ALSPAC AVON LONGITUDINAL STUDY OF PARENTS AND CHILDREN

Access Policy

v. 7.1

November 2017

ALSPAC Executive Committee
University of Bristol
Oakfield House
Oakfield Grove
Bristol
BS8 2BN

Email: ALSPAC-exec@bristol.ac.uk

Tel: +44 (0)117 331 0167

Supported by
wellcome
MRC Medical Research Council
University of BRISTOL
Avon Longitudinal Study of Parents and Children

CONTENTS

Summary

ALSPAC is a longitudinal birth cohort study which enrolled pregnant women who were resident in one of three Bristol-based health districts in the former County of Avon with an expected delivery date between 1st April 1991 and 31st December 1992. Around 14,000 pregnant women were initially recruited. Detailed information has been collected on these women, their partners and subsequent children using self-completion questionnaires, data extraction from medical notes, linkage to routine information systems and from hands-on research clinics. Additional cohorts of participants have since been enrolled in their own right including fathers, siblings, children of the children and grandparents of the children. Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee (IRB00003312) and Local Research Ethics Committees.

The purpose of this document is to describe in detail the general processes and procedures involved in accessing the ALSPAC resource (defined as data already collected and the participants themselves for the purposes of new data collection). **We encourage and facilitate data sharing with all '*bona fide*' researchers** (see the MRC definition; please contact us if you are in any doubt as to whether you fit this definition) and aim to enable researchers to conduct their studies in a transparent and streamlined manner.

For an outline of researcher responsibilities, see Section 6.

1. Data access procedure: overview

1.1 The ALSPAC resource

ALSPAC is run as resource to be used by the research community. We encourage and facilitate data sharing by all researchers from all disciplines across the world to maximise use of the resource. The process for accessing data is the same for all, regardless of research area, institution, location or funding source, provided the proposed research is **not** being carried out for personal or commercial gain.

The vast majority of data are available for immediate use on request and we do not consider the issue of potential overlap between research projects. The study website provides an up to date list of publications that have used the resource and also provides the title and a brief summary of all approved applications to use the resource since April 2011. The website also describes the resource and summarises the types of data available and is a useful place to start to give you a good idea as to whether ALSPAC would be potentially valuable in addressing your research question. The ALSPAC data dictionary provides an overview of all the data held. It contains a series of PDFs, which are fully searchable and detail information on the ALSPAC data that are currently available; it provides information on how data were collected, together with frequencies and coding details. This will be a valuable tool when preparing a proposal to access data and/or samples. The ALSPAC variable catalogue lists every single variable name and label, grouped according to data source.

Searchable indexes of ALSPAC variables are also available via CLOSER Discovery. This is a new search engine that enables researchers to explore the content of eight leading UK longitudinal studies. This detailed search tool provides metadata on phenotypic and environmental variables across the CLOSER longitudinal studies.

Access to ALSPAC research data must be requested using the formal procedures described in this document and is subject to eligibility, the ALSPAC funder's terms and conditions and University of Bristol policies and procedures.

1.2 Requesting access to data

Researchers are required to complete an online proposal form. This proposal should have clearly stated aims and hypotheses and describe the relevant exposure, outcome and confounders that will be considered, justifying the data you require. For multiple projects you must submit multiple forms; one per project. For requests requiring secondary data only, you will get a decision within 10 working days of submission to inform you of the outcome. For more complex proposals involving non-standard data, the approval process is likely to take longer, but you will be contacted within 10 working days to advise you of the progress. Once a proposal has been approved, you will receive advice on the next stages of the process. Please see later sections for timescales relating to applications for biological samples or new data collection. In some cases, approval may be refused due to the lack of relevant data or biological samples. The Executive also reserve the right to impose additional restrictions as appropriate.

The Executive reserve the right to check that all objectives in the original proposal are completed by cross reference to publications and make any additional analyses that were in the initial proposal but that have not been published via letters to journals and/or on our website, in order to avoid publication bias.

If a researcher is seeking funds for their research from a funding body, the Executive must receive the completed ALSPAC research proposal form at least **one month prior** to the submission deadline. It may not be possible to approve those received less than one month before the submission deadline in time for the deadline. It is the responsibility of the researcher to ensure compliance with their funder's terms and conditions with respect to their use of ALSPAC data and samples.

Supervisors are ultimately responsible for their PhD students in the same way that PIs are responsible for their researchers. We request that any **proposals for PhD projects are therefore submitted by the supervisor** rather than the student themselves.

Any proposals that include the collection of *new* data (see Section 5), should usually include a member of the Executive (or a Bristol-based scientist nominated by the Executive) as a co-applicant so they can act as guarantor for the proposed new data. We therefore suggest you approach the Executive with such a proposal at least **two months prior** to the submission deadline.

An amendment to your original proposal (to be submitted via the same online system as your original proposal) must be completed if any of the following change during the course of your approved project:

- Significant extension of research scope;
- Change in start or end date;
- Additional researchers accessing the data;
- Change in institution;
- Any additional data required;
- Change of funding source.

1.3 Charges for access to existing data and new data collection

ALSPAC receives funding from Wellcome, the Medical Research Council and the University of Bristol to support *core* activities. These do not extend to providing support for individual projects and researchers will be expected to meet any and all additional costs for data access and provision. All researchers accessing ALSPAC data will be charged on a cost recovery basis: This cost will vary depending on the amount and type of data. Please note, we cannot give discounts to PhD students or for any other reason. Once a proposal has been approved and the applicant informed of the cost these **are non-negotiable**.

Costs will be determined on a project-by-project basis and will reflect only the true costs to ALSPAC of providing the resources requested. Once a proposal has been agreed in principle an accurate costing will be provided. Example costings for data requests are provided overleaf (Table 1).

Data will **not** be provided until an invoice has been settled or a purchase order number is received by our finance department.

Costs are reviewed on a regular basis (although access to education data may be subject to change at short notice as ALSPAC's access to these data are subject to unforeseen changes beyond our control).

VAT will be charged in addition to the table costs, where applicable.

1.3.1 Charges

Table 1: Charges to be applied for data requests (prices correct as of 31st October 2017)

Type of data request	**Charge (+ VAT where applicable)**
Standard fee - Includes up to 50 variables (this includes individual genetic variants)	£2360
Every additional 100 variables	£170
Genetics data*	
Individual variables, e.g. a number of SNPs (these are included in the section above)	As above
GWAS dataset no imputations (available: mother and child)	£1815
GWAS imputed with HapMap (available: mother and child)	£1815
GWAS imputed with 1000 genomes (available: mother and child)	£1815
GWAS imputed with HRC (available: mother and child)	£1815
GWAS analysis done internally by ALSPAC staff (*Guide price: depends on project*)	£2250
GWAS analysis done by visiting researcher (including desk fee)	£2960
Methylation data (from ARIES project; any cohort/time point)	£1815
Gene expression	
CNV	
UK10K	
DTA administration costs	£910
Text data (to be coded by applicant) – up to 5 fields	£240
Each additional field thereafter	£45
Extracting geographical data, e.g. address data / GIS data †	£725
Other 'physical' data; scans; sound file; videos etc. [per time sweep, per participant cohort and per type of data/media]	£1265
MTA administration costs	£910
Extracting education data‡ [please see Appendix Two: Linkage data] †	£525
Amendment fee (additional data request) - Includes up to 50 variables	£1040
Every additional 100 variables	£170

* All data sets are available as one release at the prices detailed. Further permutations will be charged as per time required to formulate additional releases.

† This covers the cost of extraction only – the number of variables from these categories will still be costed

‡ All other linkage data to be costed on a case by case basis

Costs for example projects are included in Appendix One: Example data access costs.

1.3.2 Charges for project amendments

Additional data requests (i.e. data that were not included in the original proposal but subsequently required for any reason) will be subject to the costs shown in Table 1. Any projects that were approved prior to April 1st 2014 may submit one amendment at no charge; any additional amendments will be charged according to Table 1.

