudinal studies, it is expected that we will see both a wider and larger use of the longitudinal studies. It will be important to document and monitor the use of the longitudinal studies related to this funding mechanism in order to understand the impact and use of the birth cohorts and panel studies.

4.4 Training to use longitudinal data

Training includes a broad range of initiatives, many addressing much broader areas than that of longitudinal studies but remaining relevant to them. These include formal academic training in quantitative methods in the undergraduate, postgraduate and early career stages as well as training workshops, short courses open to all users, and other bespoke professional training that may be required within government departments to make use of longitudinal study data. In addition, there is training in data science, research administration, epidemiology, statistics and the social sciences.

There has been general agreement that more training in longitudinal methods is needed[83, 84, 85, 86]. Since the 2006 Strategic Review of Panel and Cohort Studies, the ESRC has launched a range of training initiatives to develop capacity nationally in quantitative analysis, and several targeted specifically at longitudinal research. It is possible that there is a need for better discoverability of the range and depth of available training. Both formal degree training and short course and workshop training are offered, and the new CLOSER learning hub and the NCRM offer training on longitudinal analysis and provide information on the workshops and short courses available over the next period of time.

The initiatives now in place for training in quantitative methods include the Q-Step programme, funded by the Nuffield Foundation, ESRC and the Higher Education Founding Council for England; and the ESRC Doctoral Training Network. The network includes studentships targeted specifically at quantitative methods, the 14 ESRC-funded **Doctoral Training Partnerships** (DTPs), and the two new interdisciplinary centres for doctoral training, one in the area of Biosocial research (co-funded by BBSRC), the other focused around new forms of data (e.g., big data, administrative linkages). Even with these initiatives, information from the consultation, workshop and individuals involved with training continued to note a dearth in the amount of researchers with advanced training in statistical methods, so it is possible that there is inadequate publicity of the most recent initiatives. Such training was noted as fundamental to a strong data science workforce that focuses on the use of longitudinal data.

4.4.1 Training materials and courses specific to longitudinal research

New opportunities for training specific to longitudinal research have been developed in the last decade. CLS, Understanding Society, the NCRM and CLOSER are all funded to provide training in a range of aspects of longitudinal studies design, collection, management and analysis.

The new **CLOSER Learning Hub** aims to provide a one-stop shop for longitudinal training information. Currently this focuses on upcoming events, workshops and short-term training, but could be expanded to include signposting for students interested in obtaining more substantive training (directing them to the DTPs for instance).

- **CLOSER, in 2017, launched a bespoke simplified dataset** based on the 1958 cohort and a new online **Learning Hub**. This is intended to be used by universities in training undergraduate, masters and doctoral students.

[83] Martin, J. et al., (2006) **Strategic Review of Panel and Cohort Studies: Report to the Research Resources Board of the Economic and Social Research Council**. London: Longview.

[84] Medical Research Council. (2014). **MRC Strategic Review of the Largest UK Population Cohort Studies**.

[85] Townsley, R. (2016). **Interim report: initial analysis of responses to the consultation**. Ruth Townsley Research

[86] ESRC. (2017). **ESRC Longitudinal Studies Review: Workshop summary report**.

- **CLS and Understanding Society have developed a range of materials to assist longitudinal study users** including: online learning materials and video introductions to each study and contributing to CLOSER. CLS has a methods strategy on the most important challenges in analysis. Many Q-Step courses draw on the birth cohorts and panel studies. For example, CLS has a joint Q-Step programme with UCL which uses material from cohort studies as examples. In this programme, students can opt for a short course on longitudinal analysis.[87]

- **The Centre for Longitudinal Studies and University College London, together with Understanding Society and the University of Essex are among the partners in the new Biosocial Centre for Doctoral Training**, (CDT), starting in October 2017. They have a grant specifically for building capacity to use the biological and clinical data linked to the social data.

- **Understanding Society received ESRC funding to offer multidisciplinary training** through a targeted grant designed to enhance interdisciplinary training (beginning in May 2015), an area quite relevant to perceived needs in longitudinal research training.

- **National Centre for Research Methods (NCRM):** The NCRM has a broad responsibility to engage with ESRC investments, co-ordinate with other training providers and provides an online training portal where longitudinal study training courses should be advertised.

