Report

closer
The home of longitudinal research

Preparing for the future: Tackling the key challenges facing the UK's longitudinal population studies

Conference report

Edited by Meghan Rainsberry

CLOSER

March 2019

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

MRC Medical Research Council

Correspondence:

Meghan Rainsberry
CLOSER, University College London
closer@ucl.ac.uk

First published in 2019 by
CLOSER (Cohort and Longitudinal Studies Enhancement Resources)
UCL Institute of Education
20 Bedford Way
London WC1H 0AL
www.closer.ac.uk

About CLOSER

The UK is home to the world's largest and longest-running longitudinal studies. CLOSER aims to maximise their use, value and impact both at home and abroad. Bringing together eight leading studies, the British Library and the UK Data Service, CLOSER works to stimulate interdisciplinary research, develop shared resources, provide training, and share expertise. In this way CLOSER is helping to build the body of knowledge on how life in the UK is changing – both across generations and in comparison to the rest of the world. CLOSER was funded by the Economic and Social Research Council (ESRC) and the Medical Research Council (MRC) from 2012-17, and by the ESRC from 2017 to present. Visit www.closer.ac.uk.

The views expressed in this work are those of the conference delegates and do not necessarily reflect the views of CLOSER, UCL, ESRC, MRC or the Wellcome Trust.

This document is available in alternative formats.

Please contact CLOSER:

tel: +44 (0)20 7612 6938

email: closer@ucl.ac.uk

Contents

Executive summary

On Thursday 6 December 2018, CLOSER hosted a one-day conference entitled *Preparing for the future: Tackling the key challenges facing the UK's longitudinal population studies,* at the Wellcome Trust in London. The event brought together around 70 delegates from across the longitudinal community, in social and biomedical sciences, including study Principal Investigators, researchers, professional staff, policymakers and funders. In a packed programme, the delegates discussed key issues in new forms of data collection, data harmonisation, data linkage, biosample management, impact, and engagement with policymakers.

The conference report documents the day's discussions, and the actions proposed by delegates to address key challenges, and fill knowledge & skills gaps. This Executive Summary provides an overview of the key points raised.

1 New forms of data collection

1.1 To drive innovation in longitudinal population studies, we must share and value learning on our failures as well as our successes, without judgement or risk to funding.
1.2 While the novelty of new technology may be attractive, longitudinal population studies need to carefully consider the biases it introduces, longitudinal continuity, and the practical and financial implications they will face. These need to be weighed against the quality, utility and scientific potential of data produced by new technology.
1.3 We need more research into what participants are willing to do, and how acceptable they find new data collection methods.
1.4 Data infrastructure and researchers' analytical skills will need an upgrade to cope with new forms of data.

2 Data harmonisation

2.1 Collaboration between studies is essential to achieving prospective harmonisation, particularly when considering new technology and methods of data collection.
2.2 There is a risk of sacrificing data quality in harmonisation, and researchers need to ensure the right balance is struck between precision and scope.
2.3 Good quality documentation is vital.
2.4 Harmonisation has important scientific benefits, but has been insufficiently resourced and incentivised. The longitudinal community needs advocates from across disciplines to promote the value of this work.
2.5 Harmonisation practice could be improved by researching new methods for retrospective harmonisation, developing more techniques for validating harmonisation outputs, and by exploring opportunities offered by calibration and emerging analytic techniques.
2.6 Strategies for overcoming the challenges of data harmonisation should be developed through cross-disciplinary collaboration, and an overall coordinating role could

prioritise measures to be harmonised and to promote the added utility of harmonised data.

3 Data linkage

3.1 Longitudinal population studies need to collectively champion administrative linkage, advocate for their needs, and proactively make the case for the value of linked data in answering research and policy questions.
3.2 The multiplicity of government data holders means longitudinal population studies struggle to navigate different application processes and requirements, including what constitutes good practice for consent.
3.3 Longitudinal population studies face specific challenges related to consent, including ensuring its validity over time and managing consent as their participants' transition from childhood to adult life.
3.4 Public perceptions of data sharing affect both participant and administrative data holders' attitude and engagement with linkage efforts.
3.5 Staffing, resource and documentation issues in data holding organisations create practical problems for the linkage and post-linkage processes.
3.6 Collective problem-solving and collaboration between studies, and between studies and data holders, could help overcome technical issues, and challenges like missingness and disclosure risk.

4 Biosamples

4.1 Longitudinal population studies need a long-term storage solution for biosamples that is low cost but secure, and is supported by a stable funding stream.
4.2 A coordinated approach to the analysis of biosamples – favouring a multiplex or omics platform – could avoid unnecessary waste of a precious, exhaustible resource.
4.3 While remote methods of biosample collection are cheaper and reduce respondent burden, more research is needed to understand how the quality and integrity of samples collected remotely compares to conventional methods.

5 Impact & policy engagement

5.1 The longitudinal community needs to broaden its working definition of impact, and crucially give more prominence to scientific impact.
5.2 The longitudinal community can maximise impact by pushing both the evidence and the data to policymakers.
5.3 There are several areas where joint action could push forward progress, including developing more sophisticated methods for capturing impact, developing best practice guidance on impact metrics, and enforcing DOIs for datasets.
5.4 Both the generation and tracking of impact need dedicated staff resources, however reward structures in academic institutions do not always place high value on this work.
5.5 Longitudinal population studies need to develop a better understanding of the policy landscape and the policymaking cycle, as well as guidance on how to cope with the mismatch of timescales between policy processes and research.

5.6 Studies need an ‘access point’ to the policy landscape, such as the Government Office for Science, network of What Works Centres, or government chief scientific advisors.
5.7 Government Areas of Research Interest are potential starting points for discussions with government departments on aligning longitudinal studies to policy needs.
5.8 Policymakers value synthesised evidence across a number of studies, and academic outputs should be accompanied by plain English translations.
5.9 Studies and policymakers require specialist, dedicated resource and training to support better engagement.

About the conference

In October 2017, the Wellcome Trust hosted a meeting to for the UK's largest longitudinal population studies to discuss and agree the criteria for evaluating the quality of these investments. During the course of the discussions, it became clear that there was a need for conference that brought together study teams from across the UK to identify key challenges facing these studies now and in the future. CLOSER offered to take this action forward as part of its mission to bring the longitudinal community together.

In July 2018, CLOSER carried out an online consultation exercise with the UK's longitudinal population studies to identify the key priorities for discussion (see Appendix B for a summary of the findings). On Thursday 6 December 2018, CLOSER hosted a one-day conference entitled *Preparing for the future: Tackling the key challenges facing the UK's longitudinal population studies*, at the Wellcome Trust in London. The conference programme was set according to the responses to the consultation (see Appendix A for the official conference programme). The following areas were chosen for inclusion in the programme:

- integrating new forms of data collection
- future directions in cross-study harmonisation
- improving access to administrative/linkage data
- developing an overarching longitudinal biosample strategy
- demonstrating the impact of longitudinal studies
- improving engagement with policymakers.

Unfortunately, there was not enough room in the agenda to accommodate every topic covered in the consultation and it should not be inferred that these are the only key challenges facing the longitudinal community. Notable omissions include data discoverability and shared metadata infrastructure.

With a strong focus on problem solving, the purpose of the conference was to identify the best ways to tackle these key challenges. Each session included an introductory presentation from the session lead, followed by small group discussions and a final plenary feedback segment. Delegates were encouraged to share their own experiences, as well as help identify potential solutions to the issues raised.