See Section 1.2 for details on how to submit an amendment.

1.3.3 Costs for grant proposals

If you are submitting a grant in order to cover the costs of the agreed research, you must include a set 'Data Management' fee of £7500 (+ VAT where applicable). If a grant is successful we request a copy of the award letter and will invoice you for the full amount before data can be provided. Please be aware that in order to take into account fluctuating exchange rates you may want to consider including a small amount of contingency if you are not seeking funds in GBP.

Any negotiations with funders regarding changes in requested funds must involve the Executive at all stages. The Executive are happy to provide a letter of support once your research has been approved and the budget has been agreed. Please note that if a grant that includes a Data Management fee is successfully funded, that project will include up to three amendments at no additional cost. If you submit more than three amendments during the life of the project you will be charged for each one as detailed above (see Section 1.3.2).

1.3.4 Costs for pre-access information

On occasion, researchers require more detailed information than can be gleaned from the website, e.g. cross-referenced data for inclusion in grant proposals. These will be dealt with on a case-by-case basis and may incur additional costs.

1.3.5 Costs for access to data previously funded by a researcher

If you have previously paid for the collection of data (e.g. questions in a questionnaire, a measure in a clinic) within ALSPAC and require access to these as part of a new project: (1) You will get a reduction of £1180 from the data fee for those variables (equivalent to a 50% discount on the data preparation aspect of the fee: reduced standard fee will therefore be £1180); (2) you will be provided with those variables free of charge (but will have to pay full costs for all other variables requested that you did not collect); (3) As the current grant holders, the co-PIs of the ALSPAC core infrastructure strategic award have free access to all variables.

In order to qualify for this reduction, the PI of the new project proposal to ALSPAC must be the original person who funded the data collection, or someone being directly managed by that person. There will be no reduction if the project is subject to a data management fee (see Section 1.3.3).

1.3.6 Costs for new data collection

Projects involving new data collection will need to be fully costed. The PI should liaise with the Executive to request a costing following approval of their project proposal, and before any grant is submitted. Questionnaires are charged at £6397.70 per A5 page per questionnaire. VAT may be added as appropriate. ALSPAC will calculate the number of A5 pages required. Costs for clinical measures and sub studies involving hands-on data collection will vary according to the requirements of the project.

1.4 Management of the resource

The team is led by Professor Nic Timpson (Principal Investigator) and Professor George Davey Smith (Scientific Director).

The Principal Investigator (PI) has overall responsibility for all areas of activity in ALSPAC. The PI is supported by the ALSPAC Executive in the execution of ALSPAC activities. Members of the Executive are Professor Nic Timpson (Principal Investigator), Ms Lynn Molloy (Chief Operating Officer), Dr Susan Ring

(Executive Director, Bioresource) and Dr Kate Northstone (Executive Director, Data). The Executive is supported by members of the Senior Management Team (SMT) for all operational activity.

The various committees governing ALSPAC are described below.

1.4.1 ALSPAC Executive

The PI of ALSPAC has overall responsibility for all areas of activity. Members of the Executive support the PI with the management of the study, which includes the approval of data access proposals and new data and sample collection projects, and meet on a weekly basis. The Executive refer to the Scientific Director (SD) for issues regarding scientific direction, the Board for unresolved issues/problems and to the Independent Scientific Advisory Board (ISAB) for data and sample access requests that it has been unable to adjudicate on. The terms of reference of the ALSPAC Executive are available on the website.

1.4.2 Senior Management Team (SMT)

Executive members are supported by members of the Senior Management Team (SMT) who manage all operational activity within the study. Membership of SMT comprises senior managers from the key operational teams (operations, administration, data access, data pipeline, data systems, data linkage and security, clinic, G2 and bioresource). The terms of reference of SMT are available on the website.

1.4.3 ALSPAC Board

The Board provides oversight of the Executive and supports the SD and PI with scientific strategy. The Board is chaired by Professor John Henderson. Members are made up of those responsible for the delivery of work programmes under the current strategic award and independent representatives of key scientific groups. The Board requires the presence of at least one practising clinician. The Board meets on a quarterly basis starting from September 2017. The terms of reference of the ALSPAC Board are available on the website.

1.4.4 Independent Scientific Advisory Board (ISAB)

The Independent Scientific Advisory Board provides advice, support and guidance to the SD and PI on ALSPAC's scientific direction, strategy and operations, and the furtherance of its scientific mission. ISAB reviews complex, controversial or disputed proposals for access to data and samples that have been referred on by the Executive. ISAB meets every six months and is chaired by Professor John Gallagher. The terms of reference of ISAB are available on the website.

1.4.5 ALSPAC Ethics and Law Committee (ALEC)

Studies involving the collection of biological samples, ionising radiation, NHS staff or premises are reviewed by the NHS Research Ethics Committee (NHS REC). All other studies, such as those involving questionnaires or qualitative research, and policy changes to the way participants are handled go through the independent ALSPAC Ethics and Law Committee (ALEC). The committee meets every two months and is chaired by Dr Mike Shere. The membership comprises clinicians, researchers, lawyers and participants. The terms of reference of ALEC are available on the website.

1.5 Intellectual property

The University of Bristol owns the ALSPAC resource: Any data generated and the biosamples collected. As such, any requests to access the data must be made through the Executive. Any data generated through an approved project must be returned to the resource to encourage ongoing use by the research community.

2. Types of data and rules governing access

A wide range of data are available through the resource. ALSPAC collects data directly through questionnaires and hands-on assessment clinics. Data is derived from biological samples and other types of media we collect such as videos, scans and audio files. We also link to data made available by various third parties.

It is worth noting that sampler sets are available on the UK Data Archive. These are available for use by bona fide researchers.

The data dictionary includes detailed documentation on the data collected via questionnaires and clinics and some other subsets of data and is freely available for download from the website. The variable catalogue shows every single variable name and label, grouped according to data source. Searchable indexes of some ALSPAC variables are also available via CLOSER Discovery. This detailed search tool provides metadata on phenotypic and environmental variables across the eight CLOSER longitudinal studies. If you have any questions regarding general ALSPAC data, please email ALSPAC-data@bristol.ac.uk (for linkage data, please see Section 2.2).

Proposals for access may be refused. Reasons for refusal include the following:

- Lack of availability of data/samples;
- Applicant not being a bona fide researcher (see Summary);
- The proposed work, in the view of the Executive, risks bringing the study into disrepute;
- The proposed work risks disclosure of identifiable information relating to any individual participant;
- In the view of the Executive, there is a conflict of interest in relation to the proposed project;
- The proposed outputs of the project are outside the scope of the ALSPAC ethical approval, funders' terms and conditions or the University of Bristol's policies and procedures;
- Access to data obtained via linkage to health and administrative records is subject to complying with the terms imposed on ALSPAC by the original data owners (see Section 2.2).
- Biological samples are a finite resource and therefore need to meet the criteria outline in Section 4.

Any challenge against a refused proposal will be considered by ISAB (described in Section 1.4.4).

2.1 Questionnaire, clinic and biosamples data

We ensure data collected by questionnaire and our hands-on clinics are made available as soon as possible after data collection is complete (normally around 6 months). Non-genetic results obtained from biosamples are made available as soon as all assays have been completed and the data have been cleaned. All potential identifiers are removed and disclosure risks are considered such that data may be grouped where appropriate. Available data are not restricted in any way.

2.2 Linkage data

ALSPAC collect data using linkage to routine health, administrative and environmental records. Linkage can either be conducted at an individual level (e.g. primary care records) or an organisational level (e.g. a school). These data are collected from external organisations. To do this, ALSPAC enter into data usage agreements with the relevant data owners. These data usage agreements specify the conditions under which ALSPAC can share these data with third parties (e.g. researchers). The data access conditions differ for each data set we link to, and they also change over time. The rules governing access for these different types of data are described in more detail in Appendix Two. Necessarily, we consider the following when we adjudicate requests to access linkage data:

- The scope of the research investigation: Linked data are provided solely for the investigation of a single research hypothesis. Requested data should be relevant to the research investigation.

