HDRuk
Health Data Research UK

Annex 4

Population Research UK – PRUK Green paper consultation response summary

Part of the PRUK scoping programme

Overview

A Green Paper setting out proposals for PRUK (Annex 3) with published for feedback from the community in July 2021. This document summarises the responses collected from three structured exercises.

The outputs of these exercises have informed the PRUK Prospectus delivered from the scoping programme.

4a: PRUK longitudinal population studies leads stakeholder consultation

Overview

Health Data Research UK (HDR UK) has been commissioned by the ESRC, MRC and Wellcome to scope a new initiative, Population Research UK (PRUK), with the aim to maximise the use of longitudinal population studies (LPS) in the UK. After initial consultation, a Green Paper was developed and published, which outlines the key recommendations for PRUK and further questions for consultation[1]. Feedback and further consultation on the Green Paper was sought in July 2021 through two parallel stakeholder workshops. The workshops were attended by 14 leaders in LPS, who were invited due to their role in the leadership and delivery of one or more significant UK-based LPS.

The aims of the consultation were to:

1) discuss and give feedback on functions and activities for PRUK as set out in the Green Paper;

2) understand barriers and facilitators of how individual LPS could participate in a new initiative and;

3) inform the scope and recommendations for the development and commissioning of PRUK.

This report summarises some key themes and topics arising from the two workshops. Stakeholder attendees were broadly positive about the potential of PRUK and the value that the initiative could add to LPS, whilst noting areas for development, clarification and refinement. The experience and expertise from the attendees on LPS provided useful and constructive feedback to inform the proposed aims, activities, and recommendations for PRUK.

The agenda was structured around three main topics: 1) Purpose, ambition and deliverability of PRUK; 2) Data discoverability, access and linkage and; 3) Engaging stakeholders, capacity building, and community. A list of workshop attendees and agenda can be found in Appendix 1 (page 8).

---

[1] https://www.hdruk.ac.uk/population-research-uk/

Theme 1: Purpose, ambition, and deliverability

At the start of each workshop, participants were given an opportunity to share initial thoughts, reflections, and any questions they had about the proposed functions of PRUK based on the Green Paper, which had been shared in advance with them.

Other resources and infrastructures: A key discussion point arose around how PRUK aimed to add value beyond or integrate with other initiatives such as CLOSER and UK Data Service (UKDS). It was evident that members of the community have familiarity with the aims and activities of some, but not all, LPS resources and data infrastructures; this reflects different cultures, approaches, levels of knowledge in different domains as well as different funder requirements.

Role and vision of PRUK: Participants suggested that there needed to be clarification of the main role of PRUK; was it to assist and enable users of LPS in their current work or to encourage making studies more discoverable and well-known? One participant suggested there was a mismatch between what is stated in the mission and what is proposed in the recommendations, e.g. how the recommendations were coupled to advancing scientific opportunity. Others agreed that there may be a tension between the ambition and deliverability with the ambition to cover all studies and all disciplines which can present several challenges.

Activities: There were some suggestions for additional areas for inclusion are not covered in the Green Paper that warrant consideration. These included:

- the involvement of census-based population studies;
- being able to retain the data and knowledge of studies that become dormant;
- facilitating access and analysis of biological samples for additional testing

International focus: It was suggested that the development of PRUK could benefit from existing large international models aiming to link studies and their data[2]. This may be particularly pertinent as international users may engage with PRUK and this would fit in with the longer-term vision of exploring where international partnerships can be built or expanded.

Incentivising collaboration with PRUK: Participants discussed that there were substantial scientific benefits in working together to address common issues (e.g. harmonisation, ethics, linkage, access), although this also surfaced challenges (distribution of resources, agreements and consent with participants, harmonisation across LPS at the expense of harmonisation within studies), as had been demonstrated through the rapid cross-study collaborations that resulted from the Covid-19 pandemic as an example. Participants suggested the benefits of PRUK for individual studies could be clearer in the paper, which might differ between studies based on their size, resources and existing partnerships and collaborations with resources (CLOSER, UKDS etc).

---

[2] For example EU Child Cohort network https://lifecycle-project.eu/

Theme 2: Data discoverability, access and linkage

Data discoverability: Discussion of the recommendations regarding data discoverability highlighted that the most ideal way for studies to be discovered would be through a single portal / catalogue. However, this was envisaged as a portal/website to other repositories which hold metadata and not represent a considerable additional burden to studies in preparing new metadata for an additional platform. There was a suggestion that UKDS and CLOSER could be extended to include studies that are not currently engaged in them. A motivation for doing this would be to reduce the burden on preparing different sets of meta-data or data for different platforms. For metadata, 5his could be achieved by creating a federated meta-data searching tool where a layer is added on top of what already exists, allowing for a single point of discovery across all LPS.

Data Access: Both workshops discussed the recommendation to access LPS data via multiple repositories/Trusted Research Environments (TREs) using a common application process. Although many studies have set up their own access systems, there may be a role for PRUK to provide guidance and standards for some new studies which are setting up data-sharing processes proactively. Differing access policies for each study and/or funder, including specific obligations made to participants, presents a challenge for the creation of a single data access portal. Furthermore, some studies may have specific requirements related to data governance and consent which would limit their capability to adopt new data sharing practices. For example, for some studies often-longstanding consent agreements with their participants set out how data may be shared. For these challenges to be overcome, there would need to be alignment across funders in addition to resources to consult and engage with study participants about new approaches to data sharing.

User and study support for discovery and access: There was discussion about the constraints on some studies to be able to participate in data sharing activities; in some cases it was suggested that there is currently not enough funding available for data managers and research assistants needed to do the required work (e.g. to prepare standardised meta-data). The resource required can also depend on the kind of data which is needed to be shared which requires highly specialised data managers. For example, LPS gather data using questionnaires, blood samples (including potentially identifiable genetic data), medical imaging, and wearable technology data which all have varying levels of complexity in terms of data formats and requirements relating to governance, consent and access. Provision of the resources to standardise all these data types, across all studies would be helpful – although mechanisms for prioritization based on scientific need and impact would be required.

In addition to facilitating data discovery and access, it is also important to provide other information and support to help ensure that data are used appropriately and effectively. For example, there could be a resource to help LPS users access complete information about the studies and search for which studies could answer a specific question. It was suggested that methodological support should not include harmonisation of data without specific scientific purposes because this can quickly become outdated, and LPS users often wish to conduct their own harmonisation specific to their research question.

Data linkage: Within the workshops there were less discussions regarding the data linkage recommendation contained in the PRUK Green paper. Across the two workshops, there was broad support for studies working together for data linkage, although it was noted that this may create a single point of failure. An important consideration is that PRUK is not seen to prohibit innovative study-led data linkage programmes.

A key challenge identified by some participants was the lack of agreement on what constitutes 'consent' for data linkage. Different studies have requested consent for linkage in different ways and some are uncertain if they have consent from participants for linkage with third party data. Additionally, there are historical studies which did not request consent for data linkage, meaning that those who have passed away or have been lost to follow-up cannot be included in any data linkage requests.

Theme 3: Engaging stakeholders, capacity building, and community

Stakeholder engagement: The participants highlighted the individuality of studies which is a strength, and there is no single solution for stakeholder engagement. For example, studies may have more local, national or international aims, studies often have their own dialogue with participants, third parties, funders or academic partners; and studies with less funding, often smaller studies, cannot engage easily with participants due to funding limitations. Therefore, PRUK may be able to support in some areas, although individual studies would need (and want) to maintain strong networks. There were several areas where it was suggested PRU might play a helpful role:

- PRUK could provide some centralised resources on how to engage with study participants which could be adapted by each study. For example, PRUK could help with common issues in LPS. Resources could include strategies to minimise attrition, to ensure findings remain representative of the population of interest and sharing good practices on how to recruit hard to reach groups. It is likely that some studies already have these tools and PRUK could help co-ordinate shared learning.
- The Covid-19 pandemic has led to not only great strides being made in study timelines (e.g. ethics approval, data access and linkage), but also in the awareness of the public about the power of data to solve real-world problems. PRUK could capitalise on this and play a role in improving public understanding of LPS research including making the case for the benefits of the research and of data linkage.
- PRUK could facilitate collaborative research projects by helping LPS users navigate the remits and limitations of proposed projects and understand how to work together across research councils. Furthermore, PRUK could work directly with relevant stakeholders to facilitate research. For example, if policy makers require more research evidence in a specific area, this could be requested via PRUK who could then help coordinate this. Similarly, PRUK could help with facilitating the communication of policy-relevant research to policy makers.

Capacity building in data management: Participants raised many points with regard for supporting skills development and capacity within studies. These included metadata preparation, data cleaning, curation, governance and ethics, and open science practices.

These required skills are often built-up in house without the recognition of formal job titles such as data manager or research software engineer. Retention and progression in these roles has traditionally been a challenge, and the levels of funding held by studies dictates the level of resources that can be dedicated to them. Furthermore, the contribution of data management work is not well captured, measured or appreciated in career achievements. Although it is now possible to submit curated data sets to the Research Excellence Framework (REF) alongside and in parity with research articles, this is not widely practised. It was suggested that PRUK could play a role in training or career accreditation to help retain staff in these important roles or that PRUK could offer training to individuals on relevant topics such as data cleaning, curation of meta-data, documentation and writing reproducible code. One practical suggestion was that PRUK might seek partnership with organisations such as the UK Reproducibility Network to provide 'train the trainer' approaches to open science practices tailored towards LPS research.

Community building: It was recognised that CLOSER networks have been very effective for technical members of staff, particularly for knowledge exchange of best practices and assistance with diverse problem solving.

PRUK could contribute to building a community of LPS users. For example, one suggestion was for PRUK to organise knowledge exchange events or bids where users can suggest ideas and conduct their own events on topics needed by the LPS community. When building this community, it is important to acknowledge and embrace the individuality of the different studies rather than over-specify and implement a one-size-fits-all approach. Diversity can lead to creativity and help people inspire each other and find novel solutions to problems.

Implications for PRUK Green paper recommendations

The informative perspectives provided through the workshops are used to inform the remit of PRUK and next steps for the initiative. The learning is recorded through the narrative above, and in addition HDR UK have noted the following points.

- Develop further clarity on how different LPS users, disciplines and stakeholder groups (e.g. users and studies) would interact with PRUK and how they would benefit.
- Considering how the recommendations and activities of PRUK can be more strongly connected to scientific opportunities. For instance, the potential of exemplar scientific projects that support prioritisation and development of infrastructure.
- Consider greater specificity in the recommendations that encourage collaboration with pre-existing resources. The current Green Paper was intentionally not prescriptive in defining the other resources that PRUK should collaborate with and work alongside; this made the role that PRUK has in the landscape difficult for the workshop participants to understand.
- There will be significant challenges in the implementation of data access recommendations due to the unique position of each LPS. Recommendations for streamlining access need to recognise current

complexity and be realistic in their deliverables, timeframes, resources required and impact that can be made.

- Capacity building and career support of study and data management roles should have greater prominence to the in final recommendations.
- Provide greater clarity on a proposed role of PRUK in management of non-traditional data (e.g. tissue samples).
- Greater recognition of the individuality of LPS studies is needed. PRUK may not need to provide a single umbrella solution for discoverability, access, linkage, stakeholder engagement, etc. PRUK could consider a second approach of providing standardised resources, leading solutions on common challenges, sharing best practices, improving public knowledge of the benefits of LPS, etc.