In total, 68 delegates attended, including representatives of 17 longitudinal population studies:

- 1970 British Cohort Study
- Aberdeen Birth Cohorts
- Avon Longitudinal Study of Parents and Children
- Breakthrough Generations Study
- Dementia Platform UK
- English Longitudinal Study of Ageing
- Generation Scotland
- Millennium Cohort Study
- Million Women Study
- MRC National Survey of Health and Development (1946 British birth cohort)
- National Child Development Study (1958 British birth cohort)
- Next Steps
- ONS Longitudinal Study
- Southampton Women's Survey
- Twins Early Development Study
- Twins UK
- Understanding Society

Integrating new forms of data collection

Session leads: Prof Annette Jackle, Understanding Society, University of Essex, Dr Andy Skinner, MRC Integrative Epidemiology Unity, University of Bristol
Report author: Dr Lisa Calderwood, CLOSER

Key learning

- To support innovation in longitudinal population studies, we must share learning on our experiences and collaborate to promote cross-study comparability.
- While new technology brings attractive opportunities for novel measurement, longitudinal population studies also need to carefully consider data quality, longitudinal continuity, and the practical and financial implications they will face.
- We need more research into what participants are willing to do, and how acceptable they find new data collection methods.
- Data infrastructure and researchers' analytical skills will need an upgrade to maximise the utility of new forms of data.
- The costs of innovative data collection methods need to be weighed against the scientific potential of new forms of data they produce – and studies should work together to achieve greater cost efficiency.

Introduction

There are many opportunities for embedding new forms of data for novel measurement into longitudinal population studies. Sensors can measure air quality and noise, and cameras can capture images and video. Innovative apps can be geo-triggered, or serve as diaries or time samples. A growing array of wearables is now available to collect GPS, air quality and movement data.

Some UK longitudinal population studies have already begun using new forms of data collection. Examples include:

1970 British Cohort Study	Accelerometry
ALSPAC	Videos of parent child interactions Wearable air monitors Images of food intake (R&D) Smart Watch to capture smoking (R&D)
Growing Up in Scotland	GPS Accelerometry
Millennium Cohort Study	Accelerometry App time use diary
Understanding Society	Images of till receipts (R&D) App spending diary (R&D)

There are lots of opportunities for new and innovative scientific measurements, for example apps to collect dietary data. However, it is important to consider these new forms of data within the context of the total survey error framework, looking at both errors of measurement and errors of representation. New data collection methods raise questions of measurement validity and reliability, and can have technological challenges in relation to usability, acceptability, implementation and scalability. Problems with coverage and participation can introduce sample bias.

New data and their collection also introduce administrative challenges, for example in data processing and storage, but also budget management, as purchasing devices, developing and maintaining apps, etc. is expensive.

The session chair put the following questions to delegates for discussion:

- What are the most important challenges?
- What are the key priorities for solving challenges?
- What are the most important practical next steps?
- What could longitudinal studies do in collaboration?

Summary of themes arising from discussions

New technology and the challenge of understanding data quality

While there was a keen interest among delegates in the use of new technology to keep longitudinal population studies on the cutting edge of science, it was recognised that there are challenges in relation to data quality, and it is important to consider and understand these.

There is potential for **sample bias** with new technology, as there are concerns about selective participation. Willingness to take part in new forms of data collection may be influenced by participants' demographic characteristics, personality, their technical skills or familiarity with technology, or their trust in technology or the study. The data may be less useful if there was a lack of diversity or representation among participants taking up the new technology.

Measurement biases are also a risk, for example in the accuracy of measures. Step counters are known to over-report activity, and GPS is not completely accurate. Delegates noted that use of multi-method approaches can minimise bias from specific collection methods. Some suggested that due to the richness of existing data in longitudinal population studies, there may be opportunities to correct or adjust for biases introduced through new data collection methods.

Longitudinal population studies also face unique scientific challenges in adopting new technology. Rapid changes in technology could lead to measurement changing at each wave of data collection, leaving little **longitudinal continuity of measurement**. There are also limitations on measurement due to the time periods and duration of data collection waves.

It was also noted that while the opportunities offered through new technology may be novel, they may not always produce **better quality or more useful data** than existing data and methods. Delegates suggested that exploration of new data collection methods should focus on **scientific areas that could benefit most from new measurement**, for example income and expenditure data. In some cases, smaller, more focused studies or sub-studies may be better able to collect some kinds of rich or specialist data. It was also noted that harmonisation across devices was important, and that calibration studies would determine how measures differ.

The delegates also emphasised the **importance of sharing successes as well as failures** when it came to experimenting with new forms of data collection. Collaboration, strong evaluation and sharing learning were seen as critical to the effective adoption of new forms of data collection.

Data processing, access and analysis

Existing infrastructure for archiving and sharing data may not be fit for purpose when dealing with new forms of data. Data collected through new technology may also necessitate a review of **safe access arrangements** and **anonymisation** challenges. **Re-identification risks** need to be assessed for certain types of data, for example those collected through GPS.

New forms of data also bring **new analytical challenges,** and there is a need for researchers to acquire new skills and techniques. Data structure can be very different. Emerging techniques like machine learning may help overcome some of the analytical challenges. Delegates noted that **standardised approaches for data processing and documentation** could also benefit end users.

Securing participant involvement

Delegates wanted to see more research into what participants are willing to share and do, and what the barriers to participation are. Some suggested involving participants as co-designers of the research.

New technology also allows for both active and passive data collection, each of which brings different concerns for participants and study managers. **Active data collection**, such as diaries, gives more control to participants but there can be low take-up. On the other hand, **passive data collection methods**, such as browser usage, are lower burden, but some delegates noted that evidence suggests participants prefer to have control and knowledge of what data they are providing. For those used to questionnaires, they are accustomed to having the choice not to answer specific questions. Passive collection methods give little transparency about what data are collected, so it is critical to find a way to inform them adequately or even let them see their own data. Giving participants feedback may boost return rate, but may also affect behaviour. It is important further investigate the role of feedback, and to determine when it becomes an intervention.

Technology must be **user-friendly** in order to be attractive to participants. Some delegates felt the technology needed to be seamless, for example by being collected through the user's own device or at little to no effort, but others pointed out that participants may prefer not to use their own devices and to have more control/knowledge of the data they are giving. Delegates also wanted to ensure participants understand what information they are being asked to give, and determine how acceptable they find the request, and their ability and willingness to provide informed consent for this, especially for passive collection where they don't actively provide the data.

Finally, the **acceptability of new technology** to different participant groups needs to be considered, for example older people for whom new technology may be source of anxiety, or viewed as a chore. However, there will be others who may very well enjoy taking part in this kind of data collection.

Costs

Longitudinal population studies face an array of practical barriers in adopting new forms of data collection. **Equipment costs** can be preventatively high, and low return rates for sensors and other kit have significant financial implications (as well as implications for sample size and bias). Delegates felt it was important to consider the costs of different data collection methods in light of the quality of data they produce. Some suggested longitudinal population studies could reduce costs through collective purchasing and bargaining, and through partnerships with industry or technology companies.

New forms of data collection also introduce **data processing challenges**, due to the volume of data, the need to calibrate, and difficulties in assessing the quality of data. Dealing with these challenges also increases the cost of data processing.

Resource and training needs

Delegates expressed an overwhelming desire for **more knowledge exchange** between studies. They are keen to know what each other are working on, and to understand the difficulties and costs. They are equally interested in the learning from successful and unsuccessful experiments.

Delegates felt collaboration at an early stage would contribute to a shared understanding of the meaning of new forms of data, and lead to **cross-study comparability of new data** and best practice guidance for data collection. Consortia working could ensure agreement at the outset of approach and thus comparability at the end.

There was also a clear need for **training in new data analysis**. Delegates also suggested sharing code and analytic approaches on open-source platforms.

Some delegates felt there may be benefit to having a **centralised body** investigating and sharing learning on:

- emerging technology
- mapping technological opportunities
- how to calibrate over different devices, measures and over time
- how to quantify bias.

