International Journal of Population Data Science

Journal Website: www.ijpds.org

Participant acceptability of digital footprint data collection strategies: an exemplar approach to participant engagement and involvement in the ALSPAC birth cohort study

Kate Shiells[1,2,3], Nina Di Cara[3], Anya Skatova[1,2,3], Oliver S.P. Davis[2,3], Claire M.A. Haworth[2,4], Andy L. Skinner[1,5], Richard Thomas[3], Alastair R. Tanner[3], John Macleod[6,7], Nicholas J. Timpson[3,6], and Andy Boyd[3,6,8,*]

Submission History

Submitted:	**10/01/2022**
Accepted:	**13/01/2022**
Published:	**16/03/2022**

[1]Medical Research Council (MRC) Integrative Epidemiology Unit, University of Bristol, Bristol, UK
[2]Alan Turing Institute, London, UK
[3]Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK
[4]School of Psychological Science, University of Bristol, Bristol, UK
[5]Integrative Cancer Epidemiology Programme, University of Bristol, Bristol, UK
[6]Avon Longitudinal Study of Parents and Children, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK
[7]NIHR Applied Research Collaboration West, Bristol, UK
[8]CLOSER longitudinal study consortium, University College London, London, UK

Abstract

Introduction
Digital footprint records – the tracks and traces amassed by individuals as a result of their interactions with the internet, digital devices and services – can provide ecologically valid data on individual behaviours. These could enhance longitudinal population study databanks; but few UK longitudinal studies are attempting this. When using novel sources of data, study managers must engage with participants in order to develop ethical data processing frameworks that facilitate data sharing whilst safeguarding participant interests.

Objectives
This paper aims to summarise the participant involvement approach used by the ALSPAC birth cohort study to inform the development of a framework for using linked participant digital footprint data, and provide an exemplar for other data linkage infrastructures.

Methods
The paper synthesises five qualitative forms of inquiry. Thematic analysis was used to code transcripts for common themes in relation to conditions associated with the acceptability of sharing digital footprint data for longitudinal research.

Results
We identified six themes: participant understanding; sensitivity of location data; concerns for third parties; clarity on data granularity; mechanisms of data sharing and consent; and trustworthiness of the organisation. For cohort members to consider the sharing of digital footprint data acceptable, they require information about the value, validity and risks; control over sharing elements of the data they consider sensitive; appropriate mechanisms to authorise or object to their records being used; and trust in the organisation.

Conclusion
Realising the potential for using digital footprint records within longitudinal research will be subject to ensuring that this use of personal data is acceptable; and that rigorously controlled population data science benefiting the public good is distinguishable from the misuse and lack of personal control of similar data within other settings. Participant co-development informs the ethical-governance framework for these novel linkages in a manner which is acceptable and does not undermine the role of the trusted data custodian.

Keywords
ALSPAC; attitudes; co-development; data linkage; digital footprint data; engagement; longitudinal research; participant involvement; safeguards

*Corresponding Author:
Email Address: a.w.boyd@bristol.ac.uk (Andy Boyd)

https://doi.org/10.23889/ijpds.v7i1.1728

Introduction

Digital footprint records – the tracks and traces amassed by individuals as a result of their interactions with the internet, digital devices and services [1], provide a wealth of ecologically valid data on their behaviours and actions. These data have great potential to enhance the existing datasets of longitudinal population studies (LPS), given that these data are generated in real time, with potentially high temporal frequency and may be less impacted by some sources of bias than traditional methods, such as questionnaires. The opportunities for developing greater understanding about how behaviours impact upon individuals and population health are growing due to increasing digitisation, such as the Internet of Things (personal, household and community devices that are connected over the internet), social media records, and widespread adoption of mobile and wearable devices, which have all produced various novel types of digital footprint data [2]. In response to these opportunities, the primary funders of longitudinal research in the UK are encouraging studies to incorporate linkage to routine records (such as health and social administrative records) and novel forms of data (such as Digital Footprint records) into study data collection strategies [3–5]; and this form of linkage features in considerations for a Population Research UK programme to help contribute to the future direction of UK longitudinal research [6]. However, linkage across datasets is associated with increased risk of privacy breaches, such as identity disclosure [7], and well-publicised data misuses and the increasing awareness of personal data being used as a commodity are resulting in the loss of trust in organisations collecting data [7, 8].

A substantial effort has been made over the past decade to understand the public/patient/participant perspective on the use of their routine records in health and social research and to involve the public in the research process in a substantive way. Eliciting evidence from the general population, deliberative workshops [9, 10] and citizens' juries [11] have explored the acceptability of using routine health and social records in research. These have identified that there is typically a low understanding of research methods and the value linked records can bring; but, once individuals understood the purpose and potential benefits of using data in research, then this is typically supported on the condition that the research is intended to benefit the public good and that sufficient safeguards are in place [9–12]. However, there is a varying degree of trust placed in different organisations for the use of data in this way, and the type of organisation deemed most trustworthy is context specific [13, 14]. Within longitudinal studies, Stockdale and colleagues' [15] systematic review on public attitudes towards the use of patient data identified a complexity of views and a low level of awareness regarding the secondary use of health and other records in research. Nonetheless, the review identified a broad willingness to share health records for secondary purposes, where research is designed to benefit the public good; but this willingness is dependent on sufficient safeguards being in place. The authors identified that many of the issues raised can be mapped back to fundamental ethical principles of autonomy, beneficence, justice and non-maleficence [15].

There is also a growing body of research exploring public attitudes towards the linkage of novel forms of digital footprint data. Clarke et al. [16] consulted with the public to explore attitudes towards sharing loyalty card data and health and fitness app data for linkage with health records. Barriers to sharing these data included fears around data security, mainly data breaches; invasion of privacy; and fears around third-party access of data. For instance, participants questioned whether supermarkets would be able to view their health records as part of the linkage process. Participants believed that their data could be inaccurate, potentially misleading researchers; in particular they shared concerns that loyalty card data could show an incomplete picture of what they actually consume, and therefore suggest that their diets are less healthy than they consider them to be. Finally, participants would require more information in order to understand the purpose and benefits of linking these types of data for research. A study conducted by the Wellcome Trust [12] which involved consulting with members of the public about a variety of different linkage scenarios revealed various concerns depending on the type of digital footprint data. For instance, concerns about sharing social media data were centred around whether location data could be used for tracking their whereabouts, and the ethical implications of data relating to third parties who had not consented to data sharing. Reflecting findings of Clarke et al. [16], the linkage of loyalty card data and health records made participants uneasy; in particular, they were concerned that they may be judged negatively by their GP for eating unhealthy foods. They also questioned the accuracy of loyalty card data and the extent to which the data would provide a true representation of what they actually consumed.