It is too early to evaluate the impact of some of these new initiatives and resources; it is difficult without further analysis to extrapolate from user downloads and publications whether longitudinal study data is being used to its full potential. It may be the case that a substantial portion of the relevant workforce lacks the requisite training to conduct methodologically sophisticated analyses using longitudinal data. Notably, a recent submission from the Royal Statistical Society in July 2017 to the ESRC review panel suggested that many of the publications arising from the birth cohorts did not use proper analytic methods and tended to ignore the central benefit of the longitudinal design by conducting cross-sectional investigations. This brief argued that the training in advanced analytic methods necessary for longitudinal study was not adequate, and put forward the suggestion that ESRC conduct an evaluation and move to implement improvement in longitudinal statistical training based on the report findings.

We were unable to establish a complete picture of the training available to students and other learners who wished to develop skills in working with longitudinal data, but it would be useful to have this provision mapped out to enhance discovery of training resources and for CLOSER, the NCRM and ESRC to better understand gaps and any redundancy.

4.5 Barriers to using longitudinal studies

4.5.1 Students

In 2014 the ESRC commissioned a report on enhancing access to ESRC longitudinal datasets by doctoral students and early career researchers.[88] In preparing the report, Gayle interviewed a convenience sample at an annual conference of ESRC-funded doctoral students.[89] The Gayle study noted that despite ESRC being a major UK funder of longitudinal studies and a substantial funder of over 500 postgraduate students per year, very few students used these datasets in their doctoral research. There is no current evidence which could estimate this proportion, though ESRC will collect this information on students who are using ESRC datasets starting in 2017.

However, our own analysis of UKDS download data demonstrated that about 40% of the UKDS users across all longitudinal studies are postgraduates, so there is no clear evidence of barriers to access. In discussion with ESRC staff, there was no current estimate of the proportion of ESRC-funded dissertations using longitudinal study data, though, again, this information is currently being collected by the ESRC. Gayle reported that students also cited the following as barriers to working with longitudinal studies:

- Late introduction of the availability of longitudinal study into their studies, thereby precluding their use

- A shortage of supervisory capacity in some institutions

- Inadequate provision of analytic training in methods specific to longitudinal study

[87] Kneale, D., et al. (2016). **Piloting and producing a map of Millenium Cohort Study Data usage: Where are data underutilized and where is granularity lost?** EPPI-Centre, Institute of Education; UCL.

[88] Gayle, V. (2015). Enhancing Access to ESRC Longitudinal Datasets (Doctoral Students and Early Career Researchers). Unpublished report, University of Edinburgh.

[89] Personal communication (2017)

CLOSER now attends the annual ESRC student conferences (to which ESRC-funded doctoral students are invited) in order to highlight to these students the potential of longitudinal studies data. Though the current review did not have the capacity to investigate these potential barriers, it would be important for the ESRC to require that the availability of specific training and supervision in longitudinal research is mapped out to avoid duplication and identify gaps.

4.5.2 Government

From the statistics on data downloads from UKDS, government users represent just 2% of the UKDS downloads of all longitudinal study data. There is somewhat greater government use of household panel data from Understanding Society and BHPS by government analysts; the difference between the use is relatively small (see **Annex L** on longitudinal study data usage 2007-2016). Government representatives, however, did express the desire to make better use of longitudinal data in all forms either through commissioning analyses or in conducting the analyses themselves, which may require additional training on longitudinal analytic approaches.

4.5.3 Developing a career in longitudinal studies research

This review could not undertake a focused review of evidence of barriers to use, but the 2006 review and the 2015 Gayle draft report each noted similar barriers to building careers working in longitudinal study, including large panel studies and birth cohorts. Gayle noted barriers to building a career focused on longitudinal analysis, citing the long time that is needed to gain the methodological skills required for such analyses.[90] EAGDA also noted more broadly that there may be disincentives to building a career in research in general, a challenge which is magnified in the area of studying large complex datasets, which can adversely impact early career researchers who are trying to develop careers in these fields.[91]

EAGDA has made a number of recommendations to improve incentives for researchers and data managers to work with longitudinal studies as well as to improve timely and effective access to longitudinal study data.[92] Gaining expertise in longitudinal analysis takes time and establishing a cohort takes time. If researchers wish to make a career in this area, the time lag in getting prospective longitudinal research funded and implemented is a challenge to career success. This is particularly true of those who maintain the infrastructure for implementing, documenting and enabling the use of longitudinal studies. Mid-career support may be necessary.

4.6 Overcoming these barriers

4.6.1 Need for mentors

Training and supporting the trainers may be an important area for investment. There may be a dearth of mid-career academics to fill the need arising from the expansion of training. This need was expressed in the meetings between panel members and stakeholders and in the Gayle report, but we have no specific data on workforce capability to deliver training across these domains that would inform a recommendation. ESRC should evaluate the availability of senior academic staff to mentor a growing number of doctoral, postdoctoral and early career researchers.