Appendix A: Workshop participants

Workshop 1

Naomi Allen, University of Oxford, UK Biobank

Louise Arsenault, King's College London, Environmental Risk (E-Risk) Longitudinal Twin Study

Jeremy Auerbach, Queen's University Belfast, Northern Ireland Census Longitudinal Study

Charlotte Clark, St Georges, University of London

Ian Shuttleworth, Queen's University Belfast, Northern Ireland Census Longitudinal Study

Oliver Duke-Williams, UCL, ONS Census Longitudinal Study

Robin Flaig, University of Edinburgh, Generation Scotland

Nicholas Timpson, University of Bristol, Avon Longitudinal Study of Parents and Children (ALSPAC)

Workshop 2

Janis Baird, University of Southampton, Southampton Women's Study

Michaela Benzeval, University of Essex, Understanding Society

Olly Butters, University of Liverpool, Children Growing Up in Liverpool (C-GULL)

Lisa Calderwood, UCL, Centre for Longitudinal Studies

Nish Chaturvedi, UCL, MRC National Survey of Health and Development Cohort /1946 Birth Cohort (NSHD/ 1946BC)

Jack Kneeshaw, University of Essex, Understanding Society

Observers:

Catherine Moody, MRC

Laura Morrell, ESRC

Bridget Taylor, ESRC

Facilitators:

James Pickett, Health Data Research UK

Jo Blodgett, Kohlrabi Consulting

Aradhna Kaushal, Kohlrabi Consulting

Workshop agenda

The overarching question for the session is 'What are the barriers, facilitators, incentives or resources that would support and sustain the engagement of studies with proposed activities for Population Research UK?'

Time	Topic	Linked consultation questions
1:30	**Welcome and introductions**	
	Presentation – Overview of population Research UK Green Paper *Participant questions and clarifications*	Q10: What activities proposed PRUK would be most beneficial to you or your organisation? Q11: What would be the barriers to you, or your organisation, being engaged with PRUK? How could they be overcome? Q12: Are there any further priorities that you consider important to advancing the LPS field that are not covered in the recommendations listed?
2:00	**Data discovery, access and linkage**	Q1: What would be most helpful to facilitate data discoverability across the full range of LPS and how could this be achieved? Q2: In your opinion, is there a need for new TREs that hold LPS data or is the need to increase the coverage of LPS in current environments? Q3: If you have experience in provisioning data to data platforms, what are the resource requirements for doing so? What are the barriers to doing so, and how may they be overcome? Q4: What further actions could be taken to facilitate the process of creating and using linked data?
2:30	**Break**	
2:40	**Engaging stakeholders, leadership and governance**	Q5: With an aim to build capacity, how can PRUK support skills development and progression of roles in data curation, management, and linkage? Q6: In what other areas could strategic working between LPS be extended? Q7: Does the work proposed for PRUK in this area provide distinct value over the engagement and involvement of LPS engaging with their participants directly? Q8: Are there other roles that PRUK should play in building connectedness across and beyond LPS? Q9: How might a leadership and governance structure be created for PRUK that facilitates a collective approach across the community to advance the recommendations set out?
3:10	**Additional discussion**	
3:30	**CLOSE**	

4b) Population Research UK Green paper online consultation

Overview

The PRUK Green Paper[3] was published on 13 July, alongside an online survey was launched promoted to a mailing list of individuals who had signed up for notifications about PRUK. In addition, the paper was promoted on social media (twitter and Linked In). The survey was open for one month.

The survey asked specific questions related to the Green Paper Recommendations, in addition to free text questions about the perceived benefits and barriers for PRUK. The feedback, in conjunction with other consultation responses received, has informed the updated PRUK Green Paper and recommendations to the funders on delivery.

20 responses were received in total. 18 respondents agreed to unattributed publication of their responses.

	No. of responses
Academic	6
Study team	5
LPS data resource	4
Industry	2
Member of the public	2
Funder	1

Implications and conclusions

High quality and actionable feedback were received from respondents, which came from a range of disciplines and sectors. Consultees provided thoughtful suggestions regarding the structure and delivery of PRUK. There were both supportive and refuting comments towards a centralized model (for instance, it could provide strong representation to the collective views/needs of studies, but on the other hand a centralised, standardised approach might reduce quality and long-term draw resources from studies). A theme was the need for how resources would be made available to studies to support participation and related the need for capacity building in data management and aligned incentives

---

[3] PRUK_Green-paper.pdf (hdruk.ac.uk)

What would be most helpful to facilitate data discoverability across the full range of LPS and how could this be achieved?

Answered: 18   Skipped: 0

#	RESPONSES
1	A common landing page for all Longitudinal Studies, linking their study pages, building on the CLOSER website and hosted by a statutory body. Standards for data discovery methods and tools to enhance the discovery potential of data sources are required for research, teaching and practice. Wellcome has funded studies e.g. to enable data discovery for archival sources. The learning from these needs to be implemented and scaled up.
2	Center on individual and first step to make decision
3	Focus on creating greater awareness and visibility of the UK TREs that hold LPS data and assets. It can be achieved working with the UK TREs to ensure they have a consistent approach to data discoverability, and to avoid LPS having to do redundant work when submitting data and metadata and data to each TRE.
4	Studies being supported and funded to improve own meta-data documentation or being closelyinvolved is important, as they know own data the best. Improved topic and question searching. Important that this does not re-invent the wheel. Many studies have this already on own searchtools or are producing this. CLOSER Discovery was funded to do this. Some data services do this. Not at all sure that another centralised effort starting from scratch again is a good idea.
5	We would welcome standardised cost models of access to datasets, as there is no standard and some are prohibitive. In addition, cohort-specific data may need a different model of access and associated costs. For example, TwinsUK Data Access Committee (DAC) also offer a statistical support package with data access to support complex twin modelling. This may need to be done in a cohort-specific environment. We would like to see the creation of a register listing approved data access requests, including researcher requestor details. This will not only offer transparency to the public and participants but also encourage collaborative research.
6	Ability to generate common data fields across different LPSs might be limited to basic data – e.g. demographic data. Suspect this requires live seminars (face-to-face or remote) with presentations by PRUK staff and LPS representatives.

7	A1: There are already multiple discovery platforms available, HDRUK Gateway, Atlas, BC Platforms and DPUK. We should be focusing on raising the profile of the LPS community within those platforms rather than creating a new one. This combined with the effective application of the TRE principle by the Longitudinal Linkage Collaboration signals the way forward for PRUK to adopt.
8	Metadata - harmonisation to enable cross-study comparison and joined-up discovery - more of it available and open for improvements on discovery platforms that rely on it. Discovery platform join-up; integration with 'look-see' analysis tools for non-sensitive data; active marketing of resources, advocacy and training. Data - Clarity on rights to access and reuse; costs to access -prepare (if appropriate); clarity on derived data and its use to enrich the originating resource
9	Not the biggest issue - see below.
10	Data discoverability implies a reasonable question or area to investigate and an appropriate analytical approach to the diversity of studies that might be captured within LPS. This requires a very careful entering into any such 'discoverability' which goes well beyond basic variables, but also ensures that those 'using' the data know what the provenance of the data are, and what the steps are to maximise the value of any analyses based on the data they have 'discovered'. Alignment of thematic areas and drawing on experience from the many others who have engaged in such exercises will be really vital. Investing in the cohorts themselves where there is no core funding to develop well annotated, curated data that are meaningful might be vital.
11	Catalog of studies and data/samples they have available, phenotypes must be matched/aligned to an ontology. Plus aligned access processes
12	adequate use of IT and all the advantages it brings
13	Implementing standards of metadata across all LPS> allowing searchability by variable type across all LPS.
14	Single coherent approach, not just PRUK but UKLLC/NHS, with proper funding and capacity provided to studies. Current system demands data discoverability and sharing but fails to address difficulties.
15	A common data model based around OMOP perhaps - wide search capability - Like Cohort Discovery tool on the HDR Gateway Innovation Portal
16	This is already being done by CLOSER and UKDS. Rather than re-invent the wheel, CLOSER should be resourced to expand their role in this area.
17	We note that there are distinct dimensions to 'discoverability' and that these can relate to awareness of studies/infrastructure, the contents of the collection, gaining insight into what that content means (detailed metadata) and also discoverability of research tools (e.g. syntax, code lists) and that the issue of 'discovery' also relates to data citation and curation. 'Discovery' is distinct from, but should interact with, data application mechanisms. Relating to Green Paper R1, we consider it is unlikely that any one resource will be optimal to address all these aspects of discovery and recommend against capital-intensive approaches to build such a resource. Our key observation is that the overwhelming priority is for the LPS community to agree standards for approaching this (e.g. whether data citation should be facilitated through the use of DOIs) and that our preferences is that existing discovery tools (HDR UK Innovation Gateway, HDR UK CALIBER Phenotype Library, CLOSER Discovery, Co-Connect, GitHub) should be utilised and resource allocated to enable studies to contribute to these (e.g. structuring metadata into a standardised format such as DDI) rather than investing in more newinfrastructure in this area. The tools provided should be required to ensure that their infrastructure/capabilities can be integrated and shared across studies/other infrastructure in a flexible way (e.g. that a 'shopping list' of variables created in a metadata catalogue can beexported to a study or data provision infrastructure).

Q2: What impact (positive or negative) do you think the implementation of common access processes will have for you or your organisation? Please explain your view.

Answered: 16   Skipped: 2

#	RESPONSES
1	A single gateway that directs you to the appropriate place for each of the UK countries (and to the authorities and specific experts for contact in other countries) would be helpful.
2	Decision making, analysis for better solutions, enhance for general well-being; detection, prevention, risk assessment, discover of new patterns, find correlation and solution within societal issues
3	Positive! At the moment each LPS have to do their own legwork on how to manage data sharing processes, application forms, approval criteria, data sharing agreements, ethics, application forms, licences, anonymisation, GDPR issues, researcher accreditation, etc , which is very time consuming and potentially inconsistent across different studies, even if they use the same data sharing methods. Having a central go-to area for solid guidance on how to manage data sharing processes would be very helpful for our LPS and the TREs that hold/share our data.
4	Data from our studies is already widely available via UKDS and other repositories. I don't see the need for common access processes across studies - though some studies for sure need to improve access to their data and this may drive them to improve this. Would also not want the common access to drive down standards of access and this is a risk. This would be challenging to implement, and unclear how this could be enforceable other than via funders, who already have existing requirements. It would likely create additional work for studies. Different studies also have contstraints on this relating to how study members were recruited/consented, many years ago often.
5	In principal, a common access process harmonised across LPS could be beneficial especially with common datasets that are commonly covered by each cohort, or in cases where all LPS have a common theme e.g. Covid impact. However, this would be very difficult to achieve in practice as LPS have not only unique data questions and data collections pertaining to their participants but also have each made different promises to their cohort participants around access and data sharing with external researchers. Each LPS currently sets their own costs for data access, and so this aspect would also need further consideration. Data access requests from an LPS often lead to requests for associated sample access, further longitudinal questionnaire collections, OMICs data that need analytical support or potentially identifiable data that may need linking to phenotypic or genetic data. Therefore a request for data access to centralised datasets will still require additional study-based actions and associated costs, and this needs to be considered further by PRUK. The Data Access Committees (DACs) for

each cohort have different remits, skill/experience levels and goals. Processes may be relatively easy to harmonise but the requirements in the application forms may be a barrier. What we would like to see from PRUK is a central triage service where genuine/bona fide researchers and valid data access request were filtered before reaching each DAC for approval of the request.