It is therefore imperative that the collection of digital footprint records takes place within an 'appropriate ethical and regulatory framework' [5]. In order for this framework to confer acceptability on this data use – to have a 'social licence' – it will need to extend beyond legal and regulatory compliance [17] and will need to incorporate the safeguards that are seen as meaningful and necessary to those whose data will be used (i.e. participants of LPS and the wider public) [18]. In order to develop such a framework, the Avon Longitudinal Study of Parents and Children (ALSPAC) longitudinal birth cohort study sought to engage and involve participants. ALSPAC carried out five qualitative forms of inquiry in order to gauge cohort members' understanding of, and elicit their views on, the use of various sources of digital footprint data for incorporation into the ALSPAC study databank for novel forms of research. This process reflects those carried out by researchers elsewhere who have engaged with biorepository participants in order to explore their views on collecting, storing and sharing genetic research data outside of local institutions [19]. ALSPAC cohort members have previously described the importance of engagement that emphasises their value and leads to them feeling less like a 'data source' [20]. Our approach employs a range of methods, which reflects some of the different options available within longitudinal studies for participant involvement. A diversity of approach – utilising qualitative studies, focus groups and a standing panel of participants with in-depth knowledge of study methods – is intended to elicit a range of views and perspectives [21].

In this paper we illustrate ALSPAC's approach to participant involvement in the design of study data collection activities. This is illustrated through a description of a series of studies involving participants in the consideration of novel

forms of data capture relating to digital footprint records. We synthesise the results of these studies and discuss the ways in which the findings will guide the development of the ALPSAC digital footprint strategy. This summarises the findings of a report prepared to the UK's Economic and Social Research Council [21] and a series of focused papers which describe some of these specific studies in depth [22–24].

Methods

About ALSPAC

ALSPAC, also known as 'Children of the 90s' (Co90s) to its participants, is a multigenerational prospective birth cohort study. ALSPAC recruited pregnant women resident in and around the city of Bristol (south-west UK) and due to deliver between 1st April 1991 and 31st December 1992. There were an initial 14,541 enrolled pregnancies comprising 14,676 foetuses (for these at least one questionnaire has been returned or a "Children in Focus" clinic had been attended by 19/07/99). These pregnancies resulted in 14,062 live births and 13,988 children alive at 1 year. From age seven attempts were made to recruit additional cases who were eligible under the original sample definition [25, 26]. By age 24 an additional 913 index children had enrolled. The total sample size for analyses using any data collected after the age of seven is therefore 15,454 pregnancies, resulting in 15,589 foetuses. Of these, 14,901 were alive at 1 year of age. Of these, 14,775 were live births and 14,701 were alive at 1 year of age [27]. The cohort has been followed intensively from birth through self-completed questionnaires and attending clinical assessment visits. The cohort is multigenerational: comprising the original pregnant women and the fathers/partners; our index participants (those due to be born 1991–1992); and now their offspring. This paper describes evidence collected from the index participants.

ALSPAC has built a rich resource of phenotypic and genetic information relating to multiple genetic, epigenetic, biological, psychological, social, and other environmental exposures and outcomes. The ALSPAC Web site hosts a data dictionary that describes the available data (https://bristol.ac.uk/alspac/researchers/our-data/) and further information can be found via the CLOSER Discovery metadata platform (https://discovery.closer.ac.uk/).

Participant involvement in ALSPAC

ALSPAC involve participants in the conceptual and operational decision making of the study (see Panel 1). A range of these mechanisms have been used to help develop ALSPAC's digital footprint strategy: with participants providing insights and contributing to research publications through a standing committee, the Original Advisory Cohort Panel (OCAP) and targeted focus group exercises. Formal ethical review and approvals have been sought from the study's faculty ethics committee, ALSPAC Law and Ethics Committee (ALEC). The methods and insights from these exercises are described in this paper; and are illustrative of ALSPAC's wider approaches to participant engagement and involvement.

Evidence from our digital footprint programme

Methods of data collection

The aims of each individual study and the methods used to collect data are summarised below.

Study 1

Aim: to elicit opinions on the acceptability of data acquisition from different types of routinely generated records.

Study 1 was an exercise conducted as part of the first of a series of three focus groups exploring attitudes and understanding relating to the use of transactional records in longitudinal research. The methodology is described in greater detail elsewhere [24]. Through a convenience sampling approach, we mailed postal invitations to a randomly sampled sub-set of 600 participants living in the Bristol area. The focus group was conducted twice, with two independent sets of participants. The focus groups were held at the ALSPAC study centre during 2018. Ten participants attended the first focus group (Focus group 1a), and six different participants attended the second (Focus group 1b). The researchers (Authors Anya Skatova, Andy Boyd) presented participants with a deck of 20 cards (Figure 1), each printed with the names of either a type of routinely generated digital footprint data, such as 'mobile phone use', more established forms of routine records, such as 'health records', or traditional categories of research data, such as 'age' and 'gender'. An explanation of each card was provided and then participants were asked to rank the cards on a sensitivity scale in sub-groups (pairs/trios) and then as a whole group, in order to produce a consensus ranking. This involved ordering the cards from most willing (least sensitive) to least willing to share (most sensitive).

Study 2

Aim: to explore attitudes towards sharing commercial transactional records (e.g. supermarket loyalty cards, bank statement data) for longitudinal research, and to understand which safeguards researchers should consider implementing when looking to request transactional data from participants.

Study 2 [24] is drawn from the same series of three focus groups as described in Study 1, although the findings of Study 2 are collected from across all focus groups and specifically relate to the use of transactional records in longitudinal research (in contrast, Study 1 describes the findings from a specific exercise relating to the use of different types of digital footprint data). Overall, 20 participants attended at least one focus group. The first focus group explored the use of different types of digital footprint data in general. The second focus group involved discussing attitudes towards sharing transactional data in more detail through numerous interactive tasks, such as participants being asked to select a 'story card', which presented the viewpoints of different fictional individuals involved in or affected by record linkage. Participants were asked to discuss whether or not they identified with the different viewpoints and discuss with the group. The final focus group in the series involved presenting participants with a conceptual framework representing the process of data linkage within ALSPAC, with participants asked to reflect on their preferred mechanisms of consent.

Panel 1: Overarching approach to participant involvement in the design and operation of ALSPAC

ALSPAC has included participant input in the design and operation of the study since its inception. This is achieved through a spectrum of formal to informal routes and settings (described below). There is particular attention on the acceptability of study activities and the impact of new work on participant trust; on participant inclusion and retention; and, the feedback of findings and consideration of future research direction. The work on involvement and retention is informed through the monitoring of response and participant attrition/withdrawal and detailed investigation of the patterns and predictors of participation and the impact this has on the heterogeneity of the study sample and the inclusion of vulnerable sub-groups [28].

ALSPAC Law & Ethics Committee (ALEC): a dedicated faculty ethics committee. ALEC was an early, innovative example of a study ethical review committee with participant membership (now 50/50 professional/participant membership). ALEC consider underlying ethical principles necessary to protect participants, proposals for new data collection and the development of policies concerning confidentiality and anonymity, consent, non-intervention and disclosure of individual results, data access and security [29].

Original Cohort Advisory Panel (OCAP): at participant age ~14 ALSPAC convened a young person's advisory panel which continues to meet. OCAP members provide advice across different aspects of the study: including reviewing and developing materials; providing participant perspectives on new study directions and data collection activities; guiding engagement activities and strategy; and helping with recruitment to participant facing roles.

