



DATA NOTE

The Avon Longitudinal Study of Parents and Children (ALSPAC): an update on the enrolled sample of index children in 2019

[version 1; peer review: 2 approved]

Kate Northstone ¹, Melanie Lewcock¹, Alix Groom¹, Andy Boyd ¹, John Macleod¹,
Nicholas Timpson ^{1,2}, Nicholas Wells ¹

¹Department of Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, BS8 2BN, UK

²MRC Integrative Epidemiology Unit, Department of Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, BS8 2BN, UK

v1 First published: 14 Mar 2019, 4:51 (<https://doi.org/10.12688/wellcomeopenres.15132.1>)

Latest published: 14 Mar 2019, 4:51 (<https://doi.org/10.12688/wellcomeopenres.15132.1>)

Abstract

The Avon Longitudinal Study of Parents and Children (ALSPAC) is a prospective population-based study. Initial recruitment of pregnant women took place in 1990-1992 and the health and development of the index children from these pregnancies and their family members have been followed ever since. The eligible sampling frame was constructed retrospectively using linked recruitment and health service records. Additional offspring that were eligible to enrol in the study have been welcomed through major recruitment drives at the ages of 7 and 18 years; and through opportunistic contacts since the age of 7. This data note provides a status update on the recruitment of the index children since the age of 7 years with a focus on enrolment since the age of 18, which has not been previously described. A total of 913 additional G1 (the cohort of index children) participants have been enrolled in the study since the age of 7 years with 195 of these joining since the age of 18. This additional enrolment provides a baseline sample of 14,901 G1 participants who were alive at 1 year of age.

Keywords

ALSPAC, Children of the 90s, birth cohort study, cohort profile, enrolment



This article is included in the [Avon Longitudinal Study of Parents and Children \(ALSPAC\)](#) gateway.

Open Peer Review

Referee Status:

	Invited Referees	
	1	2
version 1 published 14 Mar 2019	 report	 report

1 **Sarah Crozier**, University of Southampton, UK

Hazel Inskip , University of Southampton, UK
University Hospital Southampton NHS Foundation Trust and University of Southampton, UK

2 **Aisling Murray**, Economic and Social Research Institute (ESRI), Ireland

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Kate Northstone (kate.northstone@bristol.ac.uk)

Author roles: **Northstone K:** Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Writing – Original Draft Preparation; **Lewcock M:** Data Curation, Writing – Review & Editing; **Groom A:** Data Curation, Writing – Review & Editing; **Boyd A:** Conceptualization, Data Curation, Formal Analysis, Methodology, Writing – Review & Editing; **Macleod J:** Methodology, Writing – Review & Editing; **Timpson N:** Methodology, Writing – Review & Editing; **Wells N:** Data Curation, Investigation, Project Administration, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by the Wellcome Trust through project grants to ALSPAC [102215 and 086118]. The UK Medical Research Council and Wellcome [102215] and the University of Bristol provide core support for ALSPAC. The research which established the ALSPAC eligible sample and the phase IV enrolment campaign was funded by the Wellcome Trust through the Project to Enhance ALSPAC through Record Linkage (PEARL) [086118]. This publication is the work of the authors and Kate Northstone will serve as guarantor for the contents of this paper. A comprehensive list of grants funding is available on the ALSPAC website (<http://www.bristol.ac.uk/alspac/external/documents/grant-acknowledgements.pdf>).