4.6.2 Opportunities for interdisciplinary training

There is increasing evidence for the interaction of social with biological processes in determining health outcomes (epigenetics). To ensure appropriate interdisciplinary approaches to a research agenda, collaboration should be grounded in interactive training. The 2016 consultation and 2017 workshop (for this review) noted the need for cross-disciplinary collaboration on longitudinal study between the social and the biological sciences, including health. The ESRC and the funded studies have begun to take steps in this regard.

In the context of extending and deepening training in quantitative methods, the ESRC has recently funded a CDT to undertake interdisciplinary doctoral training across the biologic and social spheres.[93] They have also made strategic allocations to DTPs demonstrating strength in the use of ESRC datasets and advanced quantitative methods. In the UK and the US the concept of "Team Science" is being developed within a biomedical context[94], but by extension should include the bio-social domain and competencies should be required to facilitate this at a postgraduate level.

---

[90] Gayle, V. (2015). Enhancing Access to ESRC Longitudinal Datasets (Doctoral Students and Early Career Researchers). Unpublished report, University of Edinburgh.

[91] Expert Advisory Group on Data Access. (2014). **Establishing Incentives and Changing Cultures to Support Data Access**.

[92] Ibid., 91

[93] Soc-B Centre for Doctoral Training in Biosocial Research; **https://socbcdt.wordpress.com/**

[94] The Academy of Medical Sciences; Team Science Project; **https://acmedsci.ac.uk/policy/policy-projects/team-science**

The incorporation of the seven Research Councils, with Innovate UK and Research England, into UK Research and Innovation (UKRI) in April 2018 provides enhanced opportunities for collaboration with the other councils on developing interdisciplinary training across the biological and social sciences and humanities.

4.7 Summary

There is clear evidence that there has been a substantial increase in the utilisation of longitudinal study data from both the cohort studies and the panel studies, as measured by the UKDS downloads. Whether this increase is the continuation of an upward trend or a direct result of the ESRC initiatives taken in response to the 2006 review, as well as other broader initiatives that also support longitudinal studies, is unclear. The high percentage of users who are postgraduates makes it unlikely that there remain significant barriers to data access for students. A similar increase in the volume of publications arising from longitudinal study data is also evident. This would indicate that access is not a major challenge for academic users, though there may be underutilisation by national and local government (less so in Understanding Society and BHPS than in the three CLS birth cohorts). Some disciplines may not be well represented. The coding for disciplines (of data downloads) is clearer for the social sciences than for biological and health sciences, where the codes are fragmented and poorly categorised. Through the discovery platforms of both CLOSER and UKDS, researchers can discover the opportunities inherent in the data as a result of clear information about available variables and measures.

Training efforts to build a foundation within undergraduate degrees (Q-Step) are too recent to evaluate as are the new DTPs, particularly those aimed at quantitative doctoral degrees or interdisciplinary biosocial doctoral degrees. The new ESRC requirements for reporting will soon reveal the degree to which doctoral dissertations are making use of the longitudinal data. However, the question regarding the extent to which the appropriate analyses are used by emerging and established researchers remains unanswered. Mere use of longitudinal data does not ensure appropriate analyses. Questions remain as to the degree to which there may be gaps in the availability of mentorship at the mid- and senior-career level to mentor students and early career researchers. In addition, integration between institutions offering training and advanced methodology is essential, and better mapping online would increase discoverability in training opportunities.

Section 5: Appendices

Appendix 1: Contributors

This document lists those who were consulted by the Panel or Steering Group and contributed to the Review.

Academy of Social Sciences
Sharon Witherspoon

ADRN
Chris Dibben
David Ford
Paul Jackson

Avon Longitudinal Study of Parents and Children
Andrew Boyd

Born in Bradford
John Wright

CALLS-HUB
Fiona Cox
Oliver Duke-Williams

Centre for Longitudinal Studies
John Bynner
Lisa Calderwood
Emla Fitzsimons
Alissa Goodman
Morag Henderson
Alice Sullivan

CLOSER
Rob Davies
Rosie French
Alison Park
Natasha Wood (MRC LHA)

Dementias Platform UK
Chris Orton

Federal University of Bahia, Brazil
Marcos Barreto
Mauricio Barreto

Health Data Research UK
Andrew Morris

Institute for Fiscal Studies
James Banks
Chris Belfield
Robert Joyce
Zoe Oldfield