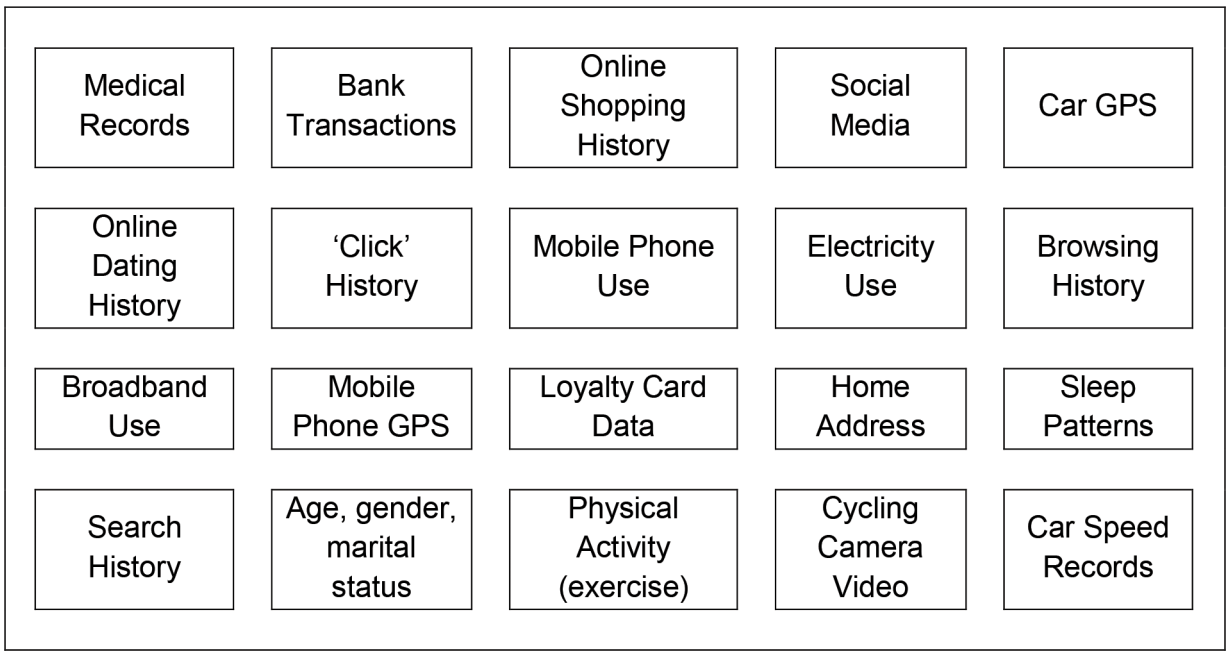
Qualitative Interview: ALSPAC recall participants into (voluntary) sub-studies using qualitative approaches. Whilst typically these relate to substantive research investigations; these have also explored attitudes to 'consent' for record linkages [30], views on ethics committee composition [31] and perceptions around engagement in longitudinal research [20].

Focus groups: focus groups enable rapid and efficient feedback on specific topics. These are frequently used to explore attitudes and understanding towards new data collection activities (e.g.,23,24).

Quantitative Survey: ALSPAC questionnaires can measure the acceptability and viability of new data collection options (e.g., the Life@26+ questionnaire sought views on linkage to banking and loyalty card records [32]).

Events & community engagement: study science engagement activities (e.g., attending local citizen science events; hosting study 'summer school' events) have included sessions designed to enable participants to feed into the future directions and priorities of the study.

Figure 1: Cards illustrating a range of routinely generated data sources



Study 3

Aim: to seek views on ALSPAC's use of residential and personal location data, whether this type of research is viewed as important and within the perceived scope of ALSPAC, and whether participants are concerned about the use of their location information, or perceived specific risks of this type of research.

In Study 3 [22], ALSPAC data managers (Authors Andy Boyd, Richard Thomas) attended an OCAP meeting in 2017. Selection and recruitment to OCAP is summarised in Panel 1. To facilitate the discussion, the data managers presented hypothetical research scenarios that described sharing approximate location (e.g. 1 km2 area), specific locations (e.g. home or school addresses) and exact location (e.g. GPS tracking). Views were collected, and OCAP members unable to attend were provided with the information and were able to submit written comments.

Study 4

Aim: to explore participant views on the acceptability and necessary safeguards needed to support the use of social media data in research.

As part of Study 4 [23], ALSPAC participants over two generations (young people (N=9) aged 26-28 and parents (N=5) aged 53-65) took part in two separate focus groups. A random sample of ALSPAC participants who lived within travelling distance were invited to take part. Researchers (Authors Oliver Davis, Claire Haworth, Nina Di Cara, Alastair Tanner) used a phrase template as an elicitation tool. The template (Figure 2) had fixed text punctuated with blank spaces; separate word cards were provided and participants were asked to use the cards to fill in the blanks (up to 108 possible scenarios) and discuss how they would expect these data to be shared and presented. Participants discussed a range of social media platforms, but were informed that access would only be to their 'visible' information (i.e. what the public or their permitted friends and family could see), not their private information or other data stored on their systems (e.g. direct messaging, or in online cloud storage associated with their social media).

Study 5

Aim: to seek views on the understanding and acceptability of collecting Ecological Momentary Assessment (EMA) data through an ALSPAC issued smartwatch.

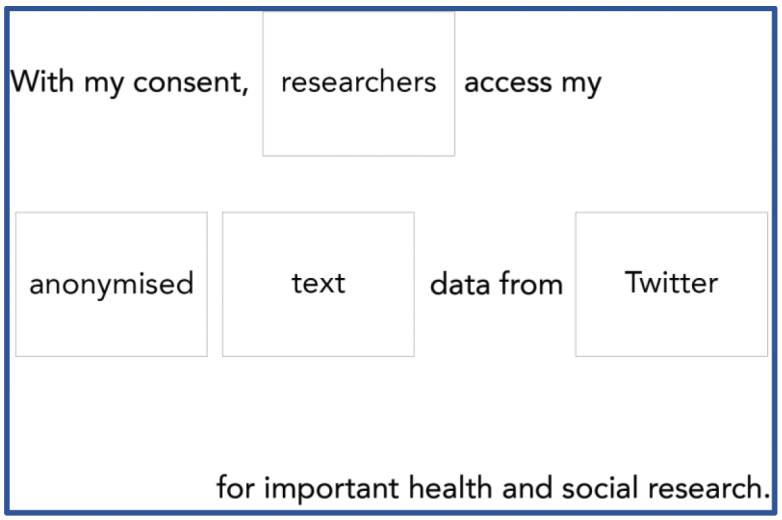
Study 5 involved holding discussions at an OCAP meeting in 2019. The discussion considered novel mechanisms of collecting EMA data, but was framed using an exemplar proposal where ALSPAC participants would be issued with a Smartwatch which collected self-reported EMA data on consumption of alcoholic drinks, worn daily for three months. There were six OCAP members present at the meeting. The meeting was chaired by a member of the ALSPAC participation team and the study PI (Author Andy Skinner) was present at the meeting to outline the project's aims and methods. An example smartwatch was brought to the meeting to illustrate the EMA system. The project was first described and this was followed by a discussion comprising questions from OCAP and answers from the author.

Due to confidentiality reasons, minutes from this OCAP meeting were not transcribed verbatim, and therefore no direct quotes from members are included in this article. However, OCAP members consented to share minutes for analysis.