Future directions for cross-study harmonisation

Session chair: Prof Rebecca Hardy, MRC Unit for Lifelong Health and Ageing at UCL
Report authors: Dr Dara O'Neill, CLOSER

Key learning

- Collaboration is essential to achieving prospective harmonisation, particularly when considering new technology and methods of data collection.
- There is a risk of sacrificing data quality in harmonisation, and researchers need to ensure the right balance is struck between precision and scope.
- Good quality documentation is vital.
- Harmonisation is insufficiently resourced and incentivised, despite its important scientific benefits.
- Harmonisation practice could be improved by researching new methods for retrospective harmonisation, developing more techniques for validating harmonisation outputs, and by exploring opportunities offered by calibration and emerging analytic techniques.
- Strategies for overcoming the challenges of data harmonisation should be developed through cross-disciplinary collaboration, and an overall coordinating role could prioritise measures to be harmonised and promote the added utility of harmonised data.

Introduction

Data harmonisation involves recoding or modifying variables so that they are comparable across research studies. It is not only about achieving greater statistical power through a pooling of larger samples, but also about opening up new avenues of cross-study and longitudinal inquiry (e.g. enabling cross-generational comparisons). Harmonised data can offer useful new insights, as is illustrated by two studies that looked at obesity and blood pressure trajectories across the life-course using cross-study harmonised data. However, while harmonisation enhances existing data, but also brings with it many challenges.

Harmonisation can be done retrospectively or prospectively. Retrospective harmonisation requires researchers to grapple with conceptual and data-related differences between data sources. Harmonisation decisions need to be driven by the scientific question to hand, but also on the studies available. This means there is a balancing act between the level of detail needed but also the types of comparisons sought. Where the hurdles faced in retrospective harmonisation are principally data-centred, prospective harmonisation poses particular challenges around inter-study collaboration and coordination. Careful preparation is required in terms of identifying and evaluating the needs and benefits of prospective harmonisation to the individual studies involved and their intended research scope, while ascertaining the practical feasibility and resource requirements of increased cross-study alignment.

Broadly speaking, more work is needed on prospective harmonisation, data depositing and the role of repositories, calibration/comparison studies, and the development of guidance, standardised documentation and meta-data.

The session chair put the following questions to delegates for discussion:

- What are the key challenges for longitudinal population studies in relation to harmonisation?
- What are the key next steps to solve these challenges?

Summary of themes arising from discussions

Divergence within and across studies

Different assessment protocols and instruments across studies and waves of data collection have led to divergence both between and within studies. **Changing technology and data collection methods** can bring greater divergence. For international comparisons, **language and cultural differences** can add to the challenges for harmonisation.

Delegates felt that **prospective harmonisation** should be considered at the beginning of the data collection processes. Decisions should be made collaboratively with other studies, and **standardised and validated measures** (e.g. questions and scales) should be promoted through a **shared question bank**.

Preserving data quality

Many delegates noted that there is a **risk of sacrificing data quality** in the process of making it more comparable.

Researchers often need to decide 'what is good enough' in undertaking harmonisation, in order to ensure the **right balance** is struck between **precision** (detail retained) and **scope** (number of studies/time periods included).

Documentation

Access to adequate and accurate detail both about the original and harmonised variables is necessary in ensuring the valid usage and interpretation of such data. It can also play a key role in promoting and guiding future prospective harmonisation efforts. As such, **good quality documentation** of harmonisation processes and harmonised datasets was seen as a critical need.

In the first instance, delegates felt additional documentation and explanation by study investigators of the **assessment and measurement decisions** made during data collection could help better guide later usage and harmonisation efforts. They also identified a need to document the **changing social and historical context** to inform researchers' understanding of the data from older collection periods.

Following harmonisation, delegates felt it was important to link harmonised data back to the **original study (source) data**, and to maintain **good metadata** that adheres to a recognised documentation standard both for the original study data and any harmonised variables derived from them. In addition, they recommended **related syntax** (data processing code) should always be made publicly available.

Some expressed a need for more guidance on how much documentation is sufficient, and how harmonised data should be made available (e.g. where it should be stored, who the data controller is).

Resource and training needs

Data harmonisation can be a time-intensive and expensive exercise, with associated costs for skilled staff, data access, software licencing, and IT infrastructure (e.g. data safe havens for special licence data). However, this work is currently **insufficiently resourced and incentivised**. Delegates felt the important benefits that such work offers to research and policy is not fully appreciated. There is a need to **convey the benefits of harmonisation more widely** to help achieve additional funding support and incentivisation for this work.

Further **training and support** resources would improve harmonisation practice. **New methods for retrospective harmonisation** should be researched (e.g. federated data analysis technologies, machine learning, etc.), and more **techniques for validating harmonisation outputs** should be investigated and promoted (e.g. calibration and simulation methods to detect and help adjust for measurement error). **Calibration** and **emerging analytic techniques/technologies** offer new opportunities for furthering the quality of harmonisation outputs and the efficacy of their creation.

The challenges that harmonisation brings were seen by delegates as cross-disciplinary, and that **greater collaboration and innovation** in developing strategies for addressing these challenges is needed. In addition, **overall coordination of harmonisation work** can play an important role by helping to prioritise which measures are harmonised and to promote to funders and the research community the added utility that data harmonisation offers.

Improving access to administrative/linkage data

Session lead: Prof Alissa Goodman, UCL Centre for Longitudinal Studies
Report author: Andy Boyd, ALSPAC, Bristol Medical School, University of Bristol

Key learning

- The multiplicity of government data holders means longitudinal population studies struggle to navigate different application processes and requirements, including what constitutes good practice for consent.
- Longitudinal population studies face specific challenges related to consent, including ensuring its validity over time and managing consent as their participants' transition from childhood to adult life.
- Public perceptions of data sharing affect both participant and administrative data holders' attitudes and engagement with linkage efforts.
- Staffing, resource and documentation issues in data holding organisations create practical problems for the linkage and post-linkage processes.
- Collective problem-solving and collaboration between studies, and between studies and data holders, could help overcome technical issues, and challenges like missingness and disclosure risk.

Introduction

Many longitudinal population studies are seeking to undertake ambitious data linkage programmes covering everything from health and mortality records, to earnings and benefits, and criminal convictions. However, progress to achieving these linkages has been mixed. While some have been achieved, others remain stubbornly stuck in negotiations with data controllers.

The landscape for linkage in the UK is also in a state of change. The new Digital Economy Act 2017 (DEA) is introducing a clearer legal gateway for linking to administrative records. The advent of novel linkages to 'digital footprint' data – such as those collected from sensors, transactional records and social media data – introduces new opportunities and new challenges. Finally, the changing nature of research infrastructure and innovations in research design impacts on data linkage strategies (e.g. the use of Data Safe Havens), increasing desire for cross-study and consortium research designs.

The session chair put the following themes to delegates for discussion:

- Managing consent: how to maximise coverage and minimise response bias; how to embed longitudinal population studies into national linkage infrastructures (such as the proposed Administrative Data Spine); how to link longitudinal population studies to locally held records (e.g. social care records, GP records); and, how to further develop linkage to novel digital sources.

- The linkage process: how to ensure stakeholder engagement in the process; how to quality assure the technical linkage process; and, how to ensure transparency in how it is carried out.
- Post-linkage data processing and use: how to clean, document and quality assess linked data; how to derive new variables consistently to ensure harmonisation across studies; and then, how to onwardly share linked records with the research community.

Summary of themes arising from discussions

Consent

Delegates identified that establishing 'consent' for record linkages was a substantial challenge and that approaches were radically different across studies. Some studies only pursue explicit consent-based linkages, while others do not seek consent, but instead offer a means to object. Here a distinction was identified between seeking **consent to establish a legal basis** for 'data linkage', and seeking **consent for ethical reasons** and to help build and retain participant acceptability for this form of data collection. However, both types of consent require communications with participants that inform, engage and build trust, and both require approval from the data owner and ethico-legal regulators in order to establish legitimacy.

