

Research feasibility and ethics in Scottish new-born blood spot archive.

David Porteous (✉ david.porteous@ed.ac.uk)

University of Edinburgh <https://orcid.org/0000-0003-1249-6106>

Sarah Cunningham-Burley

University of Edinburgh

Daniel McCartney

Archie Campbell

University of Edinburgh <https://orcid.org/0000-0003-0198-5078>

Robin Flaig

<https://orcid.org/0000-0001-8007-1279>

Clare Orange

Carol Porteous

Mhairi Aitken

Ciaran Mulholland

Sara Davidson

Selina McCafferty

Lee Murphy

Nicola Wrobel

Sarah McCafferty

Karen Wallace

David StClair

Shona Kerr

<https://orcid.org/0000-0002-4137-1495>

Caroline Hayward

<https://orcid.org/0000-0002-9405-9550>

Andrew McIntosh

The University of Edinburgh <https://orcid.org/0000-0002-0198-4588>

Cathie Sudlow

Riccardo Marioni

University of Edinburgh

Jill Pell


University of Glasgow

Zosia Miedzobrodzka

Keywords:

Posted Date: March 14th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1315662/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Version of Record: A version of this preprint was published at Communications Medicine on October 6th, 2022. See the published version at <https://doi.org/10.1038/s43856-022-00189-2>.

Abstract

New-born heel prick blood spots are routinely used to screen for inborn errors of metabolism and life-limiting inherited disorders¹. The Danish Newborn Screening Biobank (DNSB) has shown that nucleic acid, metabolic and protein assays are feasible on archived new-born dried blood spots². Denmark is currently the only country in the world where nation-level linkage of such secondary data to health records is approved². In the new era of health data science and predictive medicine, the potential value of secondary data from new-born blood spot archives merits wider ethical consideration and assessment of feasibility for public benefit. Early life exposures and behaviours set health trajectories in childhood and later life. The new-born blood spot is potentially well placed to create an unbiased and cost-effective population-level retrospective birth cohort study. Scotland has retained new-born blood spots, commonly known as Guthrie cards, for all children born since 1965, around 3 million in total. However, a moratorium on research access is currently in place, pending public consultation. We conducted a Citizens' Jury, with Jurors delivering an agreed verdict that conditional research access to the Guthrie Cards was in the public interest. The Chief Medical Officer for Scotland authorised restricted lifting of the current research moratorium to allow a feasibility study. In support of a wider public consultation, Guthrie cards from consented Generation Scotland volunteers^{3,4} were retrieved and their potential for both epidemiological and biological research demonstrated.

Study Setting

A Scottish Government commissioned review had previously considered the ethical, legal and social implications of research access to the Guthrie Card archive⁵. In Scotland, the Guthrie card archive is considered part of the medical record and subject to the provisions of the Data Protection Act 1998. The report recommended that the Guthrie collection should be treated as both tissue and data in legal governance. The report also recommended wider public engagement on the issues. There are many ongoing research studies in Scotland that have obtained consent from study participants for NHS health record linkage. Others have explicit consent to use participants' Guthrie spots. Nevertheless, a research study moratorium remains in place. The purpose of this study was to provide supporting evidence for a planned Scottish Government commissioned public consultation.

Citizens' Jury

In June 2017, we conducted a 2-day Citizens' Jury in partnership with the polling organisation, IPSOS MORI following INVOLVE guidelines⁶. The question posed was "Would research access to the Guthrie Card heel prick blood tests be in the public interest?" The responses were systemically analysed to identify substantive themes and key messages (see Methods for details and Table 1 and SI Table 1 for summary conclusions). Overall, jurors were very positive about the importance of health research and a willingness to allow their medical records to be used for such purposes. However, reservations were expressed about possible involvement of commercial actors and also about data security. After hearing from expert witnesses, the jurors were overwhelmingly positive in terms of expressing conditional support for the use of Guthrie cards in research. Jurors felt that there was a 'clear public interest' case. This case rested on the potential advancements in the identification and/or prevention of disease, development of new treatments and ultimately, improved population health. For some jurors, it seemed inherently wrong and wasteful that these blood spots could not

to be used for research. Nonetheless, questions and concerns were raised around three main themes: data protection and security; control and oversight; and the possibility of overstressing the allowed uses. Although Jurors were reassured by how NHS data were protected, they were acutely aware that mistakes do happen and that any leak or misuse could be serious. Some concern was expressed about whether or not an individual would wish to know they had been identified as at risk of a health-related condition and the obligations of researchers in this regard. A small number of jurors expressed concern about the risk of the use of the blood spots by, for example, insurance companies, and the need to restrict use to *bona fide* health research. A small number of jurors expressed concern that there would be incremental increases in use, beyond the approved types of health research. At the end of Day One, 10 out of 19 jurors considered the use of Guthrie Cards for research to be in the public interest with no concerns and 9 answered that they had minor concerns.

Day Two provided the opportunity for more focused, deliberative discussion on the considerations that Jurors felt were key in deciding whether or not research access to the Guthrie Cards was in the public interest.