Data analysis

Study 1: The ranked cards showing participants' decisions on the sensitivity of different digital footprint sources were photographed at the meetings and associated discussions were transcribed.

Figure 2: An example of a completed template from the elicitation exercise-keep in methods



Studies 1 – 5: A qualitative meta-analysis approach was adopted in order to synthesise primary qualitative findings from across the five studies and build a more complex understanding of ALSPAC participant views on digital footprint data collection strategies [33]. Thematic analysis as described by Braun and Clarke [34] was used to identify common themes amongst the studies. It is important to note that studies 1 and 2 involved the same sub-group of participants; that studies 3 and 5 will have overlap in contributing participants, and that participants in study 4 are unlikely to have taken part in the other studies.

The qualitative meta-analysis approach involved first searching the transcripts for initial codes, which resulted in nine codes, including misunderstanding, consent, and trust. These were then merged as appropriate, and final themes were developed that corresponded to the conditions associated with the acceptability of sharing digital footprint data for inclusion in the ALSPAC databank. Author Kate Shiells developed the initial coding framework, which was discussed and refined with author Andy Boyd. All authors reviewed the themes. Data analysis was carried out using NVivo Version 11.4.3.

Results

In this section, we describe the conditions associated with the acceptability of sharing digital footprint data for inclusion into the ALSPAC study bank, as specified by cohort participants.

Participant ranking of the sensitivities of data sources

Results of the data sensitivity ranking exercise in Study 1 are presented in Table 1.

This suggests that participants independently reached broadly similar conclusions as to the spectrum of sensitivity across digital footprint data. However, the most sensitive data (medical records) have been systematically shared with ALSPAC, indicating that individuals may be willing to share sensitive forms of data subject to establishing appropriate safeguards. Bank transactions were ranked as the second most sensitive form of digital footprint data in the first focus group, however the participants in the second focus group could not reach an agreement on their ranking for this form of personal data. This group of participants were unsure of the granularity of data that could be exposed through access to their bank transactions, which they also discussed in relation to other categories of data, such as mobile phone records.

The group discussions identified a broad agreement that data showing patterns of use (e.g. duration of social media use) was less sensitive than data showing specific itemised use (e.g. the content of social media posts). Data showing precise location (e.g. Global Positioning System [GPS] records from phone sensors) were also considered sensitive. An individual's understanding of the granularity of data that can be accessed through the various categories is therefore influential on their decision as to the sensitivity of their digital footprint data.

Whilst the consensus group exercises produced very similar outcomes, the different sub-groups produced some 'outlier' rankings. For instance, the sensitivity of 'click history' was ranked towards the middle in the overall consensus of both focus groups, but sub-group 3 in Focus group 1a did not see it as particularly sensitive:

Search history I wouldn't be bothered about, I think that can go over here, you've got your click history up there.

-*Yeah that's the same.* (Study 1)

This may result from personal experiences or from non-typical interpretations of the data or the proposed use of the data. For instance, the use of cycling camera video use was considered highly sensitive by sub-group 1 in Focus group 1a due to its links to location and when considering it from a third person perspective:

People can see where you're going. (Study 1)

Camera video, didn't like the idea of being recorded by other people and stuff like that. (Study 1)

Table 1: Participant ranking of the sensitivities of different data sources in Study 1

	Ranking of sensitivity*							
	Focus group 1a				Focus group 1b			
Data Source	Table 1	Table 2	Table 3	Consensus	Table 1	Table 2	Table 3	Consensus
Medical Records	5	20	20	20	8	20	18	19
Bank Transactions	5	19	19	19	9	19	20	n/a**
Online Shopping History	1	13	9	8	10	12	4	8
Social Media	2	8	6	10	12	11	1	11
Car GPS	3	7	9	16	18	14	11	14
Online Dating History	5	9	9	11	11	9	6	11
'Click' History	2	13	2	11	15	12	5	15
Mobile Phone Use	4	11	6	11	14	4	13	9
Electricity Use	2	5	14	5	4	4	1	4
Browsing History	4	13	n/a**	11	15	17	10	15
Broadband Use	4	13	14	5	4	4	13	4
Mobile Phone GPS	3	18	14	16	20	15	13	18
Loyalty Card Data	2	6	2	8	6	7	1	6
Home Address	5	12	14	16	7	16	18	10
Sleep Patterns	1	4	6	2	1	1	7	1
Search History	4	13	9	11	15	18	5	17
Age, gender, marital status etc	1	10	1	1	3	10	17	7
Physical Activity (exercise)	1	3	2	2	1	2	7	2
Cycling Camera Video	5	1	2	7	13	8	7	11
Car Speed Records	3	2	13	2	18	3	11	2

*Many groups ranked multiple data sources as having the same level of sensitivity, or clustered sensitivity into groups. The rankings expressed therefore have many tied values. The colour-coding reflects the quintile of sensitivity ranking, with the shading progressing in density as the sensitivity increases (quintile 1, lowest sensitivity coloured in a light shade, quintile 5, highest sensitivity in the darkest shade).

**Participants were unable to reach a consensus as a whole group as to the sensitivity of these data sources.

Themes associated with the acceptability of sharing digital footprint data

Participant understanding

Under the theme 'participant understanding', we address participants' need to understand the **value**, **validity** and **risks** associated with sharing digital footprint data for research purposes.

Participants made reference to the importance of understanding the **value** of using digital footprint data before providing informed consent, particularly in Study 2, where they were unclear about why researchers may want to access their transactional data:

What are you guys hoping to achieve by understanding what we're buying, and how is that going to help future generations? (Study 2)

When study aims were clearly explained to participants, they appeared more willing to donate their data:

I personally would probably give you all of that [transactional data] if I had a sheet explaining that you were going to do something with it and I was happy with the purpose. (Study 2)

Likewise, in Study 3, when the potential to improve public health was made clear, participants considered the uses of spatially indexed and location records for research purposes important:

It seems to be the reverse of what we normally do, like they normally like test our bodies and our brains and this seems to be testing something that may affect our bodies and our brains, so it seems like it's going like a little bit further, so I think it's quite good. (Study 3)

There was confusion expressed as to the degree of **validity** associated with digital footprint data for research purposes. In Study 5, participants were curious whether people would answer questions truthfully about their alcohol intake if, for instance, they had been binge drinking. Participants were also asked to discuss the option of being able to edit data. They questioned whether being able to edit answers and add missed information the next day may increase recall error.

Similarly, in Study 2, participants questioned whether gathering data about their weekly shop would be indicative of their individual behaviour if, for instance, it was their family members who consumed certain products instead of themselves:

If you've got a family, you're going to shop for like all four of you and you might not drink any of it [milk]. So actually, it's not that accurate. (Study 2)

Finally, participants require an explanation of the actual **risks** involved in sharing digital footprint data. For example, whilst based on misunderstandings about how their data would be used, some individuals were concerned that sharing transactional data may impact on the service they receive:

Could there be a chance that it might impact the deals you get from your bank, maybe? (Study 2)

Will it affect things like your credit history? (Study 2)

Likewise, in Study 3, participants questioned whether their location data could be used in legal cases, and in Study 4, had misconceptions that the extraction of their social media data could expose them to IT security risks:

Anything that can pick up words on a computer can pick up everything [...] So if they're picking up what's on Twitter then they can pick up everything else. (Study 4)

Sensitivity of location data

Digital footprint data sources were consistently considered more sensitive and potentially unacceptable to share if they revealed location, as discussed by two participants in Study 2, who feared that they could be 'tracked' through their GPS:

I would say it's because they know your routine more anything.

