

The future of panel and cohort studies: the need for a scoping study

1. Context

Through its recent renewed investment in the UK Longitudinal Studies Centre (ULSC) and via its support for the cohort studies managed by the Centre for Longitudinal Studies (CLS), the ESRC has committed itself to maintaining the UK panel and cohort studies for a further five years. This investment followed from the recommendations of the Mid-Term Review of the ULSC, which recognised that the British Household Panel Study and the Birth Cohort Studies are of the highest international quality. The main outcome of the review was further funding for the British Household Panel Study, the 1958 and 1970 Birth Cohort Studies and the Millennium Cohort Study.

2. Why we have to think about the future

Despite this more stable context, we need to start thinking now about the longer-term future of these studies. There are a number of reasons for this, outlined below.

2.1 The value of long-term panels

The major strength of long-term panels and cohort studies are the insights they offer for researching the lifecourse in ways which are not available by other means. This is already evident in UK social science research based upon the 1958 and 1970 birth cohorts (and in certain respects from the 1946 cohort, which is not part of the ESRC-funded suite of studies), and is beginning to be clear from the BHPS too (now that it has completed 13 waves). There is a need to be sure that the fullest use is being made of these resources as they have matured. In this respect it will be of interest to evaluate the experience with long-term panels from other countries.

2.2 Attrition and attenuation

Long-term panels and cohort studies experience attenuation (falling response rates) and attrition (falls in response rates that are differentiated by relevant characteristics). Attenuation is inevitable: at wave 10 (in 2001) the BHPS had a net response rate at the individual level of just over 60% of the respondents at wave 1, and the NCDS had a similar rate in 2000. Even if this attenuation is not biased, it still matters, because ultimately it reduces the power to undertake meaningful analyses, especially when the remaining samples are further reduced in size by item non-response. But where attenuation becomes differential attrition, the situation is more serious. There is then a need to explore questions such as:

- (a) the extent to which differential attrition matters: what kind of biases does it induce, both to substantive statistical estimates and to estimates of the sampling error of these?
- (b) how differential attrition comes about;
- (c) how differential attrition may be kept to a minimum;
- (d) whether, if at all, analysis could be adjusted statistically to compensate.

A succinct summary of existing knowledge on these issues would help inform decisions about the funding of future panel and cohort studies.

2.3 Refreshment

Refreshment – the inclusion of new study members – could in principle be used to deal with attrition and attenuation, but it is usually thought of in the context of threats to the representativeness of a longitudinal study which are unavoidable in the original sample design. The most obvious is in relation to immigration: new residents of the UK generally have no chance of being included in a panel that has already been established (although in the case of the British Household Panel Study they would be included if, for example, they married an existing sample member). There is a need to know what the scope is for using refreshment to maintain the representativeness of longitudinal surveys in a context of social change and what modifications would need to be made to statistical methods in order to adjust for refreshment.

2.4 ICT, survey design, conduct and management

Over the last 10 years there have been major technological changes which have impacted upon the design, conduct and management of the process of collecting data from participants in panel and cohort studies. Some of these developments have already been incorporated into the data collection methods (*e.g.* all of the major longitudinal studies in the UK use computer-assisted personal interviewing techniques). But what scope exists to utilise further these technical developments? With the growth of access to the internet, the ubiquitous spread of mobile phones and SMS services, methods of maintaining contact with participants might change. Other hardware and software innovations now make possible a variety of data collection approaches. For example, what role might be played by web surveys in the long-term maintenance of panels? Could a mixed mode of contacting respondents be developed that would use email *etc.* for those who have access to it? What would be lost and gained by this approach?

