UK STATISTICS AUTHORITY

NATIONAL STATISTICIAN'S DATA ETHICS ADVISORY COMMITTEE

NSDEC(20)xx

ESRC Early Life Cohort Pilot Study – proposed sampling and participant recruitment plan

Purpose

1. This paper sets out a proposal for the sampling and recruitment of the Economic and Social Research Council's (ESRC) new Early Life Cohort (ELC) and asks NSDEC for their advice on the proposed approach to meeting their ethical principles for this.

Timing

2. Urgent. Advice is needed at this meeting because a call specification for a team to lead a pilot for the new cohort is being prepared in summer 2020. A sampling and recruitment strategy needs to be agreed in principle before significant funds are committed to the pilot development. ESRC is keen to make progress on this, working in partnership with ONS and other UK statistical authorities, before commencing the process for appointment of a Scientific Leadership Team in autumn 2020 (see Annex A for timetable).

Recommendations

3. Members of the NSDEC are invited to:
    - Advise on the proposed sampling and recruitment approach and the ethical principles guiding this;
    - Note the timetable at Annex A

Background

4. The UK has a long history of conducting high quality, scientifically robust cohort studies on representative samples that track people from birth throughout their lives (for example, the 1946, 1958, 1970 and Millennium Cohort). To continue this series, the Longitudinal Studies Strategic Review and UKRI's Infrastructure Roadmap have shown there is a strong case for a new early life cohort. The new study must be UK-wide, post birth, embedded in an administrative data sampling frame and, as far as possible, inclusive of 'hard-to-reach' groups that are frequently under-represented in social science-led surveys.

5. The sampling frame used for the most recent study of this kind (the Millennium Cohort Study) is no longer suitable (see point 16). Pilot work in 2014 for a new UK birth cohort, planned for 2016 but subsequently withdrawn, indicated that careful attention should be given to recruitment procedures to maximise participation, particularly among 'hard-to-reach' groups, and that an opt-in mechanism was not effective (resulting in a response rate below 20%). The ESRC is therefore planning to commission a pilot study to test key aspects of the

sample design and feasibility of the full early life cohort study, including recruitment methods to maximise participation and minimise bias. An Early Life Cohort Advisory Group has been appointed by the ESRC, comprising experts who can advise on the development of the pilot, and the commissioning and oversight of a Scientific Leadership Team.

6. The ESRC intends to appoint a Scientific Leadership and Delivery Team through an open call (see Annex A) based at one or more UK universities to develop the objectives and design of the pilot study, and to conduct the pilot. The Scientific Leadership and Delivery Team would finalise the sampling specification and formally seek ethical approval from NSDEC, in accordance with the *ESRC Framework for Research Ethics*, to draw and use a sample for survey purposes. This paper sets out the high-level principles of a proposed sampling approach that will help shape the Scientific Leadership Team's more detailed work.

Discussion

7. The new cohort study will be designed to capture the impact of key **societal challenges and the interplay with individual-social factors** on new generations of children and their families including the lasting impact of the 2008 Great Recession, Brexit transitions, growing in-work poverty, complex family structures and dynamics, a changing world of work and education, the digital revolution, climate change, health and mental health, and the pervasive impacts of the COVID-19 pandemic.

8. To meet its objectives, the study needs to recruit a sample of babies and their families that is representative of the population, includes traditionally 'hard-to-reach' groups, and has minimal bias. A pilot stage is essential to establish that a suitable sampling frame can be identified, that effective recruitment mechanisms have been established and are feasible, and that participation is acceptable across groups, before the full study commences.

9. Prior discussions with ONS have indicated that the **birth register linked to NHS maternity records** would provide the most effective sample for the new ELC. In England and Wales this is held by the Office for National Statistics (ONS). The equivalent agencies in the rest of the UK (National Records of Scotland (NRS) and the Northern Ireland Statistics and Research Agency (NISRA)) will be approached to establish the feasibility of using these sources. Our approach to how using these sample frames would meet NSDEC's ethical principles is set out in this paper.

10. The recruitment model proposed will be that sampled participants will be contacted by the relevant statistical authority and invited to join the study, given the opportunity to decline to participate at that point, and then contacted in person when they will be offered further opportunities to withdraw before, during or after the interview itself.