Following further, more cautionary, witness sessions, jurors remained generally positive in their views although there was more discussion about possible commercial access and how this might be controlled. Jurors' discussion moved from 'for' or 'against' arguments to more nuanced discussion in the deliberative phase.

Jurors were asked to identify and agree on considerations they felt were key in deciding whether or not research access to the Guthrie Cards was in the public interest. Five key considerations were identified and a consensus statement achieved for the first four (see SI Table 1). The overall verdict of the Citizens' Jury was that access to the Guthrie Card heel prick blood samples would be in the public interest. The distribution of opinion recorded at the end of event questionnaire varied only slightly after the deliberation, with 8 stating that they had no concerns, 9 that they had minor concerns and 1 who had major concerns. This verdict was accompanied by a clear articulation of four specific conditions and a fifth recommendation, as summarized in Table 1.

Stakeholder meeting and feasibility study approval

The findings from the Citizens' Jury were shared at a stakeholder meeting convened by the Scottish Government's Chief Scientist Office (CSO) in January 2019. The meeting concluded that further work should be conducted before wider public consultation. That consultation would not only have to consider whether research access with suitable oversight and governance would be in the public interest, but also that the proposed research was feasible and justified. Two categories of feasibility study were recommended, epidemiological and biological. A letter approving such a feasibility study was provided by the Chief Medical Officer for Scotland in September 2019. The Caldicott Guardians of NHS Greater Glasgow and Clyde and NHS Tayside granted approval in January 2020 and March 2020, respectively. A substantial amendment to the Research Tissue Bank approval for Generation Scotland, a family and population based cohort with consent to link to health records and conduct genetic studies, to cover the feasibility study was submitted to the Research Ethics Committee and approved in March 2020.

Epidemiological potential

We sought to evaluate the feasibility of deriving a population-level e-cohort from 1965 onwards through linkage of data held on the Scottish Guthrie card archive to routine NHS Scotland health records. In the 1970's, Scotland progressively introduced the Community Health Index (CHI), a unique identifier for all persons born in

Scotland or registered with a GP in Scotland^{7,8}. Each individual's CHI is now associated with every NHS Scotland transaction. The CHI can be used to link all patient interactions, including primary, secondary and tertiary care, vaccination, laboratory results and prescriptions. All Scottish Guthrie cards have been catalogued from 2000 onwards, but not so for earlier cards. Indeed, there is no formal record of the way in which earlier cards were collected and archived. To establish the physical state of the cards and associated identifying information, NHS Greater Glasgow and Clyde Biorepository staff, who have Scotland-wide responsibility for the archive, inspected 30 boxes of Guthrie cards dating from 1965 to 1999 (see Methods, Table 2 and SI Figure 1 and 2).

Each box contained circa 4,000-6,000 cards in bundles by area health board, main hospital and date. Boxes and bundles were visually inspected for their physical state and origin. The Guthrie card archive as a whole is stored at ambient temperature, but in the early days there was no unified approach to the type of dried blood spot card used nor the storage conditions. Some batches of cards were visually damaged by damp and there was only partial correspondence between the box identifier and the contents (see Methods).

A random sub-set of Guthrie cards (circa 1% and circa 1,200 in total) were inspected in more detail. Cards were checked for feasibility of linkage to NHS health records by noting whether or not the following types of information were included: child name, date and place of birth, hospital number, maternal and paternal name. Under the terms of the Caldicott Guardian's approval, only summary data was tabulated. No personal identifying information was collected or stored. We observed significant variation in the Guthrie card information content, both over time and between different health boards. Despite these inconsistencies, for the vast majority of Guthrie cards from 1965 onwards, there was sufficient information to provide direct or high-confidence probabilistic matching to CHI-linked records. This would provide a strong basis for anonymised, population level epidemiological study, or of individual level research, subject to consent and other relevant permissions. That said, for the earlier cards to be systematically catalogued by time and birthplace, digitally searchable and linked to routine NHS health records would require substantial investment of time and effort.

Biological potential

Any proposed assay of the Scottish Guthrie card archive of dried blood spots must take into consideration the finite nature of the resource, the scientific added value and public benefit gain. Pilot studies conducted between 2011- 2013 on de-identified sample punches demonstrated that sufficient quantity and quality of DNA could be extracted for whole genome sequencing from 128 of the 136 cards tested (94%) with closely comparable success rates between the earliest (1965) and the most recent (2012) cards sampled (see Methods and Table 3).

The next objective was to evaluate the feasibility of recovering epigenetic signatures from the Guthrie card blood spots of Generation Scotland volunteers through genome-wide DNA methylation array technology (see Methods). This assay was selected because it consumes a very small amount of material and would create a multipurpose dataset of high value across a wide range of traits with predictive value^{9,10}.

Generation Scotland was the approved source of test samples for this purpose because volunteers gave broad consent for access to their medical records and to genetic and other biomarker studies^{3,4}. Genome-wide DNA methylation data from peripheral blood DNA collected at recruitment baseline (2006-2011) was available for

comparison and validation. Information on maternal smoking behaviour was also available as a test case for epigenetic validation using gene specific and genome wide scores previously linked to smoking^{11,12}.