- Consent status: Access to linked data is frequently subject to participant consent status. Participants are free, at any time, to withdraw from the study, or withdraw their consent from linking to third-party data. Once released it is acceptable that researchers can retain the data for the duration of the research investigation, regardless of consent change. However, if additional data are requested these will be filtered on *current* consent status which may result in a different sample size.
- Changing access conditions: Data access conditions can (and do) change over time, sometimes with little warning. ALSPAC, and third-party data users, are required to comply with any new data sharing conditions. This may impact research investigations in unforeseeable ways.
- Data quality: ALSPAC provide linked data on the understanding that these are routine records being used for a secondary (i.e. research) purpose. We make no guarantees regarding the accuracy of the data and have no means of verifying the data. Where possible we will document the data using information provided by the data owner and provide quality information about the linkage process where available.
- International researchers: Some data usage agreements specify that data cannot be sent outside of the UK or the European Economic Area (EEA) – these requirements are beyond our control.

The host organisation of third party researchers is required to enter a legally binding contract (Data Transfer Agreement; DTA) with the University of Bristol prior to receiving linked data. This contract commits the third-party researcher/host organisation to maintaining the conditions set out in the data usage agreement between ALSPAC and the data owner. The ALSPAC Data Linkage Team will provide an estimate for any data access costs during project negotiations.

It should be noted that we are required (as requested by the Information Commissioner's Office) to post a lay summary of all projects using linkage data on the ALSPAC website. It must remain there for 4 weeks before the project team can be provided with data, in order to give participants the opportunity to inform us if they want to opt out of any specific project.

We encourage potential users to contact the linkage team in advance of submitting their project proposal.

Please contact the Data Linkage Team (alspac-linkage@bristol.ac.uk) if you are interested in using any of the data detailed in this section. More detailed information on each type of linkage data we hold is described in Appendix Two.

2.2.1 Bespoke linkage

The ALSPAC Data Linkage Team has a strong track record of conducting bespoke linkages to both NHS and administrative records. Please contact the team to discuss your project.

Linkage to external administrative datasets is a dynamic process we use to enhance the ALSPAC data resource. This document will be updated regularly to reflect progress of this enhancement, however, for the most up to date picture of the extent and availability of linked administrative data, researchers are advised to contact the linkage team.

2.3 Genetics data

Details of genotype data available from the ALSPAC participants (currently young people and mothers) are available on our website. This includes genome wide microarray data (GWAS data), expression data and specifically genotyped polymorphisms (e.g. Single Nucleotide Polymorphisms (SNPs) and variable number tandem repeats (VNTRs)).

Whole genome sequence data for 2000 individuals has been generated as part of the UK10K project. Sequence data linked to a limited number of phenotypic variables is available from the UK10K team. Sequencing data linked to other ALSPAC phenotypes is available on application using the normal processes described in this document.

In addition, methylation data (Illumina infinium human methylation 450 bead array and whole genome bisulfite sequencing) has been generated as part of the ARIES project (Accessible Resource for Integrated

Epigenomic Studies) on 1000 mother-child pairs. Aggregate DNA methylation data can be accessed via ARIES-explorer.

If you require any bespoke genetics data not described above, please contact us *prior* to submitting a proposal to discuss your needs.

2.3.1 Accessing genetics data

Provision of genetics data requires a legally-binding agreement between the University of Bristol and your host institution. This agreement is called a Data Transfer Agreement (DTA). For University of Bristol staff based outside the Bristol Medical School, a Data Service Level Agreement (DSLA) is required. These forms differ in terms of the signatories required, not the access level received. A project-specific appendix must be agreed before the agreements are signed.

Genotype data cannot be released until fully completed forms have been received. Copies of the DTA and the DSLA are included on the website to show the information that is required. Once your proposal is approved we will provide Word files with the specific appendix for you to complete, sign and return to us.

2.4 Potentially identifying data

Some of the data ALSPAC have collected could be used to readily identify study participants. These include: personal details such as names, addresses (including postcodes) dates of birth; "free text" information that could contain identifiers; or other clinical data in a format that could readily identify a participant.

The study team will not link *any* of these data directly to the published data resource; this would breach the agreements we have in place with participants to maintain their confidentiality. Instead a split-stage process is required if a researcher wishes to make use of any potentially identifying data (e.g. to administer data collection at a specialist clinical facility, or to derive air pollution measures from residential addresses using specialist skills or equipment).

The aim of the ‘split-stage protocol’ is to create the conditions where research collaborators are able to access identifiable information, while maintaining a distinction and organisational separation between this data use and subsequent use of de-identified attribute data.

Depending on the request, the Executive may request that a researcher makes an application to ALEC to review and approve the specific request in the context of the overall project and specific hypotheses being tested. ALEC may ask the researcher to attend a meeting to explain their proposal. If a proposal requires detailed potentially identifying data (such as videotapes) to address a scientific question, then a bespoke Material Transfer Agreement (MTA) must be completed before the data will be released.

For more details about how the split-stage protocol works, please see Appendix Three.

3. Data provision

3.1 The ‘data buddy’ process

When the Executive have approved your project, a data buddy will be assigned to you. Your data buddy will assist you through the duration of your project; advising you on making your formal data request, costing your proposal, administering the necessary paperwork (such as confidentiality forms and transfer agreements) and ultimately provide you with your dataset. They will not provide statistical, methodological or other support without prior agreement and the relevant costs being covered. Standard datasets are prepared within two weeks of all paperwork being completed and invoices being settled (or purchase numbers being provided); however, some types of data may take longer (see Section 2).

3.2 Confidentiality form

Protecting the confidentiality of the study families is a primary concern of the Executive and the ALSPAC study team. This is a particular issue as ALSPAC is a regionally based study that recruited children born within a defined period. The PI, and any member of their team who will directly access the data, will be requested to adhere to a number of clauses regarding confidentiality (please see our confidentiality form). It is the PI’s responsibility to ensure that all their staff (or members of the research group) are known to ALSPAC, have completed a confidentiality form, and are made aware of, and are adhering to, their data usage responsibilities (including all relevant ALSPAC policies and procedures).

3.3 Other paperwork

If a project requires any linkage or genetics data (see Sections 2.2 and 2.3 respectively), a Data Transfer Agreement (DTA) must be completed. This requires the signature of a legal signatory in the PI’s institute. For other potentially clinically identifiable data (see Section 2.4.4) a Material Transfer Agreement (MTA) may be required.

When a paper is submitted to the Executive for approval (see Section 6.7) a papers checklist must be completed and sent to the Executive with the manuscript.

3.4 Unique project identifiers

For each project a unique set of identifiers is created. If a PI has more than one project, separate identifiers will be attached to each dataset relative to each project, thus datasets *cannot* be combined.

3.5 Secure data transfer

All data transferred electronically must be encrypted using AES-256 encryption (this can be achieved using compression tools such as WinZip or 7-Zip). Data users will be assigned a password for a project when the first dataset is provided.

4. Access to biological samples

ALSPAC has collected many biosamples since the start of the study. There are over 250 different sample types on the database which are defined by:

- Participant group (e.g. original cohort, mother, father, CoCo90s (Children of the Children of the 90s))
- Main sample type (e.g. plasma or urine)
- Additives (e.g. EDTA or heparin plasma)
- Time point (e.g. antenatal, 7yr, 17yr etc.)
- Aliquot size.

Details of available samples is provided on our website.

