ALSPAC AVON LONGITUDINAL STUDY OF PARENTS AND CHILDREN

Access Policy

v. 14.2

March 2024

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 and grandchildren using self-completion questionnaires, data extraction from medical notes, linkage to routine information systems and from hands-on research clinics. 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 both the 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.** A bona fide researcher is defined as being a person with professional expertise to conduct bona fide research; and who has a formal affiliation with a bona fide research organisation that requires compliance with appropriate research governance and management systems as described in the Data Access Agreement (DAA, see section 3.3). Please contact us if you are in any doubt as to whether you fit this definition, which includes both public and private users 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 a resource to be used by the research community. We encourage and facilitate data sharing by 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 in the public interest and is **not** being carried out for personal or commercial gain.

The vast majority of data are available for 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.

1.2 Requesting access to data

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.

Researchers are required to complete an online proposal form available via our website. 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 a fully completed submission to inform you of the outcome. Please note that in some cases we may need to complete a Data Privacy Impact Assessment (DPIA) which will affect the time taken to approve a proposal (see section 6.4). For more complex proposals, .e.g. those including 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 conditions 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 **four weeks prior** to the submission deadline. It may not be possible to approve those received less than one month before the submission deadline in time. 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.

Any proposals that include the collection of *new* data (see Section 5), should 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 **eight weeks prior** to the submission deadline.

1.2.1 Amendments to existing data requests

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 (note we may request this is submitted as a new project if the Executive deem appropriate);
- Change in start or end date;
- Additional researchers accessing the data;
- Change in institution hosting the data;
- Any additional data required;
- Change of funding source.

If a researcher wishes to re-use a dataset that has already been provided for a previous project (for example, to study a new aim or for a new student to use the data for their studies – please see section 1.3.6 regarding student projects), a new proposal is required. The proposal must include reference to the original project B number (ALSPAC reference number provided when a proposal is submitted) and justify all the topics of data required, as expected for all proposals. We require a list of all the variables to be used which will be reviewed against the proposal as usual (see section 3 Data provision). All applicants will be required to complete a new Data User Responsibilities Agreement. There are no charges, unless:

(i) The dataset is >2 years old. In this case the dataset will require 'refreshing' with the current withdrawals of consent and any other updates to the data in the main resource. Costs will apply and additional costs may apply for certain types of data e.g. linkage, genetics.

(ii) Additional data is required at any stage; A new project ID and a new Data Access Agreement (DAA) (see section 3.2) are required. This may be costed as an amendment fee (plus any additional data fees), plus the DAA administrative fee, as outlined in table 1 below. However, if a large amount of new data are required, the 'Basic Fee' (in table 1) may be applied instead due to the work involved in reviewing and generating the request.

A note regarding projects that started prior to 2018 when the Data Access Agreement (DAA) came into use (see Section 3.2); If a project has been dormant (i.e. the researchers have not been in touch with ALSPAC) for more than 2 years, it will be considered 'closed' unless an amendment stating the project end date has been approved. This policy does and can change often and therefore older projects will need to be brought in line with current practices as much as possible. This may include (i) submitting a new proposal (ii) ensuring that co-applicants are known to ALSPAC, and (iii) generation of an updated dataset as mentioned above, Researchers must only use the data provided for the aims of the approved proposal and for no other reason unless approved by ALSPAC.

Important information regarding international data sharing: For institutions that have already received individual-level ALSPAC data but now meet the criteria for the International Data Transfer Agreement which came into effect on the 24th March 2024 – please see section 3.3 'Data Access Agreements' (DAA). Please we aware that if additional data is requested, any existing agreements must be updated. If the DAA v3 (effective 28th May 2021 to 24th March 2024) has been signed already, an Addendum will be required before any new data can be shared. If a previous version of the DAA was signed, a new DAA and an IDTA will be required before any new data can be shared. These updates will be required for all institutions affected by this change who will receive the additional data. Your Data Buddy will provide the appropriate agreements for signing.

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 costs relating to data access and provision. All researchers accessing ALSPAC data will be charged on a cost recovery basis unless a grant proposal is being submitted (see section 1.3.3). **This cost will vary depending on the types of data requested and the amount of time required to facilitate your request.** Once a proposal has been approved and the applicant informed of the costs, 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. Current charges for data requests are provided overleaf (Table 1) and example costings can be seen in Appendix One. There are three types of data requests detailed in Table 1: 1) Basic costs for data that is readily available; 2) Bespoke data that requires additional work to release and 3) Amendments to the original request. We reserve the right to charge additional costs not detailed in Table 1, if a project requires more resource than typically expected. These costs will be discussed with researchers in advance as required.

Data will **not** be provided until all necessary paperwork is completed and an invoice has been settled.

Costs are reviewed on a regular basis and are therefore accurate at the time they are provided. It is the researcher's responsibility to check that costs haven't changed if there is a significant lapse in time between obtaining the quote and starting the project.