Amongst the contents of the first 30 boxes examined for record linkage potential, we identified 92 Guthrie cards from consented members of the Generation Scotland research cohort. A condition of this feasibility study approval by the Caldicott Guardians was that cards could only be sampled if one of the four blood spots was left untouched. This condition meant that we were only able to take punches from 58 of the 92 cards retrieved (60%). DNA from these 58 punch samples (dating from 1983-1989) was extracted and a genome-wide DNA methylation study conducted (see Methods). Results were compared to matched blood samples collected from the same volunteers at their baseline research clinic visit. There was excellent genetic congruence between the perinatal and adult samples (SI Figure 3 and 4, SI Table 2).

In Generation Scotland, we have data collected at baseline on smoking behaviour. Using this information in combination with genome wide methylation data from Guthrie cards showed that DNA methylation signatures at birth reflected maternal smoking, replicating previous studies¹¹ (Figure 1).

Discussion

Strengths and limitations

The Citizens' Jury strongly agreed that research access to the Guthrie cards was in the public interest. Yet, the articulation of key conditions provides clear evidence that careful and transparent regulation must be in place (Table 1). These results, based on in-depth, qualitative data, should inform future public engagement to ensure a wider mandate for research use of Guthrie cards. The findings of the Citizens' Jury illustrate how the public can engage in nuanced ways with the issue in hand and make recommendations relevant for policy. The findings, including the cautions and concerns, reflect other research that has explored public views on health data, health research, public benefits and data sharing and how these might be addressed^{14,15,16}. That said, whereas Citizen Juries are carefully balanced to enable in-depth exploration of the nuances of individual perspectives, the modest number of jurors means that they cannot fully represent the wider public. Wider public consultation is recommended.

Some of the suggested safeguards and other recommendations are covered under the respective responsibilities of the Caldicott Guardians, Public Benefit and Privacy Panel (part of the information governance structure in Scotland that scrutinizes requests for access to health and other personal data) and NHS Research Ethics Committees (who protect the rights, safety, dignity and wellbeing of participants and collectively safeguard and allow legitimate access to health data in the public interest), but others would require further action. Following on from the Citizens' Jury, the Stakeholder Meeting recognised that a bespoke and dedicated governance and oversight structure would be advisable, at least in the development phase of any proposed lifting of the moratorium on research access, despite the range of protections already in place for the use of health data¹⁷. The consensus view from Stakeholders was that a system for "opt-out" would need to be developed in line with current developments in secondary use of health data, especially for studies linked to health records¹⁷.

The follow-on feasibility study has established the potential value of the circa 3 million Scottish Guthrie card archive for a) population-level epidemiological studies, and b) DNA-based biomarker studies of people born since 1965. This provides an evidence base for further public and stakeholder consultation on establishing a managed NHS resource for research.

In the process, our study did highlight practical issues regarding the currently uncatalogued, pre-2000 archive, constituting circa 2 million cards. Significant investment would be required to retrieve these boxes from long term storage, catalogue their contents, render them searchable and return these to safe storage for future retrieval in an organized manner. Despite these caveats, the feasibility study demonstrated that the archive provided sufficient information for CHI seeding and linkage to NHS Scotland routine data sets. This could therefore provide the basis for population-level, epidemiological health data research. A fully catalogued, searchable data set with independent oversight for regulated access and use, plus obligatory return of processed data and analysis would maximise the public benefit.

The feasibility study also demonstrated the potential value of the Scottish Guthrie card archive for genetic (Table 3) and epigenetic studies (Figure 1). DNA array technologies now require very small amounts of input material and are readily scalable. If suitably indexed, a wide range of trait-defined or population-level studies would be possible. Indeed, we have shown that the epigenetic scores correlate well with circulating proteome scores and predict disease traits in Generation Scotland¹⁸. There is thus scope to investigate a wide range of health issues where ground-state knowledge at or around birth would be valuable. The small amount of material required for DNA-based studies would leave material intact for other future assays. Again, obligatory return of processed data and analysis to the archive would maximise the public benefit. Retaining an archive of extracted DNA for additional and unforeseen uses would be prudent. These might include assays for the presence of non-host nucleic acid assays indicating prevalent infection. Feasibility studies to establish the possibility of other assays such as blood proteins, metabolites or environmental pollutants should be considered. In all cases, the public benefit would have to be seen to outweigh the cost and the consequential depletion of the archive.

Taking together all elements of the study, there is now an evidence base for wider public consultation to inform decisions about whether or not to lift the current research moratorium and if so under what ethics, oversight, governance and engagement terms and conditions. The public interest will however require that public support is established and maintained through transparency of purpose and demonstration of public benefit.