The use of data has clear benefits for users and serves the public good

11. Cohort studies have played an important role in providing evidence on issues relating to child development, physical and mental health, education, social mobility and poverty, informing many areas of government policy. The Millennium Cohort Study's findings about breastfeeding and infant health have informed infant feeding guidance in the UK and internationally, demonstrating how evidence can inform practice and, ultimately, help to improve outcomes.

12. The ELC will generate data for highly innovative scientific and policy relevant research to tackle major societal and policy challenges for future generations of UK children and families capturing information related to: (a) income, (b) employment, (c) health, (d) education and skills, (e) access to housing, (f) public services, (g) family structure, (h) environmental changes, (j) spatial and intergenerational inequalities.

13. Addressing these areas serves the public good in two ways: the findings generated by this study will inform the public about a wide range of important matters affecting the lives of children growing up in the UK, and they can be used to help develop and evaluate public services for children and families.

14. The study can serve the public good only if its findings are generalisable to the population of interest and includes vulnerable groups that are of particular interest to policy makers. It must therefore aim to achieve a representative sample. This requires a suitable sample frame and effective recruitment methods. In the absence of a comprehensive population register**, birth registers linked with NHS maternity records** provide the closest alternative source. They are near universal and, unlike health records, can provide information about the socio-economic status of parents and other relevant information for sampling purposes such as whether births to unmarried mothers were registered by both parents.

15. The most recent major UK birth cohort (children born around the Millennium) was sampled using the Child Benefit register held by HMRC. This benefit is no longer paid universally to all families, so this register is no longer a suitable sample frame for a new birth cohort. NHS records are a potential source, but they lack crucial information about families' socio-economic status and fathers' registration which are crucial for helping to sample hard to reach groups. Recruitment via maternity services during pregnancy has proved effective for studies focusing on specific locations (the Born in Bradford cohort, for example) but has not been successfully applied on a UK-wide basis on the scale proposed for the new cohort.

16. **Birth registers linked to NHS maternity records** held by the UK's three national statistical institutes (ONS, NRS and NISRA) would ensure the ELC serves the public good by delivering the analytical benefits described above. These sample frames provide near **universal coverage** and contain **information**

about characteristics of interest that can be used in the sample design, such as:

- Age of mother at birth
- Parity of the birth
- Birthweight
- Gestational age of child
- Socio-economic status of registrants
- Status of registrant (s) (married/ civil partnership, couple not married/ civil partnership, single registrant)
- Parents' country of birth and/or ethnicity

Using these sample frames, with an appropriate method to approach sample members, would give the ELC the best chance of generating a representative sample UK-wide to support analyses that will serve the public good.

The data subject's identity (whether person or organisation) is protected, information is kept confidential and secure, and the issue of consent is considered appropriately

17. Established measures will be implemented to protect data subjects' identities at the point of sampling, during recruitment to the study, when sample performance is analysed, and when data are shared with researchers. In our proposed approach, sampling will be overseen by ONS, NRS and NISRA, following a specification developed with the ELC Scientific and Delivery Team. Names and addresses of selected families will be provided to interviewers to make their initial contacts, but other personal information used in the sample design will not be disclosed to fieldworkers.

18. Previous studies have already demonstrated that survey designs that require participants to opt-in before they are approached have very low rates of recruitment with heavily biased profiles of participants. This would not be a good use of public money nor would it be ethical to collect data from a skewed sample of participants that cannot be generalised to the population. The success of this study therefore depends on interviewers being able to approach all sampled families to seek their consent to participate in person, following an opportunity to opt-out. This is typically done by sending a letter to sampled families letting them know that they have been selected for the study and that an interviewer will visit unless they opt-out. The acceptability of this type of approach will be assessed in the pilot.

19. Securing a sample that does not require participants to opt-in to being approached about recruitment is vital, but the initial invitations and any approaches from interviewers must also be effective in generating a good response rate and participant enthusiasm for the study. The pilot will test different approaches to help inform the main study's recruitment materials and strategies. This might involve qualitative work with members of the Millennium Cohort Study and their families to find out what they have and haven't enjoyed about being part of a cohort study and what has motivated their ongoing participation into adulthood.

