



DATA NOTE

REVISED **Gestation at completion of prenatal questionnaires in ALSPAC [version 2; peer review: 2 approved]**

Yasmin Iles-Caven , Kate Northstone , Jean Golding

Population Health Sciences, Bristol Medical School, University of Bristol, Oakfield House, Oakfield Grove, Bristol, BS8 2BN, UK

v2 First published: 22 May 2020, 5:100
<https://doi.org/10.12688/wellcomeopenres.15976.1>
 Latest published: 07 Oct 2020, 5:100
<https://doi.org/10.12688/wellcomeopenres.15976.2>

Abstract

Enrolling a cohort in pregnancy can be methodologically difficult in terms of structuring data collection. For example, some exposures of interest may be time-critical while other (often retrospective) data can be collected at any point during pregnancy. The Avon Longitudinal Study of Parents and Children (ALSPAC) is a prime example of a cohort where certain data were collected at specific time points and others at variable times depending on the gestation at contact.

ALSPAC aimed to enrol as many pregnant women as possible in a geographically defined area with an expected date of delivery between April 1991 and December 1992. The ideal was to enrol women as early in pregnancy as possible, and to collect information, when possible, at two fixed gestational periods (18 and 32 weeks). A variety of methods were used to enrol participants. Approximately 80% of eligible women resident in the study area were enrolled. Gestation at enrolment ranged from 4-41 (median = 14) weeks of pregnancy. Given this variation in gestation we describe the various decisions that were made in regard to the timing of questionnaires to ensure that appropriate data were obtained from the pregnant women. 45% of women provided data during the first trimester, this is less than ideal but reflects the fact that many women do not acknowledge their pregnancy until the first trimester is safely completed. Data collection from women at specific gestations (18 and 32 weeks) was much more successful (80-85%).

Unfortunately, it was difficult to obtain environmental data during the first trimester. Given the time critical nature of exposures during this trimester, researchers must take the gestational age at which environmental data was collected into account. This is particularly important for data collected using the questionnaire named 'Your Environment' (using data known as the A files).

Keywords

ALSPAC, pregnancy, questionnaire administration, mothers, fathers, birth cohort study

Open Peer Review**Reviewer Status**

Invited Reviewers

1 2

version 2(revision)
07 Oct 2020
report**version 1**

22 May 2020

report?
report

1. **Lucy Pembrey** , London School of Hygiene and Tropical Medicine, London, UK
2. **Stephanie M. Eick** , University of California, San Francisco, San Francisco, USA

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



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

Corresponding author: Yasmin Iles-Caven (cdylic@bristol.ac.uk)

Author roles: **Iles-Caven Y:** Funding Acquisition, Project Administration, Writing – Original Draft Preparation, Writing – Review & Editing; **Northstone K:** Data Curation, Writing – Review & Editing; **Golding J:** Conceptualization, Funding Acquisition, Methodology, Supervision, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by the Wellcome Trust through core support to ALSPAC [217065 and 062199]. The UK Medical Research Council, Wellcome Trust and the University of Bristol provide core support for ALSPAC. This publication is the work of the authors and they will serve as guarantors for the contents of this paper. A comprehensive list of grant 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: © 2020 Iles-Caven Y *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Iles-Caven Y, Northstone K and Golding J. **Gestation at completion of prenatal questionnaires in ALSPAC [version 2; peer review: 2 approved]** Wellcome Open Research 2020, 5:100 <https://doi.org/10.12688/wellcomeopenres.15976.2>

First published: 22 May 2020, 5:100 <https://doi.org/10.12688/wellcomeopenres.15976.1>

REVISED Amendments from Version 1

The following changes have been made in response to reviewers comments:

Abstract: A sentence added at the end of the Background section to clarify

Introduction: 2nd paragraph (a) in 1st sentence 'caused' changed to 'initiated'; (b) 2nd sentence: a new reference was added.

Introduction: last paragraph: wording of last sentence changed to clarify.

Methods: Recruitment section – added the last sentence to clarify.

Methods: Questionnaire administration – 5th paragraph, 3rd sentence – slight rewording for clarity.

Methods: Second pregnancies – removed last sentence and added one in response to reviewers' query.

Methods: Women <16 years – added the last 5 words for clarity.

Reference no. 4 is new – and inserted in response to the reviewer.

Table 1 and Table 2: the titles have been changed slightly.

Table 1: Column headings have removed the abbreviated titles of the partners' questionnaires.

Any further responses from the reviewers can be found at the end of the article

Introduction

There are considerable problems in child development that have been shown to be causally related to both genetic susceptibility and environmental impact. Three key examples that raised awareness of the importance of antenatal environmental exposures are: (i) the thalidomide tragedy, which resulted in unusual malformations, mainly involving loss of limbs¹; (ii) Minamata Disease resulting from pollution of seafood with excessive amounts of methylmercury and cerebral palsy in offspring of exposed pregnant women²; and (iii) the increased prevalence of deafness and blindness following antenatal exposure to rubella³. The associations with thalidomide, mercury and rubella were all discovered because the outcomes were so unusual.

One reason for failure to spot environmental features that could have initiated adverse consequences relates to the fact that often the woman is unaware that she is pregnant in the early weeks. Importantly, at this time the developing embryo is most vulnerable to adverse effects⁴ – whether this is of a maternal infection, a drug ingested, a binge of alcohol or a traumatic event. Nevertheless, such exposures operating later in pregnancy can also have different but important consequences, particularly on the development of the brain⁵.