-Yeah, where you shop, where you work, where you go out. (Study 2)

The acceptability of sharing data associated with location was specifically explored in Study 3. The major concern here was the way in which multiple datasets could be linked through common variables, making identification more likely:

I guess people would be like familiar with this kind of data collection where apps kind of like track you, so that's kind of standard, but then I guess like the ethics are more complicated because you have information on our genes and stuff as well so like the computers that track our movement, don't have that level of information so then like that's when I think it's more challenging. (Study 3)

In Study 4, there were mixed opinions on sharing location data. One participant was more against the idea:

I think people would be more reluctant to give their location data, just from speaking to my friends and things because it's more... I don't know, creepy. (Study 4)

However, whilst other participants were less inclined to share their location data on social media in general, they did not seem to be more concerned about it being used for research than any other features of their social media profile that could be collected:

Anything I was sharing on my social media, including location, I'd be happy to share with Co90s but I do my best not to share my location with anyone. (Study 4)

Concern for third parties

Participants widely considered the sharing of digital footprint data to be less acceptable if details of third parties associated with the data could be revealed, which was linked with lack of consent from the third party:

It's all about permission [...] Because if you post that out there and they want to take it down and you haven't got evidence that you've got permission from them then it's a legal battle really from there. (Study 1)

I object to it strongly [...] my friends haven't agreed to that. (Study 4)

A participant from Study 3 was also concerned about sharing data that could reveal details of their children's location:

As a parent you kind of think every stranger is a danger to your child, but I don't know, yeah for me that's just [...] he is not old enough to know really he is part of a study or the information being recorded. (Study 3)

There were also similar concerns expressed by some participants in this study regarding the sharing of photos from social media that included third parties.

Clarity on data granularity

For many participants, the acceptability of sharing certain digital footprint sources was associated with the level of data granularity:

So it it's your phone numbers and the conversations that you're having then I wouldn't be ok with that, but if it's just, I don't know, how often I send a text or, I don't know, how long I spend on the phone, then that wouldn't be an issue. (Study 1)

Likewise, when considering the sensitivity involved in sharing bank transactions, one participant explained how:

It depends on how it's collected and how confidential the data's going to be, how anonymous the data is going to be etc, because like, if you're looking at us as a wide group and you're seeing like, one of us bought flowers on a particular date, that's not really an issue but like, if you're looking at each individual and you're seeing personal transactions, that's more confidential. (Study 1)

Whilst generally comfortable with sharing approximate location data, one participant expressed concerns about sharing precise location data, which they perceived as sensitive, particularly when linked to their child. In particular, they were concerned about the potential for identification where cell sizes were small, and felt that the sharing of multiple locations could increase the risk of identification and potential physical harms to their child:

I would be fine if it was walked past such and such [...] a tree at 8.02, if it records those things, but if it said walked past such and such a tree at 8.02 on such and such road, I personally wouldn't be happy that that detail was recorded about my child. Do you see what I mean, because that's very specific, you can see where they are, what time of day they are going past a route. (Study 3)

Mechanisms of data sharing & consent

In Study 5, participants described a number of features that could improve the design of the smartwatch for data sharing. For instance, the survey buttons on the watch face should be large, and the reminder trigger should not be disruptive. Participants stated that a 'back button' would be useful in order to correct mistakes, and that they would like the addition of a 'repeat option' that entered the same details, and a 'skip option' in order to complete the survey at a later date.

Participants across all studies also debated retrospective versus prospective data sharing, with retrospective data collection seemingly preferable, particularly for location and transactional data:

With mobile phones' GPS, like, I don't mind if it's retrospective. So, if like, they see where I've been, like, at some point in the past – fine, but if it's like, live, like where I am now... (Study 1)

So your situation might change and the last thing on your mind would be oh someone's collecting this, I should be careful about what I do. (Study 2)

There were also varying opinions expressed in regards to opt in versus opt out consent, although participants recognised that if they were asked to use a device to collect data, opt-in consent would be required:

I guess for this one it would be an opt in, because you would be asking them to wear the thing or do the thing on their phone or either way, so they would be fully aware. (Study 3)

Study 2 participants advised that consent forms should be structured in such a way as to allow participants to fine tune their consent decisions, allowing for more sensitive elements of the data to be removed:

If you just say transactional data, if someone doesn't really want online stuff within that they will just say no to the whole thing. Whereas they might have been happy for the loyalty cards stuff. (Study 2)

One participant also expressed the need to be able to opt out of sharing data at any time:

So, I think if we had something we could, like a link or something, we could click on at any point and opt out I think that would be good. (Study 2)

Trustworthiness of the organisation

The trustworthiness of the organisation utilising the digital footprint data was another factor which participants stated would play a deciding factor in whether or not they considered it acceptable to share their data. ALSPAC was consistently described as trustworthy. For instance, even though medical records were ranked the most sensitive of digital footprint types in Study 1, participants reflected on how they are willing to share this type of data for ALSPAC research:

I mean, we've got medical records as the worst one we'd share but we've all shared and done tests here and had stuff poked and prodded and filmed. (Study 1)

Participants across various studies discussed the reasons why they trusted organisations like ALSPAC with their digital footprint data:

The motivations of Children of the 90's, some policy makers, and people who are researching or looking into rare diseases, that would be, erm, I don't know, it's something about their motivations just seems more legitimate. (Study 2)

Children of the 90's is fine because I know you're not going to sell it. (Study 4)

Participants also spoke about the importance of trusting the organisation collecting data ensures anonymity at the point that researchers access the data:

I don't think I mind if it's, if it's anonymous [. . .] because it can't be attributed back to you in any way. So long as there's no way of linking it back to you, I don't really see the harm in that I don't think. If it's just a number. (Study 1)

However, participants were happy for identifying data to be available to the ALSPAC data study manager, who would need this for linkage purposes:

I am happy with it, I guess a lot of our stuff is putting our faith in Children of the 90s. (Study 3)

Finally, when researchers seek to obtain any new sources of digital footprint data, participants discussed the importance of enacting standard safeguards already used by ALSPAC, such as issuing contracts for data sharing, enforcing sanctions for misuse, and encryption of data.

Discussion

Given the rapidly evolving nature of populations today, it is crucial that longitudinal population studies adapt to collect novel forms of data representative of these populations accordingly [4, 5]. However, this should be accompanied by a consideration of the concept of 'social licence' [17], whereby it is imperative that this novel use of participant data is understood by and acceptable to individuals in order to maintain trust in the wider study.

Cohorts have a variety of methods available to them to engage with their participants in order to understand what is considered acceptable in the realm of data sharing, from established participant representatives to more formal research processes and informal community activities. This paper provides a summary of the types of qualitative methods used by ALSPAC to engage with their cohort. Participants have been able to use these exercises, amongst a wider engagement and consultation programme, to help contribute to the design of the study. The paper also provides a synthesis of insights into cohort members' views and suggestions for the collection and linkage of digital footprint data. This will provide ALSPAC with a valuable evidence base with which to develop their digital footprint data linkage strategy. Several specific meaningful impacts for this strategy have been identified and are elaborated upon below.