London School of Hygiene and Tropical Medicine
Katie Harron

Medical Research Council
Rachel Knowles
Rhoswyn Walker

MRC National Survey of Health and Development
Diana Kuh
Marcus Richards

MQ
Cynthia Joyce

NHS Digital
Garry Coleman

Nuffield Foundation
Tracey Budd
Teresa Williams

Office for National Statistics
Iain Bell
Pete Stokes

Resolution Foundation
The Rt Hon. The Lord Willetts

Royal Statistical Society
Tom King
Fiona Steele

The Children's Commissioner Office
Leon Feinstein

UCL Institute of Child Health
Carol Dezateux

UK Biobank
Sir Rory Collins
Cathie Sudlow

[94] This list does not include people who were present at meetings attended by Panel members but not interviewed by them.

UK Data Service
Louise Corti
Jon Johnson
Katherine McNeil
Margaret Ward
Matthew Woollard

UKRI
Rebecca Endean

UK Government
Chief Scientific Advisors
DEFRA
Department for Education
Department for Work and Pensions
Departmental Directors of Analysis
Heads of Policy Profession
Welsh Government

UNC Gillings School of Global Public Health
Linda Adair

Understanding Society
Michaela Benzeval
Jon Burton
Tom Crossley
Glenn Dunn
Gundi Knies
Meena Kumari
Peter Lynn
Alita Nandi
Victoria Nolan
Raj Patel
Birgitta Rabe

University of Aberdeen
Ian Diamond

University of Edinburgh
Vernon Gayle

University of Glasgow
Jill Pell

University of Southampton
Brienna Perelli-Harris

University of Warwick
Peter Elias

Wellcome
Nicola Perrin
Erica Pufall

Welsh Government
Richard Thurston

Groups whose meetings Panel or Steering Group members attended:

CLOSER Governing Board
25th April 2017

CLS Strategic Advisory Board
16th May 2017

Understanding Society Governing Board
25th May 2017

"The Value and Future of Birth Cohort Studies for Social Science and Policy"; CLS Roundtable at the British Academy
10th July 2017

"Large Longitudinal Studies: design and methodology"; Royal Statistical Society
26th October 2017

Appendix 2: Review specification

ESRC Longitudinal Studies Review 2017 - Specification and Remit
(Published February 2017)

This specification belongs to the Review Steering Group and its purpose is to inform the Review Panel, setting out the purpose, objectives of the review and the deliverables required. It is also to inform Council and other stakeholders in the review, including other funders.

Background
It is timely to carry out a strategic review of the ESRC's investment in longitudinal resources including the continued and future scientific relevance, sustainability, and contribution to the wider portfolio of social science and interdisciplinary data and resources in the UK and internationally.

Our funding environment is changing, as is the availability of data, and as we move forward into **UK Research and Innovation** (UKRI) it is time for ESRC to take stock, look forward and review the scientific need for longitudinal studies.

Longitudinal studies are often seen as a flagship component of the social science data resources we fund, and form a core part of our **strategic plan**. They have proved their worth in **addressing scientific questions of central importance** to individual and societal wellbeing and been major aids to policymakers, as the only means for studying processes of individual life-course development and dynamics, and the effects of earlier characteristics and experiences on later outcomes. ESRC makes significant annual and long term investment in longitudinal studies and related resources.

We need a clear vision for the future of our commitment in this area so that we can make the case for investment in our portfolio that remains relevant to research and policy communities, that our investments will continue to evolve and innovate to keep the UK at the global forefront of methods and research, that they connect appropriately to the rapidly developing data-for-research landscape, and that they deliver value for money.

The **last strategic review** of our longitudinal studies reported in 2006. Since then, ESRC has continued its investment in the **Centre for Longitudinal Studies** as well as investing in new studies including **Understanding Society** and **Life Study** (discontinued). Additionally, ESRC has invested strategically in other longitudinal resources including the **Census longitudinal studies**, **Next Steps**, **CLOSER**, and UK ageing studies **NICOLA** and **HAGIS**, as well as research projects that make extensive use of the longitudinal data including the **International Centre for Lifecourse Studies**. Further, we co-ordinate funding for **ELSA**, co-fund **Born in Bradford** and are continually involved in discussions with other funders about contributions to other studies. Investment has used both responsive mode research funding and capital infrastructure funding.

The context in which these large surveys are undertaken is changing in a number of respects, including changes in the research councils' funding environment and preparations for the establishment in 2018 of UKRI, shifting attitudes towards participation in such studies, survey costs, and technological and policy advances in data acquisition and linkage, including administrative data and biomarkers; examples include **ADRN**, the **Farr Institute** and **UK Biobank**. It is essential that future research data provision meets interdisciplinary needs and that the ESRC works with the MRC and other funders. The review will need to scope out the potential of data linkage, including use of administrative, biomarker and genetic data, the barriers to linkage and how, how far and when these could realistically be overcome.