6	Re 2: Each LPS has distinct operating practices which relate to assurances made to participants and local governance rules. Any common process for applying for access will need to be flexible so that it can capture information necessary for local study review or for infrastructure review with study level checks (such as UK LLC). A positive though is that over time some functionality could be addressed centrally (e.g. due diligence checks on a researcher and their institution) and that a central portal could be more efficient, particularly for cross-cohort applications. Negatives include the fact that a central portal will not integrate with study application management systems, that some centralised standard approaches (e.g. adopting the Five Safes approaches) may not be compatible with local governance requirements, and that care will need to be taken so a central portal reflects the needs of our interdisciplinary community. We believe that a common application process (Green Paper R2) is desirable but will always require some element of distributed study control in order to retain participant trust. Whilst feasible (UK LLC and DPUK provide examples) this would require substantial resource and goodwill from studies.
7	Will no doubt be useful – but would need some "hands-on" practice at accessing data and how to link between the existing different LPSs. Probably needs on-going technical support during projects.
8	A2: Common access processes is less important or realistic for established cohorts governed by their original consents and governance. Moving towards processes that ease the burden on both cohort owners and users is however important and laudable. Establishing and upholding principles for data access and sharing for new investments in LPS is however desirable and achievable. A common application for LPS is a more achievable goal.
9	not directly applicable for me in my primary role at the British Library, but writing here as an individual, an a member of the CLOSER Exec Team
10	It will help many early and mid career researchers, but is no substitute for deep understanding of what the data mean within cohorts. This needs to be a requirement of any access to more than one dataset. A common approach will be helpful, but without the support for the cohorts and the investigators that provide support and meaning to analyses and framing of questions the analyses become disembodied.
11	Currently the overhead of accessing LPS is a major barrier to us, with a few exceptions (UK Biobank). Streamlined processes would mean we used , and potentially invested in, these studies much more
12	The average patient does not really know tha all of this is going on. Most would think it to be a good move so long as none of the information is sold o to those who wold profit from it financially rather than practically.
13	Will require trust from study PIs, if they are to 'lose control' over use of their participant data. Existing relationships between study and participants are strong and crucial for continued data collection. Common access to a TRE may well work butin the hsort term it is a huge culture shift.
14	Common access processes would be very good and common consent processes
15	With the exception of certain data sets, we already have this with the UKDS. LP studies that have not yet made their data available through this service should be required (and resourced) to do so.

For LPS teams, what actions and resources would be most helpfulto increase access to study datasets you manage?

Answered: 12   Skipped: 6

#	RESPONSES
1	I don't currently manage relevant datasets but permanent infrastructure funding for datapreparation, custodianship and repositories & archives with grant funding to assist with inclusion of specific studies would be helpful. Funded training and time for teaching researchers, undergraduate and postgraduate students about longitudinal populations studies is essential.
2	NA
3	Enable dialogue among the UK TREs that hold LPS and ensure they: - implement a consistent approach in terms of researcher accreditation and application forms. - provision on guidance for Data Access Committees set set up at LPS level - represent the LPS metadata in a consistent way and enable variable-level selection
4	Access is not a problem as data is available. Improved services via UKDS especially for secure access, and more streamlined TREs so we don't have to duplicate and share data in multiple platforms, or at least it is easier/less work to do this. Another TRE to add the already crowded field doesn't seem necessary. Also unclear that bespoke infrastructure for longitudinal studies specifically is needed, better to use infrastructure for the wider community.
5	At TwinsUK, we have extensive genotyping, genetic and multiomics data which we are keen to share with researchers. These data sets are currently difficult to share or add to cross-LPS collections however due to the storage space required and the lack of appropriate infrastructure. We therefore would like to see infrastructure set up to handle and enable access to our vast genotyping, genetic and multiomics data, which has the potential to greatly enrich health research. TwinsUK is set up as a biobank, and so we also hold hundreds of thousands of biological samples, including blood, stool, urine and saliva. This unique resource could enable researchers to study specific diseases for example in conjunction with environmental or socio-economic data from other LPS, and so we would like to see PRUK boost discoverability and awareness of biological samples as well as data. Lastly, each LPS, including TwinsUK, needs a fully-funded core data team who are intimately familiar with their data collections. This is essential to curate and share our data collections successfully. It would not be suitable or possible for a centralised PRUK data team to fully understand and manage every single dataset available from the diverse LPS.
6	Re 3: We refer back to the points made about intrinsic capacity issues relating to recruitment and retention of staff with appropriate skills and the need for financial support to curate legacy datasets to contemporary data and metadata standards and formats. Re 4: Following on from #2 above, we suggest there may be value in scoping the potential for a standardised study application management software system which interfaces with a central application portal, user/project register (Green Paper R2) and metadata catalogues. We note the complexity of existing systems and how these integrate into other study functions and we remain concerned that harmonising functionality may lead to a lowest common denominator set of functions and that this could mean in practice studies are operating parallel systems which increase rather than reduce burden. We also are concerned that such a software would be expensive to develop, have a high risk of failure, high maintenance costs and may also stifle innovation. That said, the volume of applications across the sector suggest the interface between a central portal and study systems may generate substantial burden and cost if not automated and adopting a central application system would likely require such a system or at least a standardised output from the system. We recommend that at the very least HDR Gateway prioritise being able to provide their application materials in a machine-readable annotated format (e.g. JSON).
7	N/a
8	A3: Means and resources need to be made available to LPS cohorts to ease the provision of

searchable summary data by potential users. Funders do not tend to fund this core infrastructure for the LPS so that current funding doesn't cover this work, but instead only funds expansions and additions to studies. The cohort-specific and meta-data display platforms must be highly visible and easy to access for research planning purposes. In general, individual cohorts lack the internal resources to guide would-be users through all of the steps necessary to optimise study design and data provision. UKB aside, few cohorts alone have sufficient statistical power to address PRUK questions. Scoping across cohorts will become the norm. UKB however lacks the granularity of LPS cohorts and the agility to respond quickly to new PRUK priorities.

9	See CLOSER strategy on this. I would also suggest federated TREs that are not just HEI based to enable a broader user engagement with different public, policy and professional audiences
10	Taking the concerns from cohorts in terms of meaning seriously, and addressing these properly within the planning will be important. It is vital, if LPS work is to be useful to society, that the data are approached with SDGs in mind, and how evidence can be generated that add value to the way in which we can tackle these. Without thinking about diversity, disadvantage, community, locality this cannot be done. Data meta-analyses to date that MRC etc have invested in have been based mostly on an individual approach. To be fit for the future this will have to change to something that is more about collectives and how communities and localities experience health and ill-health.
11	The single biggest issue for us is NHS governance of access to linked data. This cannot be solved by PRUK alone- we need a single proper public debate on data use.
12	Global data federated with LPS would be very helpful - global reach out to similar initiatives

What further actions could be taken to facilitate the process of creating and using linked data?

Answered: 16   Skipped: 2

#	RESPONSES
1	Fund and staff the expert teams with responsibility for data linkage that are located in statutory bodies e.g. ONS, NRS and those in the NHS with responsibility for preparing and linking health and care data, a statutory body with responsbility for standards and linkage of local authority and education data. Greater availablity of training and research in datalinkage methods and in the quality assessment and analysis of linked data.
2	Standard of collects, lists of useful name of parameters for common use, templates of datasets
3	• Liaise closely with the government departments that provide data (mainly NHS Digital, DfE) toensure they provide data readily and at reasonable cost.
4	Supporting existing infrastructures - especially ADRUK - to facilitate linkage to LPS. I don't feel another body trying to do this will be at all helpful.
5	One joined-up system or protocol from start to finish that is recognised nation-wide. To link to health records for at present for example, studies must (amongst many other processes): - Apply for approval at local NHS R&D - Apply for ethics approval from NHS REC or HRA CAG, which may take a few rounds of revisions. - Apply to NHS Digital, who must give their own approval, irrespective of NHS REC/HRA CAG's approval, and so which may also take a few rounds of revisions, potentially requiring amendments to NHS REC/HRA CAG. - NHS Digital then passes the application onto IGARD for their approval. These processes are very slow and require a huge bureaucratic burden to complete. One, joined-up system which streamlines these processes and does not require multiple applications and approvals would greatly facilitate the process of accessing and using linked data. Creating a central Data Safe Haven (rather than localised ones for each LPS) would provide more security as well as cost efficiencies for confidential data supplied by NHS for data linkage. The UKLLC has shown this to be true and workable for linking COVID related data of several LPS datasets and linked datafrom NHS Digital. In facilitating future design and harmonisation of collections PRUK should consider using skills within individual LPS to trial pilot studies. TwinsUK have shown themselves to be highly agile cohort and in a leading space for prototyping innovative science and data collections which have then been rolled out across other LPS.
6	The LHW NCS has taken the unprecedented step of creating the UK LLC as an interdisciplinary, pan-UK linkage infrastructure and centralised Trusted Research Environment for longitudinal research. Uniquely, it provides the basis for highly efficient and predictable linkage across studies and across different data domains. The UK LLC has been implemented for COVID-19 research, but designed to be scalable to include any UK cohort and be generalised to any research topics: the UK LLC design is agnostic to the methods used by its researchers. The UK LLC has achieved its minimal viable product and is now an operational infrastructure containing data from 15 studies with 9 further studies currently being integrated. It has developed bespoke automated pipelines to health records and – in the spirit of NCS – has a 'Team Data Science' approach where users will ensure reproduceable research through the centralised curation of documented research tools (e.g. programming code, code lists, documentation forming components of a GitHub Library and training/tools resource). We propose that the UK LLC forms the basis of a linkage solution within PRUK and that this forms part of the legacy of the LHW NCS and takes on the roles envisaged in Green Paper R3 in a

centralised manner with appropriate mechanisms to distribute data to studies. We note that the development of the UK LLC has taken considerable resource, and that the ongoing operation of the resource will include substantial fixed infrastructure lease costs and data charges. There is therefore financial efficiency from centralising these costs as they have high fixed rates and low increased costs from scaling to include additional studies or participants. In reference to the above, we recommend that the UK LLC ensures it is effectively aligned to wider EHR infrastructure (for example, through adopting common standards and approaches for supporting analysts, replicating definitions and tools within each TRE, ensuring common access requirements). For this UK LLC model to be successful, it will be necessary to resource studies to provide data into the UK LLC and review applications/manage participant permissions. We note that many longitudinal analysts will be unfamiliar with using linked records (or the differences across the range of linked data sources) and PRUK should aid capacity building in this area. This should align with developing capacity in Data Science ways of working.

7	Build awareness of what is already possible and what might be possible in the future. Need clarity on the degree to which data is linked at the individual level and the reliability of the linkage (i.e. likely proportion of mis-matched records).
8	A4: Follow the leaders e.g. LLC, OpenSAFELY, EAVE II, CVD-COVID UK. Support the adoption of common standards and metadata methodology. Engage with data sources to ensure that administrative data is research-ready, documented, and affordable. Support the capacity to update data dictionaries cohort by cohort.
9	Robust, long-term linkages that are not reliant on individuals in orgs and govt departments being in post to maintain. Andy Boyd's work and that of the ADR point towards good practice in this domain
10	It may do more harm than good to facilitate these processes without simultaneously increasing researchers' and policy-makers' awareness of why this kind of research is more complex than it looks - see response to last question below.
11	Mass linkage for studies that are assessed for their value would be extremely helpful but again it is vital that the nature, strengths and weaknesses of such data and the sensitivity of the linkage is addressed in a deep manner, not the superficial approaches seen recently (rather polarised - either 'don't worry your little head about it' or 'dangerous scientists running away with your personal information)'). A careful conversation and dialogue with the public, which is not seen as driven by pharma/big data or the politicians and scientists aligned to these, but reflective of the concerns in the diverse communities of the UK in a mature manner recognising the validity of concerns and sharing the enormous value of publicly orientated research for benefit of communities.
12	standardised phenotypes, linkage to EHR etc
13	Building trustworthiness with public about use of data. Robust transferable data sharing agreements that can be adopted by any / all LPS.
14	Get the NHS on board. I strongly feel that the issues of linkage, and of access to linked data, are not specific to PRUK and will only be addressed when we have an informed debate about use of all health related data. Must also recognise that some LPS rely on linked data for outcome measures, and not as an optional extra.
15	The ability to access cohort data which has been searched in a cohort discovery or builder tool which would allow external ML/AI tools to analyse the data

Q5: With an aim to build capacity, howcan PRUK support skills development and progression of roles in data curation, management and linkage?