Ethical approval was obtained for all sample collections (details for specific samples available on request). This usually included consent for future research including genetic studies. Occasionally analysis will not be covered by existing ethical approval and further review will be needed before samples can be released. Where possible samples are stored in multiple aliquots to limit the need for freeze thaw cycles and thereby enable as much analysis as possible. However, the number of aliquots varies between sample types and for some samples (e.g. buffy coats, hair, teeth) only one aliquot was produced.

If you are a planning proposal to access biological samples you are strongly advised to discuss your plans with the laboratory team prior to submission who will be happy to provide further information about samples or laboratory procedures, contact Bristol Bioresource Laboratories (bbl-info@bristol.ac.uk) -*please allow 10 working days for a response*.

To use existing biological samples or to carry out specific genotyping on ALSPAC DNA you need to complete the online research proposal form describing your proposed research. You must ensure you complete the specific sections on the biological samples and genotyping including details of the type of sample required, amount needed and, in the case of DNA, the minimum concentration required.

The majority of material in the ALSPAC biorepository are finite samples, i.e. there are limited stocks available. The exception is lymphoblastoid cell lines and DNA derived from these cell lines as they can be grown to provide more material. The Executive are responsible for ensuring that samples are used for projects that maximise the amount of data obtained from available samples and that these data are subsequently made available to other researchers.

Proposals that request finite samples must satisfy the following criteria to be approved:

- Scientific strength of the proposal must justify use of ALSPAC cohort samples, i.e. the data obtained from the samples will be analysed in conjunction with other data held by ALSPAC. Requests for projects that could be carried out using samples available elsewhere will not be approved.
- The analysis proposed does not already exist for the same time point. Requests to repeat or carry our very similar analysis will not be approved unless there are compelling reasons.
- Evidence must be provided to show methodology is appropriate given the processing history of the samples, e.g. evidence from published literature or pilot data generated on samples processed in a similar manner. ALSPAC samples will not be released for method development.
- The assay test platform should have proven quality assurance measures in place.
- The methodology should include measures to ensure the quality of any remaining sample is not jeopardised and can be used in further assays which can be used on freeze thawed samples.
- The volume requested is reasonable and does not seriously deplete the resource.
- The work proposed is within the scope of the consents obtained for the specific samples.

Final aliquots of the majority of sample type will be retained for future global discovery projects. The exceptions to this are those sample types where only one aliquot was originally produced (hair, teeth, and whole placentae). In addition, white blood cell pellets (either buffy coats or peripheral blood lymphocyte preparations) are reserved for production of DNA, RNA or cell lines since they were specifically collected for these purposes. White blood cells will not be released as DNA. RNA and cell lines will be produced in-house

and managed by the ALSPAC team. Reasonable requests for DNA will be approved and an appropriate aliquot provided.

Proposals which might significantly deplete stocks of finite biological samples or have insufficient evidence of the validity of an assay will be subjected to independent peer review (by a sub group of ISAB) to ensure they meet the conditions above and that the amount of material required is acceptable. The ALSPAC Executive will notify you of this requirement where necessary. Please note this means the approval of requests to access ALSPAC samples will take considerably longer than those for data alone and we try to complete the process within 6 weeks.

If two or more researchers are requesting the same samples either to carry out the same analysis or a different assay the proposals will be referred to ISAB who will make the decision regarding which proposal (if any) offers the best use of the material.

If a request is approved samples will be supplied with the following conditions:

- Costs incurred in providing samples will be covered by the applicant. These will include costs for retrieval, additional processing necessary for the specific project, shipping costs (both out and return) and linking data. Costs will be provided on a case by case basis depending on the work involved and may be subject to VAT.
- All results generated from samples must be returned to ALSPAC for inclusion in the data resource and will be made available to other researchers.
- Where it is possible to use samples that have been thawed and refrozen, samples which have been returned from other projects will be supplied in preference to unused stock if available.
- ALSPAC reserve the right to specify where analysis will be carried out in order to ensure results obtained are comparable to existing data.
    - Illumina bead chip analysis on ALSPAC samples will be run in house (e.g. HumanMethylationEPIC BeadChip).
    - LGC genomics hold sets of ALSPAC DNA for single SNP genotyping and their service is normally quicker and more economic both in terms of DNA use and genotyping costs than supplying DNA to individual researchers. Any order to genotype ALSPAC DNA at LGC Genomics must be placed via the ALSPAC laboratory.
    - We have preferred laboratories for outsourcing biochemical and metabolomic analysis please contact the laboratory to discuss such requests.
- BBL will advise if current ethical approval will cover a project. For some work, it will be necessary to submit a new application and BBL will assist with this process.

Samples, including DNA, are provided under the terms of a Material Transfer Agreement (MTA) or in the case of University of Bristol staff a Service Level Agreement (SLA). Each agreement will include a project specific appendix detailing the samples ALSPAC will supply and the analysis to be completed. Samples will not be released until an agreement has been completed and signed. Please note that for samples that are classified as relevant material under the Human Tissue Act (HTA) the HTA MTA should be used instead. Once your proposal is approved we will provide word files with the specific appendix for you to complete, sign and return to us.

Please note if your project requires grant funding samples will be reserved until the outcome of the first funding application is known or 12 months whichever is shortest. After this time ALSPAC cannot guarantee sample availability as they may be released to other projects. Researchers will be expected to discuss plans for further funding applications with the Executive.

Note if a participant withdraws consent whilst samples are being analysed researchers may be asked to destroy relevant samples and provide evidence that this has been completed. Costs of supplying samples cannot be refunded in these cases.

5. New data collection

The ALSPAC study team collects new data from the study families using self-completed postal and online questionnaires and at ALSPAC study clinics. Data collection may be on the whole cohort or on a specific sub sample. The current plan for data collection can be viewed on our website.

Researchers are encouraged to apply for funding for data collection at least 18 months in advance of the proposed start date for data collection. The costs of new data collection must be agreed with the Executive (described in Section 1.3.6).

The following are essential requirements for new data collection to be undertaken (additional requirements may also apply on a case by case basis):

- A requirement of ALSPAC's core funders is that a member of the Executive is named as a co-applicant on grants involving any new data collection.
- All new data collection is discussed with our participant advisory groups (see Section 5.1.10) to test acceptability. We would strongly encourage researchers to provide sufficient time between submitting their proposal to the Executive and their grant deadline so that the participant advisory group may consider the measurements, particularly if they are sensitive. If this is not done, researchers must understand that approval to collect the new data is conditional on our participant advisory group agreeing with the collection or providing no evidence to suggest that a significant proportion of participants would be upset or refuse such collection.
- Ethical approval will be required from either ALEC or an NHS REC depending upon the nature of the study (Section 5.2). We appreciate that in most cases this will be sought after grant submission, but no data collection will be undertaken without appropriate ethics approval.

5.1 Prior to data collection

Approval for any questionnaire or hands-on data collection (as part of a questionnaire, main clinic or through a sub-study) must be obtained from the Executive *before* funding is sought. Once the proposal has been approved in principle by the Executive, SMT will review the proposal in more detail and assess feasibility, resourcing, timescales, recruitment, any case selection requirements, participant overload and other practical and ethical considerations. SMT will then advise the PI if there are any issues which need to be addressed at an early stage in order for the study to proceed.

5.1.1 Management of the data collection exercise

The study PI is primarily responsible for the study and ALSPAC will play a supporting and advisory role in delivering the project. ALSPAC is required to adhere to certain policies, particularly around contact with participants and the feedback of results, which protect the confidentiality of participants. All studies will be obliged to conform to these.

5.1.2 Early planning of the data collection exercise

A planning ('set-up') meeting will be arranged between the PI and the ALSPAC study team in Bristol as soon as we have confirmation that a new data collection exercise has been funded. The responsibility for project deliverables, objectives, timescales, finances and policies for the study will be agreed at this meeting. Changes following this setup meeting will need to be agreed by both parties, and major changes to the study design may need to be submitted to the Executive as an amendment to the original proposal before they can be actioned.