Aims and objectives
The aim of this review is to provide an evidence-based and challenge-led assessment of the future social and interdisciplinary scientific and policy-relevant needs for data to address the types of research questions for which longitudinal data has typically been used (or could be used), and the value of the life-course evidence from our longitudinal studies in comparison with other sources of evidence. It will provide an assessment of the feasibility and effectiveness of alternative options for delivering the evidence needed, through exploration of trade-offs between options or combinations of options, and costs and benefits of each, under certain specified conditions. The review will need to address the questions how research and policy-relevant needs could be met without continuing investment in longitudinal studies, and what of value would be missing without them. It will need to be mindful of uncertainties about the future context, including Research Council funding and key issues such as access to routine data.

The review will form a core contribution to the development of our vision for longitudinal studies and related resources the next 10 years and beyond. It will provide us with the evidence we need to support future social science investment in longitudinal infrastructure as we move forward into UKRI.

The review, which builds on the **ESRC 2006 Strategic Review of Panel and Cohort Studies**, will provide clear and realistic recommendations with reasoned priorities on how our longitudinal investments individually and collectively could best be positioned to meet the needs of future research and policy-making challenges, and criteria to guide our funding decisions. The recommendations will inform our strategy for longitudinal studies and future funding, management and commissioning decisions, including what to continue, to change, to stop and to start.

These recommendations will relate to our decisions about future support for individual longitudinal studies, and to the support for discoverability, comparability, innovation, and development of an interdisciplinary UK longitudinal studies community (currently supported by CLOSER), and to international opportunities. Recommendations could include to seek additional co-funding for particular studies, to focus support on particular studies or particular topics where longitudinal data is most valuable, or to change the shape of the portfolio in other ways such as by reducing the frequency of data collection, changing the information that is collected, adopting new modes of data collection, starting a new study, or shifting from data collection to linkage. The recommendations need to take into account feasibility of practical realisation, and provide alternatives that can be applied in different conditions.

Rationale

The review will address the following questions:

1. What are the current and likely future scientific (including and beyond traditional user communities) and policy-related needs for data, including what types of data?

2. What and where, broadly, is the value and potential of the life-course data and other evidence from our current portfolio of longitudinal investments in meeting current and future scientific needs? This will draw on learning from case studies of key impacts from the longitudinal studies. What are the strengths and weaknesses in the studies' coverage (population and topics) and data quality? Where are the key limitations and gaps? Where, if anywhere, is there duplication?

3. How do our longitudinal resources fit with the broader national and international data infrastructure, looking ahead? What is their value in comparison with other sources? What is the realistic potential for research of linkage of administrative and biomarker data within the next 10 years?

4. How do and might our investments complement other resources? What are the gaps and overlaps, and how could these be addressed?

5. What can be learnt from other studies in the UK and internationally for our existing studies, for a potential new study/studies and by the ESRC?

6. What should be our future priorities for funding longitudinal data resources? How might we change or stop existing studies or focus on particular studies or areas? Should we start new studies or other related investments, with what focus? What infrastructure is needed, broadly, to realise the potential of the studies, including in combination with other data resources. The panel is invited to set out priorities in the scenarios of a substantial increase and decrease in funding as well as with no change.

7. What should be the key considerations in commissioning a study, alone or with other funders, to scope and shape the design and commissioning of a potential new cohort study?

8. Given the changing context within which we are operating what approaches, broadly, might we take to funding that would best deliver strategically on our future portfolio and ensure sustainability, working with other funders?

The review will use and assess evidence from existing sources, including the **2014 MRC cohort review**, case studies of impacts, and from specially designed work-streams, commissioned externally or undertaken in the ESRC office, including the community consultation and a workshop focussing on needs and how they can be met. Each workstream will produce a report that can be shared publicly, as well as informing the overall review report.

Report composition, format and quality criteria

The review report will provide a concise executive summary and the Panel's recommendations, and will set out the background to the review, its objectives including the key questions addressed and the review methodology. It will present a thorough evaluation of the evidence the Panel considered along with its insights, options analysis and the recommendations arising in context. The report will be concise and written in a publicly accessible way suitable for publishing; it will use annexes for more detailed evidence as necessary.

Quality review

The draft report will be reviewed by the review Steering Group and ESRC review team leads against this specification and other guidance provided to the Panel. The final report will be signed off by the Steering Group Chair.