#	RESPONSES
1	Collaborative funding of posts in national and local public health institutes and departments, the national statistical authorities, as a requirement for funding research networks, consortia and by funding capacity building for skills development within programme and research centre grants as MRC has done.
2	update information about training, induction with experts, links for specific articles directly related to health data, software used, lists of skills needed and where.
3	• - Gather and publish information about training courses, conferences, etc organised by TREs (e.g. UKDS have excellent resources) and established data management organisations (e.g. DCC, RDA, UKAN). - Explore the creation of a data management professional qualification.
4	CLOSER communities of practice already is doing this function well. There is a fundamental problem with reward structures in universities which PRUK cannot change. Key thing is for these activities to be properly funded within studies. And for infrastructure/resource investments to be linked to science and led by science to make more attractive for researchers e.g. National Core Studies.
5	Skilled professionals in data curation, management and linkage are desperately needed. We would like to see PRUK fund a full ladder of data-related jobs for professionals, right from entry-level posts to senior managers. This way, we could develop and retain highly-trained staff over many years. We feel this approach of creating a full career ladder with progression will be more effective in the long-term and better provide the LPS community with the skilled professionals needed than one-off training schemes or fellowships. At a minimum, we would expect PRUK to provide a set of core competencies, job packs and training schemes for data professionals within LPS. We feel it is essential for each LPS to have their own core data teamwho are intimately familiar with the data collections in their LPS. It would not be appropriate or indeed possible to have a centralised data team able to adequately oversee the curation, management and linkage of all datasets collected across our diverse LPS.
6	Re 5: Offering training fellowships, designed to address or help address a specific question, would help to boost the skill base and develop the interoperable vision set out in Green Paper R4. The most challenging aspect is that data science is collaborative across multiple disciplines, funders remain focussed though on a single leader which is in opposition to the need for emphasis on interdisciplinary 'team data science' approaches. Fellowships (and equivalent schemes) should have requirements to deposit learning as training guides and research tools (e.g. programming scripts) as documented re-useable products in addition to metrics to acknowledge and reward this activity. PRUK should scope the existing options for a library of such products which can support the LPS community. Re 5: PRUK could take the role of a central hub that contributes to wider – funder led – initiatives to develop standard measures for recognising contributions, outputs and impact for data curation, management, linkage and Data Science ways of working. The LPS community – via PRUK – can provide case studies for this thinking and also represents a wide range of other areas of essential roles

and responsibilities, e.g. PPIE, security and governance. However, this is a sector wide challenge, will face challenges (e.g. what would be a metric of success for PPIE? And how could this capture work which tackles structural issues facing LPS such as lack of engagement in harder to reach communities rather than to-hand metrics, such as the number of likes on social media, resulting from interactions with engaged participants. We recommend it is tackled at a sector wide level and that this work is prioritised. Re 5: The CLOSER learning hub (https://www.closer.ac.uk/events-training/closer-learning-hub/) provides a model for developing training material content which could be replicated/enhanced in PRUK. However, this would need to be enhanced with professional training and specific skills training. This could be formal training tools or contributions from expert speakers via the themed networks (e.g. the LAND network bringing together longitudinal analysts using geospatial data). This could be enhanced with new bespoke content or by signposting to relevant content elsewhere (e.g. Turing Institute, UK Reproducibility Network https://www.ukrn.org/).

7	Establish and deliver training seminars for key staff fulfilling these roles. This looks like a considerable amount of expertise that would be built over time - i.e. probably beyond the general researcher familiar with other research methods. Would it be possible for key staff to follow a recognised career pathway supported by training at different levels?
8	A5: Provide MSc, PhD and early-career training through a network of Centres of Excellence (see A10) with special consideration given to those wishing to move from one skill area to another, e.g. epidemiological to/from informatics. Secure funding for data management positions, with the possibility of career progression to encourage the attraction and retention of staff. Avoid over-centralisation as that would risk losing the depth and breadth of local LPS knowledge, expertise and agile response that only a distributed network can provide, adding value to known and newly emerging PRUK challenges.
9	Within the LPS environment, there are a few exiting providers that you will know well, so I won't list them here. However, I think there is more to be done here on career paths and recognition cf scientific leads. Mid career fellowships are a good idea, as well as ECR training.
10	Work more closely with the NIHR Academy and its Methodology Incubator.
11	Map the pipeline opportunities but also this does not include the very real concerns about understanding population data provenance - time, geography, culture, history etc and there really ought to be a clear strategy for capacity building and training in these areas too, including ethical training.
12	Money! Centralised structure for funding or even a central structure that collaborates/deploys people to individual studies
13	Training of relevant staff, the acquisition of really good IT experts, proper and adequate overseeing by PPIE people
14	Funding.
15	Common Data Model would be good - more focus on the search rather than curation
16	Perhaps a dedicated UKRI-funded CDT on this topic.

Q6: In what other areas could strategic working between LPS be extended?

Answered: 16   Skipped: 2

#	RESPONSES
1	Build on the existing developments in CLOSER, build a shared doctoral school affiliated to all participating institutions and a community of practice that engages studies, users and participants. Host four country and international cross collaborations to consider future questions - similar to a Lancet commission -and fund the infratructure required to sustain them and to publicise the joint working that already happens.
2	Personal behaviour, food habits for obesity, alcohol consumption, emotional stress, financial difficulties, life project, business and relationships, sports, psychologic health, personal surveys about well-being and daily life.
3	Especially with smaller and newly emerging LPSs, to provide guidance with appropriate governance processes.
4	- Syntax sharing - Metadata management (technical solutions, web services) - Genomics and other omics data management, storage and data sharing
5	Several others e.g. data collection/research methods. Knowledge sharing is best done ground-up rather than being centralised. A pot of funds for knowledge sharing that studies could apply for would be my suggestion. Strategic working around policy impact may be valuable also.
6	We would like to see a central unit within PRUK who could curate core comms, PPIE materials, study protocols and template ethics submissions for use and adaptation by individual LPS. For example, we (TwinsUK) have shared protocols, ethics applications, participant information materials, PPIE consultation results and other documentation throughout the pandemic to other LPS in order to facilitate their research processes and participation in cross-cohort COVID-19 studies. As another example, due to our expertise, TwinsUK developed a long Covid questionnaire which was then shared with many other LPS for their use, and crucially, for recruitment to the NCS LH&W CONVALESCENCE study. A central unit responsible for these activities would be beneficial and more organised than the current ad-hoc approach to sharing materials.
7	Re 6: LPS share many functions and face many common challenges. We suggest it is worthwhile that all major role groups within LPS (e.g. data management, linkage, PPIE, communications) would benefit from a focused 'user community' group that can meet and share approaches, precedents, materials and insights. This has been seen to work well in the CLOSER 'communities of practice' networks which have tackled shared challenges (e.g. onward sharing of NHS records). For some groups, e.g. communications and PPIE, there is likely to be value in a central resourced team providing sector wide materials (although we note that communications and engagement is likely to need customising at a study level to reflect traditional approaches, the characteristics of the specific sample and other study level factors – e.g. geographical setting). Re 6: We also consider there is great potential for a 'policy exchange' function which raises the awareness of LPS as a resource for policy makers and enables them to pose research questions which can be taken up by the longitudinal community. This should also form a route for rapid and high impact dissemination of LPS findings, for example the recent LHW report on Long-COVID to SAGE or the Parliamentary Office for Science and Technology briefing on mental health during the pandemic. The mechanism for this is not clear and will require engagement and dialogue with policy makers to identify needs, benefits and the structure which would best support this. We recommend that this engagement is conducted through PRUK with links to wider relevant stakeholders (e.g. HDR UK/ADR UK, Parliamentary Office of Science and Technology).
8	Is there a national/international LPS conference (similar to Cochrane) – sharing experiences across the different LPSs?
9	A6: This is a UK-wide initiative but should have an international / global perspective, with a

special focus on LMIC's. The role for partnerships with tech start-ups and established data-led industries should not be underestimated, but needs to be thought through very carefully to avoid undermining the whole intent through loss of public confidence. PPIE should go beyond the new norm and embrace citizen science.

10	Looking beyond the biomedical and social sciences, and thinking in relation to enviromental, physical sciences, humanities (eg ethics, safeguarding, historical datasets) and arts (eg design, creative engagement) and more broadly culture (eg impacts of COVID are not just health, economy related). Having some stranded themes, workgroups or pump primed cross-sector research would help. The fellowships with No 10 data science team seems like a good model to embed practice into policy practice, this model might be replicated in different contexts
11	NHS Digital
12	A serious analysis of current LPS, age groups, gender, culture, diversity including mixed ethnicities and seeing how current datasets and ongoing studies are matched to SDGs and Grand Challenges. And where future investment for ongoing LPS and those that might be initiated should be focused for maximum benefit to the nation's and globe's knowledge about how we sustain, improve health and wellbeing, address inequalities and do it in a sustainable manner.
13	It has almost immeasurable potential. The more we know the better we can improve public health
14	No comment
15	Get Pharmaceutical inputs into how they would like to use this data for RWD and RWE
16	Funder requirements that all LPS participate and make their data available.

Q7: Does the work proposed for PRUK in this area provide distinct value over the engagement and involvement of LPSengaging with their participants directly?

Answered: 17   Skipped: 1

#	RESPONSES
1	It could complement it but would never replace it. An ongoing programme of co-design of future longitudinal population studies with minoritised and excluded groups (see e.g. work on Nawken culture being undertaken in Scotland) and public panels would be helpful. There are ongoing concerns about increasing closeness of research funders to profitmaking industries and about centralisation of data and decisionmaking that would need to be dispelled. Transparency and a meaningful four country approach are essential.
2	Yes
3	NA
4	Yes, the proposed activities in R6 seem very comprehensive
5	Unsure about this. Overall, my experience of this is that works best when very applied to specific contexts. Don't sense that this is a particular problem, though enhanced funding for public engagement for studies would be beneficial.
6	We would like to see PRUK support individual LPS to carry out their own patient & public involvement & engagement (PPIE) activities, for example by providing protocols for setting up PPIE groups and templates for consultations, and possibly administrative support. We feel strongly that each LPS is best-placed to engage with their own participants and carry out their own PPIE activities. LPS participants have a relationship with and trust their LPS, beyond any other larger structure to which the LPS belong (including even the universities housing the LPS). In addition, each LPS will have made a different set of promises and have unique relationships with their participants, and participants will have expectations and consents specific to their LPS. It would not be possible to move to a centralised model without a huge participant consultation, ethics amendments and reconsenting exercise, which would be incredibly burdensome to LPS and may well decrease participant trust. The relevance of broad, pan-LPS PPIE activities is also unclear – what our TwinsUK participants are happy and willing to accept may be completely unsuitable for participants in another LPS.
7	We consider that the value of PRUK is to help share insights, to help develop approaches and materials and to help target resources to particular engagement challenges (e.g. distinct sub-groups for whom targeted approaches may be beneficial). Further to this, infrastructure such as UK LLC could be used to help systematically identify patterns in participation across studies and where there are systemic gaps in coverage. We strongly believe that the studies should retain responsibility for engagement and communication with their own participants – but would benefit from the background support of PRUK in achieving this (see point #6 above). We also consider, that while more capital-intensive materials (e.g. an animation) may be produced centrally for general public awareness, or by infrastructure to reflect their operating model, a one-size-fits all approach implemented by studies would likely require significant modification, rendering the process not worthwhile or making the resulting product generalised to the point of not being meaningful.
8	Yes; raising awareness of the scope/number of LPSs and their value in improving health and well-being, independent of any specific LPS, would be very worthwhile.
9	A7: Yes, through combined effort, shared skills and increased power to make sound discoveries and minimise ill-founded conclusions.
10	The value add for LPS will likely be as an agent of change at a sector and government level,

rather than within the bi-multilateral relationships that studies have with participants

11	Should do - but see response to last question below. Who speaks on whose behalf needs much more thought than currently given.
12	It might, but depends on what it is. Think it's vital that those LPS that are based on community and locality retain deep connection with their local investigators or those with whom trust has been built over the years and decades.
13	yes, and is vital if we are to promote more joined up workin across these studies
14	only if coordinated with other national data initiatives to do wtih access to health and other potentially sensitive data. otherwise saturation and mixed messaging likely to backfire.
15	Potentially, yes, if itis part of a coherent public debate on use of publicly collected data. Could reassure participants that any additinal collaboration/ linkage would be secure and well governed. I challenge the idea that LPS deal solely with consented data and that the issues for this are therefore separate from use of unconsented data. Most LPS have not obtained (and many cannot now) explicit consent for linkage to other cohorts, admin data, etc. Where consent was obtained many years ago, its continuing relevance is debated.
16	Yes - there is significant public interest in this area and support
17	The aims and added value of PRUK are nearly identical to those of the existing CLOSER, with the exception of data linkage. This existing structure should be built on for the current purposes.