Impact of cohort views on the ALSPAC strategy

Cohort views 1

In view of the novel nature of the data ALSPAC are seeking to obtain, participants will require an explanation of how their digital footprint data will be used prior to seeking consent for incorporation into the databank. This is reflective of previous research which has found that participants often have low levels of understanding of how health records [10] and social media data [35] could be used in academic research. However, once participants from the ALSPAC cohort understood how sharing their data could benefit the public good, they were largely more accepting of doing so. It is therefore paramount that studies communicate the benefits that can be realised through using these data sources, and it would be prudent to seek insights on how best to achieve this from initiatives such as Understanding Patient Data [36].

Impact 1

Study communication materials are shaped with participant input (both insights from exercises such as these, and also direct involvement by participant advisors and contributors; e.g. http://www.bristol.ac.uk/alspac/participants/newsletters-leaflets/). In response to a previous participant involvement exercise [30], ALSPAC have developed a formal set of study principles, 'Our Commitment to You' (http://www.bristol.ac.uk/alspac/participants/our-commitment-to-you/), emphasising autonomy, the role of ethical oversight, the principle of confidentiality, and that the study's research aims to deliver public benefits and will not be used for profit. These high-level principles are used as part of 'fair processing' information describing study activities and safeguards.

In some situations, the creation of these safeguards – for example, that only study data managers are able to process sensitive identifiable information such as exact address – may be seen to hinder the research process and be a barrier to 'open research'. To help explain this to researchers, participants have co-authored a publication which outlines the participant rationale for these decisions to an academic audience [22].

Cohort views 2

ALSPAC have traditionally used questionnaires to gather a range of data from the cohort, where participants are aware of the parameters in which they are being measured and can make a conscious and informed choice about which data to share. However, requesting the passive regular sharing of large amounts of digital footprint data is accompanied by legitimate fears about what could be revealed to researchers. Participants expressed concerns about the disclosure of third-party information that could take place particularly through the sharing of transactional and social media data. Similar concerns have been raised in previous research [35], where participants were unsure whether they could consent to share their photographs from social media that included third parties. Furthermore, data granularity was associated with increased risk, in particular when sharing location data. Prior research has shown that this may be a valid concern. For instance, the granularity of social media data collected at high frequency intervals potentially allows researchers to build a very accurate picture of an individual [37].

Impact 2

The Project to Enhance ALSPAC through Record Linkage (PEARL) has established a data extraction and processing pipeline, which is built to 'Data Safe Haven' principles [38]. A central principle in this approach is that access to sensitive, identifiable information is restricted to study data managers. The approach maximises the potential for future use by allowing study data managers to curate rich, identifiable data which can be held within the Safe Haven and be processed for each approved study in order to meet their specific needs. For example, free text extracted from social media records would not be shared in its raw and disclosive form: rather, it would be typical for machine learning approaches (such as Natural Language Processing algorithms) to produce non-disclosive derived outputs [38]. These algorithms and outputs would be tailored to each use case and applied by the study data manager: although this methodology may introduce barriers to some research designs (such as exploratory Machine Learning which requires interactions between the free text and study outcome data), and this may require alternate safeguards (e.g., operating on a locked down machine in a secure setting).

Cohort views 3

Evidence collected from participants across the studies suggests that in order to maximise data sharing, participants should be provided with granular consent options, such as the choice of whether to share precise location data, or various spending categories within transactional data. Furthermore, the choice to share data prospectively was described as allowing participants to feel more 'in control' over what data is being collected. In addition, participants highlighted the importance of being able to opt out of sharing data in the future, with a quick and easy mechanism to do so.

Impact 3

As regards to consent approaches, all participation in ALSPAC is voluntary and providing data or involvement in any part of the study is discretionary. Consent for sharing of digital footprint records is likely to be based on opt-in consent, as most digital footprint linkage will require active participant involvement. For instance, participants actively need to disclose their Twitter account handle [39], or to participate in an EMA exercise. The information materials provided will be tailored to address areas of uncertainty and the safeguards participants consider necessary.

Strengths and limitations

Whilst Studies 1, 2 and 4 were conducted and reported according to the Core-Q guidelines for qualitative research [40], Studies 3 and 5 are based on discussions held during OCAP meetings and cannot be considered robust qualitative research. This is specifically true of Study 5, where minutes from the meeting were used instead of verbatim transcripts.

Furthermore, OCAP members – as with any committee – will not necessarily be representative of the ALSPAC cohort, although this is offset by the value arising from their strong understanding of epidemiological methods. However, this paper is part of a diverse group of work that is using a breadth of approaches in order to ensure the voice of the ALSPAC cohort is heard, and evidence is collected in order to shape future strategies. For instance, qualitative research with cohort members has also been conducted in order to explore their views on data linkage [30]. Furthermore, ALSPAC have conducted randomised trials with participants in order to elicit their opinions on materials to improve consent response rates [41], and to explore the effectiveness of opt in versus opt out methods of contacting participants for re-engagement [42]. Finally, participant representation on study committees helps ensure the interests of the cohort members are formally represented on an ongoing basis.

It is anticipated that the insights generated, and actions taken by ALSPAC, may also serve as an exemplar for other longitudinal studies seeking to act within an ethically-sound framework. This paper provides new evidence in an emerging research area, within the context of longitudinal research: particularly reinforcing existing findings on the importance of clear communications emphasising the purpose, benefits and potential limitations of a novel technique [12, 16]; identifying the role of study staff as trusted actors in the de-identification process [12, 16] and the importance of autonomy [16]. These findings supplement recent public dialogue exercises conducted by the ESRC 'Population Data Laboratory' programme [43] which are intended to help design a new UK birth cohort study and a focused case study within the evidence on 'social licence' in a research context. The insights will also help inform the development of appropriate data flows of de-identified data to researchers, and also centralised infrastructure such as the UK longitudinal Linkage Collaboration (https://ukllc.ac.uk/) or the UK Data Service (https://www.data-archive.ac.uk/).

Conclusion

Realising the potential for using digital footprint records within longitudinal research will be subject to ensuring that a 'social licence' for this use of personal data is achieved; and that rigorously controlled population data science delivering benefits for the public good can be distinguished and viewed separately from the misuse and lack of control of these data within other settings. Key to this will be the use of granular and clear consenting strategies and that studies take on the role of trusted data custodians and implement transparent and robust controls on data processing and use which have been co-developed with participants. This will include separating the data collection, processing and dissemination processes so that data acquisition and the processing of incoming granular data is conducted by study staff operating in a trusted role, before de-identified and filtered data are securely shared with research users in line with the reasonable expectations of study participants. Where such governance is in place and clearly communicated to participants, the evidence from our participants suggests that the collection of digital footprint records will be viewed as acceptable.

Acknowledgements

Avon Longitudinal Study of Parents and Children (ALSPAC)

We are extremely grateful to all the families who take part in the study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses.

The evidence synthesis report was funded from UK government Strategic Priorities Fund and awarded by the Economic and Social Research Council (ESRC).