Particularly during the early stages of the study design, regular meetings will be required to facilitate planning and problem solving.

Prior to commencing data collection, ALSPAC staff will work through a 'pre-start checklist' with the PI to ensure that the project is ready to begin, and implementation plans are fit for purpose. Any outstanding actions will need to be identified at this stage and agreed with the PI for data collection to begin.

5.1.3 Case selection for sub-studies

The PI will define the criteria for case selection where appropriate, based on the ALSPAC data and samples available. An ALSPAC member of research staff will produce the case selection in collaboration with the PI. The case selection criteria originally requested by the PI in their application may need adjusting, depending on the number of participants who are ultimately available to take part. However, once data collection has started changes to the case selection criteria should be avoided wherever possible.

In order to prevent overburdening of study participants, ALSPAC may seek to manage the number of sub-studies taking place at any one time.

5.1.4 Controls in sub-studies

ALSPAC has an ethical requirement to ensure that where individuals are recalled for assessment on the basis of a particular characteristic all necessary steps are taken to ensure that disclosure of this characteristic in relation to an identified individual is avoided. A study may therefore be required to include controls (or additional participants in order to protect disclosure risk in cases) in order to mask the characteristics of participants to researchers, staff and other participants. The number of controls/additional participants required and any additional steps needed to maintain confidentiality will vary between studies, and will be determined in conjunction with ALEC.

5.1.5 Electronic data collection systems

A database for the collection and storage of data will be set up by ALSPAC staff in conjunction with a member of the PI's team who can then input variables into the framework system provided. This will ensure that the variables collected are labelled in a consistent and systematic manner, comparable with any previous studies, and ultimately allow the integration of the data collected into the main ALSPAC resource.

5.1.6 Administration

The data collection exercise will be set up on the ALSPAC participant contact management and booking system ('Arcadia') to allow for the printing and mailing of invites to participants, phone call reminders and the like. It will also provide a record of all contacts participants receive for the study.

5.1.7 Participant documents

ALSPAC will work with the PI to put any specific study documents (invitation letters, information sheets, consent forms etc.) into the ALSPAC 'house style'. This ensures that the design, wording and information given in each study are consistent. The ALSPAC Communications and Participation team and OCAP must sign off any paperwork prior to submission for ethical approval.

5.1.8 Ethical approval

ALSPAC will advise the PI on their ethical application and where to submit their proposal, although the application must be completed and submitted by the PI. Those studies which do not involve human tissue, use of NHS staff or premises or ionising radiation should be submitted to ALEC, others to an NHS REC of the PI's choice. The REC with most experience of ALSPAC study proposals is South West – Central Bristol. A University of Bristol Research and Enterprise Development (RED) research registration checklist is available which should also be submitted for University of Bristol governance purposes for those proposals seeking REC approval.

5.1.9 The Original Cohort Advisory Panel (OCAP) and other participant advisory forums

Depending on the nature of the data collection exercise it may be appropriate to submit the study protocols, participant documents and so on to OCAP or other participant groups as required. ALSPAC will advise on whether this is necessary, and PIs may be asked to attend an OCAP meeting to explain their study. ALSPAC will work with PIs to incorporate feedback from participant advisory forums into their study design.

5.1.10 Piloting

ALSPAC strongly advise that a pilot takes place for each new data collection exercise, which must be funded by the PI. ALSPAC will work with the PI to organise and conduct this pilot. The pilot will ensure that the study design, methodology and data collection systems are fit for purpose.

5.1.11 Collection of new biological samples

Any request to collect new samples from ALSPAC study participants must be approved by the Executive. The Executive will provide details of the costs which need to be covered, these will include costs of contacting participants, collecting the samples, processing and storing samples with the Bristol Bioresource Laboratories (BBL), shipping to external laboratories and any in house analysis by BBL. Sufficient funds must be in place before sample collection can commence. Sample collection may be part of a main clinic sweep or a sub-study.

In order to maximise the value of the ALSPAC resource residual samples will be stored for future use. Priority will be given to the analysis outlined in the original proposal, but sample remains will be aliquoted, added to the ALSPAC biorepository and made available for further analysis.

Appropriate ethical approval must be obtained for all sample collections. Where the sample collected is classed as "relevant material" under the Human Tissue Act (HTA), a favourable opinion must be obtained from an NHS Research Ethics Committee (NHS REC).

Consent forms and participant information sheets must be approved by the Executive Director, Bioresource (who is also the HTA designated individual) before submission for ethical approval. These should include a statement asking for generic consent for future non-genetic research and, if applicable, a separate statement for future genetic research. Participants should have the option not to consent for these, in which case residual samples will not be stored in the biorepository for future use.

Before collection begins the collection and processing protocols must be agreed with the Executive Director, Bioresource and the ALSPAC Clinic Manager. This will include details of how samples will be labelled to ensure samples entering the biorepository are uniquely identifiable.

Some sample analysis can be carried out in house by BBL or samples can be shipped to other laboratories for analysis. An MTA will be set up with receiving laboratories before samples are released (examples of an agreement for HTA relevant materials and other samples are available on the ALSPAC website). All data derived from samples from ALSPAC participants will be incorporated into the main ALSPAC resource.

5.2 During data collection

5.2.1 Response rates

Response rates are unpredictable and cannot be guaranteed, especially amongst the study young people. Attention should be paid to ensuring that the case selection is large enough relative to the target numbers. ALSPAC will advise the PI on measures which may be taken to increase the response rate, which will need to be included in the funding coming to ALSPAC if not included in the original costings. The ALSPAC cohort profile papers give historical response rates for various projects. More recent response rates are available through the relevant documents in the data dictionary.

5.2.2 Personal data and confidentiality

Under no circumstances will ALSPAC provide the PI or their study team with personal details (names and addresses) of participants unless the participants have consented to take part in their particular data collection exercise, in which case such details will be given for administrative reasons only (for example if it has been agreed that the PI's staff will make telephone bookings with participants who have responded to a study invite). All data collected must be recorded against an anonymous study ID. Mailings to participants will be sent via the ALSPAC administration team, unless agreed otherwise.

Any participant personal data must be stored and transferred in a secure fashion, in line with our data management policy. Any personal data used should be returned to ALSPAC on completion of the study.

5.2.3 Dealing with participants

ALSPAC participants are valuable long term committed study members who may withdraw from the study if they have a bad experience. The ALSPAC study team will always make initial contact with participants, even if the study is being conducted by an external team. PIs are reminded that they have a wider responsibility to ALSPAC when they or their staff are dealing with participants. If there are any problems or queries raised by participants, they should be referred to ALSPAC immediately.

The ALSPAC Participation Team will meet with the appropriate staff and provide an ALSPAC background briefing including some frequently asked questions. This is important because participants do not discriminate between different studies, and would expect any person contacting them on behalf of the study to be aware of all other relevant aspects of ALSPAC's work.

5.2.4 Disclosure and Barring Service (DBS) checks

Any staff who will be dealing with participants face-to-face will require a DBS check. The PI must ensure that any members of their team who are meeting participants are approved.

5.2.5 Sending study invitations

It is important that all data collection exercises are coordinated alongside other activities taking place within ALSPAC, particularly with respect to any contact made with participants. ALSPAC will agree a mailing/invitation schedule with each study, which may be affected by other activities taking place concurrently within ALSPAC.

5.2.6 Feeding back contact information

If new or updated contact information or other administrative information is obtained during the course of the new data collection exercise by the study team, this should be referred back to ALSPAC to update its participant contact management system. A record of all contacts made with participants is required, which will need feeding back to ALSPAC at regular intervals. This will include any phone contacts with participants.

5.2.7 Feedback of incidental findings

The ALSPAC Disclosure Policy states that information shall not, as a general rule, be disclosed to participants. This general policy should only be set aside when it is reasonably certain that the benefits of disclosure clearly outweigh any possible risks to the participants or their families. Feedback of incidental findings will be discussed at the new data collection planning stage.