Longitudinal Studies Review Steering Group

The role of the Steering Group has been to represent the interests of the ESRC in signing off the specification for the review and the final the report from the Panel, to provide effective links with the ESRC's governance structures and partner organisations, to provide the ESRC office team with independent strategic advice during the progress of the review, and to offer their expert knowledge, in particular on the UK context and issues, to the independent review Panel.

Membership:

Anna Vignoles (Chair), University of Cambridge, ESRC Research Committee

Mary De Silva, Wellcome

Jane Falkingham, University of Southampton, ESRC Council

Gordon Harold, University of Sussex, ESRC Capability Committee

Debbie Lawlor, University of Bristol, MRC Population Health Sciences Group

Nyovani Madise, University of Southampton, ESRC Capability Committee, MRC Global Health Group

Joe McNamara, MRC

Tom Wells, Government Office for Science

More information about the Steering Group can be found on the **Longitudinal Studies Review website**.

Appendix 3: Glossary

This appendix provides alphabetically a glossary of acronyms, studies and organisations that appear in the report with links to their relevant websites.

ADRN: Administrative Data Research Network

ADT: Administrative Data Taskforce

ALSPAC: Avon Longitudinal Study of Parents and Children

BBSRC: Biotechnology and Biological Sciences Research Council

BCS70: 1970 British Cohort Study; held at CLS

BHPS: British Household Panel Survey

BIB: Born in Bradford BIB

Brazilian 100 Million Cohort: A new resource based in Brazil, built around the Cadastro Único database

British Library

Cancer Research UK

CelSIUS: Centre for Longitudinal Study Information and User Support (England and Wales)

CDT: Centre for Doctoral Training

CLS: Centre for Longitudinal Studies at University College London Institute of Education

CALLS-Hub: Census & Administrative Data Longitudinal Studies Hub

CFPS: China Family Panel Studies

CLOSER: Cohorts & Longitudinal Studies Enhancement Resource

CLOSER Learning Hub

CLOSER Discovery

CNEF: Cross-National Equivalent File

A Comparative Study of Life Course and Family Dynamics

Department for Digital, Culture, Media & Sport

Digital Economy Act 2017 DTPs: Doctoral Training Partnerships

Department for Education

Department of Health and Social Care

DWP: Department of Work and Pensions

EAGDA: Expert Advisory Group on Data Access

ECR: Early Career Researchers

ELSA: English Longitudinal Study of Ageing

ESRC: Economic and Social Research Council

Farr Institute

Gateway to Global Ageing Data

Growing up in Scotland

GSOEP: German Socio-Economic Panel

Health Data Research UK

HAGIS: Healthy Ageing in Scotland

HCS: Hertfordshire Cohort Study

Higher Education Founding Council for England

HRS: Health and Retirement Study

HM Treasury

Institute for Fiscal Studies (IFS)

IoE: UCL Institute of Education

Life Sciences Industrial Strategy

Life Study ('2012 birth cohort')

LMIC LPS Directory: Low and Middle Income Longitudinal Population Study Directory

MCS: Millennium Cohort Study; held at CLS

METADAC: Managing Ethico-social and Technical issues and Administration Data Access Committee

MRC: Medical Research Centre

MRC National Survey of Health and Development

NCDS: 1958 Birth Cohort/ National Child Development Study; held at CLS

NCRM: National Centre for Research Methods

NLCS: Netherlands Cohort Study

Next Steps: 1989-90 cohort; held at CLS (formerly known as the Longitudinal Study of Young People in England - LSYPE)

NHS Digital

NICOLA: Northern Ireland Cohort Longitudinal Study of Ageing

NILS-RSU: Northern Ireland Longitudinal Study Research Support Unit

NKPS: Netherlands Kinship Panel Study

Nuffield Foundation

PSID: Panel Study of Income Dynamics

Q-Step: programme designed to promote a step-change in quantitative social science training in the UK

RSS: Royal Statistical Society

Secure Anonymised Information Linkage Databank (SAIL Databank)

SDAI: Secondary Data Analysis Initiative

SLS-DSU: Scottish Longitudinal Study Development and Support Unit

SWS: Southampton Women's Survey

UCL: University College London

UK Biobank

UKDS: UK Data Service

UKDS Discover

UK Office for Life Sciences

UKRI: UK Research and Innovation

UK Statistics Authority

Understanding Society: The UK Household Longitudinal Study

Wellcome

Welsh National Pupil Database: Welsh Pupil Level Annual Schools Census and Pupil Attainment dataset

Appendix 4: Bibliography

Administrative Data Taskforce (2012). **The UK Administrative Data Research Network: Improving Access for Research and Policy**.

