ESRC Longitudinal Studies Review 2017

Further analysis of responses to the consultation

Paper 4:

Use and accessibility of longitudinal resources

April 2017

Report author:

Ruth Townsley, Independent Researcher

The views represented in this report are from those who responded to the consultation and do not represent the views of ESRC
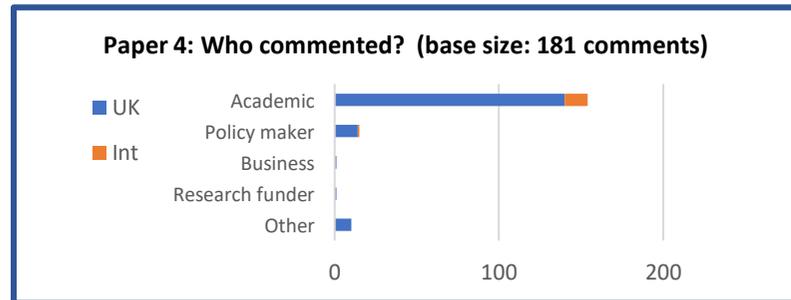
E·S·R·C
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL

Introduction

The ESRC Longitudinal Studies Review 2017 is exploring the current and future scientific and policy-relevant need for longitudinal research resources. The review started in October 2016 and will report to ESRC Council early in 2018. An open online consultation in autumn 2016 sought input broadly, resulting in 637 completed responses from UK (83.4%) and international (16.6%) respondents. Respondents were predominantly from the academic sector (80%) as well as government, civil society and business sectors (20%). The main findings of the consultation were published in December 2016 in an initial report. This report is supplemented by short briefing papers that examine key themes from the consultation data in more detail.

Paper 4: Use and accessibility of longitudinal resources

This paper examines respondents' comments about the use and accessibility of longitudinal resources. A total of 181 comments were examined from a number of sources: (1) further analysis of existing coded comments relating to documentation and dissemination (coded for the questions relating to methodological and technological issues); (2) wildcard searches using the strings use* and access* across the whole dataset.

Paper 4: Who commented? (base size: 181 comments)

Awareness and accessibility of longitudinal resources

Respondents commented on a range of issues relating to the awareness and accessibility of current longitudinal resources, reflecting concerns from both experienced and less frequent data users.

> Greater awareness and wider use of the range of longitudinal resources available and how to access them – there was some evidence that non-academics may be unaware of the quality and relevance of publicly available longitudinal data
> Access to training and support for non-expert users – e.g. training for policy makers, or support for schools, to use the data to develop analyses of interest to them
> Increased and improved online access tools for non-expert users – particularly for those who do not have access to (or the skills to use) advanced statistical analysis software
> Development of simplified datasets - to help non-expert users get started and to determine which variables are of potential interest
> Further work to ensure there is consistency of measures to make data easier to use
> Further data harmonisation and updating to make data more useable and cross-cohort research easier – e.g. partnership histories in BCS70 after 2008
> Easier and faster access to sensitive data – VML access may be limiting the use of certain data and having potential impact on science
> Further collaborations with and among data archives (e.g. between UKDS and biomedical data archives) to provide easier discovery and access to data across different repositories
> Cost-free access – although access to ESRC-funded surveys is free via the UK Data Service, there was a plea for free access to some paid-for

resources, such as ALSPAC and Hospital Episode Statistics, by certain groups (e.g. early career researchers and PGRs)

> Ensuring that new forms of data are promoted and can be accessed, understood, explored and integrated - innovative instruments are collecting new and often high-dimensional data (e.g. through bio samples, wearables and smartphone apps) which need documentation and interpretation to make them accessible to researchers.

"Longitudinal studies remain an untapped resource despite the SDAI programme, CLOSER and the sterling efforts of the core research teams. Policy makers need to be better engaged, through initiatives such as the Policy Unit at ISER and the Knowledge Transfer Programme at TILDA." (ID 16)

"More focus on user friendliness and more funding for proper data interface, documentation and useable data for a wide variety of users. This is even more essential for innovative biosocial data that combines genetic data." (ID 449)

"It is imperative that the new data – highly detailed and technical in many cases - are made accessible to researchers, to facilitate research to improve our understanding of existing research questions and to open up new important areas of enquiry." (ID 394)

Documentation of paradata and metadata

Many comments relating to use of longitudinal studies reflected concerns about the effective, regular and on-going documentation of the processes and decisions by which data are collected, handled and analysed. Key issues raised were:

> Information on extent and handling of missing data, such as details of time taken and computational power/software needed to undertake missing data analyses

> Information on sample weights - where these have been used to correct for missingness and non-response, some researchers queried the implications for survey representativeness and requested clearer documentation on methodology used to derive sample weights, and examples of their use in longitudinal analyses

> Complete and consistent variables documentation for some of the cohort studies and, ideally, the production of a new integrated data set every time a new wave of data collection is integrated into a longitudinal study

> Documentation of data transformation for older survey sweeps

> Considering more use of life-cycle data documentation that is standardised across longitudinal studies – e.g. the Data Documentation Initiative (DDI) metadata standard, which can be used to help studies to document their data (and associated methodology)

> Other areas where standardised information was felt to be needed included guidelines for privacy/non-reportable data and for reporting computational models

> Reducing duplication of effort in the cleaning, processing and analysis of data by means of sharing work by researchers to create new measures using multiple variables, undertake variable derivation, or create new code/syntax for analysis.

"Missing data and attrition are becoming a key issue and while in theory suitable tools exist they need to be made routinely available in a form social science users can access." (ID 576)

"The format of the data is often challenging to interpret and variables may require derivation (e.g. education attainment at a particular age, or total household income). It would be helpful to reduce duplication of effort by ensuring those who run longitudinal studies are sufficiently resourced to enable the sharing of such variable derivation/syntax depositing, or hosting a website where such syntax could be readily shared by the entire research community. (ID 98)

Disseminating data and findings

A small number of comments related to improving the dissemination and impact of data and findings:

> More accessible feedback to study participants through non-traditional means – e.g. lay summaries, online dissemination, social media posts and broadcast content.
> Engagement and dissemination events for participants throughout the study process – to value their contribution and to enable them to see the difference their participation is making to wider society
> Targeted outputs for policy impact and knowledge transfer, through improved access of longitudinal data to a wider range of data users
> Consideration of whether it would be useful to set up a register of researchers who have the training and expertise to readily access and interpret data from longitudinal resources
> The importance of planning for the unknown and unspecified future uses of longitudinal data.

"I think we should make a bigger effort to feed-back findings of the study to participants - so they have better understanding of the value and contribution they are having. I would push academics to publish lay person summaries for participants - which could be passed onto the participants with a note of thanks. People are often happy to participate if they know how valuable it is." (ID 311)

"If a wider set of potential 'users' of longitudinal data are discovered through better sharing of evidence and potential, this will also lead to a better return on investment in data collection." (ID 17)

"Longitudinal studies by their nature are unique and future unspecified uses should not be underestimated!" (ID 560)