2.5 Population diversity

There is probably greater interest in analysing diversity in the UK population now than there was when some of the longer-running studies were established. This has been recognised through the additional funds made available for enhancements to the BHPS in Scotland, Wales and Northern Ireland. Similar enhancements, funded by departments of the devolved administrations, were added to the Millennium Cohort Study from the start. For several of the large cohort or panel studies, there are likely to be further proposals along these lines, for example in response to the need for information relating to the Greater London area. These changes are not without methodological problems. In the BHPS, the adding on of extra new samples in Scotland and Wales has already raised some issues of consistency of design, and the MCS is finding it difficult to reconcile the common framework of the whole study with the divergent policies and needs of funders in the different nations of the UK. There is a need for information about the effects of such changes on the scope for valid and reliable analysis of diverse groups.

2.6 International comparability

The high reputation of the UK's panel and cohort studies has helped to ensure that the issue of international comparability remains firmly on the agenda, and there have also quite recently been some valuable projects to harmonise variables for use in comparative research¹. There is a need to know whether further investment is needed to ensure continuing comparability.

3. Subject matter of surveys

3.1 Topics covered

The cohort and panel studies have always tried to cover a broad range of topics, but inevitably some fields are less thoroughly dealt with than others. For example, it has been suggested that there is not enough information relating to health issues in some of the studies. Some changes to the evolving design of the questionnaire have been made in response to such concerns: for example, in recent sweeps of the BHPS there has been rather more on electoral topics than there was in the 1990s. Each study has approached this important issue by different means (user group meetings, topic reviews and/or via consultation with sponsors). There is a clear need for a more coordinated approach to the choice of topics covered by the various studies and the development of these topics within the specific sweeps or waves of each study.

3.2 Short-term versus long-term dynamics

There is no doubt that the most visible uses of the BHPS have been in relation to short-term dynamics, or even to its status simply as a high-quality cross-sectional survey with a large amount of life-history data. Attracting funding from government sources is likely to require that some attention is paid to relatively short-term change. Yet the main argument for investing in cohort and panel studies is their long-term value, as noted in 2.1 above. How can the needs for both short-term and long-term dynamics be reconciled?

3.3 Questionnaire design and the frequency of data collection

Arising from both 3.1 and 3.2 is the specific issue of how to balance continuity and change in the content of questionnaires. Some components of the questionnaires for the panel and cohort studies are now included regularly but not annually. In the case of the BHPS others remain annual. But the rationale for the allocation to annual or less frequent collection is not entirely clear. For example, it has been suggested that important aspects of job history could not be reliably collected at intervals greater than one year. Yet some topics that are included at intervals of greater than a year are known from other panel surveys to vary at individual level yearly or more frequently: an example is 'national identity', which the British Election Panel Study shows to vary much more from year to year than is commonly supposed. There is a need for better information on which subject areas may be dealt with in this less frequent way without significant loss of data quality.

¹ See for example <http://www.ceps.lu/CHER/CHERPRES.htm>

<http://www.nessie-essex.co.uk/dataResources.asp>

3.4 Bio-medical data

A very specific set of ethical and other issues arise in connection with adding bio-medical data to panel studies. There is a need for further information about the particular methodological and ethical issues raised by adding such data to panel or cohort studies. Some studies, such as the Avon Longitudinal Study of Parents and Children, have collected much information relevant to the psychological condition of study members. What can be learnt from this experience? Should more such data be collected in other longitudinal studies, alongside biomedical and ecological data?

4. Linking to administrative and other contextual data

There are two kinds of information that could be used to enhance the value of cohort and panel studies. One is the matching to individuals of information held about them in administrative records, for example school examination results or hospital records. The other is contextual information, for example information about the characteristics of schools or hospitals which they attended. Individual-level data can often be obtained from respondents themselves, although perhaps less reliably than from administrative records, but contextual data would usually have to be obtained by other means.

What specific issues arise in this respect for panel studies? For example, could administrative data be added in retrospectively? What permissions are required to allow linking to take place? Even where data at the individual level have already been collected (directly or by matching) is there scope for retrospectively matching contextual data? If so, what funding would be needed to do that for the early waves of the oldest cohorts, for example for administrative information about schools or hospitals in the 1958 cohort?