The challenges in managing consent related to three core areas:

1) the difficulty in establishing **consensus on what constitutes 'good practice'** consent wording and structure due to the multiplicity of data controllers (both in terms of government departments and across the UK home nations) and their differing requirements
2) the challenge of maintaining the **longitudinal validity of consent wording** when data controllers change their requirements
3) how to ensure that the **'consenting' process is accessible and informative** to all participants, no matter which form it takes.

These challenges are being compounded by increased complexity as longitudinal population studies move to adopt digital-footprint linkage strategies, as well as the **increased social awareness of data use and misuse** resulting from data breaches, data misuse scandals, media attention relating to data use and changes in legislation (particularly the EU General Data Protection Register [GDPR]). The challenge of managing the **transition to GDPR** was seen as particularly interesting given GDPR's general public profile of strengthening citizens' data rights, yet also containing powerful exemptions to some of these rights for research in the public interest.

While many of these challenges are generic to population data science, some have **specific longitudinal dimensions**. In addition to maintaining the perceived legitimacy of consent over time, studies were particularly concerned with managing the transition in consent from child participants to adults; how to manage withdrawal of consent; reconsenting to accommodate changes in study data management or access, and – in relation to 'digital footprints' linkages – the relevance of specific consents (e.g. as participants start to use new social media platforms). Longitudinal population studies also need to ensure the onward-sharing of collected and processed data to users.

Delegates were also interested in methodological research into the **impact of different modes to collect consent**, in terms of response rates, consent rates, and effectiveness for securing participant understanding and trust. It was also identified that research should be conducted to help identify if particular groups are being systematically marginalised from the research as a result of how consents are collected or structured.

The bulk of the discussion related to how to identify and share good practice, as well as insights into successful applications (precedents). Delegates discussed how acting together increased their chances of pressing for answers and changes to data access processes that are more conducive to longitudinal research. They also expressed a need to **map application processes and requirements** across data owners and across the UK home nations. Finally, delegates wanted a better understanding of **public and participant attitudes to data sharing, privacy and trust in government**. In particular, they wanted to understand public views of the level of sensitivity of different data types, and attitudes to underdeveloped linkage opportunities, such as student loan company records and UCAS university admissions records.

The linkage process

Engaging data holders with individual applications also presented challenges. That there is a **multiplicity of holders**, some of whom are known and accessible to the longitudinal community, and some decision makers who are not known and are not accessible (often the legal advisors). In some organisations, there is little clarity as to who to approach and a **lack of a formal 'data linkage' access procedure**. Other organisations have a more visible access procedure, but are **under-resourced** to handle increasing levels of demand and complexity.

Frequent staff changes at the data holders often involves a lengthy process of bringing new staff 'up to speed' with the complex longitudinal population study applications. Developing longitudinal expertise within data holder organisations was seen as important. There can also be a **lack of authority for middle management staff** to facilitate data sharing within the data holder organisation. In these cases, engaging middle managers may not result in successful applications as the decision-making power lies elsewhere.

However, delegates noted that there were higher-level issues that affected engagement with data holders. More effort is needed to **ensure broad acceptance of the research use of routine records** – and the longitudinal community needs to do more to demonstrate the value of longitudinal research and reciprocal benefits arising from linkage. One suggestion was that it could be effective to frame the value of linkage in terms of the particular research interests of the data holder organisation. Others suggested that participants could be key voices when dealing with data holders (an example of this is the Aberdeen Children of the Nineteen Fifties). Funders and ministers were also seen as key influencers, and some suggested identifying and developing 'linkage' and 'longitudinal champions' within data owner organisations would be helpful.

Delegates noted that data holders' perceptions of public attitudes to data sharing affected their engagement with longitudinal population studies. One suggestion was that studies could **involve data holders in the design of engagement strategies** to ensure their concerns were addressed, and to offer clarification on how their actions and responsibilities were viewed.

Delegates raised **transparency of data linkage processes** as an area for improvement. They recommended transparency through the whole of the data linkage pipeline, including which citizens are included in the data (and which are systematically under-represented), how the data are collected and processed, and the data linkage process itself including algorithm development. Transparency would help a wide range of stakeholders, and could develop public trust and support longitudinal population studies' data science strategies.

A wide range of **specific technical issues** were also raised:

- challenges arising from the dynamic nature of routine records across time, including dynamic population coverage; inconsistent collection of data and data collection protocols; and, versioning and archiving so historical data are not lost
- the range (and potential impact) of the use of different software in the data linkage pipeline
- risks relating to disclosure of participant identities, but uncertainty over how to mitigate these risks without disproportionately impacting on the potential for accurate and informative analysis
- linkage error may introduce systematic bias
- linkage error is compounded by the lack of a universal unique citizen ID, or the lack of a persistent ID linking the major ID number systems
- lack of technical expertise across providers and a lack of resources to enhance linkage infrastructure (NHS Digital's linkage development programme, which involves academic experts, including those from the longitudinal community, was noted as good practice and an encouraging development)
- longitudinal population studies need to work with their host institutions to implement core information security and infrastructure requirements (e.g. secure servers, departmental information governance certification to standards such as the NHS Data Security & Protection toolkit).

Delegates felt technical challenges could also be overcome through collaboration and coordinated efforts to engage data owners on key issues. A regular **forum for key data linkage stakeholders** could lead to a more efficient resolution of problems, innovation across the system, and progress on more complex problems.

It was recognised that some system changes that would improve linkage accuracy (e.g. requiring individuals to prove their identity before receiving services) would be highly politically and socially contentious, and the focus for improvements should lie in different solutions.

Post-linkage data processing and data use

Delegates discussed the issues relating to processing extracted linked data into a research ready format. A key issue, particularly for the UK-wide studies, relates to **challenges integrating data extracted from different systems and to different standards** across England and the devolved nations. Where change has occurred over time, across systems and across devolved nations and authorities, administrative datasets have become fragmented, making post-linkage processing more complex. Different legislation and regulation – or different interpretation of these – across different jurisdictions mean governance requirements differ, which can also impact on post-linkage processing.

There is a **lack of detailed and rigorous metadata** for administrative datasets (particularly historical records), and gaining access to institutional understanding of the data is difficult. Delegates wanted to incentivise data holders to provide metadata and expertise on data interpretation. They felt post-linkage routines for longitudinal population studies should be done in consultation with the producer. More 'data resource profiles' should be produced through academic/data owner partnerships to provide detailed and readily accessible information on datasets.

Each study needs to correct for **missingness** in data, and more could be done to share understanding, code and best practice. However it is difficult to understand fundamental issues relating to the linkage sample denominator, representation and the possibility of systematic missingness.

Delegates noted that mechanisms for the **onward sharing** of linked longitudinal data were evolving, and there was a need to agree a standardised framework for onward sharing. There was also a growing awareness of **disclosure risk**, but no clear strategy for addressing this. However, delegates noted that this is challenging in itself as the risk is predominately context specific and initiatives such as the UK Anonymisation Framework have already developed a framework for controlling risk with this in mind. Once again, delegates called for more **transparency of post-linkage processing** and emphasised the need to engage participants and research users in the development of processing models.

There were several suggestions relating to **secure research environments**. Delegates felt secure environments needed enhancing to enable cross-study assessments and to enable 'omics based investigations. They also noted that some secure services require researchers to visit physical locations, and there is a need to move to remote secure access or more access points to fixed access secure servers. Some also called for a 'research passport' system that allows trained and legitimate researchers to demonstrate their credentials, skills and legitimacy in an efficient and consistent manner.

Developing an overarching longitudinal biosample strategy

Lead: Prof Nish Chaturvedi, MRC Unit for Lifelong Health and Ageing at UCL
Report author: Dr Natasha Wood, CLOSER

Key learning

- Longitudinal population studies need a long-term storage solution for biosamples that is low cost but secure, and is supported by a stable funding stream.