Bann, D., Johnson, W., Li, L., Kuh, D. & Hardy, R. (2017). Socioeconomic Inequalities in Body Mass Index across Adulthood: Coordinated Analyses of Individual Participant Data from Three British Birth Cohort Studies Initiated in 1946, 1958 and 1970. *PLoS Med* 14(1): e1002214. **https://doi.org/10.1371/journal.pmed.1002214**

Brimblecombe, N., Evans Lacko, S., King, D., Maughan, B., Takizawa, R., Arseneault, L., & Knapp, M. (under review). Long term economic impact associated with childhood bullying victimisation.

Brown, D., Anda, R., Tiemeier, H., Felitti, V., Edwards, V., Croft, J., & Giles, W. (2009). Adverse childhood experiences and the risk of premature mortality. *American Journal of Preventative Medicine*, 37(5), 389-96.

Bryson, C., Purdon, S., Skipp, A., Barlow, A., Ford, T., Hunt, J., Kiernan, K., low, H., McKay, S., Miles, M. & Trinder, L. (2017). **Understanding the lives of separating and separated families in the UK: what evidence do we need? Summary report**. The Nuffield Foundation; London.

Buck, N. & McFall, S. (2011). Understanding Society: design overview. *Longitudinal and Life Course Studies*, 3(1), 5-17. DOI: **http://dx.doi.org/10.14301/llcs.v3i1.159**

Centre for Longitudinal Studies. (2017). **The value and future of birth cohort studies for social science and policy: Summary of Round Table meeting at the British Academy, July 10th 2017**.

CLOSER. (2017). Additional information from CLOSER (Cohort & Longitudinal Studies Enhancement Resources) produced for the Longitudinal Studies Review Panel.

Dawn Teare, M. & Barrett, JH. (2005). Genetic linkage studies. *Lancet*, 366(9490):1036-44. DOI: **10.1016/S0140-6736(05)67382-5**

Digital Economy Act 2017

Dolton, P.J. and Vignoles, A., 2002. The return on post–compulsory school mathematics study. *Economica*, 69(273), pp.113-1. DOI: **10.1111/1468-0335.00273**

Dwork, C. & Yekhanin, S. (2003). Database Privacy. Retrieved from: **https://www.microsoft.com/en-us/research/project/database-privacy/?from=http%3A%2F%2Fresearch.microsoft.com%2Fresearch%2Fsv%2Fdatabaseprivacy%2F**

ESRC. (2017). **ESRC Longitudinal Studies Review: Workshop summary report**.

Evans-Lacko, S., Takizawa, R., Brimblecombe, N., King, D., Maughan, B., Knapp, M., & Arseneault L. (2016). Childhood bullying victimisation is associated with use of mental health services over 5 decades: A longitudinal nationally-representative cohort study. *Psychological Medicine*, 47, 127-135.

Expert Advisory Group on Data Access. (2014). **Establishing Incentives and Changing Cultures to Support Data Access**.

Expert Advisory Group on Data Access. (2017). Additional Information from EAGDA produced for the Longitudinal Studies Review Panel.

Gayle, V. (2015). *Enhancing Access to ESRC Longitudinal Datasets (Doctoral Students and Early Career Researchers)*. Unpublished report, University of Edinburgh.

Goisis, A., Özcan, B., & Myrskyla, M. (2017). Decline in the negative association between low birth weight and cognitive ability. *Proc Natl Acad Sci U S A,* 114(1): 84-88. DOI: **10.1073/pnas.1605544114**

Harold, G. T., Acquah, D., Chowdry, H., & Sellers, R. (2016) What works to enhance inter-parental relationships and improve outcomes for children. Department for Work and Pensions (DWP), Ref: ISBN 978-1-78425-719-4, Ad hoc research report 32.

Harron, K., Gilbert, R., Cromwell, D. & van der Meulen, J. (2016). Linking Data for Mothers and Babies in De-Identified Electronic Health Data. *PLoS ONE*, 11(10): e0164667. **https://doi.org/10.1371/journal.pone.0164667**

HM Government (2011). **Opening Doors, Breaking Barriers: A Strategy for Social Mobility**.