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2019 Northstone K *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Northstone K, Lewcock M, Groom A *et al.* **The Avon Longitudinal Study of Parents and Children (ALSPAC): an update on the enrolled sample of index children in 2019 [version 1; peer review: 2 approved]** Wellcome Open Research 2019, 4:51 (<https://doi.org/10.12688/wellcomeopenres.15132.1>)

First published: 14 Mar 2019, 4:51 (<https://doi.org/10.12688/wellcomeopenres.15132.1>)

Introduction

The Avon Longitudinal Study of Parents and Children (ALSPAC) is a geographically defined, longitudinal birth cohort that recruited pregnant women with an estimated date of delivery between April 1991 and December 1992^{1,2}. During the initial recruitment campaign (between 1990 and 1992) women with a total of 14,541 pregnancies were enrolled (some women had more than one eligible pregnancy during this period). The children resulting from these pregnancies and their parents have been followed up ever since, primarily via questionnaire, hands-on measurement at clinical assessment visits (called 'Focus' clinics) including the provision and assaying of biological samples and through linkage to routine data. ALSPAC is now a three-generational study, comprising 'G0': the cohort of original pregnant women, the biological father and other carers/partners; 'G1': the cohort of index children and 'G2': the cohort of offspring of the index children. The study website contains details of all the data that is available through a [fully searchable data dictionary](#) and [variable search tool](#).

The 2012 G1 cohort profile paper¹ describes in detail how the families have been followed up until the age of 18 (G1) and describes how recruitment was extended to include members of families who were eligible to have taken part (using the original eligibility criteria), did not initially enrol but who subsequently wanted to take part. Families were enrolled into the study in *phase I* between 1990 and 1992 during pregnancy and shortly after birth. Further enrolment occurred systematically in *phase II* in 1999 at the age 7 assessment clinic (child mean age: 7.5 years) and then opportunistically in *phase III* from 1999 until 2012 (child mean age: 17.8 years). Since 2012, ALSPAC has conducted an additional systematic recruitment drive and have continued to recruit additional families through opportunistic contacts and through recruitment of the G2 offspring of the index children (in *phase IV* recruitment). This data note provides an update on recruitment *phases II* and *III* and reports specifically on *phase IV* recruitment: that is, all those G1 participants who enrolled in the study from the age of 18 up to and including the clinic assessment held at age 24 (mean age 24.5 years).

Methods

Through the Project to Enhance ALSPAC through Record Linkage (PEARL), ALSPAC has retrospectively defined the study's eligible sampling frame through linking study recruitment records to NHS delivery records and child health records¹. This means that the study has a record of the identities (including given and family names; date of birth; NHS ID) of most index children who are eligible to participate. This has allowed ALSPAC to respond to enquiries from both participants and the wider public, for example around eligibility to participate and use of an individual's data in research projects. For *phase IV* recruitment, since the age of 18 years, there were three ways in which eligible G1 index children (now adults) could enrol into ALSPAC. Firstly, all participants who could be traced were systematically invited to enrol in a postal exercise conducted by PEARL. This was part of their campaign to provide fair

processing information to G1 as they reached legal adulthood and as ALSPAC started to systematically link to their health and routine administrative records. Secondly, during recruitment of the G2 participants, the study came into contact with individuals - either as the partner of an enrolled participant or through screening by midwives - who were eligible but had not previously enrolled. Thirdly, eligible participants proactively requesting to enrol are always welcomed and enrolled into the study.

Study numbers

A total of 20,248 G0 pregnancies (resulting in 20,505 potential G1 participants) are eligible to take part in the study. Of these pregnancies 116 had an unknown birth outcome. Of the eligible pregnancies, the G0 mothers of 14,676 G1 participants enrolled during the original recruitment campaign (i.e. enrolled in *phase I*). It should be noted that the G0 mothers of 69 pregnancies with unknown outcome have historically always been considered as *Phase I* enrollees, this is appropriate for the G0 cohort profile but not for G1 (since it is not clear how many foetuses were in each of these pregnancies and resulting offspring have not taken part in the study and therefore cannot be adequately quantified). G0 mothers of 456 G1 participants enrolled during the systematic campaign at age 7 (i.e. enrolled in *phase II*). A further 262 G1 participants enrolled through opportunistic contact between the ages of 8 and 18 (i.e. enrolled in *phase III*). This results in a total of 15,394 G1 participants who enrolled by the age of 18 (see [Figure 1](#)). Please note, this is a slight increase since the original cohort profile (1) was published.