Are there other roles that PRUK should play in building connectedness across and beyond LPS?

#	RESPONSES
1	There is a requirement for an arms length body to engage with industry so that LPS primary commitment remains to their participants and the indepdendence of their work is maintained.
2	Publication of datasets available, visualization of outcomes for prevention by example.
3	NA
4	- To engage with the UK TRE - To explore the international dimension of data sharing: EU TREs, release of data outside of the EU, etc
5	As above, this could work best ground-up. Policy focus would be good but then also some studies are funded to do this themselves, and CLOSER is funded to do this.
6	As discussed above, PRUK should play a key role in connecting and streamlining the process of applying to local NHS R&D, NHS REC, CAG, NHS Digital and IGARD to enable LPS to access linked data – much like the UK LLC has been able to achieve.
7	We refer back to our comments above regarding PRUK helping establish the interface between the LPS community (Studies and infrastructure) and EHR and other databases and also – in terms of connectiveness – with ethical/approval bodies such as HRA REC, HRA CAG and NHS Digital (to support Green Paper R3). In terms of Green Paper R6 we want to emphasise the benefits arising from the NCS are in part resulting from close working between Data & Connectivity NCS (including NHS and NHS Digital, HDRUK, ADRUK, ONS) and the resources within LHW NCS (including studies, OpenSAFELY, UK LLC, BHF DSC and EAVE II). The UK LLC and D&C NCS 'interlock' could form a prototype relationship where there is strong coordination to enable LPS to inform wider Data Science challenges (e.g. optimising NHS data sharing approaches) and the wider community to help inform specific LPS challenges and to help ensure alignment with wider infrastructure and approaches (e.g. onward sharing of linked records). The PRUK record linkage function could help broker a clearer and more predictable relationship between LPS (as a community, as centralised infrastructure, as individual studies) and groups such as NHS Digital IGARD. The central PRUK hub – with Team Data Science leadership (see #9 below) – would maximise connectedness with policymaking through the policy exchange function (see #6 above) and central lobbying (see #12 below).
8	Yes; there is a "chicken/egg" dilemma as it is difficult for research teams new to LPS to know what research questions to ask without knowing what data are available. PRUK could run "hands-on" seminars detailing what data are available and use exemplar projects to illustrate how LPS data has been accessed/used successfully.
9	A8: Core support for LPS cohorts and capacity building – see also A12. Also more effort must be made to figure out how to share best practice, and resources developed across multiple LPS. For example, there is no resource holding the questions asked by studies by topic so that validated questions are available to all in an easy resource. This is just one example of the work that needs to be gathered together and made accessible across the LPS by PRUK.
10	see my response to Q6

11	Better involvement of NIHR.
12	See answer to Q5&6
13	Clear link to UK genomics strategy
14	Work with UKLLC and NHS.
15	Global reach out to global LPS

How might a leadership and governance structure for PRUK be created that facilitates a collective and impactful approach across the community to advance the recommendations set out?

#	RESPONSES
1	Has to be a coordinated network with distributed leadership, transparency at all levels, population wide participation, accountability to professions and public set out. A clear conflict of interest policy, commitment to integrated impact assessment to ensure equity interests are addressed, training and development of peer researchers and participants.
2	visualization, conference, social media to share the founding of the project. Involved patients and persons at risk: include in the community
3	NA
4	This will be very challenging, and should avoid a centralised model of everyone having to follow same approach and being very top down. Not all of the bits of what is proposed need to be done by same organisation. Strongly recommend that commissioning allows bidding for part of the work and not all of it together.
5	We strongly feel that any governance structure must include representatives in key areas from each cohort, such as: Study design and management Data management Research PPIE Governance Pilot trials Sample processing/biobanking Clinicians
6	We provide the leadership and governance structure for LHW NCS as an example: within LHW NCS there is co-leadership drawn from senior academics with expertise in LPS and in EHRs. Decision making is conducted via an Executive which is drawn from interdisciplinary LPS PIs relevant domain experts, funder representatives, and policy makers (DHSC, NICE) and infrastructure leads: the diversity in the Executive membership is key to the 'Team Data Science' approach, ensuring the effective interface between LPS and EHR database approaches, and effective dissemination of findings and 'policy exchange'. The LHW Executive reports to a Scientific Advisory Board on a regular basis. We recommend that the senior team for PRUK (e.g. a PRUK Executive) has similar diversity of membership and its terms of reference is explicitly interdisciplinary and has a Team Data Science emphasis and this will help facilitate the interoperable and connected vision set out in the Green Paper (e.g. R4 and R6). That this group does not seek to be inclusive of LPS PIs but interacts with a separate forum of these. The analytical activities are taken forward by interdisciplinary working groups including analysts from LPS and Data Science backgrounds and where junior analysts are offered leadership opportunities working alongside experienced mentors. Infrastructure development is lead by each infrastructure grouping (e.g. UK LLC) but ensuring cross-working with members of the infrastructure teams sitting in analyst groups to ensure coordination and effective needs assessment. We also note the importance of the 'Vanguard' group of data managers to the development of the UK LLC and the success this has generated in creating good will across studies and where the benefits of collective experience are contributing to the overall UK LLC model.
7	It would be good to establish a living/learning forum – with dedicated PRUK staff that work alongside some selected researchers to see what issues/practical problems they face – and then how these can be overcome for others likely to face the same issues.
8	A9: A network of Centres of Excellence – see A10. Ensure that intrinsic (including structural, legal, sociodemographic and ethnic) differences between the four nations are taken into account.
9	Look to trusted articulate leaders across all stakeholder groups, including public and patients, as well as academia, technology, public health and social care. MoUs and public commitments at the highest level may help, if backed by investment and a credible plan.
10	Include the people who understand the areas mentioned above - interdisiciplinarity, diversity,

	inequalities, ethics, history, culture as well as the current investment into those who can handle data
11	Include all parties that should be included. Do not forget patients
12	Work and coordinate with HDR and also our Future health project
13	Involve leadership across all LPS equally. Build on existing role of CLOSER.

Which proposed recommendations for PRUK would be most beneficial to you?

Answered: 13   Skipped: 5

#	RESPONSES
1	public trust and support
2	Many of the recommendations could have a positive impact but these are broad-reaching plans and therefore there is a need for prioritisation.
3	R1: Development a metadata strategy and discovery standards across UK LPS R1: Create greater awareness and visibility of the totality of UK LPS data and assets R2: Develop a strategy for future wider use of data platforms and TREs for LPS access R2: Create a centralised register of LPS data access requests R3: Support the increased use of linked data resources and readiness of linked data for analysis R4: Capacity building R5: Implement a public-facing communication strategy R6: Provide a collective perspective of UK LPS into the R&D landscape - A coordinated network of LPS data platforms and infrastructures
4	Uncertain what benefit will be. Would like funding to improve discoverability on our studies, but unclear/don't have confidence that PRUK best way to achieve this. Also would like linkages to be unlocked, but again would prefer to build on what we have e.g. ADRUK rather than start again.
5	TwinsUK would prioritise: - Recommendation 3 (facilitate linkage) - Recommendation 4 (Expand opportunities for aligned activity and enhanced interoperability of LPS) - Recommendation 1 (enhance discoverability)
6	Re 10: we strongly recommend that PRUK forms a mechanism for the legacy of LHW NCS in terms of Team Data Science ways of working in longitudinal analysis (open and reproduceable research, LPS working with whole population resources) and to continue to provide core support for the UK LLC.
7	"Hands-on" seminar(s) illustrating what data are available, trial run at formulating specific research questions (according to our interests) and then beginning to access data to answer the questions. This would help us to know how we could use LPSs to advance our own research interests.
8	A10: Relative to other cohorts and HEI's, we are at the vanguard of richly annotated open-access cohort studies and have strength and depth in the skills and expertise required to deliver PRUK ambitions. However, much of this is dependent upon relatively short-term PI-led response mode funding. Cross-disciplinary partnerships are ad hoc and short-term funding dependent. There is no headroom for capacity building or resource allocation for data archiving, quality assurance or sharing beyond the immediate needs of our internal investigators. This is a structural issues for all of the leading centres. PRUK might best deliver for the UK through a limited number of Centres of Excellence that would provide know-how, templates, technology evaluation, methods development, skills training and capacity building for onward sharing and PRUK impact.
9	All the recommendations are sensible. As a specific beneficiary that is hard to say. On other specifics I am conflicted because of a) CLOSER b) DataCite involvement
10	Having a central approach that really does incorporate a true sense of what it is to create and use population relevant data would be valuable for the reasons stated in the question above.
11	I think they are all important and dependent. EG, improving discoverability will have little impact unless data access is also improved
12	Financial and training support for data curation, and anything which helps solve the problem of barriers to sharing of NHS linked data
13	Please reach out and work with Pharmaceutical companies who would be willing to support you financially and developing a sustainable model

What would be the barriers to you or your organisation being engaged with PRUK? How could they be overcome?

Answered: 16   Skipped: 2

#	RESPONSES
1	Establishment as a public body
2	no barriers with updates and follow up.
3	NA
4	Time constraints on already overstretched staff
5	How this fits with CLOSER. How this fits with own funding. How this fits with other infrastructureinvestments. How our participation will be funded, what it will crowed out and how it will benefit us and not be burdensome
6	A common, historical issue with data access has been that researchers and grants require periods of exclusivity of data collected by LPS before outward sharing. This is essential for motivating PI's and researchers in writing grants. PRUK should consider adding new collections to a discoverability platform with restrictions on access – either to a timeframe or after keynote publication. However this should also include information to contact the LPS directly in the meantime for potential collaboration on data analysis.
7	Re 11: Team Data Science is constrained by staff and skills shortages (as discussed above), and this threatens the roll out and adoption of this model. The UK LLC has been specifically designed to provide a linkage solution for 'PRUK' and the challenges to this therefore relate to continued funding and the continued facilitatory environment for data science that has emerged in the UK in response to the pandemic.
8	Lack of knowledge of what research questions could be answered by accessing the LPSs. Seminars led by PRUK/LPS leads with exemplar projects would be helpful. Would benefit from some initial engagement so that the seminars would be designed to be directly relevant to advancing our research interests.
9	A11: None that I can see. We are already in 'PRUK' mode, but without the essential core support and capacity building funding.
10	Clarity on where this all sits cf ESRC/UKRI/Other funders plans on the R&D infrastructure roadmap for the UK
11	When an area of work is being expanded, gaps in multidisciplinary expertise soon appear, but it is very hard to create established posts in new areas and fixed-term funding does not attract staff able to look beyond the length of their contract. Early and predictable 'wins' are preferred by employer and employee in the current climate, so PRUK will need to offer rewards of other kinds - prestige, profile, membership of this and that, etc.
12	Finance, lack of investment into primary data collection infrastructure in an enduring manner. The sense that many think UKB can answer all questions relevant to health and wellbeing now.
13	studies currently have widely different attitudes/approaches to dealing with industry
14	If I were certain that the ethics of this transformation were adequately understood and acted upon and I felt the public were sufficiently informer I would see no logical barrier. There is always the potential for the system of anonymisaton to break own somehow - PRUK must be really secure
15	Capacity- need support. Data management is expensive. Concern over handing study data to

any other platform and losing control of data access. Existing platforms allow individual study control but this is not a model which works well for wider use.

| 16 | Do not reinvent the wheel - make use of tools available from non or commercial companies |

Are there any further priorities that you consider important to advancing the LPS field that are not covered in the recommendations listed?