Wider implications

Our study has wider implications. Birth cohorts are a bedrock of much epidemiological research but are limited by cost, attrition and bias. Unbiased routine data collections have proved crucial in measuring the impact on health of environmental exposures, such as smoking, poisoning, pollution and infection, and policy changes introduced to mitigate these exposures. They do not however take account of biological factors. We now know that genetic factors influence resistance or vulnerability to all of these exposures to a greater or lesser extent. Population-level, new-born blood spot archives are uniquely positioned to integrate exposure data with biological data that is not compromised by selection bias. It would be possible to design well-powered studies against biological (including genetic) and non-biological (sociodemographic and environmental) criteria. The Scottish Guthrie card archive has the potential to pioneer such an approach. In so doing, best practice could be

shared, hypotheses tested, and published findings arising from conventional cohort studies replicated or refuted. For public support, robust and transparent governance must be put in place, with onward data sharing for legitimate use to maximise public benefit.

Methods

Citizens' Jury

In June 2017, we brought together a small, diverse group of citizens to address the question: "Would research access to the Guthrie Card heel prick blood tests be in the public interest, and, if so, under what conditions?" A Citizens' Jury is a well-established method to enable public participation in policy making, allowing informed deliberation on an issue and the provision of recommendations. Our Citizens' Jury followed best practice for such deliberative public engagement⁶. First, we convened a steering group to provide oversight of the materials prepared for the jurors and to identify a range of expert witnesses to give evidence. Next, we consulted a Patient Participatory Involvement and Engagement (PPIE) panel to review these materials. The academic researchers did not involve themselves directly in the Jury process, other than to provide evidence or observe. The recruitment, facilitation and analysis was conducted by Ipsos MORI, Scotland, to preserve neutrality. Jurors met together for two day-long sittings to hear evidence (neutral, for and against), deliberate and reach conclusions.

Using a quota sample, a representative pool of the adult public was recruited in terms of sex, age, working status and social grade. Additional quotas were set to ensure sufficient representation of people with children under the age of five, and people with a family history of a medical condition. An attitudinal quota was set to ensure inclusion of people with varied levels of trust in public, private and third sector organisations, as previous research has found this to be a significant factor underpinning views of data sharing and use¹⁴. A total of 20 were recruited of whom 19 actually participated on day one, and 18 returned for the second day a week later. Jurors were given monetary recompense for taking part in each sitting.

Day One started with a warm up session sharing their views on health research and health relevant information, followed by evidence from various experts to stimulate discussion of issues around research use of the Guthrie Cards. Day Two included further expert witnesses but more time for jurors own deliberations and to arrive at a conclusion on the key question. Facilitation tools, such as speed dating techniques, meant that all jurors could express and reflect on their own views as well as those of others, building up to group and plenary deliberations. The first day focused on more general discussion and information sharing; the second day involved detailed deliberation of the key question and delivering of the verdict. Both days were audio recorded and transcribed for subsequent analysis. A short questionnaire was administered at the end of each day to gauge individual level views and thinking.

Background information was provided by authors SCB and DJP during the morning of Day One. Over the two days, six witnesses were called to give evidence and answer jurors' questions. These comprised health care professionals and scientists, a Caldicott Guardian (a person responsible for protecting the confidentiality of people's health and care information and making sure it is used properly) and a Genewatch spokesperson (representing not-for-profit groups that monitor developments in DNA technologies). Jurors also had access to

a short video and other written information about comparative deliberations and policy in California and Denmark. Having heard on Day One some of the health research opportunities uniquely possible if access were granted, an introduction to some of the social, ethical and legal issues, and how the NHS protects privacy, Jurors heard more opposing and critical views on Day Two, with one witness cautioning against allowing research access and another the importance of not compromising the newborn screening programme itself.

Guthrie card documentation and sampling

Epidemiological feasibility

The Scottish Guthrie card archive is stored in around 900 boxes each containing circa 3,000 cards in a single secure location under the authority of the Director of the Scottish Newborn Screening laboratory and custodianship of the NHS Research Scotland Greater Glasgow and Clyde Biorepository. A unified digital record is in place from 2000 onwards, but the content and consistency of information available from older cards was not known at the outset. The design and information content of the Guthrie card was known to have varied over time, but not documented. We were aware that some cards had suffered water damage prior to assembling the nation-wide archive. We also know that for a period of time some cards had been autoclaved before storage. For the vast majority, the Guthrie cards have simply been stored at room temperature. The unified Guthrie card archive management database has limited, high-level summary information on the contents of each box.

We obtained permission from the Caldicott Guardians of NHS Research Scotland Greater Glasgow and Clyde and of Tayside to retrieve representative boxes from the NHS Research Scotland secure archive. Permission was given for a) examination and documentation of a sub-sample from each box to provide a snap-shot of the information attached that might be required from the purposes of linkage to other routine health records and b) sampling of cards corresponding to consented members of Generation Scotland.

Thirty boxes representing each decade from 1965 to 1999 were retrieved and examined by NHS Scotland staff at the NHS Research Scotland Greater Glasgow and Clyde Biorepository. Only summary information was shared with the rest of the study team.

The organisation and information content by box varied over time (SI Figure 1). In some boxes there were unlabelled bundles of cards (SI Figure 1D), but most had 4 labelled foolscap sub-boxes (SI Figure 1B) with multiple, date-labelled bundles of cards (SI Figure 1C).