Since the age of 18, a total of 195 additional G1 participants have enrolled (i.e. enrolled in *phase IV*). Of these, 177 were recruited through systematic postal invitations during 2014. An additional 14 were recruited through the ALSPAC-G2 recruitment campaign: This means that an eligible member of the cohort enrolled after either: i) becoming pregnant, ii) their partner becoming pregnant or iii) they were already a parent. Finally, 4 enrolments resulted from *ad hoc* approaches from eligible participants during this period (see [Figure 1](#)). As with previous phases, *phase IV* enrolment included two twin pregnancies where only one twin enrolled. It should be noted that at the time of writing, *phase IV* participants may not have contributed any data to the resource.

A total of 913 (456, 262 and 195 recruited during *Phases II*, *III* and *IV* respectively) G1 participants have enrolled who were not in the initial study sample (i.e. *phase I* enrolment). The total sample size for analyses is therefore 15,454 pregnancies, resulting in 15,589 foetuses. Of this total sample of 15,589 potential G1 participants, 14,901 were alive at 1 year of age (it should be noted that self-reported data are not available prior to the time of recruitment for those enrolled in *phase II* or later) and are considered the baseline sample for reporting purposes (14,888 excluding triplets and quadruplets - their data are not generally released for confidentiality purposes). [Table 1](#) summarises the active cases within the study and the potential data available at the time of writing.

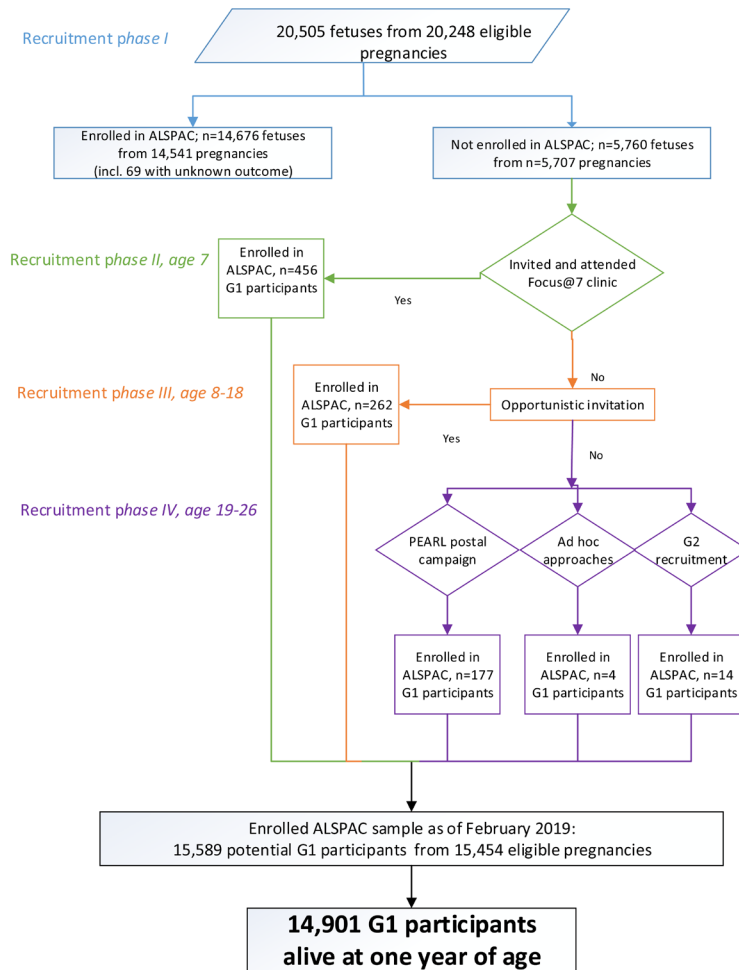


Figure 1. The ALSPAC enrolment campaign flow diagram, illustrating enrolment phases I to IV. (adapted from Boyd *et al.* 2013¹; Oxford University Press, [Creative Commons Attribution Non-Commercial License](https://creativecommons.org/licenses/by-nc/4.0/)).