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 January 2022)

	Type of data requested	Charge (+ VAT where applicable)
1.	**Basic fee** – (this includes the cost of the PI's Data Access Agreement (DAA) and any data available in the data warehouse [e.g. clearly described in the ALSPAC data dictionary). Large datasets (>10,000 variables) will be charged an additional £400 per 10,000 variables to cover the additional time required to review and create Additional DAA*	£1925 £390
	Omics data – this includes a bulk data release including *any* of the omics datasets outlined in the Omics Catalogue. The cost is per release, e.g. three datasets may be requested initially and the access fee listed will cover all three. If an additional omics dataset is requested at a later date, the access fee will apply again.	£825
	MRI scans - this includes a bulk release of *any* MRI scans as outlined in the Data Dictionary. There a number of sets of MRI scans collected from different sub-studies and the access cost is per release e.g. two MRI scan sets may be requested initially and the access fee listed will cover both, if an additional set is requested at a later date the access fee will apply again.	£330
2.	***Bespoke data (these charges are in addition to basic fee)***	
	G2 (Children of the Children) data**	£1375
	Other 'physical' data; scans; sound file; videos etc. [per time sweep, per participant cohort and per type of data/media] †	£1375
	Adding other data **not available in ALSPAC data dictionary** as part of the core datasets e.g. linkage, variables obtained from other researchers, variables derived as part of on project to be used in a new project etc‡	£275 per data type
	Data requiring **split stage protocol** (see Appendix 3)	£825
	Linked Health and Education Records	See the linkage data user guide
	Other linkage data (e.g. linked geographical data, other third party data– see Section 2.2)	
3.	**Amendment fee** – to add additional data available in the data warehouse [clearly described in ALSPAC data dictionary] to the dataset at a later date. Additional charges will apply as above for other data types The 'amendment fee' will be waived if you are asked to submit an amendment to justify additional data *prior* to your dataset first being shared with you. If the data requires a re-fresh (see section 1.2 for details), the charge will be £250	£550

* A DAA is required for all sites receiving data and the cost is *per* additional site.

** This fee is to cover the additional work currently required to link G2 data and undertake disclosure controls (see section 2.1.1)

† This fee is to cover the cost of de-identifying the physical data to send out, handling derived variables and incorporating them into the dataset

‡ This covers the additional work this entails

Costs for example projects are included in Appendix One: Example data access costs. Please note that access to biological samples will incur additional costs (see section 4 for further information).

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. Administrative amendments will not be charged unless a new DAA is required as the result of adding a co-applicant who will be using the individual level dataset in a different Institution, the administration fee is described in Table 1 above.

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 (including personal fellowships and PhD studentships where possible) for secondary data analysis, you must include a set **'Data Management Fee' (DMF) of £8250** (+ VAT where applicable). This includes researchers based at the University of Bristol.

This fee covers access to existing data (any data available in the data warehouse and clearly described in the ALSPAC data dictionary and the omics catalogue) and the administration of your proposal only; if you are proposing to collect new data or request biological samples, these will be costed separately (see sections 4 and 5). Similarly, if you require any non-standard or linked data, appropriate costs will be added onto the DMF.

The DMF includes the basic data access costs as detailed in section 1 of Table 1, one DAA, plus up to three data request amendments for standard data (the data readily available in the data warehouse and data dictionary). If you submit more than three amendments during the life of the project, or you request data that is non-standard (see Table 1, section2), you will be charged for these as detailed above (Section 1.3.2) Additional DAAs will be subject to the DAA administration fee outlined in Table 1. If you require data described in section 2 of Table 1, these costs will need to be added to the DMF included in your grant. Please contact alspac-data@bristol.ac.uk to discuss an accurate costing prior to submitting your grant application if needed.

Please provide sufficient time to enable us to accurately cost your project. We request you submit your project **four weeks prior** to submission to the funder (if you are collecting new data we require a minimum of eight weeks – please see section 5).

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.

If your grant does not require secondary data analysis, e.g. it is a qualitative study, a pilot study for new data collection techniques or similar, then a fixed fee of **£770** (+ VAT where applicable) must be included. This covers the administration of your proposal *only*. Additional costs will be provided separately depending on your requirements.

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 for you to include with your grant proposal once your ALSPAC proposal has been approved and the budget has been agreed.

If a grant is not successful but you resubmit elsewhere at a later date, please ensure that you obtain up to date costs. You should submit an amendment to your original project indicating a change in funder.

1.3.4 Costs for pre-access information

On occasion, researchers require more detailed information than can be gleaned from the ALSPAC data dictionary, e.g. cross-referenced data for inclusion in grant proposals. These will be dealt with on a case-by-case basis and will almost certainly incur additional resource and time, which may need to be paid for. Any requests must therefore be made to the Executive to ensure sufficient time to agree requirements and/or release data under our usual agreements.

1.3.5 Costs for new data collection

Projects involving new data collection (see section 5) will need to be fully costed. The PI must liaise with the Executive to request a costing following approval of their project proposal, and before any grant is submitted. Questionnaires are currently 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, biological sample collection and sub studies involving hands-on data collection will vary according to the requirements of the project. We require a *minimum* of **eight weeks' notice** before the grant deadline in order to provide accurate costs and obtain agreement from the University of Bristol finance department. We regard accurate costings and appropriate planning as critical for providing the best support for new data and we reserve the right to decline new data collection activity requests if we do not have sufficient notice. For new data collection projects seeking grant funding and wishing to use ALSPAC data for analysis, our Data Management Fee will also apply, see section 1.3.3 above.