We drew up a list of potential information that the Guthrie cards might carry from which to judge the feasibility of conducting epidemiological studies by linkage to NHS Scotland routine medical record and potentially additional consented data from research subjects.

For each box, we collected the following information: Box ID; Area Health board; Hospital; Type of card; Date of test; Child forename; Child surname; presence or absence of Community Health Index identifier. No personal information was recorded. Each box took 2 members of staff working in tandem approximately 2 hours to document.

Next, circa 1 in 100 cards from each box were examined in detail and the presence or absence of the following features documented: Child DOB; Additional comments on card; Number of bloodspots; Size of spots; Mother's CHI; Mother's forename, surname and birth name; Mother's date of birth; Address; Postcode; Whether the cards had been autoclaved prior to archiving; Any other comments.

Cards from 1965 had very little information on them and in many cases did not even record the sex of the baby. Information content increased progressively over time. By the 1990s, the sex of the child and home address were generally recorded.

Retrieval of cards from consented volunteers

Over 24,000 Generation Scotland (GS) volunteers were recruited as adults between 2006 and 2011³. All were born before 1993, before the digital recording of Guthrie cards began. Consent for linkage to medical records was optional, but was given by 98% of volunteers. They were asked to give information about their place of birth (country and council area). A total of 8,703 volunteers with linkage consent were born in Glasgow or Tayside area health boards between 1965 and 1992. A list of names, birthplace and date of birth was extracted from the GS database and sent to the NHS Greater Glasgow and Clyde Biorepository to look for matching cards in selected boxes. Pseudonymous ids were added to the list so that any samples from matching cards could be labelled and linked back to the GS database after genotyping. Ninety-two matching cards were identified amongst Guthrie cards from Tayside. Of these, 58 were usable for punching having fulfilled the prerequisite of leaving one spot intact (SI Figure 2). Six to ten punches were taken from each card, placed in vials labelled with a pseudonymous ID for matching to the samples donated at baseline by each Generation Scotland volunteer, and couriered to the Edinburgh Clinical Research Laboratory Genetics Laboratory for DNA analysis.

Biological feasibility

Genetic analysis of de-identified cards

Unlike the Danish Newborn Screening Biobank (DNSB), the Scottish Guthrie card archive comprises a variety of paper types and storage conditions, particularly for older cards. To establish the effect this might have on the recovery of analysable DNA, a pilot study in 2011-2013 was undertaken on a de-identified set of 136 Guthrie cards dated from 1965-2012 (Table 3). The study was mandated by the Scottish Chief Scientist Office, following a favourable opinion from the Scottish Legal Office and Research Ethics Committee. De-identified cards were provided by the Scottish bloodspot screening laboratory. DNA was extracted from 3mm punches using the Sigma ENA kit. Yields varied from sample to sample, but there was no significant effect of date of birth and sufficient material was obtained for Sanger DNA sequencing in 94% of samples (Table 3).

Sampling of cards

Of the 92 Guthrie cards matched to GS volunteers, 58 (63%) had sufficient dried blood spot material remaining to take 3 mm diameter punch samples, whilst leaving at least one spot intact. These 58 were from Generation Scotland research volunteers born between 1983-1989 (i.e. 32-38 years between collection and profiling). DNA was extracted from between 6 to 8 blood spot punches using the QIAamp DNA Investigator Kit (Qiagen; cat. no. 56504), following the manufacturer's instructions. The concentration of the DNA samples was measured using

a Qubit 2.0 fluorometer and the Qubit dsDNA HS assay (Thermo Fisher; cat. no. Q32854). Total yield isolated was between 196ng and 1177ng of DNA. Up to 500ng DNA (range 160ng – 500ng) underwent bisulfite conversion (Zymo EZ-96). DNA methylation was profiled using the Infinium HumanMethylationEPIC v1.0 BeadChip (Illumina Inc.; cat. no. WG-317-1001), according to the manufacturer's protocol (in batches of 8 samples, 56 in total). Arrays were scanned on an iScan and analysed using GenomeStudio v2011.1.

DNA methylation analysis and statistical methods

DNAm profiles were obtained from the 56 individuals using the Illumina MethylationEPIC beadchip, measuring approximately 850,000 CpGs across the genome. In addition to standard quality control measures such as probe filtering and normalisation, additional checks were performed to ensure Guthrie card and baseline samples matched with regard to predicted sex and genotype (Supplementary information, SI Figures 3 and 4 and SI Table 2).

Maternal smoking status at the time of sample collection was derived from smoking status at GS baseline, and the “years stopped” variable for former smokers, where both mother and baby are in GS. DNA methylation-based estimates of smoking status were obtained, using previously validated methods^{11,12}.

Sample Checks

Confirmatory analyses were performed using Guthrie card DNA methylation data to ensure predicted sex (using X-chromosome data) and genotype (using “rs” control probes on the EPIC array) were consistent with peripheral blood based genotyping and DNA methylation data on samples collected at baseline recruitment (2006-2011) (SI Table 2). There was 100% concordance with predicted sex and between predicted and hard-called genotypes. Detailed information on sample checks is presented in Supplementary Information and SI Figure 3 and 4).