Table 1. Activity of G1 (the cohort of index children) and illustrative data availability in ALSPAC.

G1 Activity as of February 2019	n
Currently active in study*	13,286
Known address	9,856
Can send Qs	8,964
Can send clinic invites	9,306
Linked to Primary health care data	11,810
Linked to Secondary health care data	12,700
At least one DXA** measure between the ages of 7 and 24	9,144
Genomics data	8,952
Methylation at multiple time points	1,003
G2 parent ³	548

DXA - Dual-energy X-ray absorptiometry, G2 - the cohort of offspring of the index children

*still alive, not withdrawn, not explicitly opted out

**Body composition assessed using Dual-energy X-ray absorptiometry

Table 2. Cohort Profile Data File (Version 2b at the time of writing).

Variable name	Variable Label
in_alsp	Enrolled in ALSPAC
in_core	Enrolled as part of original core sample, Phase I
in_phase2	Enrolled as part of phase II, during focus@7
in_phase3	Enrolled as part of phase III, after focus@7 up to age 18
in_phase4	Enrolled as part of phase IV, >=19 & <=24
tripquad	Participant is a triplet or quad
kz011b	Participant was alive at 1 year of age
kz021	Participant sex

Data availability

ALSPAC data access is through a system of managed open access. The steps below highlight how to apply for access to the data included in this data note and all other ALSPAC data:

1. Please read the ALSPAC access policy (http://www.bristol.ac.uk/media-library/sites/alspac/documents/researchers/data-access/ALSPAC_Access_Policy.pdf) which describes the process of accessing the data and samples in detail, and outlines the costs associated with doing so.
2. You may also find it useful to browse our fully searchable research proposals database (<https://proposals.epi.bristol.ac.uk/?q=proposalSummaries>), which lists all research projects that have been approved since April 2011.
3. Please submit your research proposal (<https://proposals.epi.bristol.ac.uk/>) for consideration by the ALSPAC Executive Committee. You will receive a response within 10 working days to advise you whether your proposal has been approved.

Cohort profile data file

The variables described in the cohort profile data file (ALSPAC reference: cp_2b) are provided as a matter of course with all data requests (see Table 2). The denominator used in this file is the 20,505 G1 eligible sample. The cohort profile data is provided in all data extracts, even where participants have formally withdrawn from the study or who are at high risk of disclosure, though it should be noted in these cases their data are suppressed.

If you have any questions about the data or how to access it, please email alspac-data@bristol.ac.uk.

Consent

Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics

Committees. Informed consent for the use of data collected via questionnaires and clinics was obtained from participants following the recommendations of the ALSPAC Ethics and Law Committee at the time. Children were invited to give assent where appropriate. Study participants have the right to withdraw their consent for elements of the study or from the study entirely at any time. Full details of the ALSPAC consent procedures are available on the study website (<http://www.bristol.ac.uk/alspac/researchers/research-ethics/>).

Grant information

This work was supported by the Wellcome Trust through project grants to ALSPAC [102215 and 086118]. The UK Medical Research Council and Wellcome [102215] and the University of Bristol provide core support for ALSPAC. The research which established the ALSPAC eligible sample and the phase IV enrolment campaign was funded by the Wellcome Trust through the Project to Enhance ALSPAC through Record Linkage (PEARL) [086118]. This publication is the work of the authors and Kate Northstone will serve as guarantor for the contents of this paper. A comprehensive list of grants funding is available on the ALSPAC website (<http://www.bristol.ac.uk/alspac/external/documents/grant-acknowledgements.pdf>).

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgements

We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses.