#	RESPONSES
1	A longterm commitment to longterm funding of training in methods, studies large enough to contribute nationally and internationally, archiving and communication, engagement and co-design at community level.
2	
3	Knowledge sharing is strong focus
4	At TwinsUK, we require researchers accessing our data to return any derived data back to us. This is important to enhance the LPS resource. We therefore feel strongly that PRUK needs to consider how this would work in a more centralised model. It may be that PRUK needs to consider the mechanism for this but likely derived data should come back directly to individual LPS, who can pass to PRUK. We also feel that PRUK needs to consider the frequency of data refresh updates to PRUK, and what this means for associated resource of staff cost within individual LPS to manage this process. PRUK needs to consider whether this would be a standardised process e.g. once a year, or flexible according to LPS data collections and staff resource.
5	Re 12: The LPS lacks a strong and consistent voice in wider Data Science developments. Single studies have a diluted voice in consultations and are frequently not directly consulted; further to this, the awareness of relevant consultation activities and coordinating responses is resource intensive and individual studies either lack such resources or there is an opportunity cost to targeting these to consultation activities. Where the community is consulted, this is frequently targeted at UK Biobank which has distinct circumstances and is not reflective of the requirements of the wider LPS community (e.g. it is a fully consented study, its scientific model does not emphasise the need for representativeness and retention of harder to reach groups) and does not attempt to lobby for the wider community. PRUK needs to be able to channel the needs of many (all) LPS into consultations so it can effectively lobby for the community as a whole.
6	The exemplar research questions in Fig 3 are useful – but are quite general (e.g. what specific health outcomes of UK migrants could be assessed from current LPSs?). Also, it would be useful for PRUK to host some "blue sky" sessions involving many different disciplines aimed at generating a broad range of key research questions. These would then guide the further development of PRUK into the future.
7	A12: Sadly and surprisingly, nothing is mentioned about core, long-term support for LPS cohorts, nor of capacity building for the skills and know-how essential to design, build, run and sustain LPS cohorts.
8	It needs to say a lot more on international data sharing, access and discovery. Particularly in a post-Brexit context if access and research in the UK is going to thrive and benefit
9	1. Much greater clarity about the legal issues surrounding consent and the increased disclosiveness of linked data. 2. Better training of statisticians, and researchers more generally, about the provenance of data sources, the limited generalizability of consented datasets, and the very serious implications of e.g. collider bias and the magnitude of some of its effects. 3. Better social science of research participation, and how to resolve the conundrum that some of the people most in need of evidence-based care - the 'under-served' - are, for a variety of reasons, the least likely to be included in research datasets.

10	The recommendations and the synthesis are thoughtful and the quotes from researchers in this field illustrate the input. I have mentioned a few areas in response to the questions above that I think must be core and central. The first paragraph of the report talks about inequalities and societal challenges with the assumption that our LPS can address these - can they really, and if yes which? And which for which key areas?
11	Genetics/genomics should be discussed
12	The whole of the document is concerned with allowing more use of the data, not of allowing better use. The techniques to analyse these data are not deployed well yet even in single study analyses - what would this mean for aggregation? Alongside, or ideally integral to PRUK strategy, should be one of improving the general ability of people to use these data with the most appropriate analytical and statistical techniques. Otherwise, the aggregration of bias will not be accounted for and results of studies potentially more harmful than beneficial.
13	public particiation is mentioned, but I wuld stress that to me this is the single most imortant priority. We have a great opportunity, with the public more aware than ever of the benefits of data sharing, to establish a new social contract. We need to make sure PRUK is part of a coherent health data plan. At the moment, public concern about GP data- sadly, in my view, now justified, given the poor communication from government - risks tarnishing the reputation of all data use.
14	Global and Pharma reach out

Please use this box for any further commentsyou wish to make regarding the proposals for PRUK.

Answered: 6   Skipped: 12

#	RESPONSES
1	The below is an introduction and background to the answers provided above from NCS LH&W PRUK Green Paper: written evidence submitted by Longitudinal Health & Wellbeing National Core Study 1. Evidence summary 1.1 This submission is in response to the PRUK Green Paper which is considering the shared challenges faced by UK Longitudinal Population Studies (LPS) and making recommendations as to how certain challenges can be addressed through improved interactions between studies and resources, the centralisation of some functions and the consideration of new ways of working. 1.2 This response seeks to make a number of observations as a contribution to the review and then explicitly answers the questions set out in the PRUK Green Paper. 1.3 Our key observations and recommendations are summarised here and explained in the document below: • The Longitudinal Health & Wellbeing National Core Study is demonstrating success through new ways of working based on team data science, an effective interface between LPS and national EHR databases, new centralised linkage infrastructure for LPS and effective interface between LPS leads and policy makers. • We strongly recommend that PRUK forms a mechanism for the legacy of LHW NCS in terms of Team Data Science ways of working in longitudinal analysis (open and reproduceable research, LPS working with whole population resources) • We support the distributed model set out in the Green Paper and that PRUK resources would be better directed to supporting data management and curation and use of existing infrastructure rather than being focused on new infrastructure development • We propose that the UK LLC forms the basis of a linkage solution within PRUK and that this forms part of the legacy of the LHW NCS. • We suggest that the LHW NCS leadership structure provides a model for PRUK taking a Team Data Science approach in alignment and with involvement from LPS leads and domain experts. • The clear skill shortage in data science expertise and capacity represents a substantial barrier to realising a PRUK vision. • We strongly believe that the studies should retain responsibility for engagement and communication with their own participants – but would benefit from the background support of PRUK in achieving this. 2. Background to the Longitudinal Health & Wellbeing National Core Study 2.1 The Longitudinal Health & Wellbeing (LHW) National Core Study (NCS) is one of six studies commissioned by Sir Patrick Vallance as a major component of the UK's research response to the COVID-19 pandemic. The LHW NCS is designed to contribute insights only available from using longitudinal analytical approaches. This utilises a number of strengths from the LPS community: 1) the studies have rich pre-pandemic data which provides the potential to interpret how health and social factors have changed during the pandemic (e.g. mental health status) and to track trajectories emerging from the pandemic; 2) they have rich phenotypic and genomic datasets which contain behavioural data and other information not captured in routine records can inform assessments 2.2 To do this, the LHW NCS uses existing and new data from UK Longitudinal Population Studies (LPS) in conjunction with whole population databases held in national/devolved resources (e.g. OpenSAFELY) and Trusted Research Environments (e.g. Welsh SAIL Databank, NHS England British Heart Foundation Data Science Centre and the Scotland-wide EAVE II platform). This model allows the triangulation of data across sources so patterns identified within whole population records can be investigated in richer LPS data; and, self-reported status and behavioural LPS data can be used to asses quality, coverage and recording patterns in whole population databases. 2.3 The LHW NCS – along with wider NCS programme, particularly the Data & Connectivity NCS – is interdisciplinary and takes a 'Team Data Science' approach to research. This involves experts in different aspects of the research ecosystem, and from different sectors and scientific disciplines, working together to deliver research in a robust, open, reproducible and transparent manner. The openness of this process is designed to improve research efficiency, facilitate the enhancement of resources and to enable reproduceable research. 2.4 The value of UK LPS is maximised through linking participant-provided data and assayed biosamples with participants' routine health, administrative and environmental records. LPS funders are united in setting the strategic objective for studies to establish such linkages, but recognise the need for this to be done in a manner that is legal, secure and maintains participant trust. To enable this across the LHW

NCS studies (and wider UK LPS), the LHW NCS is taking the unprecedented step of developing the UK Longitudinal Linkage Collaboration (LLC) as a national Trusted Research Environment (TRE) for longitudinal research. The UK LLC hosts, integrates and provides managed access to de-identified data from many UK Longitudinal Population Studies (LPS) which are systematically linked to a wide range of health and non-health routine records. 3. Observations on the PRUK green paper 3.1 We recognise and welcome the emphasis on PRUK being truly interdisciplinary and inclusive of studies of all sizes. This is likely to maximise the value of LPS as a collective resource, improve the heterogeneity of participants (through inclusion of targeted studies with specific membership criteria), and enable studies to be enhanced by innovations developed within larger studies/centres and centralised resources. It would also support studies that are reaching end of life/funding to deposit data for long-term curation and analysis. However, accommodating 50-100 studies within a network may be an unwieldy task and dilute focus. Some existing resources – such as UK Data Service (UKDS) and UK Longitudinal Linkage Collaboration (UK LLC) are designed to scale and accommodate many studies and are less likely to be impacted by this than other aspects of PRUK. 3.2 We note with concern that standalone whole/devolved population resources are out of scope within the green paper. Whilst we recognise and support that the focus and resources in the PRUK initiative are aimed at LPS only, we strongly believe that the PRUK vision should include an interface with whole population resources (either Electronic Health Record (EHR) resources, administrative datasets such as those being developed by ADR UK and Office for National Statistics, or integrated resources such as SAIL Databank) in a manner that has been successfully demonstrated within LHW NCS. The NCS overall have demonstrated substantial innovation and accelerated progress. While a good part of this can be attributed to the galvanising impact of the pandemic response, the value resulting from close working across diverse data owners, infrastructure providers and research studies and analysists to a shared goal should not be overlooked. There is considerable enthusiasm within NCS to ensure a legacy for this aspect of 'Team Data Science' and a defined interface to this should form part of PRUK. 3.3 The LHW NCS can provide case studies to illustrate these interactions to help emphasise their value to stakeholders and to explain the rationale for this way of working with the public. For example, analysts using LPS data warned of low intention to accept vaccination by some population groups, including using LPS behavioural driver data which are not present in EHR resources to describe the rationale for refusal, while EHR analysis in this case demonstrated that intention identified in the LPS data was translated to action at a population scale. The interaction between these analyses was vital to the UK's pandemic policy response. 3.4 The green paper does not sufficiently recognise the legacy data challenge for existing studies to convert their data and documentation to a common standard. The work for each cohort is likely to be substantial – reflecting decades of under-investment, and also data complexity. This is especially true of long-duration cohorts, with dense data beyond surveys (e.g., wearables, biomarkers, genetic and other omic data and imaging). Resolving this will require data management and researcher time both at a study level and any infrastructure level (see CLOSER Discovery as an example of this). We do however support that PRUK should be a forum to help identify standards to work towards and – again, reflecting on innovations found during the pandemic – help coordination of data collection content and capture across studies as was seen with the WT funded COVID-19 questionnaire work. 3.5 The clear skill shortage in data science expertise and capacity is well recognised and is restricting the ability of the sector to deliver outputs, particularly in time pressured situations such as the pandemic. This represents a substantial barrier to realising a PRUK vision. We note the following critical points: • Capacity, skill shortages and a lack of recognition of data management and data science outputs and skillsets within university career progression paths is making recruitment and progression highly challenging – this is a major disincentive for potential recruits to choose to work in LPS; • Short term funding of some LPS infrastructure and initiatives forms barriers to recruitment and generates retention issues in the run up to renewal decisions; • Further, skills needed to manage LPS data are different to those for 'omics or imaging – and the move to utilise novel data sources (e.g. images, social media data, sensors) will generate new challenges (such as masking facial features on head images) that may require skills not currently found within the LPS community. Depending on complexity, each cohort will need time from multiple skill sets. This work has never been well recognised or funded; • Excluding the EHR/wider data science community from the PRUK solution may undermine the attractiveness of the LPS community as a research focus given the strong emphasis placed on EHRs resources and data science programmes within government, the research community and industry. This may hamper recruitment into the LPS sector and reinforces the value described above in PRUK having an interface to EHRs/wider data science community. 3.6 We therefore strongly recommend that those developing the PRUK vision ensure integration into wider data science initiatives - we understand that UKRI and HDRUK have already identified

this issue - to refresh the University progression system and recognise an interdisciplinary 'Data Science' track and to ensure this has clear and appropriate metrics of success and a defined progression pathway with grading commensurate to the responsibilities and the contribution to Government initiatives for this sector. 3.7 While training and capacity building and staff retention were identified as issued in the PRUK green paper (R4), the recommendations do not emphasise the potential value of centralised training and capacity building specifically for the LPS community: including longitudinal analysis, new ways of working (data science methods, analysing new sources such as linkages) and for data management and study management skills. We recommend that more focus is placed on this issue. 3.8 Harmonisation should not be prioritised as an activity for PRUK . The processing to harmonise is largely dependent on the specific question – and for many investigations both the granularity of the underlying data (which can be eroded through harmonisation) and explanatory in-depth covariates are needed. Harmonisation is of most value when the data are about to be used. It is therefore an applied researcher activity and is not fruitful as an infrastructure activity, although we recognise the value that harmonisation can bring to research projects and that infrastructure should be able to accommodate harmonisation activities. Once data have been harmonised, PRUK should ensure that they are catalogued, curated within a repository (and returned to source studies) and made available to future researchers via the HDR Gateway. 3.9 There are sector-wide challenges that need addressing regarding the integration of data across systems and geographies (e.g. how can English and Scottish health records be used together?). This challenge is broader than LPS but LPS should feed their requirements and experience into wider efforts to resolve these issues and to identify standards. PRUK could form a focal point for this.