1.3.6 Student projects

ALSPAC is committed to enabling and supporting education through the use of study data to answer research questions via PhDs (including the Bristol system of mini projects), Masters and undergraduate dissertations. As described above the provision of data from ALSPAC is through a cost recovery model (as the data access team are not core funded) but we appreciate that not all studentships may provide sufficient funds to cover this. In particular, we wish to promote the notion of access to ALSPAC being a benefit for studying in Bristol.

1.3.6.1 ALSPAC Pledge to PhD projects

We consider funding for these (both UoB and externally based projects) using the following hierarchy:

1. Where these are funded by a grant application. We request the data management fee as detailed section 1.3.3.
2. If there is a cap on research costs available for the studentship, we will calculate costs based on the cost recovery fees listed in section 1.3.1.
3. If limited funds are available, we will consider reducing the cost recovery fee by up to 50% as funding allows.
4. Where no funds are available, we will support up to three PhD projects per academic year (defined as starting on 15th Sept) at no charge for use of data included in part 1 of Table 1 in section 1.3.1. These are on a first come-first served basis with no single institution being awarded more than one per year.

For Bristol-based students we will continue to support our existing model for primary PhD projects to facilitate the development of independent researchers and to provide additional benefit to being Bristol based; fees may apply.

1.3.6.2 ALSPAC Pledge to Masters and undergraduate projects.

For Masters and undergraduate projects that require bespoke datasets we request that the student has an experienced Bristol-based data user as part of their supervisory team, regardless of whether internal or external. It is then the supervisor's responsibility to put together the required dataset to be processed by the data access team, with the associated charges. In the absence of such support, the usual cost recovery charges will apply. If insufficient funds are available, we will discuss with the relevant programme directors on a case-by-case basis.

1.3.6.3 ALSPAC Pledge to Teaching.

A number of synthetic and 'frozen' datasets are freely available for classroom-based teaching. If these do not meet requirements, we will consider additional requests on a case-by-case basis.

1.3.6.4 Submitting student projects

Supervisors are ultimately responsible for their students in the same way that PIs are responsible for their researchers. We request that any **proposals for projects that are to be undertaken by students are therefore submitted by the supervisor** rather than the student themselves. Each student project (by this we mean research that will in part or in whole contribute to a student's submitted work regardless of whether it is postgraduate or undergraduate) must be submitted separately, regardless of how similar multiple projects might be and whether or not they fall under the remit of a researcher's existing approved project. Student

projects should be identified as such when submitting the proposal by ticking the appropriate box on the online proposal form. If the student will re-use an existing dataset, refer to 1.2.1.

1.4 Management of the resource

The team is led by Professor Nic Timpson (Principal Investigator).

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. Other members of the Executive are Ms Lynn Molloy (Chief Operating Officer), Dr Susan Ring (Executive Director, Bioresource) and Professor 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, in addition, Professor George Davey Smith acts as Scientific Director (SD).

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 and strategic oversight of the study, which includes the approval of data access proposals,new data and sample collection projects and manuscripts 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 and security, data linkage, clinic, 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 PI and SD with scientific strategy. The Board is chaired by the Head of Population Health Science at the University of Bristol (Professor Matt Hickman) which is comprised of senior faculty, clinical and institution staff. The Board requires the presence of at least one practising clinician. The Board meets on a quarterly basis. 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 Tim Frayling. 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 study participants. The terms of reference of ALEC are available on the website.

1.4.6 Clinical Advisory Group

The ALSPAC Clinical Advisory Group provides advice and support on clinical matters to the ALSPAC Executive. The Group can provide clinical advice to researchers planning new data and sample collection grants via the ALSPAC Executive email account alspac-executive@bristol.ac.uk

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 (please see Section 1.1 for details of how to determine what is available). ALSPAC collects data directly through questionnaires, other remote methods and via face to face assessment. 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.

In rare cases, 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;
- 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 outlined in Section 4.

Any challenge against a refused proposal will be considered by ISAB (described in Section 1.4.4). It is important to note that the Executive do not judge the science of the proposed work nor the overlap with any existing approved projects.

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), with the exception of that collected from G2 (see section 2.1.1). Non-genetic results obtained from biosamples are made available as soon as all assays have been completed and the data have been cleaned and documented. All potential identifiers are removed, and disclosure risks are considered such that data may be grouped where appropriate on return.

2.1.1 G2 data

Data from G2 (please see here for a summary of the data being collected) will be provided subject to additional costs. Currently, there are additional requirements around linking datasets and issues of disclosure that have to be considered. These depend on the data requested and will therefore be considered on a project by project basis.

If you are planning a proposal to access data from G2 you are strongly advised to discuss your plans with the data team prior to submission. The team will be happy to provide further information about data collection, availability and future plans. Please contact: alspac-data@bristol.ac.uk and the team will get back to you as soon as possible.

G2 data from any individual data collection will be first released once 500 participants have been achieved. From that point on, updates will be made annually each September. This is non-negotiable: if you request data from G2, you will only receive the current released version.

