ISRCTN16910438 https://doi.org/10.1186/ISRCTN16910438

Exploring the public's opinion on the use of their data in biobanking, linked to routine medical records, and the use of artificial intelligence in care and medical research

Submission date 06/01/2026	**Recruitment status** Not yet recruiting	[X] Prospectively registered [_] Protocol [_] Statistical analysis plan
Registration date 14/01/2026	**Overall study status** Ongoing	[_] Results [_] Individual participant data
Last Edited 09/01/2026	**Condition category** Other	[X] Record updated in last year

Plain English summary of protocol

Background and study aims

Technology is developing quickly. This study funded by NIHR Newcastle Biomedical Research Centre aims to understand what people think about three important topics in modern medical research:

Biobanking: This involves collecting and storing biological samples, like blood or tissue from an operation or a biopsy. This can be collected as part of research or when people have tests or treatments as part of routine NHS care.

Data from routine healthcare: The NHS routinely stores some samples sent for tests for up to 30 years. With the development of electronic health records and increasing linkage between GP records and hospital records, this creates complex medical data which is a powerful potential source of new research. We will study how people feel about this type of research including how it is collected, analysed and who has access to it.

Artificial intelligence (AI) in research: AI is developing quickly and has great potential but also ethical considerations about how it is used in research settings. This is particularly when real-world data is required to ‘train’ AI technologies for it to work properly. This is especially the UK government is encouraging implementation of AI in the NHS via a single point of access (NHS Federated Data Platform).

With the plans of accessing personal health data to support AI development and its use in healthcare settings, this research study seeks to explore not only whether people consent to data use, but also whether they perceive these practices as legitimate and aligned with their expectations. Understanding these perceptions is essential for shaping policies and strategies that foster trust and ensure the sustainability of data-driven research.

One of my primary aims in this research is to broadly explore the public attitudes towards the use of their personal data in biobanking linked to EHRs, and the role of emerging AI technologies in healthcare and clinical research.

My secondary aim is to identify how socioeconomic status works to impact these attitudes.

Who can participate?
Anyone over the age of 18 years, resident in the UK, who can understand English, has internet access for completion of the online survey and can be approached for participation in interviews.

What does the study involve?
The study involves participants to volunteer to complete an anonymous survey online. These are aimed at the general public and in particular, those with a pre-existing diagnosis of one or more chronic inflammatory conditions such as inflammatory bowel disease, rheumatoid arthritis, psoriasis, or autoimmune hepatitis.
The second part of the study involves the selection of participants to take part in semi-structured interviews to talk in detail about their views on the issues listed above. Participants can express an interest to take part in interviews. The researcher will select a total of 30 from those who belong to underrepresented groups in the survey and the general populations, e.g. those from ethnic minority background, extremes of age groups etc.

What are the possible benefits and potential risks of participating?
While there are no immediate benefits for participating, we hope that this research will be able to guide researchers, clinicians and policymakers in best practice for health data sharing. We will be discussing sharing health data for research purposes including for AI. Some participants may find this upsetting, for example, if they have been a victim of data leakage in the past. For those who take part in the interviews, participants can be sent some topics for discussion beforehand, before committing themselves.

Where is the study run from?
The study is being run online for the anonymous surveys so participants can be based anywhere in the UK, and can complete the survey from the comfort of their own home or wherever that’s convenient for them. The interview will be conducted either online or in person which would be around the North East of England.

When is the study starting, and how long is it expected to run for?
March 2026 to April 2028

Who is funding the study?
This study is funded by NIHR Newcastle Biomedical Research Centre (UK)

Who is the main contact?
Dr Chaonan Dong, chaonan.dong2@newcastle.ac.uk

Contact information

Type(s)
Principal investigator, Public, Scientific

Contact name
Dr Chaonan Dong

ORCID ID
https://orcid.org/0000-0001-5337-5779

Contact details

William Leech Building
Newcastle University
Newcastle upon Tyne
United Kingdom
NE2 4HH
+44 (0)7533150691
chaonan.dong2@newcastle.ac.uk

Additional identifiers

Study information

Scientific Title
Perceptions of the public on the use of biobanking, large datasets from routine health records, and artificial intelligence in healthcare and research

Acronym
PUBLIC

Study objectives

Ethics approval required
Ethics approval required

Ethics approval(s)
notYetSubmitted

Primary study design
Observational

Secondary study design
Cross sectional study

Study type(s)

Health condition(s) or problem(s) studied
Mainly focused on but not exclusive to chronic inflammatory conditions: GI (inflammatory bowel disease), musculoskeletal (inflammatory arthritis such as rheumatoid arthritis, ankylosing spondylitis), skin (psoriasis, eczema, hydradenitis), and liver (autoimmune hepatitis, primary sclerosing cholangitis, primary biliary cirrhosis).

Interventions
This research will use a mixed-methods approach where an exploratory sequential design will be adopted.

Quantitative phase:
The first stage of this research will involve quantitative data collection in the form of mass surveys. A scoping review of available literature will be conducted focusing on public perception

/attitude towards use of personal data in biobanking linked to data from routine healthcare, and usage of AI technologies in healthcare and research. This will guide questions asked in the survey of the public.

The population to be focused on in this research study is patients with immune mediated inflammatory diseases (IMIDs) including: inflammatory bowel disease (IBD), autoimmune liver diseases, inflammatory arthritis, and chronic inflammatory skin diseases. These conditions often overlap with each other, and patients are increasingly recognized stakeholders in healthcare and research due to increasing prevalence, complexity of the disease course, regular contact with healthcare services, increasing research interests and diverse datasets generated when involved in research.

