ALSPAC Direct Users: User Guide

1. Rules governing access

Welcome to the ALSPAC resource. It is vital that you adhere to the rules you have signed up to by completing the online Data User Responsibilities Agreement (DURA). Please remember that you risk losing your access to the resource if you fail to adhere to these rules.

Before proceeding please read the ALSPAC access policy: (http://www.bristol.ac.uk/alspac/researchers/data-access/).

We include a list of the clauses you signed up to when agreeing to be a direct user at the end of this document as a reminder (see Appendix 1).

2. Accessing the data

Access requests can be submitted by following the process outlined here: http://www.bristol.ac.uk/alspac/ieu-info/

The vast majority of data (questionnaire and clinic-based data) are accessible via the R:\data directory. Once you have access you can map this into your computer using the pathname: \\ads.bris.ac.uk\Filestore\SSCM ALSPAC\Data. There are five top level folders within the R:\data directory:

Archive: This is where files that have been superseded with a newer version are placed (newer versions may have been created to correct errors, to add more cases or to add more variables)

Current: This is the area where the majority of data (i.e. the core resource) is stored. A separate file is available per data sweep (i.e. per questionnaire or per clinic sweep). Each data file is stored in both SPSS and STATA formats. The directory is broken down by clinic, questionnaire, external or other.

- The **clinic** folder is broken down by child and adult (where you will find mother - and father-based clinic data).

- The **questionnaire** folder is broken down by child based (completed by the mother about the child), child completed (completed by the child), mother, partner, puberty, schools and COVID.

- The **external** folder used to contain data provided by external sources – such data is now provided only by the linkage team due to the rules governing access demanded by the third parties. Please contact alspac-linkage@bristol.ac.uk.

- Finally, the **other** folder contains data files that do not fit elsewhere but contains one of the most important files – the **cohort profile**. Please see section 4 below for further details. Briefly, this supersedes the KZ file in the Sample Definition folder. The MZ file (sample definition) forms the baseline for all mother-based analysis. Other data held here includes that abstracted from obstetric records, all biosamples assays, social class data derived from occupation information and geodata (a number of deprivation scores based on postcodes)

Documentation: Please see section 3 below. Simply, this is where the data dictionary is stored.

Syntax: This is where you will find templates of syntax (for SPSS users) and do files (for STATA users) to assist you with putting datasets together. These *must* be used to ensure that participants who have withdrawn consent are removed from any analyses. These templates call up separate syntax/ do files ensuring that the withdrawal of consent is always up to date – please be aware of this if you recreate a data file for an ongoing project. These also ensure you have the correct baseline figure in your dataset.

Useful_data: This is where you will find additional data that has yet to be formally released (such as recently derived variables). If the data you are looking for are not in this directory or the current directory then it is likely embargoed or simply unavailable (i.e. still being cleaned or checked). Note that from 2019 we have been working hard to clear this directory and add data into the main resource (i.e. moved to the relevant file in the current directory).

Genetic data is not held in R:\data due to its size. To access complete GWAS or methylation data on **BlueCrystal**, please ensure you have a BlueCrystal account (you can

apply here: www.acrc.bris.ac.uk/login-area/apply.cgi) and have checked the box requesting access on the online DURA.

Note when you apply for a BlueCrystal account, that the project number is NOT your ALSPAC B number. If you are unsure of what your project number is, please ask your supervisor/line manager. You will be advised when this access has been granted. The data can then be accessed at /panfs/panasas01/shared/alspac/studies/latest/alspac.

IMPORTANT: *Any* data you access **must** have exec approval *prior* to access. If you require any additional data which are not covered in your original approval, you must submit an amendment through the Online Proposal System (https://proposals.epi.bristol.ac.uk/).

3. About the cohort

Before you look at any data please read the cohort profile papers available on our website: http://www.bristol.ac.uk/alspac/researchers/cohort-profile/ . Please note, at the time of writing the Father's cohort profile is in preparation. We refer to the original parents of the study as G0, the offspring as G1 and their children as G2 (known to the participants as COCO90s – children of the children of the 90s).