2.2 Linkage data

ALSPAC collect data using linkage to routine education, health, administrative and environmental records. Linkage can either be conducted at an individual level (e.g. primary care records) or an institutional level (e.g.

a school). These data are collected from external organisations and details of existing linkages are available in the linkage user guide. 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.
- Trusted Research Environment (TRE): Projects requesting linked health and education records will have data provided via ALSPACs Secure eResearch Platform (SeRP). The SeRP platform provides full functionality for epidemiological investigations with increased governance oversight. This comes at a direct cost to ALSPAC (as we lease this infrastructure from Swansea University) but is necessary to maintain access to data. All analysis will take place in the SeRP server and it is not possible for research users to export anything from the server. Instead, all outputs come via the ALSPAC linkage team who assess them for disclosure risk. It should be noted that SeRP does not currently have the performance capacity to support the analysis of omics data when using linked data.
- 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 ALSPAC Data Linkage Team will provide an estimate for any data access costs during project negotiations. Details of the costing structure can be found in Appendix Two.

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 linked health 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 NHS, administrative and other digital 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.

2.2.2 Secure Research Servers

The ALSPAC Data Linkage Team have established two secure research servers. We have an analysis secure server, and a separate server that is primarily used for data curation.

The primary secure server for research analysis is the ALSPAC UK Secure eResearch Platform (UKSeRP). This is provided as a service to ALSPAC by the University of Swansea. This is a remotely accessible and effectively anonymous research environment. All analysis of linked health and education records for non-University of Bristol users will take place on the UKSeRP, this will also apply to most University of Bristol users.

ALSPAC holds cancer and mortality records for which there is no current agreement to transfer this to the UKSeRP or to permit access beyond University of Bristol staff for permitted projects. For a limited number of University of Bristol users, it may be possible to permit access to a secure University of Bristol server. The ALSPAC data linkage team are seeking permissions to move these data to UKSeRP and open access more widely.

For both servers, the protection of our participants confidentiality and the security of these sensitive data is paramount. It is not permitted for users to directly extract any data from either system (either physically or by copying information from the screen). All extracts of *population data* must go through the ALSPAC data linkage team who will conduct disclosure control checks. The extraction of *individual level* data is unlikely to be permitted (if you have a need for this, please discuss it with the ALSPAC Data Linkage Managers). It is the responsibility of the user in the first instance to ensure the data are non-disclosive (it is mandatory that all users undergo training in this before access is provided). There is a cost-recovery charge levied on the disclosure checks process (this will be described to you by the ALSPAC Data Linkage Managers as part of the application process).

2.3 High throughput array/omics data

Details of genotype data available from the ALSPAC participants are available on our website and in the data dictionary. This includes genome wide microarray data (GWAS data) and expression data.

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) 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 the mQTL Database.

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

2.4 Potentially identifying data

Some of the data that ALSPAC have collected could be used to readily identify study participants. These include but are not limited to: 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 and may enable a study participant engaged in research to identify themselves. 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 this ‘split-stage protocol is to create the conditions where research collaborators are able to access identifiable information, while maintaining a distinct 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. Table 1 provides details of the costs for data that require the split-stage protocol.

3. Data provision

3.1 The ‘data buddy’ process

When the Executive has approved your project, and a project is funded, a data buddy will be assigned to you. Your data buddy is a member of the Data Access Team and will assist you throughout the duration of your project; advising you on how to make your formal data request, costing your proposal, administering the necessary paperwork including the Data Access Agreement (DAA) and the Data User Responsibilities Agreement (DURA), and ultimately providing 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; however, some types of data may take longer (see Section 2). The Data Access Team are not core-funded therefore the costs in Section 1.3 primarily support the team in providing data and supporting researchers with their access requirements.

3.2 Data requests

We are only able to share data that has been requested, justified, and approved via a submitted proposal or amendment. Please ensure all proposals/amendments submitted clearly outline all data types and topics, participant groups and timeframes you wish to access.

A data buddy will be assigned once your proposal is approved and funded. They will provide details on how to make your formal request which should include every variable you wish to use. It is your responsibility to ensure your request is complete.

Your data buddy will carefully review the request to ensure that all data types and topics, participant groups and timeframes are covered by your proposal/amendment and will ask for an amendment if these have not been adequately described. The full data request is included in the DAA (see below) and is an important record of the variables to be built into a bespoke dataset to be shared with you.

Please be aware that it is vital we review and agree your completed data request, as represented by your proposal; this can take time. We strongly recommend that sufficient time is taken to ensure your data request form is as you want it before it is returned for review. Please ensure it reflects the topics included in the approved proposal. Changes to the request form throughout this setup period can cause delays as amendment submissions may be required and the review process may need to start again.

3.3 Data Access Agreements

Protecting the confidentiality of the study participants 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.

Data User Responsibilities Agreement (DURA)

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 in our DURA. It is the PI’s responsibility to ensure that all their staff (or members of the research group) are known to ALSPAC (via an approved proposal or amendment), have completed a DURA, and are made aware of, and are adhering to, their data usage responsibilities (including all relevant ALSPAC policies and procedures). A sample version of the DURA is available on our website.