- A coordinated approach to the analysis of biosamples – favouring a multiplex or omics platform – could avoid unnecessary waste of a precious, exhaustible resource.

- While remote methods of biosample collection are cheaper and reduce respondent burden, more research is needed to understand how the quality and integrity of samples collected remotely compares to conventional methods.

Introduction

Longitudinal studies collecting biosamples face many of the same challenges, and there is value in studies working together to develop an overarching strategy. It was noted that this is rapidly developing scientific field, and one of the challenges relates to keeping up to date with emerging opportunities. This session covered some of the issues in relation to collection, storage and analysis of biosamples. Other important topics in this area include governance, ethics and participant feedback.

The session chair put the following questions to delegates for discussion:

- Should there be a common set of analytes which are collected across cohorts? This would enable cross-cohort analysis and reduce costs.
- In regards to sample preparation methods, is it possible to streamline processing and storage?
- How should funding for analysis be granted? Should it be for a subsample? For a whole cohort? Or cross-cohort?
- How can we future-proof sample storage and management?
- Collecting whole blood is the gold standard, but what about other sampling strategies? For example, blood spot/dried urine in post are cheaper and easier to collect from participants.
- Are there any other factors longitudinal population studies should be considering?
- What are the potential solutions?

Summary of themes arising from discussions

Sample storage

There was much discussion surrounding the burden of storing samples. Some samples are kept for long periods of time, which is costly. In terms of how long samples are stored, delegates felt there was a tension between waiting for the best panel of novel or existing analytes, and coping with the costs of storage and risk of sample loss.

As many samples are stored for long periods of time, it can be problematic to find adequate storage space, which is in high demand, and there are also problems of maintaining freezers. One solution proposed was to have a **cross-study storage centre** in one location (this would have to be in a location with low land prices). However, central biorepositories already exist and many are not used due to the high costs and challenge to meet annual costs from unreliable grant funding streams. Instead, many studies come up with their own solutions by storing freezers onsite. While cheap, the disadvantages are that such arrangements are unstable, and liable to lead to sample loss due to power outages and freezer breakdown. A co-ordinated strategy across cohorts with funders to devise an optimal, low cost yet secure strategy is needed.

Inefficient sample use

Historically, sample aliquots have been analysed just one or a few analytes at a time, which wastes a precious exhaustible resource.

Delegates suggested a **coordinated approach to the analysis of samples** across cohorts, favouring multiplex or omic platforms – with a dialogue with funders as to how researchers apply for such costs.

Delegates also discussed a Scottish initiative, whereby 'left over' sample analysed by hospital labs from patient testing is stored for future use. This would be of value to certain patient groups (for example, those with certain conditions or pregnant women). However, there are issues concerning consent and maybe quality issues with the blood.

Remote methods of sample collection

Remote methods of sample collection were discussed by many of the groups. Remote methods of sample collection comprise collecting dried blood spots, dried urine or collecting saliva. Such samples can be collected independently in respondent's own home and posted by respondents to the lab, therefore not requiring a clinical visit or a nurse visit to the respondent's home. Using this form of remote method of sample collection is appealing as it is **cost-effective** and potentially **reduces respondent burden**. However, a number of challenges were raised, including challenges with the quality of the samples collected, stability of measures, and problems with transport via the postal system. It is a **cost/quality trade-off** and research is needed to compare the quality and integrity of these remote methods of sample collection, with the conventional forms of sample collection.

Some recommended a **standardised protocol for sample preparation** across studies, which would be particularly valuable as many national studies employ a home or GP visit approach to sample collection. This would ensure sample quality and cover how samples were collected and how they are stored.

Demonstrating the impact of longitudinal studies

Session lead: Prof Anna Vignoles, University of Cambridge and ESRC
Report author: Meghan Rainsberry, CLOSER

Key learning

- The impact of longitudinal studies has been considerable over the decades, and has the potential to contribute even more to social, economic and health policy development in the coming years.
- The longitudinal community needs to broaden its working definition of impact, and crucially, give more prominence to scientific impact.
- We need to push both the evidence and the data to policymakers in order to maximise impact.
- There are several areas where joint action could push forward progress, including developing more sophisticated methods for capturing impact, developing best practice guidance on impact metrics, and enforcing DOIs for datasets.
- Both the generation and tracking of impact need dedicated staff resources, however reward structures in academic institutions do not always place high value on this work.

Introduction

There is much to be gained from showcasing the scientific, economic and societal impact of the UK's longitudinal population studies. Not only do impact examples help make the case for continued funding, but they can also help retain participants and win public support for the use of data in research.

Longitudinal population studies have had considerable impact. The ESRC's recent Longitudinal Studies Review showcased some important examples:

- Millennium Cohort Study (MCS) has shown that breastfed babies face fewer health risks than those not breastfed at all, evidence that changed NHS guidance to new parents.
- Understanding Society and MCS have informing Department for Work and Pensions policy and interventions on worklessness by showing that having a parent out of work is detrimental for children's development.
- Longitudinal population studies have collectively shown that mothers who go back to work are not disadvantaging their children, evidence that changed common perceptions and informed government welfare to work policies.

Despite these examples, we still have insufficient information on the impact the studies have achieved. Demonstrating impact remains a flawed, misunderstood and under-resourced process. Several factors contribute to this challenge. First, funders' definitions of impact are too narrow and prescriptive, meaning that a lot of impact 'doesn't count'.

Second, the research community lacks effective methods for tracking and recording impact. Finally, there is increasing recognition that efforts to generate impact need to go beyond dissemination of findings.

The session chair put the following questions to delegates for discussion:

- How can we better demonstrate impact from our longitudinal population studies?
- How can we record all uses of the data and hence better track its impact? Does technology offer opportunities (example: Mills and Rahal, *Communications Biology*)?
- Is it enough to record outputs or do we need to dedicate more resource to showcasing impact?
- Do we need to 'push the data' to policymakers?

Summary of themes arising from discussions

Definitions of impact

There was a consensus among delegates that funders' definitions of impact were naïve and reflected a poor understanding of both research and policy development processes.

In particular, delegates felt that there was little understanding that good policy development relies on a body of evidence, covering multiple papers and studies. Current definitions of impact used, for example by the REF, suggest an unrealistic linear relationship from single papers or analyses to government policy interventions.

Some felt that the longitudinal population studies should be more confident in how they express their own impact, and promote their role in informing on broad social problems as well as any specific policy solutions they may offer (however, many felt the latter is not the role of researchers, but of policymakers).

Delegates were keen to use their collective influence to improve definitions of impact used by different funders.

Scientific impact

While much of the debate around impact focuses on impact outside academia, delegates emphasised the enormous scientific impact of the UK's longitudinal population studies.

Delegates felt it was important for funders to recognise that some research is better for policy use than others, and that potential policy relevance is not the only measure of a study's importance or scientific value. For example, it was felt that exploratory research should not need to explain its impact, as it does not translate into policy. Negative findings were also stressed as intrinsically valuable without it being important for policy impact.

Longitudinal population studies also have scientific impact through the scientific resources they produce, and the role they play in supporting the career development and progression of scientists across a range of disciplines.

Methods for generating impact

Much of the small group discussions focused on methods and challenges in generating impact outside of academia.

Following on from the morning session on engaging with policymakers, there was a clear desire among delegates to explore **better ways of responding to calls for evidence** from policymakers, recognising that this should be seen as highly strategic work. The preference was to put forward systematically synthesised evidence that illustrates the breadth and multidisciplinary nature of the studies, however this would need to be balanced against the need to respond rapidly to calls and answer the specific questions asked. While delegates were enthusiastic about the potential of responding to government and parliamentary calls for evidence, they did not feel they had a reliable notification system in place to know which calls were open and relevant to their work. Finally, it was noted that policy engagement requires a good understanding of the policy audience and the appropriate translation and presentation of findings to suit their needs.