5.2.8 Return of data

Any data generated through an approved project must be returned to the resource to encourage ongoing use by the research community. The PI is required to provide ALSPAC with a copy of *all* data collected and/or generated form ALSPAC participants, including scanned images, recordings etc., to be archived for future use. This may be required during the data collection period or once data collection has finished; this will be agreed before data collection begins. Since data are collected using an anonymous ID unique to each study, it is impossible to link to any other data from the main resource until they are returned.

5.2.9 Backing up data

ALSPAC will arrange for data collected to be backed up on a regular basis on those studies where ALSPAC staff are collecting the data. In studies where the PI's team is collecting data backing up is their responsibility. However, ALSPAC will seek to take periodic copies of all data to reduce the risk of data loss. Clear arrangements to carry this out must be agreed before the start of the study if data collection is to be carried out off site or remotely to ALSPAC's standard data collection drive.

5.2.10 Feedback of project statistics/reporting to PIs

ALSPAC will provide regular reports on the progress of each study, the frequency of which will be agreed at the set-up meeting. ALSPAC will indicate at the earliest opportunity where there are risks or issues which need to be addressed.

5.3 After data collection

5.3.1 Participants' personal details

Once data collection is complete, all participant contact information a researcher may hold must be destroyed, having returned any updated contact details to ALSPAC.

5.3.2 Data collected

The PI is required to return a copy of the raw, unedited data collected to ALSPAC for archiving. Any processed datasets/derived variables and the syntax/scripts used in their creation should be returned to us once produced for inclusion in the resource. IDs will be converted to our centrally held ID.

5.3.3 Additional variables

Once data collection is complete any additional variables required to be added to those collected as part of the study can be arranged with an ALSPAC data buddy, who will be assigned to the study (see Section 3).

5.3.4 Data cleaning and built files

If data cleaning costs have been included in the grant award, ALSPAC will clean all data collected and provide a clean, built file for subsequent use. The delivery date of this file will be agreed in advance with the PI. ALSPAC will usually provide this service for data collected using ALSPAC staff on a PI's behalf. If the PI wishes to use their own staff to conduct data cleaning, it is their responsibility to provide ALSPAC with both raw and clean data files, along with the appropriate documentation in a timely manner.

5.3.5 Restricted data access

New data collected through a specifically funded exercise *may* be made available only to the PI for a period of exclusive access *only* on request to the Executive. Such exclusivity should not be assumed. Up to a year may be requested. If this is agreed, other interested researchers may access these data only after consultation with the PI.

6. Summary of researcher responsibilities

This section summarises the main responsibilities of any researcher wishing to work with the ALSPAC resource. The same rules apply to *all* researchers regardless of whether they are a member of ALSPAC staff, a new collaborator or a long-term collaborator.

6.1 Project proposals

It is important to note that the Executive do not consider overlap when approving projects. It is up to the researcher to determine whether any project they are proposing is not already being worked on by any other researcher and to be aware of any other researchers who may be working in their area of interest. The Executive may suggest possible collaborations, but the researcher is under no obligation: this is a suggestion rather than a pre-requisite of project approval. Summaries of all approved projects are available online.

6.2 Funding

All projects must be appropriately funded.

Please submit your online proposal at least one month prior to any funder's deadline date: Our finance team need sufficient time to be able to provide you with any appropriate costings. We request that any negotiations with funders MUST include the Executive at all times

A researcher must send the Executive a copy of the final submitted grant, the award letter and any other relevant documentation when it is received; the Executive will arrange a set-up meeting once funding is approved to agree the objectives, timetable and staff required to meet the grant commitments. This will be followed by annual review meetings to ensure the milestones are being met. The PI must make every effort to attend these meetings. It is the researcher's responsibility to ensure there is no conflict between their funder's terms and conditions and the ALSPAC DTA/MTA where applicable.

6.3 Data access

Researchers must adhere to the ALSPAC Access Policy and confidentiality form at all times. Researchers must also comply with the terms of the ALSPAC DTA/MTA where applicable. Current and future access is at risk if any researcher is found to be breaking these rules. In particular, data must NOT be shared with any other researchers without going through the Executive and the data buddy team. Serious breaches of data access rules will be prosecuted to the full extent of the civil or criminal law.

6.4 Confidentiality/security breaches

Any breaches of data security must be reported immediately to the Executive who will pass the issue on to the ALSPAC Data Security Manager for investigation. Examples of data security breaches include (but are not limited to):

- Any unauthorised person (i.e. has not signed a data access agreement for the relevant data set) gaining access to ALSPAC data;
- Sharing ALSPAC data with unauthorised persons;
- Failing to ensure data are sufficiently encrypted during transport;
- Sharing login details that permit access to ALSPAC data.

6.5 Derived variables

Any derived variables (such as data obtained as part of a new data collection exercise or newly derived variables coming from secondary analyses) created as part of any research project must be returned to your data buddy with appropriate documentation; these will be incorporated in to the main resource made available to all researchers. If you fail to return any derived variables and we receive a specific request from another researcher to access that data, we will follow up with you again. We would expect you to return the relevant data to your data buddy (in order for the IDs to be replaced) within 4 weeks. Failure to produce derived variables at this point may risk your future access to the resource.

6.6 Publication

6.6.1 Peer reviewed papers and other research output

All full papers must be sent to the Executive for approval along with a completed papers checklist *prior* to journal submission. Please note that if there are any **significant** changes to the paper after Executive approval, re-approval must be sought. This includes any research output being placed in the public domain (for example working papers or non-peer reviewed papers). The Executive will process all papers within two weeks of receipt. The Executive read all papers to check confidentiality is protected and to ensure that the paper will not bring the study into disrepute. The Executive reserves the right to require that any paper which could potentially breach the confidentiality of any ALSPAC participant(s) be withheld from submission for publication. The Executive will work with the authors to overcome such breaches. If the researcher submits such publications regardless, the Executive will attempt to prevent publication.

The Executive also provide advice and feedback to authors where we feel this may be helpful, but their role is not to provide formal peer review: The applicants/authors are not duty bound to follow the advice provided. Under all circumstances the ALSPAC Executive reserve the right to submit letters or papers for publications in response to any paper to explain study procedures or to express a coherent scientific argument.

A checklist of requirements for ALSPAC papers along with some accompanying notes explaining these requirements and containing appropriate text to insert is available with the papers checklist. A completed checklist must be included with each paper submitted for approval. Please note, this also applies to working papers. Researchers should let the Executive know when a paper is accepted and send through an electronic copy of the final published version.

A list of publications arising from the study can be found on the study website.

Any posters presented at conferences or other meetings using data from ALSPAC must include an acknowledgement to the study and the participants along with the study logo. Please see the completed papers checklist for suggested wording.

For papers using data gathered from participants at 22 years and onwards, you should also include a citation to REDCap, as the tool that ALSPAC have used to collect the data. Please see the REDCap website for details.

6.6.1.1 Open access

6.6.1.1a Papers

ALSPAC fully supports the Wellcome and RCUK policies on open access. In summary, this means that if a) the specific research presented in a paper is wholly or partly funded by **Wellcome** or b) any contributing author is wholly or partly funded by Wellcome (via salary or fellowship/studentship) any publication must be made open access. It is the senior author's responsibility to ensure that any papers published comply with this policy. It is the responsibility of the grant-holder under part a) above, or the individual author(s) under part b) above to cover the costs of making a publication open access. Please see the Wellcome website for more information. If your research is wholly or partly funded by the one of the research councils in the **RCUK** you

are required to make your research paper Open Access either via PURE, Bristol's open access institutional repository, or by publishing in a compliant journal. Please see the RCUK website for more information.