Surveys will collect information including age, occupation, education level, ethnicity, gender, and their medical backgrounds. This will help to determine if perceptions of the issues are specific and/or related to demographic status. The surveys will be distributed with the aim to capture diversity. Efforts will be made to target underrepresented populations in later stages.

Surveys will be disseminated via platforms such as NIHR Be a Part of Research, charities for the above conditions such as Crohn's & Colitis UK, British Arthritis Society, and British Skin Foundation. There are two local patient groups – PIMS (Patient and Public Involvement and Engagement in Musculoskeletal reSearch) and Liver North that the surveys could also be disseminated by.

Following dissemination of the survey, preliminary analysis will be conducted using SPSS/R to shape interview guides in preparation for qualitative phase of the research study. The aim of the survey is to capture diversity on views and attitudes found, then uses semi-structured interviews to plug any gaps.

Patient and public involvement:
This research study is within the theme of Informatics and Precision Care for an Aging Population (IPCAP) at NIHR Newcastle Biomedical Research Centre (BRC). The public contributor group from the IPCAP theme will review pilot draft of the survey first, providing feedback before rolling it out at a wider scale. This group will also be approached for providing feedback on interview guides for qualitative phase of the study.

Qualitative phase:
In England, people in the most deprived neighbourhoods, certain ethnic minority (Indian, Pakistani, Black African, Black Caribbean, other Asian, and mixed), and inclusion health groups getting multiple long-term health conditions earlier than the least deprived communities, spending more years in ill health and dying sooner. People from low socio-economic backgrounds and certain ethnic minorities are more likely to have worse outcomes. Equally, these groups of people are less likely to participate in both healthcare and biomedical research. It has been recognised and emphasised the importance of involving these underserved groups in research, therefore purposive sampling for the qualitative part of my research will be focused on these populations.

Interview details:
Participants eligibility: over 18s, live within North East of England, able to speak English
Participants will be offered in-person interviews. All interviews (n=30) will be audio-recorded and semi-structured to obtain rich qualitative data. Interview topics are due to be determined – they will be based on survey analysis.

Sampling:
Potential participants will be sourced through various communities. For example, charities that support vulnerable groups of people such as Roseworth Big Local in Stockton-on-Tees, Youth Matters Gateshead, Sunderland Interfaith Forum, or Building Futures East in Newcastle upon Tyne. Efforts will be made to target populations who are of ethnic minorities, extremes of age groups, and low socio-economic backgrounds.

Data analysis:
Interpretative Phenomenological analysis (IPA) will be undertaken for the qualitative part of the research, as it seeks knowledge and attitudes formed through the individual's recounts of their lived experiences, influenced by factors such as cultural backgrounds, education, age, or personal experience with AI.

After completing each participant interview, the researcher will record their reflections to capture any initial thoughts following the interview. All transcriptions will be done by the researcher to allow for full data immersion. Transcripts will be checked for accuracy and any identifiable information deleted. Each participant will be assigned with a code to avoid de-identification. Data collection and analysis will occur simultaneously to allow emerging themes to inform subsequent interviews.

The researcher is a clinician therefore care will be undertaken when conducting interviews to note any power dynamics between potentially vulnerable groups and the researcher.

Data will be analysed during reflective thematic analysis. NVivo15 software will be used to support this process.

Integration points
The integration of quantitative and qualitative data of the research study will occur at three points. First, using an interactive approach, the researcher will use initial data from surveys to develop an interview topic guide to address any unexpected findings or outlying data. Second, using the same approach, the researcher will use initial interviews to refine the topic guides for subsequent interviews. Third, themes generated from interviews will be defined and named, and linked to survey findings. This would identify points of convergence and divergence with the aim of using qualitative data to add nuance and explanation to the quantitative results.

Intervention Type
Other

Primary outcome(s)
1. A nuanced understanding of the key concerns and trust factors of the public on the key issues raised, measured using descriptive statistics from survey answers, then thematic analysis of interview transcripts. These will be integrated together where the qualitative data should add depth and richness to quantitative data, at descriptive statistics at 6 months after anonymous survey goes live, then every 5 or 6 interviews for thematic analysis

Key secondary outcome(s))

Completion date
20/04/2028

Eligibility

Key inclusion criteria
1. Aged 18 years or above
2. Resident in the UK
3. Able to understand and speak English

Healthy volunteers allowed
Yes

Age group
Mixed

Lower age limit
18 years

Upper age limit
100 years

Sex
All

Total final enrolment
0

Key exclusion criteria
1. Less than 18 years of age
2. Not a resident in the UK
3. Does not understand English

Date of first enrolment
02/03/2026

Date of final enrolment
31/12/2027

Locations

Countries of recruitment
United Kingdom

England

Study participating centre
NIHR School for Public Health Research at University of Newcastle upon Tyne
Newcastle University
Claremont Road
Newcastle upon Tyne
England
NE1 7RU

Sponsor information

Organisation
NIHR Newcastle Biomedical Research Centre

ROR
https://ror.org/044m9mw93

Funder(s)

Funder type

Funder Name
NIHR Newcastle Biomedical Research Centre

Alternative Name(s)
Newcastle Biomedical Research Centre, Newcastle NIHR Biomedical Research Centre

Funding Body Type
Private sector organisation

Funding Body Subtype
Research institutes and centers

Location
United Kingdom

Results and Publications

Individual participant data (IPD) sharing plan

IPD sharing plan summary
Not expected to be made available