The UK Medical Research Council and the Wellcome Trust (Ref: 217065/Z/19/Z) and the University of Bristol provide core support for ALSPAC. The ALSPAC evidence in this report project draws on methodologies developed in the Project to Enhance ALSPAC through Record Linkage (PEARL) – a Wellcome Trust award (WT grant reference: WT086118/Z/08/Z). A comprehensive list of grants funding (PDF, 330KB) is available on the ALSPAC website. For the research work that informed this synthesis: AB and RT were funded by the UK Natural Environment Research Council (R8/H12/83/NE/P01830/1) as part of the 'Enhancing Environmental data Resources In Cohort studies: ALSPAC exemplar' (ERICA) award; AS was funded by a University of Bristol Vice Chancellors Fellowship; KS, AS, OSPD, CMAH were funded by the Alan Turing Institute under the EPSRC grant EP/N510129/1; OSPD, CMAH and AB were funded by the ESRC (ES/R011583/1) as CLOSER Work Package 21; AB, JM, OSPD and CMAH were funded by an MRC Mental Health Data Pathfinder award (Ref: MC_PC_17210). ALS is supported by Cancer Research UK (C18281/A19169 and C18281/A29019); OSPD and ALS work in a Unit that is supported by the UK Medical Research Council (MC_UU_00011/6 and MC_UU_00011/3) and University of Bristol. CMAH is supported by a Philip Leverhulme Prize. NJT is a Wellcome Trust Investigator (202802/Z/16/Z), is supported by the University of Bristol NIHR Biomedical Research Centre (BRC-1215–20011), and works within the CRUK Integrative Cancer Epidemiology Programme (C18281/A19169). The views expressed in this report are those of the authors and not necessarily any funding body. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the report.

Statement of conflicts of interests

None to be declared.

Ethics statement

Ethical approval for the ALSPAC project was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees. Study 1 and Study 2 (ALEC reference 63603, Title: 'A framework for linking financial and retail data with ALSPAC to uncover causes of mental health illness and routes to wellbeing'); Study 4 (ALEC reference number 78784, Title: "A framework for linking and sharing social media data for high-resolution longitudinal

measurement of mental health across closer cohorts"); Studies 3 and 5 involved OCAP participant advisors whose role is to contribute insights to the design of ALSPAC data collection and communication strategies (the establishment of the OCAP was approved by ALEC). The use of direct quotes in Study 3 was based on OCAP members' consent.

References

1. Lambiotte R, Kosinski M. Tracking the digital footprints of personality. Proceedings of the IEEE. 2014;102(12):1934–1939. [6939627]. https://doi.org/10.1109/JPROC.2014.2359054.

2. Bidargaddi N, Musiat P, Makinen V, Ermes M, Schrader G, Licinio J. Digital footprints: facilitating large-scale environmental psychiatric research in naturalistic settings through data from everyday technologies. Mol Psychiatry. 2017; 22:164–169. https://doi.org/10.1038/mp.2016.224.

3. Pell JP, Valentine J, Inskip H. One in 30 people in the UK take part in cohort studies. Lancet. 2014 Mar 22;383(9922):1015. https://doi.org/10.1016/S0140-6736(14)60412-8.

4. Davis-Kean P, Chambers RL, Davidson LL, Kleinert C, Ren Q, Tang S. Longitudinal studies strategic review: 2017 report to the Economic and Social Research Council. ESRC. 2017. https://esrc.ukri.org/files/news-events-and-publications/publications/longitudinal-studies-strategic-review-2017/.

5. Wellcome Trust. Wellcome's Longitudinal Population Studies Working Group: Longitudinal Population Studies Strategy.2017. https://wellcome.org/sites/default/files/longitudinal-population-studies-strategy_0.pdf.

6. HDRUK. Green paper Consultation on recommendations for developing Population Research UK. 2021. HDRUK. https://www.hdruk.ac.uk/wp-content/uploads/2021/07/PRUK_Green-paper.pdf

7. Pagliari C, Cunningham-Burley S. Public Acceptability of Cross-Sectoral Data Linkage: Deliberative Research Findings. 2012; Social Research series, Scottish Government, Edinburgh. https://www.research.ed.ac.uk/portal/files/17210075/Davidson_Pagliari_et_al_2012_Public_Acceptability_of_Cross_Sectoral_Data_Linkage.pdf.

8. Information Commissioner's Office: Information Rights Strategic Plan: Trust and Confidence. 2019. https://ico.org.uk/media/about-the-ico/documents/2615515/ico-trust-and-confidence-report-20190626.pdf.

9. Cameron D, Pope S, Clemence,M. Exploring the public's views on using administrative data for research purposes. Ipsos Mori. 2014. https://esrc.ukri.org/files/public-engagement/public-dialogues/dialogue-on-data-exploringthe-public-s-views-on-using-linked-administrative-data-for-research-purposes/.

10. Ipsos MORI. The One-Way Mirror: Public attitudes to commercial access to health data. 2016. https://wellcome.ac.uk/sites/default/files/public-attitudes-tocommercial-access-to-health-data-wellcome-mar16.pdf.

11. Tully MP, Hassan L, Oswald M, Ainsworth J. Commercial use of health data-A public "trial" by citizens' jury. Learn Health Syst. 2019;3:(4). https://doi.org/10.1002/lrh2.10200.

12. Wellcome Trust. Summary report of qualitative research into public attitudes to personal data and linking personal data: summary report/Wellcome Trust. Wellcome Trust. 2013. https://wellcomelibrary.org/item/b20997358#?c=0&m=0&s=0&cv=0.

13. Jones LA, Nelder JR, Fryer JM, Alsop PH, Geary MR, Prince M, Cardinal RN. Public opinion on sharing data from UK health services for clinical and research purposes without explicit consent. https://www.medrxiv.org/content/10.1101/2021.07.19.21260635v1.

14. ODI. 'Nearly 9 in 10 people think it's important that organisations use personal data ethically', ODI Blog. 12 November 2019. Open Data Institute. theodi.org/article/nearly-9-in-10-people-think-itsimportant-that-organisations-use-personal-dataethically/.

15. Stockdale J, Cassell J, Ford E. "Giving something back": A systematic review and ethical enquiry into public views on the use of patient data for research in the United Kingdom and the Republic of Ireland [Version 2; peer review: 2 approved].Wellcome Open Res. 2019;3(6). https://doi.org/10.12688/wellcomeopenres.13531.2.

16. Clarke H, Clark S, Birkin M, Iles-Smith H, Glaser A, Morris MA. Understanding Barriers to Novel Data Linkages: Topic Modeling of the Results of the LifeInfo Survey. J Med Internet Res. 2021;23(5). https://doi.org/10.2196/24236.

17. Carter P, Laurie GT, Dixon-Woods M. The social licence for research: why care.data ran into trouble. J Med Ethics. 2015;41(5):404–409. https://doi.org/10.1136/medethics-2014-102374.

18. Gulliver P, Jonas M, McIntosh T, Fanslow J, Waayer D. Qualitative research: Surveys, social licence and the integrated data infrastructure. Aotearoa New Zealand Social Work. 2018;30(3):57. https://doi.org/10.11157/anzswj-vol30iss3id481.