Smoking

An individual's smoking status can be reliably predicted using composite DNA methylation-derived smoking scores, and effects have also been observed in the offspring of mothers who smoked during pregnancy¹³. Moreover, information from a single probe in the aryl-hydrocarbon receptor repressor gene (*AHRR*; cg05575921) can serve as a robust marker of smoking, with lower DNA methylation levels associating with current smoker status. A composite score for smoking status (EpiSmokEr) was obtained using Guthrie sample DNAm data and, along with cg05575921 DNA methylation levels, was plotted against maternal smoking status at the time of sampling^{11,12} (Figure 1). Consistent with previous literature, a higher overall value was observed for the EpiSmokEr score in the offspring of current smokers whereas a lower overall value was observed for the offspring of never smokers (Figure 1A; ever smoker $\beta=0.82$; $P=0.016$). DNA methylation levels at cg05575921 were also consistent with the literature, with lower overall levels in the offspring of current smokers relative to never smokers. (Figure 1B; ever smoker $\beta=-0.73$; $P=0.05$)

Declarations

Acknowledgements

We thank Dr Tom Barlow, Scottish Government Chief Scientists Office for guidance and input into the feasibility studies described here; Dr Sarah Smith, Director of the National Screening Service for facilitating access to the Guthrie card archive and sanctioning the work of the Greater Glasgow and Clyde Biorepository; Alexis Smith, Research Nurse at the Greater Glasgow and Clyde Biorepository; Caroline Clark, Mark Davidson and the NHS Grampian Genetics laboratory for technical support in the early feasibility studies, Dr Joan Mackenzie, NHS Scotland blood spot screening service and the Danish Newborn Screening Biorepository for technical advice and support for the pilot DNA sequencing of de-identified Guthrie card blood spot punches.

Funding

The Citizens' Jury was funded by a Wellcome Trust Institutional Science Support Fund award to the University of Edinburgh (SCB and DJP). Generation Scotland received core support from the Chief Scientist Office of the Scottish Government Health Directorates [CZD/16/6] and the Scottish Funding Council [HR03006] and is currently supported by the Wellcome Trust [216767/Z/19/Z] award to DJP. Baseline DNA methylation was funded by the Wellcome Trust (Wellcome Trust Strategic Award "STratifying Resilience and Depression Longitudinally" (STRADL) Reference 104036/Z/14/Z and Investigator Award 220857/Z/20/Z to AMcl).

The feasibility study was supported in part by HDRI Health Data Research UK substantive hub award and Chief Scientist Office Scotland. CH and SK are supported by the Medical Research Council University Unit award to the MRC Human Genetics Unit, University of Edinburgh, grant number MC_UU_00007/10, Programme MC_PC_U127592696.

Research ethics

The original Sanger DNA sequencing study was mandated by the Scottish Chief Scientist Office, following a favourable opinion from the Scottish Legal Office and Research Ethics Committee. A letter approving the inspection and documentation of Guthrie cards and selective sampling of GS cards for methylation analysis was provided by the Chief Medical Officer for Scotland on 4th September, 2019. The Caldicott Guardians of NHS Greater Glasgow and Clyde and NHS Tayside granted approval on 30th January 2020 and 3rd March 2020, respectively. A substantial amendment to the Research Tissue Bank approval for Generation Scotland to cover the feasibility study was submitted to the Research Ethics Committee and approved on 13th March 2020.

Competing Interests

Ciaran Mulholland is a paid employer of Ipsos MORI.

Sara Davidson is a paid employer of Ipsos MORI.

Riccardo Marioni has received speaker fees from Illumina and is an advisor to the Epigenetic Clock Development Foundation.

Lee Murphy has received payment from Illumina for presentations and consultancy.

Andrew McIntosh has received research funding from the Sackler Trust and speaker fees from Janssen and Illumina.

All other authors declare 'no competing interest'.

Materials and correspondence

To access Generation Scotland data, please go to www.ed.ac.uk/generation-scotland/for-researchers/access

Data availability

To access Generation Scotland data, including the data derived in the feasibility study described here, please go to www.ed.ac.uk/generation-scotland/for-researchers/access

Data from this study has been deposited here: <http://dx.doi.org/10.7488/ds/2277>

Author Contributions

Designed the study: DJP, SCB, JP, ZM, RF, SK, REM, AC

Raised the funding: DJP, SCB, JP, ZM, CS

Designed and conducted the Citizens' Jury: DJP, SCB, CP, MA, CM, SD

Conducted the examination of Guthrie cards: CO, SMcC

Processed and analysed the Guthrie card dried blood spots: KW, LM, NW, SMcC, DMcC, REM

Drafted the paper and prepared the study for submission and peer review: SCB, DJP