Hoek, G., Brunekreef, B., Goldbohm, S., Fischer, P. & van den Brandt, P. A. (2002). Association between mortality and indicators of traffic-related air pollution in the Netherlands: a cohort study. *Lancet*, 360(9341), 1203-1209. DOI: **http://dx.doi.org/10.1016/S0140-6736(02)11280-3**

Hughes A, McMunn A, Bartley M., & Kumari M. (2015) Elevated inflammatory biomarkers during unemployment: modification by age and country in the UK. *J Epidemiol Community Health*, 69(7):673-9. doi: 10.1136/jech-2014-204404. Epub 2015 Feb 19.

Kneale, D., Patalay, P., Khatwa, M., Stansfield, C., Fitzsimons, E., &Thomas, J. (2016). **Piloting and producing a map of Millenium Cohort Study Data usage: Where are data underutilized and where is granularity lost?** EPPI-Centre, Institute of Education; UCL.

Lacey RE, Kumari M, & Bartley M. (2014). Social isolation in childhood and adult inflammation: evidence from the National Child Development Study. *Psychoneuroendocrinology*, 50:85-94. doi: 10.1016/j.psyneuen.2014.08.007. Epub 2014 Aug 22.

Lacey, R.E., Kumari, M., & McMunn, A. (2013) **Parental separation in childhood and adult inflammation: The importance of material and psychosocial pathways**. *Psychoneuroendocrinology*, 38(11), 2476-2484.

Martin, J. et al., (2006). **Strategic Review of Panel and Cohort Studies: Report to the Research Resources Board of the Economic and Social Research Council**. London: Longview.

Medical Research Council. (2014). **MRC Strategic Review of the Largest UK Population Cohort Studies**.

Mills, M. C. & Rahal, C. (under review). The Anatomy of GWAS.

Moser, C. (1999). **Improving Literacy and Numeracy: A Fresh Start**. London: Department for Education and Employment.

OECD. (2016). Research Ethics and New Forms of Data for Social and Economic Research. OECD Science, Technology and Industry Policy Papers, No. 34. OECD Publishing; Paris. **http://dx.doi.org/10.1787/5jln7vnpxs32-en**

Patalay P & Fitzsimons E. **Mental ill-health among children of the new century: trends across childhood with a focus on age 14**. September 2017. Centre for Longitudinal Studies: London.

Ploubidis, GB., Sullivan, A., Brown, M. & Goodman, A. (2017). Psychological distress in mid-life: evidence from the 1958 and 1970 British birth cohorts. Psychol Med, 47(2):291-303. DOI: **10.1017/S0033291716002464**

Research Project Management. (2016). **Administrative Data Research Network (ADRN) Mid-Term Review Report: November 8, 2016**. Toronto: Research Project Management.

Sakshaug, J. W. & Antoni, M. (2017) Errors in Linking Survey and Administrative Data. In Biemer, P., Eckman, S., Edwards, B., de Leeuw, E., Kreuter, F., Lyberg, L., Tucker, C. & West, B. (eds.). *Total Survey Error in Practice*. John Wiley & Sons Ltd. DOI: **10.1002/9781119041702.ch25**

Takizawa, R., Danese, A., Maughan, B., & Arseneault, L. (2015). Bullying victimization in childhood predicts mid-life risks for cardiovascular disease: A 5-decade birth cohort study. *Psychological Medicine*, 45, 2705-2715.

Takizawa, R., Maughan, B., & Arseneault, L. (2014). Adult health outcomes of childhood bullying victimization: Evidence from a 5-decade longitudinal British cohort. *American Journal of Psychiatry*, 171, 777-784.

Townsley, R. (2016). **Interim report: initial analysis of responses to the consultation**. Ruth Townsley Research

Townsley, R (2017a). **Further analysis of responses to the consultation. Paper 3: The capacity and infrastructure for longitudinal research**. Ruth Townsley Research.

Townsley, R (2017b). **Further analysis of responses to the consultation. Paper 4: Use and accessibility of longitudinal resources**. Ruth Townsley Research.

UK Cabinet Office, Government Digital Service and The Rt Hon Ben Gummer. (2017). **Government Transformation Strategy 2017 to 2020** [Policy Paper].

UK Office of Life Sciences. (2017). **Life Sciences: Industrial Strategy report**.

Understanding Society. (2017). Additional Information from Understanding Society produced for the Longitudinal Studies Review Panel.

Van den Eynden, V., Vlad, A & UK Data Service. (2017). **Towards Open Research: what ESRC-funded researchers tell us**.

Waldfogel, J. & Washbrook E. (2010). **Low income and early cognitive development in the UK. A Report for the Sutton Trust**.

Wellcome. (2017). **Longitudinal Population Studies Strategy: Wellcome's Longitudinal Population Studies Working Group**.