19. Lemke AA, Wolf WA, Hebert-Beirne J, Smith ME. Public and Biobank Participant Attitudes toward Genetic Research Participation and Data Sharing. Public Health Genom. 2010; 13(6):368–377. https://doi.org/10.1159/000276767.

20. Ochieng C., Minion JT, Turner A, Blell M, Murtagh MJ. What does engagement mean to participants in longitudinal cohort studies? A qualitative study. BMC Med Ethics. 2012;22(77). https://doi.org/10.1186/s12910-021-00648-w.

21. Boyd A, Shiells K, Di Cara N, Skatova A, Davis OSP, Haworth CMA, Skinner AL, Thomas R, Tanner AR, Macleod J, Timpson NJ. (2019). Participant acceptability of 'digital footprint' data collection strategies: evidence from the ALSPAC birth cohort study. Bristol, UK: University of Bristol.

22. Boyd A, Thomas R, Hansell AL, Gulliver J, Hicks LM, Griggs R, Vande Hey J, Taylor CM, Morris T, Golding J, Doerner R. Data Resource Profile: The ALSPAC birth cohort as a platform to study the relationship of environment and health and social factors. Int J Epidemiol. 2019;48(4):1038-9k. https://doi.org/10.1093/ije/dyz063.

23. Di Cara N, Boyd A, Tanner A, Al Baghal T, Calderwood L, Sloan L et al. Views on social media and its linkage to longitudinal data from two generations of a UK cohort study. Wellcome Open Res. 2020;5(44). https://doi.org/10.12688/wellcomeopenres.15755.

24. Skatova A, Shiells K and Boyd A. Attitudes towards transactional data donation and linkage in a longitudinal population study: evidence from the Avon Longitudinal Study of Parents and Children [version 2; peer review: 2 approved]. Wellcome Open Res. 2021;4:192. https://doi.org/10.12688/wellcomeopenres.15557.2.

25. Boyd A, Golding J, Macleod J, Lawlor DA, Fraser A, Henderson J, Molloy L, Ness A, Ring S, Davey Smith G. Cohort profile: the 'children of the 90s'—the index offspring of the Avon Longitudinal Study of Parents and Children. Int J Epidemiol. 2013;1;42(1):111–27. https://doi.org/10.1093/ije/dys064.

26. Fraser A, Macdonald-Wallis C, Tilling K, Boyd A, Golding J, Davey Smith G, Henderson J, Macleod J, Molloy L, Ness A, Ring S. Cohort profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. Int J Epidemiol. 2012;16;42(1):97–110. https://doi.org/10.1093/ije/dys066.

27. Northstone K, Lewcock M, Groom A, Boyd A, Macleod J, Timpson N, Wells N. The Avon Longitudinal Study of Parents and Children (ALSPAC): an update on the enrolled sample of index children in 2019. Wellcome Open Res. 2019; 4. https://doi/org/10.12688/wellcomeopenres.15132.1.

28. Cornish RP, Macleod J, Boyd A, Tilling K. Factors associated with participation over time in the Avon Longitudinal Study of Parents and Children: a study using linked education and primary care data. Int J Epidemiol. 2021;50(1):293–302. https://doi.org/10.1093/ije/dyaa192.

29. Birmingham K. Pioneering ethics in a longitudinal study. Policy Press; 2018.

30. Audrey S, Brown L, Campbell R, Boyd A, Macleod J. Young people's views about consenting to data linkage: findings from the PEARL qualitative study. BMC Med Res Methodol. 2016;16(1):1–3. https://doi.org/10.1186/s12874-016-0132-4.

31. Audrey S, Brown L, Campbell R, Boyd A, Macleod J. Young people's views about the purpose and composition of research ethics committees: findings from the PEARL qualitative study. BMC Med Ethics. 2016;17(1):1–0. https://doi.org/10.1186/s12910-016-0133-1.

32. ALSPAC. LIFE@26 Questionnaire. http://www.bristol.ac.uk/media-library/sites/alspac/documents/questionnaires/Life@26YPQuestionnaire.pdf

33. Timulak L. Meta-analysis of qualitative studies: A tool for reviewing qualitative research findings in psychotherapy. Psychother Res. 2009;19(4–5):591–600. https://doi.org/10.1080/10503300802477989.

34. Braun V, Clarke, V. Using thematic analysis in psychology. Qual Res Psychol. 2006;3(2):77–101. https://doi.org/10.1191/1478088706qp063oa.

35. Williams ML, Burnap P, Sloan L, Jessop C, & Lepps H. Chapter 2: Users' Views of Ethics in Social Media Research: Informed Consent, Anonymity, and Harm. Advances in Research Ethics and Integrity. 2017;27–52. https://doi.org/10.1108/s2398-601820180000002002.

36. Understanding Patient Data. Patient Data. Finding the best set of words to find. Summary of data. 2017. https://understandingpatientdata.org.uk/sites/default/files/2017-04/Data%20vocabulary_Good%20Business%20report%20March%202017_0.pdf.

37. Kosinski M, Stillwell D, Graepel T. Private traits and attributes are predictable from digital records of human behavior. Proceedings of the National Academy of Sciences. 2013; 110(15):5802–5805. https://doi.org/10.1073/pnas.1218772110.

38. Burton PR, Murtagh MJ, Boyd A, Williams JB, Dove ES, Wallace SE, Tasse AM, Little J, Chisholm RL, Gaye A, Hveem K. Data Safe Havens in health research and healthcare. Bioinformatics. 2015;31(20):324. https://doi.org/10.1093/bioinformatics/btv279.

39. Al Baghal T. Linking survey and social media data. Understanding Society Working Paper Series. No. 2020-04, March 2020. https://www.understandingsociety.ac.uk/sites/default/files/downloads/working-papers/2020-04.pdf.

40. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care. 2007;9(6):349–357. https://doi.org/10.1093/intqhc/mzm042.

41. Boyd A, Tilling K, Cornish R, Davies A, Humphries K, Macleod J. Professionally designed information materials and telephone reminders improved consent response rates: evidence from an RCT nested within a cohort study. Journal Clin Epidemiol. 2015;68(8):877–87. https://doi.org/10.1016/j.jclinepi.2015.03.014

42. Bray I, Noble S, Boyd A, Brown L, Hayes P, Malcolm J, Robinson R, Williams R, Burston K, Macleod J, Molloy L. A randomised controlled trial

comparing opt-in and opt-out home visits for tracing lost participants in a prospective birth cohort study. BMC Med Res Methodol. 2015;15(1):52. https://doig.org/10.1186/s12874-015-0041-y

43. UKRI. Innovation and development in longitudinal studies: outputs from the 'UK Population Lab' programme. https://esrc.ukri.org/news-events-and-publications/publications/corporate-publications/innovation-and-development-in-longitudinal-studies-outputs-from-the-uk-population-lab-programme/.

Abbreviations

ALEC:	ALSPAC Law & Ethics Committee
ALSPAC:	Avon Longitudinal Study of Parents and Children
EMA:	Ecological Momentary Assessment
GPS:	Global Positioning System
LPS:	Longitudinal Population Study
OCAP:	Original Cohort Advisory Panel
PEARL:	Project to Enhance ALSPAC through Record Linkage