Contributed text and/or figures: DJP, SCB, CO, SMcC, KW, ZM, DMcC, REM, AC

Reviewed drafts and approved the manuscript for submission: all authors

References

1. Newborn blood spot test <https://www.nhs.uk/conditions/baby/newborn-screening/blood-spot-test/>
2. B Nørgaard-Pedersen, B. & Hougaard, D.M. Storage policies and use of the Danish Newborn Screening Biobank. *J Inherit Metab Dis* 30, 530-536. (2007)
3. Smith, B. H. *et al.* Cohort Profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. *International Journal of Epidemiology*, 42, 689–700 (2013).
4. L B Navrady, L. B. *et al.* Cohort Profile: Stratifying Resilience and Depression Longitudinally (STRADL): a questionnaire follow-up of Generation Scotland: Scottish Family Health Study (GS:SFHS). *International Journal of Epidemiology*, 47, 13–14g (2017)
5. Laurie, G., Hunter, K., Cunningham-Burley, S. The use of new-born blood spots for research raises ethical, social and legal issues <https://www.gov.scot/publications/guthrie-cards-scotland-ethical-legal-social->

- issues/pages/11/ (2014).
6. INVOLVE Citizens Jury <https://www.involve.org.uk/resources/methods/citizens-jury#:~:text=A%20Citizens'%20Jury%20is%20a,of%202%20to%207%20dayshttps://www.ipsos.com/ipsos-mori/en-uk>
 7. National Health Service Central Register https://en.wikipedia.org/wiki/National_Health_Service_Central_Register
 8. NHS Data Model and Dictionary https://www.datadictionary.nhs.uk/attributes/community_health_index_number.html
 9. Beyan H. *et al*/Guthrie card methylomics identifies temporally stable epialleles that are present at birth in humans. *Genome Res* 22, 2138-2145 (2012)
 10. Ramagopalan, S., V. & Rakyan, V. K. The promise and challenges of blood spot methylomics. *Epigenetics* 8, 775–777 (2013).
 11. Bollepalli, Sailalitha, Tellervo Korhonen, Jaakko Kaprio, Simon Anders, and Miina Ollikainen. 2019. “EpiSmokEr: A Robust Classifier to Determine Smoking Status from DNA Methylation Data.” <https://doi.org/10.2217/Epi-2019-0206> 11 (13): 1469–86. <https://doi.org/10.2217/EPI-2019-0206>.
 12. McCartney, D.L. *et al* Epigenetic prediction of complex traits and death *Genome Biology* 19, 136 (2018)
 13. BR, Joubert, Felix JF, Yousefi P, Bakulski KM, Just AC, Breton C, Reese SE, *et al*. 2016. “DNA Methylation in Newborns and Maternal Smoking in Pregnancy: Genome-Wide Consortium Meta-Analysis.” *American Journal of Human Genetics* 98 (4): 680–96. <https://doi.org/10.1016/J.AJHG.2016.02.019>.
 14. Davidson S., *et al*. Public Acceptability of Data Sharing Between the Public, Private and Third Sectors for Research Purposes <https://www.gov.scot/publications/public-acceptability-data-sharing-between-public-private-third-sectors-research-purposes/documents/> (2013).
 15. Aitken, M. *et al*. Public responses to the sharing and linkage of health data for research purposes: a systematic review and thematic synthesis of qualitative studies. *BMC Med Ethics* 17, 73 (2016).
 16. Aitken, M., McAteer, G., Davidson, S., Frostick, C., Cunningham-Burley, S. Public Preferences regarding Data Linkage for Health Research: A Discrete Choice Experiment. *Int J Popul Data Sci.* 3:429. (2018)
 17. Wilkinson A. If you were born in Scotland between 1965 and 2003 then, more likely than not, a sample of your blood is being stored by NHS Greater Glasgow and Clyde Health Board. <https://www.nuffieldbioethics.org/blog/what-should-be-done-with-the-guthrie-cards> (2019)
 18. Gadd, D. *et al*/Epigenetic scores for the circulating proteome as tools for disease prediction *eLife* 11:e71802 (2022)

Tables

Table 1: Citizens Jury conditions and recommendations

Citizens Jury conditions and recommendations
Condition 1: The ultimate purpose of the research should be to advance medical research in ways that could potentially benefit public health.
Condition 2: Any organisation requesting access to the data should provide a clear rationale for their research, which should be approved by an ethics committee. Access should be granted solely for the use outlined in their research request.
Condition 3: There should be appropriate regulation and oversight of the process by an independent body - that includes an ethics committee - with the authority to issue sanctions for misuse.
Condition 4: Consent should be based on an opt-out system, with the option to opt-out within one year of the child being born, and again when the child turns sixteen.
Recommendation: Mechanisms should be put in place (in the form of a central website) that enable members of the public to check on the types of research being undertaken with Guthrie Card data and the outcomes.