Please note that secondary analyses of ALSPAC data that is not funded by Wellcome nor has any contributing author supported by Wellcome does not need to comply with the Wellcome policy, however, ALSPAC would encourage this wherever possible.

A number of charities provide open access support. Please refer to the Charity Open Access Fund for up to date information.

6.6.1.1b Journals

A number of journals request that datasets used in a publication are deposited in publicly available resources. Our data management policy will not permit this, and we therefore request that the following statement be used by researchers in such cases: "Data used for this submission will be made available on request to the Executive (alspac-exec@bristol.ac.uk). The ALSPAC data management plan describes in detail the policy regarding data sharing, which is through a system of managed open access.

6.6.1.1c Grant applications

Some funders are also requesting that data be made publicly available. We recommend the same statement above is used in such cases.

6.6.2 Theses

We request that we are provided with an electronic copy of any theses that use ALSPAC data as soon as possible after a degree is awarded.

6.6.3 Reports and other publications

We request that we are provided with an electronic copy of any reports and other publications that use ALSPAC data as soon as possible.

6.6.4 Conference proceedings

ALSPAC do not need to see submissions to conferences.

6.7 The media

All press releases on research arising from the study must be written in conjunction with our communications team. We reserve the right to publish press releases on certain articles and expect the lead author of the article to agree the press release with the public relations team and to be available to deal with media enquiries and interviews. We may also ask authors to prepare a précis of important papers and/or lay-summaries to include in reports to funders and future applications for future core funding.

Appendix One: Example data access costs

Note that VAT will be charged where appropriate.

Example 1:

Total of 500 variables

Standard access Fee (includes 50 variables):	£2360
Additional 450 variables	£850
Total Cost	£3210

Example 2:

Total of 1200 variables which includes 30 SNPs and education data, so DTA required

Standard access Fee (includes 50 variables):	£2360
Additional 1150 variables	£2040
Education data	£525
DTA Administration	£910
Total Cost	£5835

Example 3:

Total of 940 variables plus ARIES methylation data so DTA required and address data

Standard access Fee (includes 50 variables):	£2360
Additional 890 variables	£1530
ARIES methylation data	£1815
DTA Administration	£910
Address data	£725
Total Cost	£7340

Example 4:

Total of 250 variables plus access to digital images so MTA required

Standard access Fee (includes 50 variables):	£2360
Additional 200 variables	£340
Digital images	£1265
MTA administration	£910
Total Cost	£4875

Appendix Two: Linkage data

Here we provide more detailed information on each of the different types of linkage data we currently hold.

1. Schools Data (provided by the Department for Education)

- This category includes data sourced from the National Pupil Database or from Local Authorities.
- Schools data are provided at a pupil and establishment level.
- Data are anonymised before release.
- Data are available for participants who have been sent a privacy notice describing our use of their data and have not dissented.
- If you require tailored data (e.g. specific subject grades at GCSE) additional disclosure controls will be required and such requests will be costed on a case by case basis.
- Schools data can only be used to investigate hypotheses which will promote the education or well-being of children

2. Further and Higher Education Data (provided by the Department for Business, Innovations and Skills)

- This category includes data from the Individual Learner Record dataset (further education) and the Higher Education Statistics Agency (HESA) (higher education).
- Data are anonymised before release.
- Data are provided for consenting index children only.
- This data can only be used to investigate hypotheses which will promote the education or well-being of children

3. NHS (National Health Service) and ONS (Office for National Statistics) data

- This category includes data from the Death and Cancer registries.
- Currently there is no agreement to share these data with third party researchers.
- The ALSPAC Data Linkage team is able to provide *aggregated* data which has been subject to statistical disclosure checks (e.g. a count of how many ALSPAC mother have a registered breast cancer).
- The ALSPAC Data Linkage team may undertake analysis of these data 'in house'. This is subject to agreement, capacity and the recovery of direct costs.

4. NHS Primary Care Data

- Primary care data, via Electronic Patient Records (EPRs) is available from ALSPAC or through ALSPACs existing linkage to the Central Practice Research Datalink (CPRD).
- CPRD: ALSPAC has linked to the CPRD database. Coverage is limited (by CPRDs population coverage) to ~500 enrolled index children. CPRD expect to increase this coverage over time. Access to these data is conducted using the CPRD Safe Setting (in London), this requires all analysis to take place in London. The ALSPAC Data Linkage team will arrange for a dataset of ALSPAC variables to be 'merged' into a copy of CPRD. This process is subject to statistical disclosure control checks which may lead to changes being made to the data to reduce the possibility of disclosing the identities of study participants/CPRD cases.
- ALSPAC Collected EPRs: ALSPAC are currently collecting the index children's EPRs directly from participants' practices. We are currently processing the data prior to making them available to researchers. Records are collected for consenting participants only. These data are subject to stringent anonymisation processes which will likely result in significant data change in some or all variables.
- The ALSPAC Data Linkage team may undertake analysis of these data 'in house'. This is subject to agreement, capacity and the recovery of direct costs.

5. NHS Secondary Care Data

- ALSPAC have linked to the Hospital Episodes Statistics (HES) dataset.
- These data are available for consenting young people only.
- These data are subject to stringent anonymisation processes that will likely result in significant data change in some or all variables.
- The ALSPAC Data Linkage team may undertake analysis of these data ‘in house’. This is subject to agreement, capacity and the recovery of direct costs.

Appendix Three: Split-stage protocol

Split-stage protocol for derived data

The split-stage protocol comprises of the following stages:

1. ALSPAC send the potentially identifying data (e.g. postcodes) in an encrypted file to the approved researcher group. This will be a stand-alone file with a unique case identifier, but unmatched to any other data.
2. The researcher then derives any new variables using their specialist skills or equipment (e.g. modelling postcode to satellite collected air pollution data). The researcher will ensure the derived values are less specific than the source variables and they could not be used as a proxy-ID to identify an individual. These new variables are then encrypted and returned to ALSPAC, along with accompanying documentation describing the derivation method(s) used to the appropriate data buddy.
3. When ALSPAC have confirmed they have received and opened the file, the researcher is then required to delete all copies of the information provided in stage 1.
4. ALSPAC will remove the potentially identifying information (e.g. postcode), then combine the derived variable with the remainder of the information in the data request. ALSPAC reserve the right to further process the derived data to ensure an appropriate level of disclosure control (ALSPAC will discuss this with researchers in the event of this occurring). This new data file will have a different case ID to the file issued in stage 1. ALSPAC will encrypt and send this file to the researcher after receiving written confirmation that stage 3 has been completed.

The extract mechanism by which the split-stage protocol is administered may vary, depending on which potentially identifiable information is being requested. Examples are provided below for illustration. The exact mechanism implemented is determined per project, at ALSPACs discretion.

IMPORTANT: Researchers are not permitted to join, or attempt to join, the information provided in stage 1 with the information provided in stage 4.

Example 1: Free text data

Both questionnaires and clinic datasheets contain free text fields, where additional information not contained within the given tick box responses can be provided by the respondent. These data are not routinely coded and included in built/release files as different researchers may choose to use the data in different ways according to their specific needs. However, there is a risk that participants have also included personal identifiers in these free-text fields.

Researchers, subject to approval by the ALSPAC Executive Committee, may be provided (there is a standard cost-recovery charge per text field; see section 1.3.1) with these data using the Split-Stage Protocol. It will be provided once the appropriate charges have been paid.

(Stage 1) the researchers will be provided the text data with a unique ‘TextID’
(Stage 2) the researcher will code the text variables
(Stage 3) return an encrypted copy of the derived – coded – variables to ALSPAC. Upon receipt, the researcher will destroy all copies of the free-text data.
(Stage 4) ALSPAC will combine the coded variables with the remaining requested data, add the project ‘Collaborator ID’ case identifier and send this data file to the researchers.