4. Getting help

Access to the resource is granted on the basis that you will consult with your immediate colleagues, line manager or supervisor in the first instance if you need help with data. If they cannot help then please use the alspac-data@bristol.ac.uk email for general ALSPAC data queries. Specific queries can be directed to the following emails:
Data linkage (e.g. education data): alspac-linkage@bristol.ac.uk
Genetics/GWAS: alspac-genetics@bristol.ac.uk
PR/Communications: alspac-media@bristol.ac.uk
ALSPAC Executive: alspac-exec@bristol.ac.uk
Data security: alspac-infosec@bristol.ac.uk
Any other data issues: alspac-data@bristol.ac.uk

5. Documentation

The fully-searchable documentation covering the released data is available in R:\data\documentation. (pathname: \\ads.bris.ac.uk\Filestore\SSCM ALSPAC\Data\documentation). Note it is also freely available for download should you choose to share it with a collaborator: http://www.bristol.ac.uk/alspac/researchers/data-access/data-dictionary/

Everything is accessible through the **main page**. The documentation is split into two main sections – "documentation for released data" and "questionnaires". The former provides the details of data that is currently available (including frequencies, though note these are not

necessarily up to date if changes have been made to the data file). The latter provides questionnaires as they have been sent to participants.

Navigation through the questionnaires and documentation can be aided greatly by the use of **indexes** within the advanced search options of acrobat (detailed information on how to do this is provided in the documentation via a link from the **main page**). By connecting the indexes you can search for keywords across the whole set of documentation in one go.

You may also find the variable search tool of use: http://variables.alspac.bris.ac.uk/

At this point we strongly recommend that you read **"The Guide to ALSPAC Data"** which will provide you with basic information on the format of the data and the cohort profile which you **must** be familiar with (see section 4).

6. Cohort profile

In order to prepare your data files and identify the baseline numbers as detailed in **"The Guide to ALSPAC Data"** there is a data file in R:\data\current\Other\Cohort Profile. This provides the indicator variables required. Note, this file contains ~20,000 cases as it contains all cases who were originally *eligible* to enrol but did not join for whatever reason. Other basic variables such as birthweight, gender and maternal age can be found in the KZ and MZ files in R:\data\current\Other\sample definition [Please note that at the time of writing the MZ file is being updated to create a mother's cohort profile dataset and a father's cohort profile dataset is being put together].

The syntax/do file templates that we have provided will ensure that the baseline number of cases in any data file you create will adhere to those displayed in the cohort profile and the description we provide on the papers checklist that must be completed when you submit a paper to the exec for approval. They also ensure that the appropriate withdrawal of consents are applied (it should be noted that the scripts will only work for numeric data only, string variables need to be handled separately). It is therefore vital that you use them.

7. Data Security

ALSPAC data must only be accessed directly through your office PC or through the faculty's remote server, please see the IT Services web page here: https://www.bristol.ac.uk/it-services/locations/fits/healthsciences/fmdysecureremotedesktop.html. **ALSPAC data must never be stored locally** (e.g. on your hard drive) on any desktop or laptop computers. Your access will be revoked immediately if we discover that data is not being accessed correctly.

All data transferred electronically **must** be encrypted using AES-256 encryption (this can be achieved using compression tools such as WinZip or 7-Zip – see below). If you wish to transfer data please use FLUFF (https://fluff.bris.ac.uk/fluff/), however you **must** ensure the

---

data are password protected before you FLUFF them. The password must be sent to a different email address than the FLUFF link. In addition, do **not** send the password via the same method as the FLUFF link!

To encrypt a file right click on it in 'My Computer', hover over 7-zip, then click on 'Add to archive…' In the bottom right-hand corner of the box that comes up you are asked to enter the password twice – do so and then click OK. In My Computer you will then find the zipped file.