Data Access Agreement (DAA)

The DAA is a legal document laying out the terms and conditions under which ALSPAC shares data and is **non-negotiable**. It must be signed by the Principal Investigator (PI), the recipient institution’s legal signatory and ALSPAC’s legal signatory. If the study includes researchers at different institutions who will be receiving

the data, a DAA will need to be completed for each institution. A sample copy of the DAA is available on our website. We recommend you share a copy with your legal signatory prior to applying for access to our data to ensure they are willing to sign up to the terms. This is particularly applicable to institutions in the United States – we would encourage any prospective data user in the US to check the terms of the DAA prior to applying for access. No data will be released until these documents have been signed and returned.

If for some reason an Institution cannot sign the DAA there are other options that could be considered to facilitate your research:

1. An additional co-applicant is added to the proposal whose institution can sign a DAA and who can perform the analysis on your behalf. It is the responsibility of the research team to identify an appropriate co-applicant to collaborate with. This co-applicant would be able to share summary data with other members of the project team but cannot share individual-level data. This is the simplest and most cost-effective option.
2. Data is released via UKSeRP under an amended DAA which has reduced information security requirements and allows more flexibility as the datasets are not physically shared. Instead data will remain in UKSeRP and analyses will take place via remote access to the secure environment, with restrictions and controls on data upload and download from the environment. This option may not be suitable depending on the why the agreement cannot be signed.
   Additional charges apply in this case. UKSeRP costs are significant and example costs can be found in Appendix 2 section 7 as it is primarily used for linked data. Each project will be costed on a project-by-project basis and the cost depends on several factors including:
   - Length of project (time)
   - Number of users requiring access to the data
   - Which statistical software is required in UKSeRP
3. As a last resort only and in very pressing situations which have been discussed and approved by the Executive; a researcher may come to work on ALSPAC premises to analyse the data or pay for a member of ALSPAC staff to analyse the data on their behalf. These options depend on desk space and capacity within the team and is therefore unfortunately often not feasible. Significant costs would apply for staff time and desk space

International Data Transfer Agreement (IDTA)

For some institutions outside of the UK, an IDTA will also be required. The IDTA came into effect on the 24th March 2024 and is a requirement of UK law. An IDTA is required when an 'adequacy regulation' is **not** in place between the UK and the country in which the institution resides. You can find out more on the Information Commissioner Office (ICO) website here.

The IDTA replaces the Standard Contractual Clauses (SCC) that were included in our DAA v3, in use from 28th May 2021 to 24th March 2024).

We use the ICO IDTA template. This agreement cannot be changed and must be signed before any data can be shared. Your Data Buddy will share a copy for signing along with the DAA.

For institutions that have already received individual-level ALSPAC data but now meet the criteria for the IDTA: Please we aware that if additional data is requested, any existing agreements must be updated. If the DAA v3 has been signed already, an SCC Addendum will be required before any new data can be shared. If a previous version of the DAA was signed, a new DAA and an IDTA will be required before any new data can be shared. These updates will be required for all institutions affected by this change who will be received the additional data.

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. If you wish to move data between projects you must submit an amendment to request the variables under one of the proposals so they can be shared with the same identifier and include a justification, fees will apply.

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. The passwords and datasets must never be shared via the same method as each other. Data is only shared with applicants identified as data users on an approved proposal or amendment.

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.G0, G1 or G2))
- 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 and on the UKCRC Tissue Directory. There may be a cost associated with identifying sample availability, particularly if the selection criteria requires specific associated data. Researchers will be informed if this is applicable to their proposal (also see section 1.3.4 Costs for pre-access information).

Ethical approval was obtained for all sample collections and usually included consent for future research including genetic studies. The participant information sheets and consent forms used for biological sample collection sweeps can be found on the ALSPAC website here. The ALSPAC biorepository is an NHS Research Ethics Committee (REC) approved Research Tissue Bank (RTB), REC reference 23/SW/0058.

For UK applicants: biological material may be provided under the RTB's generic ethical approval provided the proposal falls within the remit of ALSPAC. If a proposal is outside of the remit then separate ethical approval will need to be sought.

For non-UK applicants: biological material may be provided where the proposal falls within the remit of ALSPAC. It is the responsibility of the applicant to ensure that prior to the export of samples the applicant is satisfied that the samples will be handled appropriately and that the required standards of the recipient institute and country have been met.

Where applications fall outside of the remit of the ALSPAC RTB, proof of alternative ethical approval may need to be provided before samples are released (i.e. copies of ethics application forms and/or approval letters).

The opinion of ALEC and/or other ALSPAC committees may be sought regarding any ethical issues that could arise from sample analysis, if deemed necessary by the ALSPAC Executive.

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 planning a 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 genetic analyses on ALSPAC DNA samples 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 so that the sample can be utilised for assays which are able to use 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. 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, request longitudinal 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 eight 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. For example:Illumina bead chip analysis on ALSPAC samples will be run in house (e.g. HumanMethylationEPIC BeadChip).

- 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 a local agreement if samples remain in the University of Bristol.. Each agreement will ensure projects meet Human Tissue Act (HTA) compliance requirements and 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. Once your proposal is approved we will provide word files of the relevant agreement for signature.

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.