Many delegates supported the idea of **'pushing the data' to policymakers** – that is, promoting the opportunities within the data to answer policy questions, in addition to putting forward the existing longitudinal literature on the issues. Resourcing this type of responsive analysis would generate findings that spoke more directly to policy needs. Holding **trials within cohorts** was also seen as a potentially effective means for studies to generate concrete impact, by evaluating specific interventions and rolling them into practice. Delegates recognised there was potential to **use government Areas of Research Interest (ARIs) to align work to policy priorities**, but there was uncertainty about how to do this, and low levels of resource to do any additional analysis required.

Relationships were seen as critical to generating impact. Some studies were using formal **impact partnerships** with organisations outside academia to co-design research, building in the potential for impact right at the beginning. Others noted that learned societies (e.g. SSM, IGS, third sector organisations) could provide a channel to policymakers. One group suggested that there was a potential to have impact with business, but building relationships and interest was challenging.

Delegates mentioned that there can be a conflict between properly qualifying findings and offering clear policy and practice messages or recommendations. It was also mentioned that distinguishing between associations and causation is critical, as is expressing the value and relevance of associations for policy and practice consideration. Finally, the mismatch of timelines for longitudinal data collection and analysis and policy development was noted as a barrier. It can often take years to collect, clean, document, deposit and analyse data, whereas policy development moves at a much faster pace.

Impact metrics, tracking methods and reporting requirements

Delegates recognised the need to track emerging impact from the data, to publications, to policy, but that there was little guidance on how to track or what metrics to use.

All studies rely on users to tell them how they are using the data in order to begin tracking impact. The challenges in tracking data use differ depending on the access model of the study. Studies that require users to apply and/or pay for the data are able to enforce **reporting requirements for users**. In open access models, users agree to the terms of an end user licence, which includes reporting requirements, however these are much more difficult to enforce. These studies can track, for instance, the number of downloads from the UK Data Service, but they miss a great deal of information on ultimate use and resulting publications. One archive in Canada requires users to complete an annual return before they can use it again, which includes questions on publications. CLOSER colleagues reported they are currently working with the British Library on improving **data citation and implementing DOIs for datasets**.

Delegates suggested more could be gleaned from **grant applications**. There should be a requirement that researchers name their data sources on applications, and it would be useful to determine the total monetary value of successful applications based on longitudinal data resources.

Delegates reported more success in **capturing reach** than impact. For example, web analytics and Altmetrics offered through journals provided a good insight into the reach of the findings, but that this was not classed as impact by funders' definitions. Some delegates explained press coverage and reach were hard enough on their own to achieve, and to have this visibility count for very little is frustrating.

Researchfish, a reporting system used by the Research Councils, was widely seen as not an appropriate mechanism for tracking impact. Complaints included that the reporting categories did not surface scientific impact effectively, and that it is not kept up to date beyond a certain period after the grant despite the impact of the grant carrying on. It was also noted that a huge amount of effort was going into impact reporting for the **REF**, and that it is important to make as much use of this content as possible for other purposes.

Studies who did have resources dedicated to investigating impact are primarily using more rudimentary, desk research methods rather than **making use of new technology**. Delegates were interested in the potential use of scraping or text mining methods to gather evidence of impact, and any tools that could automate the collection of evidence.

Resource and training needs

Delegates reported that in the majority of cases, tracking impact is ad hoc, not staffed by dedicated roles, and often falls to younger researchers. Several larger studies did dedicate resource to impact tracking, however staffing structures differed dramatically – from a requirement on all teams to produce metrics, to dedicated impact posts working across the whole of the organisation.

Several groups noted that policy outreach activity required expertise, time and thus dedicated staff with skills in science communication and public affairs. Some also expressed support for more 'rapid response' units who could conduct analyses in direct response to policymakers' questions.

Overall, there was a disagreement on how to deliver and monitor impact. Smaller studies welcomed the idea of policy engagement and similar outreach being coordinated centrally across studies, as many of them lacked the funding and resource to achieve the impact they would like. They valued the role CLOSER currently plays in this area, and suggested a body like CLOSER could continue to provide this support in the future. Larger studies, however, felt this activity should be carried out by the investments themselves as the evidence needed to be translated by someone closer to the science to communicate messages correctly.

There was consensus among delegates that academic reward structures do not pay sufficient attention to impact activities, which discourages researchers from taking on these responsibilities. Almost all delegates thought more training was needed for researchers to engage in impact activity, and that more knowledge sharing opportunities on what works in tracking impact would be valuable. However, some urged caution in placing impact responsibilities on academics who are busy enough managing the studies and carrying out research.

Improving engagement with policymakers

Session lead: Dr Grant Hill-Cawthorne, Parliamentary Office of Science and Technology
Report authors: Rob Davies, CLOSER

Key learning

- Longitudinal population studies need to develop a better understanding the policy landscape and the policymaking cycle, as well as guidance on how to cope with the mismatch of timescales between policy processes and research.
- Studies need an 'access point' to the policy landscape, such as the Government Office for Science, the network of What Works Centres, or government chief scientific advisors.
- Government Areas of Research Interest are potential starting points for discussions with government departments on aligning longitudinal population studies to policy needs.
- Policymakers value synthesised evidence across a number of studies, and academic outputs should be accompanied by plain English translations.
- Studies and policymakers require specialist, dedicated resource and training to support better engagement.

Introduction

A critical part of the impact process is engaging with policymakers. This could include a wide range of stakeholders, but this session mainly considered UK Parliament (which holds the government to account and makes laws), and the UK Government (which runs government departments and public services and is accountable to Parliament). The differences between the two is important when considering how to engage.

The session chair gave an overview of Parliament's engagement with research. There are a number of bodies in the UK Parliament that use academic research, including Select Committees, Libraries, Public Bill Committees, MPs' and Peers' researchers, the Parliamentary Office of Science and Technology (POST), and All-Party Parliamentary Groups (APPGs). The newly formed Knowledge Exchange Unit supports the exchange of information and expertise between researchers and parliamentarians, and the UK Parliament website hosts a dedicated page with information, guidance and advice on how to engage Parliament as a researcher.

When engaging with Government, the Government Areas of Research Interest (ARIs), which set out the most important research questions facing each department, are key documents for researchers to consider.

A number of issues need to be factored in when engaging with policymakers. These have been articulated and explored extensively in academic literature and include differences in speed of production and language, access to research (in particular to journals), and the ability of policy makers to select and appraise research. POST has investigated how research evidence is conceived and used within the UK Parliament and also produced a briefing note providing an overview of research impact in legislatures.

The session chair put the following questions to delegates for discussion:

- What are the opportunities and routes for translating longitudinal studies into policy?
- Imagine talking to a policymaker – what is your hook?
- How can you fit your research into the policymaking cycle?
- How can big picture findings become actions?
- How could you interact with the Trials Advice Panel, Chief Scientific Advisors network, etc.?
- What needs to be on the policy dashboard?

Summary of themes arising from discussions

Policy development and research cycles

Scientific evidence is only part of the policy decision-making process. Delegates recognised the importance of **understanding the policymaking cycle**, and in particular how, when and who to engage. Unfortunately, there is often a **mismatch of timelines** between policy processes and research. Delegates reported being unclear about when researchers should report their findings, particularly if only preliminary findings are available at a timely stage in the development of policy.

Funders' requirement that **impact statements** be written at the application stage did not reflect the reality of either research or policy development cycles. Instead, delegates suggested that impact statements would be more usefully collected as part of a mid-term review of a grant, when preliminary findings were available and could be more realistically and appropriately lined up with policy development.