Table 2: Guthrie card archive examination and sample selection for analysis

Step 1	Information recorded (Boxes)
	Box ID; Area Health board; Hospital; Type of card; Date of test; Child forename; Child surname; presence or absence of Community Health Index identifier.
Step 2	Information recorded (Cards)
	Child DOB; Additional comments on card; Number of bloodspots; Size of spots; Mother's CHI; Mother's forename, surname and birth name; Mother's date of birth; Address; Postcode; Whether the cards had been autoclaved prior to archiving; Any other comments.
Step 3	Sample selection
	<ul style="list-style-type: none"> • Pseudonymised ID list of consented Generation Scotland (GS) participants provided to Greater Glasgow and Clyde Biorepository (GGCB) • GGCB staff look for matches between Guthrie cards and GS IDs • 92 matching cards identified and checked for usability • 6-10 punches of 3mm diameter taken from 58 cards and couriered to the Edinburgh Clinical Research Facility for DNA analysis
Step 4	Sample analysis
	<ul style="list-style-type: none"> • DNA methylation data from Guthrie cards compared to DNAm in peripheral blood at time of recruitment (2006-2011) • DNA methylation data from Guthrie cards analysed for smoking signatures and correlated with recorded maternal smoking status

Table 3: Guthrie card blood spot DNA extraction and Sanger sequencing

Year	Guthrie cards sampled	Successful DNA extractions	Extractions that failed Sanger sequencing	Net yield for analysis
1965	17	15	0	88%
1975	32	31	1	94%
1985	17	16	1	88%
1995	18	17	0	94%
2005	16	16	1	94%
2012	36	36	0	100%
Sum	136	131	3	94%

Figures

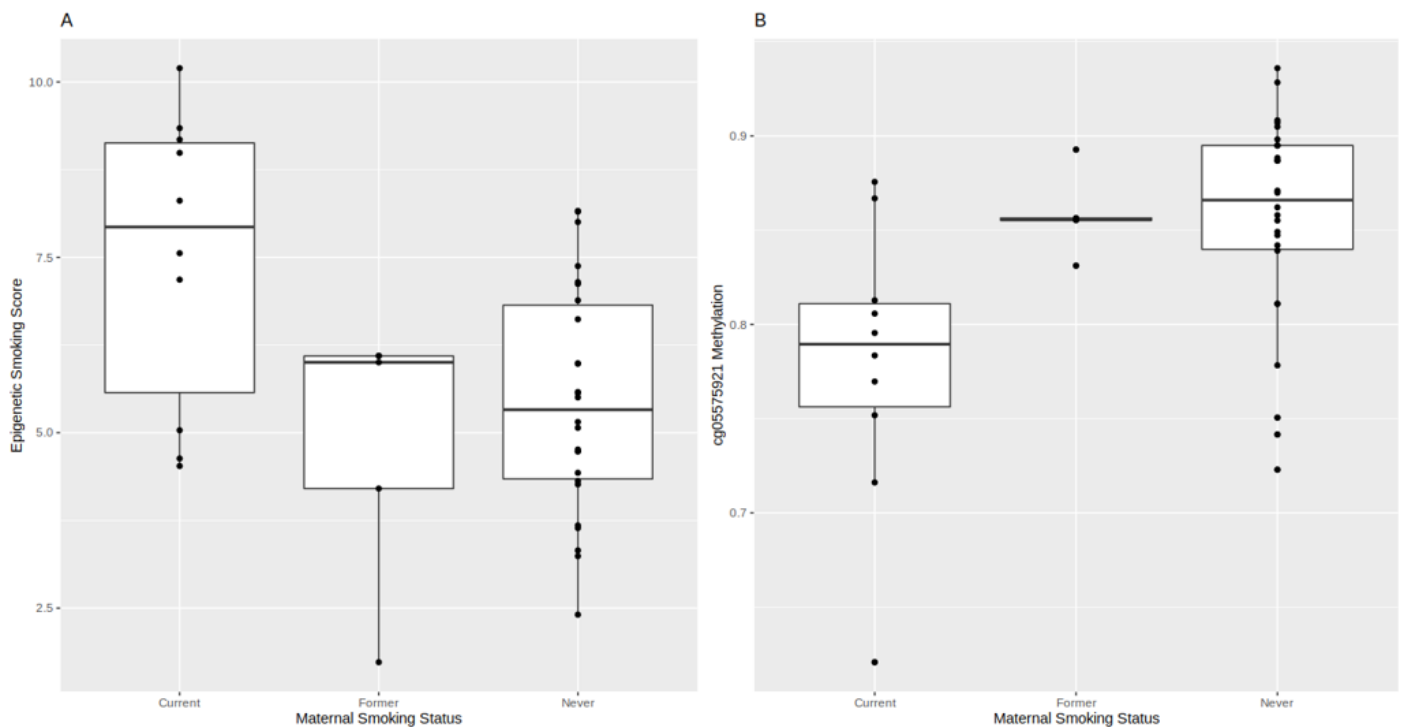


Figure 1

EpiSmoker score (A) and *AHRR* DNAm (B) plotted against maternal smoking status

Legend: Methylation-derived smoking scores from Guthrie card DNA (y-axis) plotted against maternal smoking status (current, former, never) at time of birth. Results are shown for the EpiSmoker score, a composite measure comprised of multiple CpG sites (Panel A), and DNA methylation levels at a single CpG (cg05575921) in the *AHRR* gene (Panel B).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplementarymaterial.docx](#)