2	Overall, it is not clear to what extent a "general researcher" could develop sufficient expertise to capitalise on LPSs. This seems to be a higher level of expertise than, e.g. doing a systematic review. Success is likely to depend on supporting researchers in realising the potential that LPS data has for their own interests and generating novel research questions - and then technical support from skilled experienced staff in accessing and using data.
3	I have alluded to conflicts already regarding my involvement with CLOSER and DataCite. I should also mention that I instigated MRC data sharing strategy many years ago when I was working at the research council
4	I cannot emphasise too much the need that ordinary people must know what is happening to data that is stored about them and that they feel securely that it will support Tham and their needs
5	Great project - Business would support you with RWE and LPS
6	While I applaud the role proposed for PRUK, I am concerned that public money will be wasted re-inventing structures that currently exist and could work well to play this role if expanded.

Additional response

General points:
The document presented is clear and outlines well the challenges in optimal use of population based resources in the UK, but what is the PRUK offering to be – i.e. in manifest form? There are lots of very good ideas coming together about this in the research community, but the conversion of the apparent finance commitment and now niche for PRUK into an actionable set of tasks seems absent.

There is a lot of emphasis around access and ease of access. However, this is a part of many population resources which is not funded usually and sometimes can't be standardised. As a consequence, whilst discoverability can be delivered for any asset, harmony around access process may be some way off.

There is a push towards linkage. This is great, though the query is as to whether this is this to be the way forward for PRUK explicitly or will this be a strategy of integrating cohorts as being done in the LLC.

Re. the recommendations made:
First three of these are ok and seem to be novel contributions. Further, the expansion of interoperability is great and there are prospective mechanisms to do this, but this will necessarily be tied to cohort agendas. The last two could be considered as usual business and the notion of this not happening in studies/cohorts could be potentially alarming.

**KEY– who are these developments for and do they have an impact on the nature of them as deliverables:

(i) For users – light touch and based on discoverability
(ii) For studies (which ultimately will be for researchers, but down the line) – then there are bespoke solutions and jobs to do which are about making studies better (access/linkage/prospective alignment clearing house)
(iii) The inclination would be to add in a third and important thing – a "bi-directional clearing house" i.e. a place for findings to go and a place for policy to ask/look. This has been successfully deployed in Manitoba (with a link to central funding) and has become a form of usual practice with many studies during the pandemic.

The vision section – this is based on an aim and an action rather than a manifest deliverable – e.g. actual linkage, an actual front end, etc. This is then followed immediately by the "anticipated impacts". There is not much here on actually what the manifest PRUK will be.

Timing and the general population/participants – work now post COVID-19 – there is a moment to really drive home the importance of research based on population data.

Specific comments:
R2 - Streamline data access processes and mechanism - Processes need to take account of individual consents and recognise that consenting procedures etc have changed over the years - what was acceptable in terms of sample consent in 1992 is still legally valid but doesn't necessarily meet contemporary standards and could mean won't fit in a standard mechanism etc.

R4 - Strategic Working - A recognised career framework with accreditation - this is starting to happen for biobanking but the registration is expensive and therefore not being taken up.

R3 - Facilitating Linkage - Could be expanded to make it easier to access remains of clinical samples.

Recommendations should ideally be done in collaboration with already existing resources rather than a top-down approach. The paper states that this is expansion not duplication of work which I welcome, and I would like to see more about how this collaboration with existing resources will work. Clarification is needed on how much funding is available or what mechanisms might support the next steps. There is a query as to which funders are supporting this and how this will link in with the existing LPS programmes of funding?

Clarification on the remit of PRUK would be welcome. Will the funders insist that we join PRUK? Which studies are in scope? Will UKBB be involved? Will studies be obliged to move to new ways of working e.g. data access model, including lay representation on data access models.

There was a feeling that that this is weighted towards biomedical studies e.g. data access model. This could potentially alienate these studies and there will be a split between MRC/Wellcome funded studies and ESRC studies. This would be a shame given the work that we have all been trying to do bring these studies together.

(**Despite the point above, samples are not mentioned.)

Who is PRUK to be? The partnership group can be found on the website but I'm not convinced they are representative of all LPS and their needs.

There is an emphasis on broadening/building on existing platforms etc but the current platforms are where they are for a reason. How would much of this work with international LPS. ALSPAC is part of many international consortia hosting similar studies which have had millions invested.

It could be considered that there is some conflict in the current document; e.g. p12 second bullet point under 'awareness/visibility' acknowledges that a new single platform is not recommended. Surely that defeats the whole point of recommendation 1?

Figure 3 related to linkage: concern here that the first two exemplar questions at least won't actually include any LPS participants by their nature and could be carried out using only administrative data –this therefore should not be pushed by PRUK.

P 9, first para refers to retaining and furthering the expertise of staff yet many of the plans will result in staff redundancy – 2nd para. There has to be a stepwise change in many of the initiatives and given my point above about the necessary input of senior staff I am concerned about how any of this will take place on top of the work we already have to do.

4c) PRUK public perspectives workshop

Overview

Population Research UK (PRUK) is a new initiative which will seek to increase the insights, innovations and research efficiency of the UK's wealth of social and biomedical longitudinal population studies (LPS) by bringing studies and data together. PRUK aims to address high-impact research questions that single studies cannot address alone. Health Data Research UK (HDR UK) has been commissioned by the MRC, ESRC and the Wellcome Trust to scope this new initiative. After an initial consultation phase, a Green Paper was developed and published in July 2021.[4] This outlined the key recommendations for how PRUK could be developed and identified further questions for consultation. A scoping review on best practices and recommendations for engagement and involvement with the public and study participants was published in July 2021.[5]

To address gaps from the scoping review and ensure that the public perspectives are well understood, a workshop was held in August 2021 with 16 public participants to discuss the proposals and recommendations for PRUK (see Appendix 1 for the agenda). 13/16 particants were recruited through HDR UK Voices – a network of individuals signed up to receive opportunities to contribute to involvement and engagement opportunities. 2/16 were recruited from People in Research website, a National Institute of Health Research for Public and Participant Engagement and Involvement opportunities. Characteristics of the attendees were not collected. During the workshop 4 attendees volunteered that they were participants in a longitudinal population study or cohort.

A public perspectives survey which had 37 responses over a 4-week period (see Appendix 2 for the summary and link to responses) was also run in parallel.

The general public perception was very positive about PRUK's potential, acknowledging the huge benefit that could come from this initiative. Discussions centred largely around public benefit, transparency, governing access to data, data linkage and how to engage diverse public groups. This report summarises the public workshop perspective of the four key recommendations for PRUK activities that were discussed in the workshops:

Streamline data access processes and mechanisms (R2);

Facilitate linkage between LPS data and health, administrative and further types of data (R3);

Sustain public trust and transparency in LPS data sharing (R5);

Maximise the connectedness of LPS with the research, innovation and policymaking ecosystem (R6).

These four recommendations were chosen for discussion based on alignment with topics identified in the scoping review as needing further insight and through previous PPIE work which identified these topics as having higher potential for damage to public trust. Adding to the findings from the scoping review, the public

---

[4] https://www.hdruk.ac.uk/population-research-uk/

[5] https://www.hdruk.ac.uk/wp-content/uploads/2021/07/Annex-2-PRUK-PPIE-Scoping-Review_final.pdf

perspectives workshop and survey identified four key implications that will inform the final PRUK recommendations. These were

i. Create a national register of PPIE representatives, or utilise an existing one, to facilitate public contribution and perspective in all PRUK activities. This can improve the relationship between the public and PRUK, encourage meaningful contribution from diverse public groups and may be a more efficient approach compared to recruiting PPIE representatives only when needed (Topic 1).
ii. Support the development of transparency on the data linkage process to educate and safeguard against existing public concerns; this include recording and communicating any PRUK-facilitated data linkage to the public (e.g. via website) (Topic 2).
iii. Early engagement with diverse public groups (e.g. schools, all ages, diverse groups) to educate about the benefit of LPS and PRUK (Topic 3).
iv. Conduct further consultation to understand how concerns about commercial companies can be addressed. There is substantial scepticism on government and industry partners and no potential benefit was acknowledged. If PRUK moves forward without addressing this issue, this could impact the perception and public trust in PRUK (Topic 4).

Narrative of discussions

Topic 1: Streamline data access processes and mechanisms (R2)

Workshop participants recognised the advantages of PRUK's involvement in streamlining data access as it could reduce time and effort for all parties. Individuals queried many aspects about data access, suggesting that there is a gap in public knowledge on how researchers currently access data. Queries included: anonymisation, which LPS exist, if approval must be given in order to access the data, how researchers can demonstrate their legitimacy before gaining access and how PRUK fits with other meta-data activities (e.g. HDRUK Gateway). Participants expressed a view that the public should be involved in the data access and approval process. This could take a number of different roles: ensuring data access is legitimate, sharing lived experience, upholding standards around transparency to eliminate any public suspicion and involving people from a range of backgrounds.

It was suggested that PRUK should seek to build a sustainable and long-term relationship with public representatives. One way to do this could be the creation of a national electronic database of PPIE representatives, where public members can outline their interests, experience and desired involvement and contributions. This approach, in contrast to PPIE involvement where organisations rush to have 1 or 2 representatives to fulfil a requirement, was considered more collaborative and could build better trust between PRUK and the public. Participants highlighted that advertising on the PRUK website and other networks would limit the breadth of participants to those who are already involved in PPIE roles. To reach a

more inclusive segment of society, local health boards and local authorities could help advertise roles and PRUK could collaborate with local community organisations to provide short-term internships. It was proposed these might be fixed-term full or part-time positions that would allow public representatives to have a larger, more substantial role in contrast to participating in one-off workshops. There was some discussion around whether public representatives should be renumerated or not, with a general consensus that the public deserve to be paid for their contributions.

Topic 2: Facilitate linkage between LPS data and health, administrative and further types of data

There was a positive and enthusiastic response about the potential of data linkage. However, there were some immediate concerns that reflected three key findings from the scoping review: the desire to understand data linkage governance and consent, the fundamental need for all PRUK research/linkage to be for public benefit and transparency in what data is being linked.