Communication between policy and scientific communities

The policy landscape is complex and fragmented with no central access point. Some delegates felt the **Government Office for Science** could be encouraged to simplify the process. The **What Works Centres** were identified as a valuable network with which researchers could engage, or contacting members of the **House of Lords** directly, many of whom are scientists. The suggestion was floated that a **government chief scientific adviser** (CSA) could sit on the CLOSER Executive Committee to ensure a two-way flow of information between the policy and longitudinal communities. On the researchers' side, there was a suggestion that a **portal or dashboard of research projects and researchers** for parliament to contact would be helpful, in particular if it harmonised university-based portals.

Some regional studies are able to attract the attention and interest of their local MPs by providing information about their constituencies, however the **lack of constituency-level data** in the national cohort studies was noted as a challenge. Others suggested highlighting the **spending/financial implications** or gains to get the attention of policymakers.

It was noted that policymakers value **synthesised evidence** across a number of studies, and academic outputs should be accompanied by **plain English translations** of findings and ethics.

Others suggested using **mainstream and social media** and routes to policymakers and the public. However, some delegates reported challenges of 'over-claiming' when working with their university press offices. There was also a suggestion that using study participants to engage with Parliament may be more powerful, however this raises a range of ethical issues.

Ensuring the relevance of longitudinal population studies to policy in the long-term

Co-designing future study sweeps with policymakers was offered as a potential option for ensuring they address specific, long-term policy questions, as was using common question sets across studies to enable direct comparisons.

Targeting **civil servants** rather than MPs for longer-term issues may be more fruitful, but the Civil Service churn makes it difficult for studies to establish long-term relationships. The importance of ARIs for informing future research was recognised, and delegates noted they were a potential starting point for discussions with government departments on how **existing longitudinal data could be used to answer new policy questions** rather than collecting new data. Once again, the value of **linking survey data with administrative or other government data** was noted.

Resource and training needs

Impact and policy influence demand resources and time, and delegates recognised a need for more **training** (both for researchers and policymakers) and a dedicated, specialist resource to support UKRI longitudinal population studies in their policy engagement activities. As local university policy departments are quite small, some delegates suggested a **central, shared resource** across UKRI longitudinal population studies could be established to either engage with policymakers on behalf of studies, or provide studies with guidance and advice on how to do this themselves.

Delegates also noted that the **incentive structure** for academics to engage with policymakers is weak. Citations in White Papers are not as valued as journal publications for academic advancement, and there is no institutional support or reward system in place that incentivises 'impact' work.

Proposed actions

The following section summarises the suggestions given by delegates for how to tackle the key challenges discussed during the course of the conference.

New forms of data collection

1. Cross-study collaboration on new methods of data collection, in order to investigate:
    - emerging technology
    - mapping technological opportunities
    - how to calibrate over different devices, measures and over time
    - how to quantify bias
    - standardised data processing arrangements and documentation for new forms of data.
2. Exploration of new data collection methods should focus on scientific areas that could benefit most from new measurement, for example income and expenditure data.
3. Support innovation among longitudinal population studies by encouraging and facilitating experimentation with new data collection methods.
4. Establish a forum for sharing learning on new forms of data collection, including failures as well as successes, and exploratory testing that might otherwise go unpublished.
5. Fund research into how acceptable participants find new data collection methods.
6. Review suitability of existing data infrastructure for new forms of data, taking into account safe access arrangements, anonymisation challenges and re-identification risks. Fund upgrades where necessary.
7. Prepare to fund new forms of data from collection through to analysis and storage, including equipment costs, increased costs of processing, development of new analytical techniques, and upgrading of infrastructure. Encourage studies to collaborate to achieve efficiencies.
8. Give careful attention to the quality, utility and scientific potential of data collected through new technology in the context of the potential biases, practical challenges, and financial implications it introduces.
9. Develop partnerships with industry or technology companies, or coordinate collective bargaining and purchasing of equipment, to reduce the costs of new methods of data collection.
10. Make use of new analytical techniques, including machine learning, for data collected through new methods.
11. Develop training in methods for analysing new forms of data.

Data harmonisation

12. Consider prospective harmonisation when planning new data collections. Make decisions about data collection methods in collaboration with other studies, and opt for standardised, validated measures wherever possible and appropriate.
13. Establish an interdisciplinary collaborative group to investigate:
    - providing further support for efforts to coordinate harmonisation work and promote its value
    - developing strategies for overcoming harmonisation challenges
    - promoting use and continued development of standards for harmonised data documentation, metadata, access arrangements and sharing syntax
    - working with academic institutions and funders to incentivise researchers to carry out harmonisation work
    - developing a central record of the changing social and historical context in which longitudinal data has been collected.
14. Fund research into new methods for retrospective harmonisation, techniques for validating harmonisation outputs, and calibration and emerging analytic techniques.

Data linkage

15. Build sustainable and transparent relationships between longitudinal population studies and data controllers
16. Champion issues, giving longitudinal research a voice in decision making processes, and lobbying for a more conducive environment for linkage
17. Develop a portfolio of strong research examples, preferably with policy impact and highlighting research governance good practice, that illustrate the benefits arising from data linkage in longitudinal population studies
18. Investigate the potential for a universal unique citizen ID number or a systematic and persistent linkage infrastructure.
19. Establish transparent and accurate costs for enabling data linkage strategies, and develop a sustainable funding structure.

Biosamples

20. Investigate the feasibility of a long-term investment in a cross-study storage facility for biosamples that is low cost but secure.
21. Develop a standardised protocol for biosample preparation across studies, to ensure sample quality and cover how samples are collected and stored.
22. Commission research into the quality and integrity of biosamples collected remotely compared to conventional methods.

Impact and policy engagement

23. Establish an interdisciplinary working group made up of scientific and professional experts to investigate:
    - more sophisticated methods for capturing impact (including investigations of new technology)
    - best practice guidance on impact metrics
    - promotion and enforcement of DOIs for datasets
    - promotion the impact of longitudinal population studies.
24. Develop a more appropriate definition of impact, in particular one that is broader and gives due credit for scientific impact.
25. Move the requirement for impact statements from the initial grant application to the mid-term review, when preliminary findings are available and can be more realistically aligned to policy development.
26. Work with academic institutions to establish reward structures that support impact and policy engagement work.
27. Develop a policy engagement training programme for longitudinal population studies and their users.
28. Develop specialist support unit(s) for policy engagement, to be either centralised across UKRI longitudinal investments, or based within individual studies or smaller groups of studies.
29. Build impact relationships with organisations outside academia and involve them in the design and dissemination of research.
30. Involve policymakers in the design of future waves of data collection to ensure studies can address specific, long-term policy needs.
31. Develop a longitudinal evidence training programme for Parliamentary staff and civil servants.
32. Promote the opportunities within the data to answer policy questions, and align work to Government Areas of Research Interest.
33. Fund ‘rapid response’ units to analyse existing longitudinal data in response to current policy questions, to be either centralised across UKRI longitudinal investments, or based within individual studies or smaller groups of studies.
34. Develop a notification system to alert longitudinal population studies to relevant calls for evidence.
35. Provide systematically synthesised evidence across studies in response to calls for evidence from government or Parliament, ensuring findings are translated into plain English.