Example 2: Date of birth

Complete dates of birth and other dates (e.g. clinic date or questionnaire completion date) are not usually made available; only month and year are released as standard. The age of any data collection sweep is always computed and made available. Exact dates of birth will not be given to research collaborators. We recognise that there are times when this information is important for deriving variables such as season of birth variation in a particular measure.

In these circumstances, we will work with the researcher to produce their derived variables using a modified version of the Split-Stage Protocol, as follows:

(Stage 1) the researcher will be provided with a limited dataset containing a collaborator ID, pseudo date of birth, and any other essential data
(Stage 2) the researcher will use this dataset to write syntax that correctly generates any derived variables
(Stage 3) the researcher will send this syntax to the study team (in SPSS or Stata format)
(Stage 4) the ALSPAC data team will run the syntax using the genuine full date of birth to create the derived variables.

In situations where researchers provide a strong justification for needing complete date of birth throughout their main analyses (e.g. as a primary exposure, outcome or key co-variable) and the justification is agreed by the Executive, a similar process to that described above will apply. Here the researcher would have to provide complete syntax for the all their analyses in a form that the ALSPAC team can use and the researcher would be provided with a full set of results. The real costs of doing this additional work will have to be paid in advance by the researcher and will vary on a case by case basis depending on the amount of work required by the ALSPAC team.

Example 3: Clinically interpretable data

ALSPAC collects a variety of 'clinically interpretable data', for example: DXA bone mineral density scans, liver scans and brain MRIs. In many cases, these interpretable data may be potentially identifying in their raw format yet require the specialist expertise of a clinician to interpret them properly. The specific nature of the Split-Stage Protocol is likely to differ from project to project; however, the principles – where potentially identifiable information is processed in isolation from other ALSPAC information – is likely to be applied.

Example 4: Spatial data

Complete postcode data are not usually made available; rather the very broad first digits of postcodes are released, or information derived from these (e.g. household quintile of Indices of Multiple Deprivation at the time of data completion). However, we recognise that there are times when this information is important for deriving variables, such as for spatial research projects.

In these circumstances, we will work with the researcher to produce their derived variables using a modified version of the Split-Stage Protocol, as follows:

(Stage 1) the researcher will be provided with a limited dataset containing postcode and any other essential data. To protect the identities of participants the genuine participant postcodes will be masked by including other, randomly selected, genuine postcodes and synthetically created essential data.
(Stage 2) the researcher will use this dataset to write syntax that correct generates any derived variables
(Stage 3) the researcher will send encrypted copies of the derived variables to the study team, and, upon receipt, delete all copies of the original Stage 1 data
(Stage 4) the ALSPAC data team attach the derived variables to the remaining requested ALSPAC information, change the case ID and return this file to the researchers. The derived variables will be checked for disclosure risk and may be processed to a less granular level (the means to achieve this will be discussed and agreed in advance).

IMPORTANT note for projects requesting spatial data:

- The ALSPAC team will **not** provide exact address or complete postcode data under any circumstances, due to issues of identifiability. Instead a range of derived administrative boundary variables are available as outlined in the data dictionary. Each project proposal will be judged uniquely on its own merits and disclosure risk.
- Previous provision of, and the availability of, geographical data are not a guarantee of data provision.
- Requests for specific geographies may be denied in cases where it is believed participants' disclosure may be at risk.
- As a condition of submitting a proposal that includes ALSPAC spatial data a researcher will be required to include detailed information on the reasoning and methodology behind the requested geography to justify the choice, and specify why the selected spatial resolution is appropriate for the research question, for instance, in the case of high resolution geographies being requested, the Executive require justification as to why smaller resolution data are not acceptable.
- It is important to note that at the highest resolution held by ALSPAC (Lower Super Output Area) many data cases will be reverted to missing due to low unit population counts, therefore selecting variables with the highest resolution possible so that further information can be derived when unnecessary can be counter-productive to research.
- The ad-hoc method of address data management has permitted a database with extremely high temporal accuracy. However due to historical database errors, and individual level differences in reporting address movement, there will inevitably be a small number of cases that have no address data at certain time points. These missing cases should not greatly affect research that uses additional ALSPAC data as there is understandably a very high correlation between address accuracy and questionnaire/clinic responses.

Split-stage protocol for study administration

In some exceptional data collection instances, it may be necessary for non-ALSPAC staff to be involved in the administration of the data collection. In turn, this may mean researchers accessing participant names and contact details (e.g. a qualitative researcher arranging interviews that they will conduct, and subsequently analyse). In these circumstances, a modified version of the split-stage protocol will apply with important additional stages:

Stage 1: Following project approval and negotiation with ALSPAC staff, the researcher (with input from ALSPAC) will be expected to produce a fair processing information and consent pack for study participants (this is likely to need ethical review and approval). The information must explicitly describe the fact that the researcher does not work for ALSPAC and to seek opt-in consent for ALSPAC to provide contact information to the researcher.

Stage 2: ALSPAC staff will administer the consent campaign (subject to cost recovery fees being paid).

Stage 3: Subject to required contractual agreements being entered into, ALSPAC will provide the researcher with the names and contact details of consenting participants. These will be used solely for the purposes agreed in the project proposal.

Stage 4: The researcher will collect the information required for the project. The researcher will ensure that research information is kept securely, and separately from the names and contact details (it is required that these are kept within separate electronic file systems and only accessible on a need to know basis to staff named in the project proposal). ALSPAC can provide a case ID to assist with this data separation. Once data collection is complete, the researcher will process the data and then return an encrypted copy, along with sufficiently detailed documentation, to ALSPAC. Upon confirmation of receipt, the researcher will securely destroy all copies of the name and contact information (ALSPAC will maintain a key to link participants back to their data if required).

Stage 5: ALSPAC will combine the collection information with other ALSPAC information, assign a new collaborator case ID number and will then send an encrypted copy of this final dataset to the researcher.

The above description is for illustration only; details are likely to vary on a project by project basis.

Appendix Four: Policy updates

This appendix will detail the changes made to this policy since the release of v5.0 in April 2014.

v.5.1 Released May 2014

Statement added to Section 1.3 clarifying that costs for education data are subject to change

Project amendments clarified in Sections 1.3.2 and 1.3.3

Sections 1.3.4 and 1.3.5 have been added

Final bullet point added to Section 2

Paragraph added to Section 2.2 regarding posting lay summaries of projects using linkage data on the ALSPAC website

Statement added to Section 5.1.8 regarding paperwork for new data collection must be signed off internally

Statement added to Section 6.1.1.1 regarding journals and depositing data.

Section 6.5 – Clarification added for returning derived variables

v.5.2 Released September 2014

Clarification on open access policy in Section 6.6.1.1

Statement added to Section 6.1.1.1 regarding funders and depositing data.

Including all research output in Section 6.6.1

Clarification of obtaining sequencing data in Section 2.3

v.5.3 Released October 2014

Clarifying charges in Table 1 for non-standard data

v.5.4 Released December 2014

Adding details on other charities supporting open access publication – Section 6.1.1.1a

v.5.5 Released February 2015

Clarifying the use of schools and higher education data – Appendix Two

v.6.0 Released April 2015

Updated cost recovery charges

Clarifying the data management fee and exchange rates in Section 1.3.3.3.

v.6.1 Released June 2015

Updated to accommodate new online project proposal system.

Updated to include costs for new data collection in Section 1.3.6.

v.7.0 Released September 2016

Updated cost recovery charges

Clarified the scope of set-up meetings in Section 5.1.2

Changed Wellcome Trust to Wellcome to reflect their new name

v.7.1 Released November 2017

Updated Section 1.4 to reflect changes to the management of ALSPAC

Updated Section 2 to give information about the various ways of interrogating our data

Changed Sections 4 and 5.1.11 to refer to recently formed Bristol Bioresource Laboratory (BBL)

Added information in Section 6.6 about referencing study on posters, including logo and citing REDCap

Addition of Appendix Three on the split-stage protocol