For proposals also wishing to use secondary data alongside data generated from samples, this will follow our split-stage process (see section 2.4) whereby the new data will not be linked to the secondary data until fully cleaned and returned to ALSPAC with appropriate documentation unless explicitly agreed by the ALSPAC Executive.

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). For projects seeking grant funding, see Section 1.3.3. Researchers are requested to submit a proposal through the online system, clearly indicating new data collection, as soon as possible to initiate the costing process.

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 (or a Bristol-based scientist nominated by the Executive) is named as a co-applicant so they can act as guarantor for the proposed new data.
- All new data collection is discussed with our participant advisory group (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. Before a proposal is considered by the Executive it is reviewed by a relevant member(s) of the Senior Management Team to assess feasibility, resourcing, timescales, recruitment, any case selection requirements, participant overload and other practical and ethical considerations. The Executive will consider this 'technical review' and advise the PI if there are any issues which need to be addressed at an early stage in order for the study to proceed. The timing of new data collection is subject to both participant and staff burden. We reserve the right to negotiate the timing of any new data collection once it has been funded.

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'. Note information sheets and consent for collection of biological sample must be approved by the Designated Individual for the Human Tissue Authority Licence covering ALSPAC activity. This ensures that the design, wording and information given in each study are consistent. The ALSPAC Communications and Participation team and our ALSPAC Participant and Public Advisory Panel (APPAP, See Section 5.1.9) 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. However, any proposals for ethical approval covering collection of Human Tissue should be discussed with the Designated Individual for the Human Tissue Authority Licence covering ALSPAC activity. The REC

with most experience of ALSPAC study proposals is South West – Central Bristol. A University of Bristol Research Governance team research registration checklist is available which should be submitted for University of Bristol governance purposes for before seeking REC approval.

5.1.9 The ALSPAC Participant and Public Advisory Panel (APPAP) 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 APPAP or other participant groups as required. ALSPAC will advise on whether this is necessary, and PIs may be asked to attend an APPAP 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. Sufficient time must be built in to run the pilot, assess it and adapt the protocol if necessary.

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. 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

DBS checks are carried out in line with the University of Bristol's DBS Policy. Professional Services roles are assessed as to whether they need a DBS check at the point the role is created in conjunction with HR and the requirement is captured in the job description. Academic staff who will be dealing with participants face-to-face require a determination on a case-by-case basis and should consult with the executive who is turn will discuss with the relevant University of Bristol HR Business Partner to determine if a check is required.

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 research 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.

5.2.8 Feedback of incidental findings

Researchers are expected to comply with the ALSPAC Incidental Findings policy. Feedback of incidental findings will be discussed at the new data collection planning stage.

5.2.9 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. For proposals also wishing to use secondary data alongside data generated from new data collections or new samples, this will follow our split-stage process (see section 2.4) whereby the new data will not be linked to the secondary data until fully cleaned and returned to ALSPAC with appropriate documentation unless explicitly agreed by the ALSPAC Executive.

5.2.10 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.

5.2.11 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. The data at this stage will include a unique ID e.g. study ID or clinic ID.

5.3.3 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.

5.3.4 Additional variables

Once data collection is complete and the dataset has been cleaned, 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 and at this stage the new data will be matched with the unique proposal ID (see Section 3).

5.3.5 Exclusive data access

Where a researcher (member of the ALSPAC team or an external collaborator) has secured funding for the collection and analysis of new data, they may apply for a period of exclusive access for a period of **up to 6 months** from the point at which a cleaned dataset is made available to them (please use the form available on our website for this purpose: exclusive data request form) The request must demonstrate a planned analysis associated with the new data collection that must be delivered to the specific funding scheme which is responsible for the collection of the data referred to.

Bearing in mind that researchers who collect new data will invariably have lead-in time associated with the project (as snapshots of the data can be provided prior to release), **the reason for exclusive access must be clearly defined and justified in relation to the funded research** and should not be based simply on the fact that funding was obtained to collect the data., The Executive will consider each request on a case-by-case basis. We will only grant a period of exclusive access if there is a clear reason for withholding data from other data users as describe above. Exclusive access is subject to a limited time period and will made available to others as quickly as possible, in order to comply with our funder's position on exclusive access. We therefore recommend that researchers with approved exclusive access ensure all other aspects of their project are in place (e.g. DAAs) in order to receive the data in a timely manner.

If a request is approved, then during the exclusive access period the ALSPAC Executive will still consider requests for access to the restricted data, but permission must be sought from the researcher who funded the data collection to release the data or to explore the potential for collaborative analysis. If the funding researcher declines to share/collaborate, the restricted data will not be available to others until after the period of exclusive access. After the embargo period all ALSPAC data will be released and made available to external researchers (within the usual constraints related to scientific legitimacy and disclosure risk).

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, except where a request is made for finite samples. 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.

Any submitted proposal must be completed in full and include all requested details about the applicants, funders, and data requested.

6.2 Funding

All projects must be appropriately funded. See section 1.3 regarding costs and grant funding.

If seeking grant funding, please submit your online proposal at least *four weeks (eight weeks in the case of new data/sample collection proposals)* 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.Funders are reviewed as part of our 'due diligence' checks when reviewing new proposals.