20. Explicit consent for the collection of data from subjects will be sought, and for their details to be retained for future study waves. Consent to data linkage will also be sought specifically and explicitly. All sampled families will be provided in advance of their decision to participate with Participant Information Leaflets, stating how their data will be used, how their consent will be sought and clarifying their right to withdraw from the study at any time. Information about the purpose of the study and the ways it aims to serve the public good will also be provided.

21. We believe this is a proportionate and appropriate approach to consent at the recruitment stage that balances the need to ensure the data collected are representative of the population with sample members' rights to decline to participate.

22. The pilot will seek secure access to anonymised data about non-participants who were invited to join the study. The purpose of such analysis is to examine for biases in response to help evaluate the performance of the sample frame and recruitment processes. This information can also be used to develop procedures to address biases caused by non-response.

23. There will be a strict separation of personal data (names, addresses, dates of birth, places of birth) from the analytical data files containing survey data. To prevent any possible disclosure from the analytical data files, via data matching or the exploitation of combinations of data, the release of data to researchers will follow the Five Safes Framework. Where appropriate, data will be safeguarded using secure access facilities such as those provided by the UK Data Service, ONS Secure Research Service, SAIL Databank (Swansea), Research Data Scotland and the NISRA Research Support Unit.

The risks and limits of new technologies are considered and there is sufficient human oversight so that methods employed are consistent with recognised standards of integrity and quality

24. The new ELC will drive methodological and technical innovation and growth in longitudinal cohort studies and develop future generations of world-leading data scientists. Although new technologies are not currently expected to be deployed in the sampling and recruitment for the pilot, the impact of the COVID-19 pandemic on traditional survey fieldwork methods may well require new approaches to be developed that harness technological innovations. Specific proposals for these kinds of developments will be outlined in future NSDEC applications.

Data used and methods employed are consistent with legal requirements such as Data Protection Legislation[1], the Human Rights Act 1998, the Statistics and Registration Service Act 2007 and the common law duty of confidence

25. The ESRC is working with ONS to ensure that the proposed sampling and recruitment methods, including any use of third-party fieldwork suppliers, are

[1] "Data Protection Legislation" means the full, applicable data protection framework as set out in the Data Protection Act 2018. This encompasses general processing (including the General Data Protection Regulation and the applied GDPR).

consistent with their legal responsibilities as data controllers for the sample frame. Established models for this exist with ONS subcontracting fieldwork to other agencies. Similar discussions will be held with NRS and NISRA over the summer.

The views of the public are considered in light of the data used and the perceived benefits of the research

26. The ESRC is in the process of completing a study of the public acceptability of using birth registration records linked to NHS maternity records for the purpose of developing a new birth cohort study. The views of the public were sought via dialogue workshops held in England, Wales and Scotland. The findings from this work have informed our approach to date and will continue to shape our public engagement work around the potential benefits and public good of the research findings based on these data.

The access, use and sharing of data is transparent, and is communicated clearly and accessibly to the public

27. The Scientific Leadership and Delivery Team will be responsible for ensuring that there is clear and accessible communication to the public about how data will be used, how personal privacy will be protected, how data security will be achieved, who will have access and why, and how research findings will be disseminated. The Five Safes Framework will be used to help shape these communications, as appropriate. The minimum for publication/dissemination would be that the pilot findings are openly available, and are communicated to key stakeholders (including participants themselves) to ensure impact.

Conclusion

28. The paper is intended to provide NSDEC members with early oversight of the proposed approach to sampling and recruitment for a new Early Life Cohort for the UK. NSDEC's valuable feedback will be used to inform the ESRC's continuing discussions with ONS, NRS and NISRA, and to help the Scientific Leadership Team's detailed sampling specification and recruitment plans once appointed.

Catherine Bromley, Deputy Director for Data and Infrastructure, Economic and Social Research Council, 1 June 2020

List of Annexes

Annex A – Pilot development timetable

Annex A – Pilot development timetable

Timetable

1. ESRC's timetable is largely driven by the need to appoint a Scientific Leadership Team. The expected process to do this, as of March 2020, is as follows:

April/May 2020: develop draft outline specification for the early life cohort pilot study.

June 2020: publish draft outline specification, invite expressions of interest for the Scientific Leadership Team.

September 2020: publish final outline specification. Open 'call for proposals' to award a grant to the Scientific Leadership Team.

April 2021: Scientific Leadership Team is in place. Develops full specification for the early life cohort.

April 2022: fieldwork begins.