When using removable media, only USB flash drives with hardware AES-256 encryption must be used. Removable media should be transported securely and not be left unattended at any stage. If you lose any data please report this immediately to ALSPAC's information security email (alspac-infosec@bristol.ac.uk).

8. Keeping you informed

When data access is provided you will be added you to our direct user mailing list and whenever a new dataset is released or changes are made to existing datasets you will be notified via email. We will also email with any other important pieces of information and for the annual review (usually in the spring/early summer).

9. Keeping ALSPAC informed

It is vital that you only use data that are covered by your approved project proposal – please talk to your PI/supervisor to get a copy of the proposal if necessary, if you were not involved in preparing the original proposal. ALSPAC has ISO27001 accreditation and we therefore take information security very seriously. Direct access is a privilege given to few and we keep a very clear paper trail for our annual audits to maintain accreditation.

You **must** inform us immediately of any changes to your original project. This can usually be done by submitting a project amendment form through the online system (https://proposals.epi.bristol.ac.uk/). For example, if a) you require any additional data that does not come under your original proposal (even if variables are just being used as confounders); b) a new member of staff joins the project; c) you wish to change the end date of your project if your contract gets extended. Please note for a) that we reserve the right to request that you submit a new project proposal if we feel the request is more than an amendment to the original.

If you have any queries about anything to do with the study, it is always better to check with us if you cannot find the information in the various places highlighted here or in the access policy. Not doing so, often creates significant work for the team.

10. Costs for data access

On the whole you receive access to the resource for free (unless a fee has been included in the grant you are working on, your PhD fellowship etc). It is really important to note that the data buddy team do not receive core funding. They are funded through our 'cost recovery' system whereby external researchers have to pay to access the data – hence your privileged position. In order to support the team we do ask that if you are preparing any kind of funding bid you include the appropriate costs. You should *always* submit a proposal to the Exec before you submit a funding bid to ensure the corrects costs are included.

11. Sharing data externally

There may be occasion when you have an external collaborator who you wish to share data with. Ideally, they would pay the full cost recovery fee, however, we appreciate this may not always be possible. Any data sharing can ONLY be done via the ALSPAC data buddy team and we therefore have to charge a small cost associated with administering the necessary paperwork and performing and ID swap. We would expect you to prepare the relevant dataset (containing only those variables necessary for the work being undertaken – this will be rigorously checked!). You will do this using the appropriate syntax to ensure all withdrawals of consent are taken into account and the appropriate baseline is provided. We will simply return the request back to you if you have not done this properly. Each dataset held externally requires an ID that is unique to that particular project, the data buddy team will perform the ID swap and send the data to the collaborator once all paperwork has been returned and costs have been paid. There will be a charge for each dataset that has to undergo an ID swap. Therefore, we recommend you thoroughly check that all data is included before sending through to us as you will be charged again if you 'just forget a variable or two'.

12. When you leave

Your access to the resource will end when your contract/study period ends. You need to contact us at least one month in advance to let us know if your contract has been extended and you will be continuing to work on ALSPAC data – if this is the case we will extend your direct access provided you can prove that access is still required. However, if you are leaving the University of Bristol you **cannot** take ALSPAC data with you when your contract/study ends unless you have prior permission from the ALSPAC Exec. This is regardless of you obtaining honorary status as that does not come with any binding contract.

If permission is granted for you to take data with you then you will be treated as an external collaborator and you will provide us with the data that you wish to take so that we can attach a new ID which will be unique to your project. Your new Institution will need to sign a Data Access Agreement before we can share the data. At least one month's notice (preferably more) is required to sort this out. Please talk to your PI/supervisor before you leave if you will require a dataset to take with you. Once you have left the University your

access will be revoked and you will have to pay full cost recovery charges to continue using the data.

Thankyou for your time
ALSPAC Executive

Appendix 1: Data User Responsibilities Statements

The information obtained by ALSPAC has been given by the study participants on the understanding that it will be treated confidentially and anonymously. The ALSPAC Access Policy (https://bit.ly/2YJK2yi), provides more detail and is referred to in the following agreement. Having direct access is a privileged position - your obligations as a direct user are detailed below. Please read the following statements and indicate that you hereby agree to adhere them.