References

1. Boyd A, Golding J, Macleod J, *et al.*: **Cohort Profile: the 'children of the 90s'--the index offspring of the Avon Longitudinal Study of Parents and Children.** *Int J Epidemiol.* 2013; **42**(1): 111–127.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
2. Fraser A, Macdonald-Wallis C, Tilling K, *et al.*: **Cohort Profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort.** *Int J Epidemiol.* 2013; **42**(1): 97–110.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
3. Lawlor DA, Lewcock M, Jones LR, *et al.*: **The second generation of The Avon Longitudinal Study of Parents and Children (ALSPAC-G2): a cohort profile [version 1; referees: awaiting peer review].** *Wellcome Open.* 2019; **4**: 36.
[Publisher Full Text](#)

Open Peer Review

Current Referee Status:  

Version 1

Referee Report 10 April 2019

<https://doi.org/10.21956/wellcomeopenres.16511.r35067>



Aisling Murray

Economic and Social Research Institute (ESRI), Dublin, Ireland

Overview

This data note provides a brief summary of later additions to the ALSPAC G1 cohort (i.e. the original children). As I understand from the note, these were individuals who would have been eligible to participate but did not actually enrol in the study until a later date. Summary details are given on the number of individuals, the broad stage at which they joined and how they were recruited – particularly given the most recent joiners between the ages of 18 and 24 years.

Opinion

The main points of the note are generally clearly described and benefit from the inclusion of the diagram in Figure 1, which breaks down the numbers joining and at what stage. There are two exceptions: (a) in Table 1, it is not clear whether the n for 'known address' and 'linked to Primary health care data' etc. are subsets of the 13,286 described as 'currently active' in row 1 or the full 14,901 participants in the final cell of Figure 1; (b) in the abstract, the term "additional offspring" implies later children who were born to the same mother could enrol but in the main note text that does not appear to be the case.

I have no particular criticisms of the note, and it should be a useful resource for those using (or considering using) the ALSPAC data. There are some additional points that would be 'nice to know' however. Perhaps the authors could consider either including them either (briefly) in the note or adding a reference to direct the reader to the relevant information elsewhere? They are as follows:

- Do we know why the late joiners did not originally participate? Had their mother initially refused or were they not identified as eligible at the start? The note suggests the latter might have been the case for some.
- A brief reminder of the eligibility criteria at the start would be useful. I assume, for example, that only people initially born in the Avon area be able to join the study later and not those who immigrated there at a later date, but I am not certain from reading the note in isolation.
- How similar is the demographic profile of the late-joiners to the originally enrolled sample (e.g. gender, ethnicity, social class, health status)? And have the study weights been adjusted to account for the characteristics of the new joiners?
- The note makes brief mention of cohort 'G0', the mothers of the original children known as 'G1'. If a data user were considering using G0 data, it would be good to know if (a) any of the mothers of the late joiners also entered the study in their own right and (b) if they stay in the study even if their G1 children leave.

Is the rationale for creating the dataset(s) clearly described?

Yes

Are the protocols appropriate and is the work technically sound?

Yes

Are sufficient details of methods and materials provided to allow replication by others?

Yes

Are the datasets clearly presented in a useable and accessible format?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: I am the Senior Research Officer for Growing Up in Ireland, the Irish national longitudinal study of children and youth. I have worked on the study for 13 years and I am currently a member of the executive committee of the Society for Longitudinal and Lifecourse Studies. I have a PhD in Psychology and my personal research interests are in child development, particularly cognition.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Referee Report 03 April 2019

<https://doi.org/10.21956/wellcomeopenres.16511.r35066>



Sarah Crozier¹, **Hazel Inskip** ^{2,3}

¹ MRC Lifecourse Epidemiology Unit, University of Southampton, Southampton, UK

² MRC Lifecourse Epidemiology Unit, University of Southampton, Southampton, UK

³ NIHR Southampton Biomedical Research Centre, University Hospital Southampton NHS Foundation Trust and University of Southampton, Southampton, UK

This is a useful paper for anyone using the ALSPAC resource, setting out how the additional participants have been included at different stages. The innovative methods have been reported in detail and we anticipate that the availability of this Data Note will enhance understanding. We do however have some suggestions that might aid clarity:

We found the use of the term 'index' somewhat confusing. 'Index' to us rather implies the first children recruited. This paper describes the additional recruitment of children after the original (index) children had been recruited. Maybe stick to the term G1, where necessary, after defining it?