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. For new data collection projects 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 ALSPAC's formal paperwork (Data Access Agreement, Data User Responsibilities Agreement and Material Transfer Agreement, where applicable).

6.3 Data access

Researchers must adhere to the ALSPAC Access Policy and data user responsibilities at all times. Researchers must also comply with the terms of the ALSPAC DAA, DURA, IDTA, and 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.

We are only able to share data that has been requested, justified, and approved via a submitted proposal or amendment. Please ensure all proposals/amendments submitted clearly outline all data types and topics, participant groups and timeframes you wish to access. Please see section 3.2.

6.4 Information Security

Data users must maintain sufficient information security controls to protect the data from unauthorised use or access. The minimum information security requirements are described in Schedule 2 of the DAA. In

particular, we are looking for a strong institutional commitment to Information Security and a commitment to store and analyse the data in a secure system (as evidenced by a comprehensive Information Security policy), and staff requirements to maintain confidentiality (as evidenced by an employment contract). These terms (and others) are agreed to be met by signing the DAA, Where we have any concerns around security, we reserve the right to complete a Data Privacy Impact Assessment (DPIA) which may increase the time required to approve a project.

Data users must not share, or provide access to, data with anyone outside of their approved user group (as defined in your project approval). If you need to add additional data users, please make an amendment to your project proposal through the online system.

If a data user should move institutions and wishes to take ALSPAC data to their new institution this must be agreed in advance with ALSPAC via an amendment to the proposal as we will need to ensure a new DAA is in place

It is possible that disclosure of a study participant may occur accidentally; researchers must not attempt to identify an individual. If a researcher does inadvertently identify a participant they inform the Executive as quickly as possible.

6.5 Confidentiality/security breaches

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

- Any unauthorised person (i.e. someone who has not signed a DAA or DURA 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.

Please refer to Schedule Two of the DAA for further information on our information security controls.

6.6 Destruction of data and return of derived variables

One month prior to the project end date provided on the proposal, your data buddy will be in contact to check if you require an extension. If not, you will be expected to confirm deletion of your supplied dataset and return a copy of any variables you have derived from the data along with appropriate documentation (see Schedule 1 of the DAA). Failure to confirm deletion of the original dataset within 4 weeks of the end of your project may risk your future access to the resource.

As the University of Bristol owns the ALSPAC resource (see schedule 1 of the DAA), any derived variables (such as data obtained as part of a new data collection exercise, new sample analyses, or newly derived variables coming from secondary analyses) created as part of any research project will be incorporated in to the main resource and made available to all researchers. We would expect you to return the relevant data to your data buddy (for the IDs to be replaced) within 4 weeks of the end of your project. Failure to produce derived variables at this point may risk your future access to the resource.

6.7 Publication

6.7.1 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 ensure that all papers are read to check that participant confidentiality is protected and to ensure that the paper will not bring the study into disrepute. The Executive does not judge the scientific content of a manuscript nor do they act as guarantors for any particular content. However, they do reserve 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.

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 or scientific 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 and papers from research consortia. 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.

6.7.1.1 Open access

6.7.1.1a 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 (for University of Bristol staff), or by publishing in a compliant journal. Please see the RCUK website for more information.

Please note that as Wellcome is a core funder of the study, any secondary analyses of ALSPAC data should also comply with the Wellcome open access policy. We therefore request that all articles are published in open access journals, if this is not possible, e.g. due to a lack of funds, please ensure articles are made freely available online through your Institution's website or other open access repository. You may wish to consult the Journal Checker tool to identify journals that are compliant with open access.

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

6.7.1.1b Journals

A number of journals request that datasets used in a publication are deposited in publicly available resources. The informed consent obtained from our participants and the data management policy will not permit this. We therefore request that the following statement be used by researchers in such cases: "The informed consent obtained from ALSPAC participants does not allow the data to be made freely available through any third party maintained public repository. However, data used for this submission can be made available on request to the ALSPAC Executive. The ALSPAC data management plan describes in detail the policy regarding data sharing, which is through a system of managed open access. Full instructions for applying for data access can be found here: http://www.bristol.ac.uk/alspac/researchers/access/. The ALSPAC study website contains details of all the data that are available (http://www.bristol.ac.uk/alspac/researchers/our-data/)".

We fully support the publication of all code and any other supporting information related to a publication.

6.7.1.1c Grant applications

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

6.7.2 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.

6.7.3 Policy briefings

We encourage researchers to produce policy briefings wherever possible. We ask that these or similar documents are shared with us prior to release. This ensures we are aware of ongoing work and means our communications team can work with you to ensure details are correct and provide advice and support where necessary. Please submit any documents to alspac-exec@bristol.ac.uk.

6.7.4 Reports and other publications

Any reports to funders that will be made publicly available and describe the data must follow the same approval proposal as full papers (See section 6.6.1) and must be approved *prior* to sending to the funder. The reporting guidelines provided in the papers checklist must be adhered to for reports. We request that we are provided with an electronic copy of any reports and other publications that use ALSPAC data as soon as possible.

6.7.5 Conference proceedings and other presentations

ALSPAC do not need to see submissions to conferences unless this is the primary mode of publication in your discipline. For example, computer science. In such cases we request that abstracts are submitted to the ALSPAC Executive for approval as per peer-reviewed papers (see section 6.6.1). However, we request that any talks or 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.