Access to the resource

1. I have read the latest version (dated xx/06/2020) of the direct user's user guide (formally the direct user overview), here: http://www.bristol.ac.uk/alspac/ieu-info/.

2. I will only access and use those variables that are required to carry out the specific research for which I have approval from the ALSPAC Executive.

3. If I wish to test further hypotheses that lie outside the remit of the original research proposal/s I will submit an amendment or a new project proposal for approval.

4. If the scope of any project I work on is extended extensively, I understand this may be considered as a new project and I may have to resubmit as such.

5. I understand that I can only access ALSPAC data through the provided network path when in the office or through the secure remote desktop when offsite.

6. I will not keep any subsets of data on a personal device (e.g. laptop, home computer, memory stick, external hard drive).

7. I will exclude any participants known to have withdrawn their consent to use their data by using the latest versions of the scripts provided to put datasets together (in R:\Data\Syntax)

8. If I need to return data to the ALSAPC data team I will only do so using password protection (encrypted to AES-256 standards (this can be achieved using compression tools such as WinZip or 7-Zip) and FLUFF (https://fluff.bris.ac.uk/cgi-sso/fluff/fluffauth.pl). I understand I must send the password to/from a different email account to that through which the FLUFF link is shared.

9. I understand that I will not receive any direct support from ALSPAC but will use the email alspac-data@bristol.ac.uk for specific data queries if I can't find the answer elsewhere.

10. If my dataset requires linkage to data from linked third party records, I will comply with any additional instructions as provided by ALSPAC.

Sharing data and information security

11. I will not share *any* data directly with *any* researcher. ALSPAC data may only be accessed via the current data directory or through the data buddy team.

12. I understand that any (derived) data that I have generated can only be shared with another researcher via the ALSPAC data team (and may be subject to costs).

13. I will not try to identify any study participants. I will notify ALSPAC immediately if I inadvertently identify an individual and I will not attempt to contact that individual.

14. I confirm I have read, understand, and will comply with the University of Bristol Information Security Policy (http://www.bristol.ac.uk/infosec/policies/)

15. I confirm I have undertaken the University of Bristol's mandatory training on information security as part of MyReview (http://www.bristol.ac.uk/infosec/training/) and will provide proof on request.

16. If my project involves potentially identifying data such as free text or postcodes, I understand that the data will only be released using ALSPAC's split stage protocol. I confirm I have read and understand the procedures for this process stated in the Access Policy (see Appendix Three): https://bit.ly/2YJK2yi

The end of your project and leaving the University

17. I will notify ALSPAC *in advance* if I am leaving the University of Bristol whilst still working with ALSPAC data. In the event, I will prepare a dataset for ID conversion and will be assigned a data buddy.

18. I understand my access to the resource will be revoked if I leave the University of Bristol, regardless of having an honorary contract (this is not sufficient to retain access).

19. Prior to submission of any papers for publication, I will complete a papers checklist (https://bit.ly/2N3MFpj) and submit it, along with the manuscript to the ALSPAC executive for approval. I will do this at least two weeks prior to journal submission.

20. I will securely destroy any ALSPAC datasets when my approved project ends.

21. I understand that the University of Bristol owns the ALSPAC resource and any derivations from it (see Appendix Five of the Access Policy). I will return any derived variables to ALSPAC on request, together with the scripts/syntax and relevant documentation required to generate derivations. The documentation will be sufficient for someone else to understand and replicate my analyses.

Declaration

22. I understand that the above rules apply to any *current* and *future* research projects using ALSPAC data with which I may be involved over the coming 12 months (at which point I will be asked to complete a new form).

23. I understand that if I do not comply with any of the rules detailed above I will lose my access to the resource and that this may invoke disciplinary action.
24. By submitting this from I agree that I have fully read and understood all of the above statements.