5. Methods of analysis

In principle we now have very powerful statistical tools to analyse panel and cohort data, and statisticians have developed a thorough understanding of how the appropriate models for such data work. But there always remain possibilities for further development. For example, two of the major innovations in the last 10-15 years in this field are multi-level models and mixed Markov latent class models, but so far there does not exist software that would allow the two to be brought together (so that, for example, we could understand the influence of social collectivities such as schools, labour markets or electoral constituencies on the dynamics of a panel). No doubt this specific gap will be filled in due course, but more generally there is a need to know whether further investment is required in the development of methods for analysing panel and cohort studies.

Apart from analytical tools, there may be a case for the development of mixed methods of data collection. The panel and cohort studies have used fairly traditional methods of data collection – relying heavily on well-tested closed questions. But could existing data could be augmented or new topics covered, by qualitative approaches and well as the more traditional quantitative approaches? Such developments could be linked to the latest innovations in qualitative analysis software, which facilitates switching between quantitative and qualitative analytical techniques.

6. Strategic positioning of UK longitudinal studies

The UK has a wide and varied range of longitudinal data sources, many of which now span a long period of time. These include the birth cohort studies, the Census Longitudinal Study, the English Longitudinal Study of Ageing and the British Household Panel Study. Other longitudinal resources are currently under development (e.g. the Department of Work and Pensions Longitudinal Study, the Wealth and Assets longitudinal study, the ESRC Longitudinal Study of Minority Ethnic Groups, the Department of Education and Skills Longitudinal Study of Young Persons in England, the Scottish Executive's Early Years Survey). On the face of it, this appears more 'an embarrassment of riches' than a potential future problem. Yet there are some important decisions to be made soon which will have ramifications for the research potential of these resources. The two most important matters to be decided relate to the future of the BHPS and the development of a new birth cohort, in 2012 or earlier?

There is a need to position a new or revitalised panel study within this landscape. This will involve delineating the need for the kind of information that a new or revitalised panel study will provide and identifying the issues it will need to address. There is also an important debate to be held about the size of any new panel study. Is it the case that the so-called 'crown jewels' in the portfolio of data resources funded by the ESRC are more appropriately labelled 'little gems'. Is a panel of 5,000 households big enough to address the increasingly complex, detailed and diverse set of questions posed by our society?

The cohort studies followed a 12 year cycle of renewal from 1946 to 1970. There then followed a 30 year gap before a new cohort study was launched. Two other cohorts have potential to fill the gap. The ALSPAC study covers children born in 1990 and 1991 in the Avon and Bristol area. Should this cohort now follow the lines of the 1958 and 1970 cohort? Alternatively, it has been argued that, for England at least, the Department of Education and Skills Longitudinal Study will fill this gap, focussing upon a 'cohort'² of young people born in 1990/91. But the Department currently has no plans to follow this cohort after 2015. And if we see value in cross-cohort comparisons, should we be planning now for a new birth cohort to be started in the period from 2010 to 2012?

7. Next steps

It is proposed that a scoping study should be undertaken to consider the issues outlined above in detail. The scoping study should concern itself particularly with the scientific case for the development of panel and cohort studies, rather than with issues

² This study is not strictly a birth cohort. The study has an issued sample of approximately 21 thousand young people in Year 9 (aged 13-14). This includes an ethnic minority boost of 5 thousand and a boost for pupils in schools in deprived areas. The sampling is a two-stage process, first sampling schools stratified according to known characteristics (with disproportionate sampling for those schools in the highest quintile of proportion of pupils entitled to free school meals) and then sampling pupils from school lists (with a disproportionate sampling of pupils in the six main ethnic minorities: Black African, Black Caribbean, Bangladeshi, Indian, Pakistani and Mixed). Because both stages of the sample come directly from the Department's Annual School Census, data are available on all sampled schools and in the maintained sector, at individual level for all sampled pupils.

relating to the future funding situation. It should consider in depth the strategies and approaches adopted in other countries. This will involve a wide and far-ranging consultation, not confined to the social sciences and engaging with the potential users of the information generated by long term panels and cohort studies.

Lindsay Paterson
Peter Elias

14th January 2005