6.7.6 Other output

Any other output that will be made available in the public domain must be approved by the ALSPAC Executive prior to it going live. For example, online tools for exploring data, metadata on third party websites etc.

6.8 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. Please see Appendix 2 for additional costs incurred for linked data.

Example 1:

Standard request of data available in the data warehouse and an additional DAA

Standard access Fee	£1925
DAA for one additional Institution	£390
Total Cost	**£2315**

Example 2:

Standard request of data available in the data warehouse, two additional DAAs, genetics and methylation data and 10 text data fields for coding

Standard access Fee	£1925
DAA for two additional Institution	£780
Genetics and methylation data	£825
10 text fields: Split stage protocol	£825
Total Cost	**£4355**

Example 3:

Standard request of data available in the data warehouse, G2 data, DXA images and child and maternal medication data (not standard data for disclosure reasons)

Standard access Fee	£1925
G2 data	£1375
Digital images (e.g. DXA scans at age 9)	£1375
Maternal and child medications	£550
Total Cost	**£5225**

Example 4:

Amendment to original request (any number of variables), to add GWAS data from mother and child

Standard amendment Fee	£550
GWAS data	£825
Total Cost	**£1375**

Appendix Two: 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 Three: Access 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

v.8.0 Released May 2018

Updated throughout to reflect the introduction of the Data Access Agreement, including the addition of Appendices Four and Five.

Updated section 5.3.5 regarding exclusive access to data.

v.8.1 Released June 2018

Updated section 6.6.4 and added section 6.6.5 to reflect alternative outputs produced by researchers

v.8.2 Released July 2018

Updated section 5.3.5 regarding exclusive data access

v. 8.3 Released February 2019

Updated sections 1.4.3 and 1.4.4 to reflect changes to the Board and ISAB

Updated the weblink to ARIES summary data (section 2.3)

Updated section 5.2 regarding feedback of study findings

Updated section 5.3.5 regarding exclusive data access

Updated section 6.6.3 regarding reports to funders

Updated section 2.2 removing reference to a Data Transfer Agreement, since the DAA now applies to all data sharing

v 9.0 Released April 2019

Added section 2.1.1 on G2 data

Updated costings table to remove charges for geographical data, amended charges for imaging data and added charges for G2 data

Added section 6.4 on Information security

Added section 6.7.3 on Policy briefings

v. 9.1 Released July 2019

Updated section 1.3.6 regarding notice periods for submitting grants for new data collection.

Added costs to section 1.3.3 for grants that do not require access to existing data

Clarified in Section 1.1 that research must be in the public interest

v.9.2. Released October 2019

Added costs to Table 1 for individual DAAs and ID swaps for linkage and other data

Updated section 1.2 to note that a DPIA may affect approval times

Updated section 6.4 to clarify information around security requirements and note that a DPIA may be required.

v.10.0. Released January 2020

Updated section 1.3.3 to clarify Data management fee

Updated section 3.2 to provide options when the DAA cannot be signed

Providing more detailed information on student projects in section 1.2

v.11.0 Released April 2020

Overhaul of cost recovery charges – Section 1.3, Table 1 and Appendix 1 updated

v.12.0 released June 2020

Updated section 1.3.1 to include details of data in hand required for new project

Section 1.3.5 removed following restructure of cost recovery charges

Added section 2.2.2 on secure remote servers

Section 5.3.5 Clarifying the position on requesting exclusive access to new data

v.13.0 released February 2021

Updated costs and details around accessing linked health and education records

Updated broken data dictionary links

Section 1.3 Clarifying the responsibility of researchers to check costs after grant submission

Section 5.3.5 Clarifying exclusive access requests.

Section 6.7.1.1 Reflecting the change in Wellcome's open access policy

v.13.1 released February 2022

Increase in cost recovery charges resulting in changes to Section 1.3, Table 1 and Appendix 1

Addition of ALSPAC's pledge to education (section 1.3.6)

Section 6.7.1.1b: Added support of code publication

Minor edits throughout the document

v.14.1 released April 2023

Section 2.2 Added link to linkage user guide

Section 3: A number of changes/clarifications made throughout

Section 4 - Clarifying that charges might apply for identifying sample availability. Removed examples of specific sample analysis provider requirements. Simplified the description of the MTA/contract process.

Sections 5.1.7, 5.1.8 - Reinforced the requirement for approval by the HTA Designated Individual for projects collecting Human Tissue.

Section 5.2.4 Update of information relating to DBS checking

Section 6.7.1 Update regarding exec not judging scientific quality of manuscripts

Throughout – updated OCAP to APPAP

Removal of Appendix Two – replaced by links to the linkage data user guide

Removal of Appendices Three and Four – referring the reader to the DAA instead.

Other minor edits throughout the document.

V.14.2 released March 2024

Summary – Added definition of bona fide researcher

Section 1.2.1 – Further clarification regarding ‘data re-use’; Details of IDTA requirement for existing collaborators.

Section 3.3 – Details of new IDTA requirement.

Section 4 – Providing further detail regarding ethical approval for biological samples