First, participants were concerned about the dangers of re-identification of participants, and possible misuse or exploitation of data. To avoid these risks, participants suggested that PRUK must develop a robust governance and consent model for data access and linkage. There was curiosity about the ethics and logistics of linking, including what standard operating procedures and data safety measures would look like and whether other big data organisations were already doing what PRUK has proposed. For example, it was questioned how PRUK could merge two or more datasets with different consent models and if data linkage could take place within Trusted Research Environments (TREs) instead of researchers being provided with physical copies of the data. Participants suggested that members from LPS should be involved in open consultation with PRUK to develop these procedures and ensure understanding of what consent means in data linkage. This could help participants retain a sense of control of their data. There was a brief discussion around whether consent for data sharing should be one-time acquisition and how ill-health or age-related cognitive decline may impact consent.

Next, several participants reinforced that any data linkage must be done for the benefit of the public. Failure to do so could create public mistrust in PRUK, and individual participants may wish to withdraw their consent for PRUK to share their data. One tangible suggestion to improve trust was for PRUK to effectively communicate the governance regulation and the processes involved in data linkage. This transparent approach could be beneficial in pre-empting any concerns that the public may have about linkage and could create more opportunities for the public to get involved. Several participants spoke about the need for this communication to be delivered in a common language that people understand. It was recommended that individuals with a wide range of backgrounds be sought to create accessible materials and terminology.

Topic 3: Sustain public trust and transparency in LPS data sharing (R4)

Consistent with the findings of the scoping review and observations on access and linkage above, transparency of data governance and ensuring the public benefit of any research were recognised as the most crucial aspects in building and sustaining public trust in PRUK. Three pathways were identified that could help build and sustain this trust: accountability, avoiding misinformation and reaching diverse public audiences. First, PRUK must ensure accountability by continual evaluation to ensure their work adheres to the original vision and aim. There must be a willingness of PRUK to constantly adapt and evolve so that the public benefit is at the forefront. Involving public representatives and publicising results from evaluations will help build and sustain trust.

Next, the potential of misinformation and the resulting repercussions were discussed. Misinformation can stem from ambiguity and thus PRUK should aim to educate the public about its aims, governance processes and research from its inception. PRUK can learn from the stark difference in public perception between the NHS (high trust) and the government (high scepticism and mistrust), with some participants suggesting that PRUK has the advantage of being seen more similarly to the NHS than the government. It is important that PRUK publicises themselves as an organisation who cares about the people and make it very clear its main aim is to support research that will help the public. On the other hand, participants cautioned that PRUK could immediately, and possibly permanently, lose public trust if there is any misuse of data (e.g. the Cambridge Analytical scandal was given as an example).

Finally, there were valuable discussions across both breakout groups about how PRUK can reach diverse and marginalised groups. Participants largely agreed that public involvement and engagement can be seen as a 'secret club', where only select individuals can share and contribute their perspectives. To reduce this stigma, prioritise transparency and reach more diverse groups, PRUK must carefully evaluate how information is disseminated. One suggestion was that material can reflect the same key information, but should be altered to reflect the target population (e.g. social media for younger adults, non-digital mediums for older adults). There were several suggestions that early engagement in primary school environments could be a sustainable was to fulfil the long-term strategic goal of PRUK reaching all parts of society. As an upstream approach, in-school activities in early years could embed the benefits of population research and PRUK for each new generation. Another way to target young people was to provide short-term internships or opportunities for those aged 16-17, who are old enough to have some responsibility, and may be able to meaningful contribute to PRUK.

Topic 4: Maximise the connectedness of LPS with the wider research, innovation and policymaking ecosystem (R6)

The possibility of linking LPS data to government or third-party operations was met with a high level of concern, with no positive discourse. Consistent with differing public perceptions of the NHS and the

government above, participants were reluctant for data from LPS to be connected with the government. The reaction to the potential of government and third-party involvement highlighted a sense of mistrust of the motives of those groups, removal of control from the public, and a lack of data science standards. This contrasted the clear trust that participants have in scientists, doctors and academics. Many expressed concern with commercial organisations, such as pharmaceutical companies, private health providers and health insurance companies, who could misuse or interpret LPS data in a way that could be used against the public (e.g. deny health insurance, raise commercial prices, etc.). Sharing data without investigating the ethical standards of partner organisations and their intended use for the data was suggested as bad practice. When connecting LPS data with other organisations, PRUK should consider the same issues discussed above: for public benefit, clear governance processes in place and transparency. Specifically, participants wanted assurance of where their data would be used, clarity on the integrity of the data and clear boundaries in its use. It was recognised that PRUK could build on recent initiatives related to public data (e.g. GDPR, NHS Digital, Covid-19 related research) that have raised public awareness on how data is used and shared across platforms.

Implications for PRUK recommendations

Many of the concerns and recommendations discussed in the workshop and raised in the survey were previously summarised in the scoping review and incorporated in the Green Paper. These include: transparency throughout the entire life cycle of data use, ensuring public benefit, involving public and participant representatives in all stages of PRUK's work and educating and empowering the public about LPS. There were several additional recommendations and insight from the workshops that should be further considered for the final PRUK recommendations. These include:

i. Create a national register of PPIE representatives, or utilise an existing one, to facilitate public contribution and perspective in all PRUK activities. This can improve the relationship between the public and PRUK, encourage meaningful contribution from diverse public groups and may be a more efficient approach compared to recruiting PPIE representatives only when needed (Topic 1).

ii. Support the development of transparency on the data linkage process to educate and safeguard against existing public concerns; this include recording and communicating any PRUK-facilitated data linkage to the public (e.g. via website) (Topic 2).

iii. Early engagement with diverse public groups (e.g. schools, all ages, diverse groups) to educate about the benefit of LPS and PRUK (Topic 3).

iv. Conduct further consultation to understand how concerns about commercial companies can be addressed. There is substantial scepticism on government and industry partners and no potential benefit was acknowledged. If PRUK moves forward without addressing this issue, this could impact the perception and public trust in PRUK (Topic 4).

Appendix 1. Agenda – public perspectives workshop

Time	Item	Lead	Paper
1.30pm	**Introduction to PRUK and** Presentation on the background of Population Research UK (PRUK), and the aims and objectives of this research initiative Time to ask any clarifying questions and share immediate thoughts.	Alice Dowden (Public Engagement and Involvement Officer, HDR UK) and James Pickett (Hubs Development Manager, HDR UK)	
1.45pm	**Green Paper discussion** Move into breakout rooms to discuss recommendations for PRUK in detail. Attendees will be presented the information and are encouraged to refer to their glossary as needed. Discussing overarching questions on: • What are their concerns about PRUK? • What do the public need to have confidence in the activities of PRUK? • What do the attendees see as the public benefit of PRUK?	Alice Dowden (Public Engagement and Involvement Officer, HDR UK) and Sinduja Manohar (Public Engagement and Involvement Manager)	• PRUK information slides • PRUK glossary
2.50pm	**Group discussion on key conversations** Facilitators to share highlights from between the breakout groups Attendees are welcome to share any other additional thoughts or comments	Alice Dowden (Public Engagement and Involvement Officer, HDR UK) and Sinduja Manohar (Public Engagement and Involvement Manager)	
3.00pm	**Next steps and close of meeting**		

Appendix 2. Summary of PRUK public perspectives survey and links to responses

Of the 37 survey respondents, over 10% indicated that they had been part of a cohort or longitudinal study, and nearly 50% indicated that they had participated in research, for example a clinical trial. PRUK public perspectives survey. Summary graphs and proportions for the multiple choice questions on prior PPIE involvement and the level of support and concern for each of the four topics can be found at: https://www.hdruk.ac.uk/wp-content/uploads/2021/08/Public-perspective-survey-results-multiple-choice-questions.pdf. All individual responses (including multiple choice questions and free text) can be found at: https://www.hdruk.ac.uk/wp-content/uploads/2021/08/Public-perspective-survey-results-open-text-answers-included.pdf.

Topic 1: Streamline data access processes and mechanisms (R2)

When asked whether they could see the benefit to the public in PRUK streamlining access to datasets for research, just under three quarters of survey respondents said they strongly agreed with the statement. Only one respondent indicated that they somewhat disagreed. When asked to indicate whether they are concerned about streamlining access to datasets for research, 25% of respondents said that they somewhat agreed with this statement, while just under 60% said that they neither agreed nor disagreed or that they somewhat disagreed, showing that there is some public cautiousness around streamlining access.

The follow up question of 'what might help you and the wider public have confidence in PRUK streamlining and improving access to datasets for research', survey respondents said that there needs to be accessible and clear educational resources available which explains how the data will be used, the safeguards in place and what is meant by streamlining data access. The need to improve public awareness on the benefits of research in general was also highlighted, and that all activities around data access and improving access should be transparent and easily accessible to the public. Finally, respondents also suggested that PRUK helps to reduce duplication of work, and therefore any linkage should be clearly recorded and communicated to the public and the research community.

Topic 2: Facilitate linkage between LPS data and health, adminstrative and further types of data (R3)

Respondents were asked whether they could see the benefit to the public in PRUK helping to facilitate linkage of LPS data and other types of data, including health and administrative data for research, with just over 90% of respondents indicating they either somewhat or strongly agreed with the statement. 33% of respondents said they had concerns about linkage between different types of datasets for researchers compared to nearly 60% who did not.

The follow up question of 'what might help you have confidence in PRUK facilitating linkage of LPS data with other types for research' focused on the importance of tracking all data sharing rigorously and ensuring that researchers and organisations are held accountable for the safe and legal use of data. They also highlighted that the public may be concerned about linkage enabling individuals to be identified, so PRUK should make sure that this is not possible, and that this is clearly communicated to the public. Respondents also raised concerns about data standardisation and stated that work needs to be done to ensure that different datasets are compatible for linkage. A strong theme in this area was the importance of a clear communications strategy around why PRUK is carrying out its activities.

Topic 3: Sustain public trust and transparency in LPS data sharing (R5)

Respondents were asked whether they could see the benefit to the public in PRUK playing a leading role in building and sustaining public trust and transparency in LPS data sharing and hearing how the public want to be involved. Over 90% of respondents strongly or somewhat agreed with this statement, and when asked whether they were concerned with PRUK playing this role over 55% somewhat or strongly disagreed. Just over 20% indicated that they somewhat or strongly agreed that they concerned about PRUK playing this role.

When exploring what might help the public have confidence in PRUK playing this role in building and sustaining public trust and transparency in LPS data access, many people stated that PRUK needs to ensure proper governance of data use and clearly articulate its goals and strategy for achieving them to the public. It was also stated that PRUK should show that they understand the lack of public trust in data sharing due to previous scandals such as the GPDPR dataset by listening to concerns and engaging with diverse groups and communities to help build trust. Providing information on who can accesss data and why, and ensuring that there are plain English summaries and public facing information will be vital if PRUK wants to position itself as a trusted voice in LPS data research. One suggestion for supporting this goal was to develop a patient data ambassador role at PRUK, where members of the public are involved from day one with the project.

Topic 4: Maximise the connectedness of LPS with the wider research, innovation and policymaking ecosystem (R6)

Respondents were asked whether they can see benefit to the public in PRUK working to maximise the connectedness of LPS with the wider research, innovation and policymaking ecosystem, with over 85% stating that they somewhat or strongly agreed with this statement. When asked whether they were concerned about PRUK working to maximise connectedness of LPS with wider research, just over 30% agreed wit hthis statement, while over 50% disagreed, indicating that there is a difference of opinion on PRUK playing this role.

We asked respondents what PRUK could do to help them have confidence in PRUK playing this role, and a clear theme that came through was the importance of explaining who they are working with, how and why. It was also raised that PRUK should be transparent about how industry partners might benefit from working with PRUK i.e., whether they or we would benefit financially. There is a strong mistrust in commerical organisations being involved, as the public are concerned about misuse of data and any conflict of interest as a result of working with companies. Some survey respondents always raised that this type of activity will require large capacity and resources within PRUK, so stressed that this should be properly considered and funded. Finally, it was clear from the survey that the public feel that patients and the public should be front and centre in this area to ensure that these relationships are explained to and considered by public representatives, and that PRUK should regularly communicate about these relationships and that they are for public benefit above all else.