Appendix A: Conference programme

Preparing for the future: tackling the key challenges facing the UK's longitudinal population studies

Date & time: Thursday 6 December 2018 / 10:00 – 16:30
Venue: Wellcome Trust, Gibbs Building, 215 Euston Road, London, NW1 2BE

Programme

10:00 Registration and refreshments

10:30 **Welcome to the conference**
(Prof Alison Park, CLOSER)
Main conference area

10:35 **Funders perspectives on future priorities for longitudinal population studies**
(Gavin Malloch, MRC, Erica Pufall, Wellcome Trust, and Bridget Taylor, ESRC)
Main conference area

10:45 **Improving access to administrative/linkage data**
(Prof Alissa Goodman, UCL Centre for Longitudinal Studies)
Main conference area

11:45 Short refreshment break

11:55 **Improving engagement with policymakers**
(Dr Grant Hill-Cawthorne, Parliamentary Office of Science and Technology)
Main conference area

12:55 Lunch

13:55 Parallel sessions

Integrating new forms of data collection
(Prof Annette Jäckle, Understanding Society, University of Essex, and Dr Andy Skinner, MRC Integrative Epidemiology Unit, University of Bristol)
Main conference area

Future directions in cross-study harmonisation
(Prof Rebecca Hardy, MRC Unit for Lifelong Health and Ageing at UCL)
Breakout area

14:55 Short refreshment break

15:05 Parallel sessions

Demonstrating the impact of longitudinal studies *(Prof Anna Vignoles, University of Cambridge)*
Main conference area

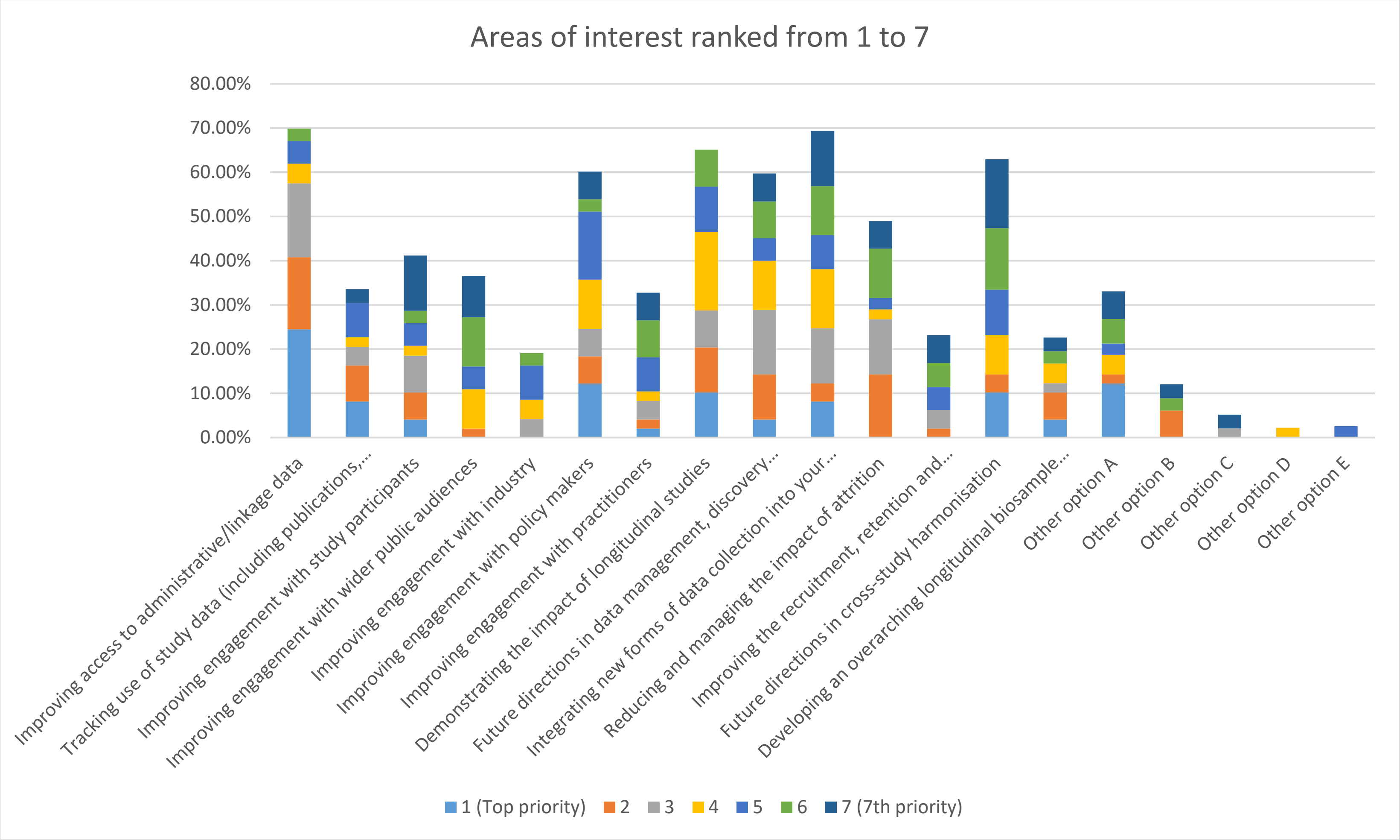
Developing an overarching longitudinal biosamples strategy *(Prof Nish Chaturvedi, MRC Unit for Lifelong Health and Ageing at UCL)*
Breakout area

16:05 **Feedback and final reflections**

16:30 Conference ends

Appendix B: Needs analysis survey findings

1. We want to know what you, the longitudinal study community, would find most useful to cover on the day. We have identified a number of areas we think might be of interest. Please choose up to seven and rank them in priority order.



2. If you selected one or more of the ‘Other option’ choices above, please provide more details about the area you are interested in us covering.

‘Other’ options ranked as a **top** priority:

- Exploring the potential for 'offshoot' studies among sub-groups
- Engaging fathers including those with part-time overnight care of child cohort members
- Providing long-term continuing funding for UK longitudinal studies
- Developing novel and efficient approaches to study design, e.g. building on local studies such as BIB and learning lessons from Life Study
- Just having a proper data dictionary that covers all waves, and all CLOSER datasets, would be a good start. It is astonishing that there is not one.
- Improving the documentation of data

‘Other’ options ranked as a **second** priority:

- Use of biosamples - coordinated proposals for generation of omics data
- Study design/ questions reflecting increased family diversity including separated families
- Secure lab access does not provide powerful enough computing resources, or the ability to import programs which need compilation. This severely restricts analysis of what are complicated data
- Improving the documentation of data

‘Other’ options ranked as a **third** priority:

- Improving the documentation of data

‘Other’ options ranked as a **fourth** priority:

- Knowledge management and documentation
- funding sources
- Improving the documentation of data

‘Other’ options ranked as a **fifth** priority:

- Negotiating the challenges of using healthcare data
- Improving the documentation of data

‘Other’ options ranked as a **sixth** priority:

- Managing response bias
- Ensuring that large enough budgets are available for the amount/complexity of data collection required to make it viable for data collection agencies to continue to bid
- The challenges associated with regulations for conducting research in the NHS and how this impacts specifically on longitudinal research

‘Other’ options ranked as a **seventh** priority:

- Increasing consent rates for data linkage
- Big Data storage for large Omic datasets
- Sampling
- what to do when regulations change during a longitudinal study that impacts on retention strategies
- how to obtain funding to maintain a cohort

3. Please tell us here if there is anything else you would like to see at the event.

	Comments:
1	Survey companies with proposals for new developments in collection methods
2	how to maximise the impact of longitudinal studies
3	Data documentation
4	gaining easy access to the datasets
5	Update on potential for new ways to collect data. e.g. using apps, wearable devices and so on
6	Options for live demo of software not just talks.
7	Communication strategies between scholars working on similar topics across studies -- Sharing code, ideas, plans
8	International participants
9	Linkage to primary care medical records (CPRD)
10	Statistics - causal inference
11	Senior UKRI presence.
12	More information on joining ongoing or forming new collaborative partnerships in which longitudinal data is being used.
13	Value of feedback for survey participants - do they want more / less, etc.
14	Networking, sharing of data management/analysis tips, stuff on structural equation modelling, cross-cohort analyses
15	International LCS working (post Brexit)
16	facilitating future collaborations between participants
17	Latest findings from cohorts that have impact for policy