Also we felt that there is a confusion in the title in that it implies that the additional children were all enrolled in 2019. Could it be reworded to say "a 2019 update on the enrolled sample of index children"?

The abstract could do with a little clarification. The words 'additional offspring' is slightly confusing as it sounds as though it's additional children born to the original mothers. As eligibility hasn't been defined at this point, it's not easy to follow what this means. Maybe the authors could define eligibility in the abstract and then make it clear that those children who would have been eligible, but who were missed originally, have later been included in the study.

Also in the abstract it would be helpful if the word 'additional' could be inserted before 'recruitment' in the sentence: "This data note provides a status update on the recruitment of the index children since the age of 7 years with a focus on enrolment since the age of 18, which has not been previously described".

When you say that the original eligibility criteria were applied retrospectively, did you extract the expected date of delivery for all them to check this, or was it based on date of birth?

It would be helpful if reference to Figure 1 was moved to near the top of the Study numbers section as it is useful to look at it while reading the subsequent paragraph.

In the first paragraph on study numbers, there are these two sentences: "G0 mothers of 456 G1 participants enrolled during the systematic campaign at age 7 (i.e. enrolled in phase II). A further 262 G1 participants enrolled through opportunistic contact between the ages of 8 and 18 (i.e. enrolled in phase III)." At age 7, presumably the G1 participants were enrolled in addition to their mothers? In the enrolment between ages 8 and 18 were no G0 mothers enrolled or was it just the children? Also were any G0 participants enrolled in phase IV?

In the sentence "As with previous phases, phase IV enrolment included two twin pregnancies where only one twin enrolled." it is not clear whether both twins were ever included in any phase. The flowchart implies not, but it would be helpful if this could be clarified.

We found the numbers in the top three boxes of the flowchart somewhat confusing. The 14,676 + 5,760 fetuses do not sum to 20,505 in the top box, whereas the 14,541+5,707 pregnancies do sum to the 20,248 in the top box. The 69 are included in the pregnancies but not in the fetuses in the left-most box, and this isn't clear.

At the end of the second paragraph of the Study numbers, it might be helpful to add something like "but will be included in subsequent data collection waves".

The following sentence was initially confusing: "The cohort profile data is provided in all data extracts, even where participants have formally withdrawn from the study or who are at high risk of disclosure, though it should be noted in these cases their data are suppressed." It seems odd to say that the profile data are provided for everyone but then to say that the data are suppressed for particular participants. It might help to modify the end of the sentence to say "... in these cases other data are suppressed". (Please note that "suppressed" has been typed incorrectly).

Minor comments:

There seems to be an extraneous "(G1)" in the second line of the second paragraph of the introduction.

At the end of the first paragraph of Study numbers section, reference number 1 should superscripted.

The sentence in the last paragraph of the introduction that starts: "Since 2012, ALSPAC has conducted an additional systematic recruitment drive and have continued to recruit additional families through..." makes ALSPAC singular and plural. We suggest that 'have' should become 'has'.

Similarly there is inconsistent use of data as plural and data as singular in the last sentence of the first paragraph of the Cohort profile data file section.

In Table 1, it's pretty obvious that Qs stands for questionnaires but there is room to spell it out in full.

Is the rationale for creating the dataset(s) clearly described?

Yes

Are the protocols appropriate and is the work technically sound?

Yes

Are sufficient details of methods and materials provided to allow replication by others?

Yes

Are the datasets clearly presented in a useable and accessible format?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: We are both statisticians working in epidemiology

We have read this submission. We believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.