There has been increasing awareness of the fact that exposures resulting in common adverse outcomes would be difficult to spot, unless one used a study that collected information prospectively during pregnancy and followed the offspring during childhood, adolescence and into adulthood. The aim of this paper is to describe the structure of the data collection used by the Avon Longitudinal Study of Parents and Children (ALSPAC),

a population based, pregnancy cohort that collected data during pregnancy. We then aim to describe the problems which arose in order to inform a) future studies and b) all researchers using the pregnancy-based data from ALSPAC.

Methods

Study overview

ALSPAC was designed to assess the ways in which the environment interacts with the genotype to influence health and development^{6–8}. Pregnant women resident in the study area in south-west England with an expected date of delivery between 1st April 1991 and 31st December 1992, were invited to take part. About 80% of the eligible population did so. The initial ALSPAC sample consisted of 14,541 pregnancies; of these initial pregnancies, there was a total of 14,676 fetuses, resulting in 14,062 live births and 13,988 children who were alive at one year of age. Information on the cohort parents and their offspring was collected using a variety of methodologies including self-completion questionnaires sent to study women, their partners, teachers, and from the age of five, the study child. The study also used direct examination under standardised conditions and linkage to educational data from the school system and other administrative records. Please note that the study website contains details of all the data that is available through a fully searchable data dictionary and variable search tool (<http://www.bristol.ac.uk/alspac/researchers/our-data/>).

Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee (ALEC; IRB00003312) and the Local Research Ethics Committees⁹. Detailed information on the ways in which confidentiality of the cohort is maintained may be found on the study website: <http://www.bristol.ac.uk/alspac/researchers/research-ethics/>. 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.

Recruitment

The study was designed to identify all eligible women based on (a) their area of residence (the former county of Avon); and (b) their expected date of delivery (April 1991 to December 1992 inclusive). The study area encompassed the city of Bristol and its surrounds, which includes the large coastal town of Weston-super-Mare, 27 miles from Bristol, and a mixture of rural and semi-urban areas. At the time a variety of prenatal health systems were in place within the area. Those pregnancies deemed to be at low risk were primarily managed by general practitioners in their local general practices. Consultant obstetricians were based at three hospitals (two in Bristol and one in Weston-super-Mare). High-risk women were transferred for antenatal care to the Bristol hospitals where the neonatal baby units were available if required. Thus, there were a number of different local health services in the area that may have had contact with eligible women at different gestations. It was therefore decided to use a variety of strategies to contact the women (both using the health service but also encouraging women to self-refer); this methodology is reflected in the wide range of gestations at which the women had first contact with the study.

Table 1. The instructions as to when the different questionnaires were to be sent out according to the gestation at enrolment and the name of the questionnaire.

Gestation at enrolment (weeks)	ENV	HAB	YP*	AY	YHL
<11	<11	18	32	14	-
11–14	11–14	18	32	23	-
15–18	22	18	32	26	-
19	24	19	32	28	-
20	24	20	32	28	-
21	24	21	32	28	-
22	28	22	32	36	-
23	28	23	32	36	-
24	-	-	29	33	24
25	-	-	30	34	25
26	-	-	31	35	26
27	-	-	32	36	27
28	-	-	32	36	28
29	-	-	32	36	29
30	-	-	33	37	30
31	-	-	31	X	34
32	-	-	32	X	35
33	-	-	33	X	36
34	-	-	34	X	37
35	-	-	35	X	37
36	-	-	36	X	38
37	-	-	37	X	39
38	-	-	38	X	40
39	-	-	39	X	41
40	-	-	40	X	41

X = administered four months post-delivery; ENV = Your Environment; AY= About Yourself; HAB= Having a Baby; YP= Your Pregnancy; YHL= Your Home and Lifestyle. *For women enrolled after 40 weeks, a short questionnaire 'Filling the Gaps' was sent 12 months post-delivery.

in particular to identify those features of the early environment that might be responsible for detrimental effects on the foetus.

The fourth questionnaire (D Files) 'About Yourself' was mainly concerned with the woman's past medical, social and environmental history, and consequently the time during pregnancy at which this was administered was relatively unimportant (majority received in the range 14–37 weeks). If necessary, therefore, this questionnaire was sent out postnatally.

For questionnaires administered during pregnancy the reminder and follow-up phase was fairly intensive. If a response had not been received within seven days, a reminder letter was sent. If the questionnaire had still not been received after a further 10 days, a second reminder letter was sent. Finally, if no response had been received after one month, a member of the study team either telephoned the woman or visited the home, and encouraged, or assisted, them in completing the questionnaire.

For women who did not enrol until six weeks after the 18-week contact, the 'Having a Baby' questionnaire was not likely to be valid. Much of the detail in that questionnaire was concerned with attitudes, activities and emotional well-being at that *particular point* in pregnancy. Nevertheless, for those women who had enrolled late, there was a certain amount of valid information concerning the environment and lifestyle that could be collected. For these women, therefore, the appropriate information, otherwise obtained from the questionnaires 'Your Environment' and 'Having a Baby', was combined into a single questionnaire 'Your Home & Lifestyle'.

For a variety of reasons (including very preterm delivery) some women did not receive the questionnaire 'Your Pregnancy' which included questions on ethnic origin, educational, social, and occupational levels. It also included questions on early sexual experiences. All these questions that were not specific to the third trimester of pregnancy were therefore included in a short questionnaire entitled 'Filling the Gaps', which was administered 12 months post-delivery.

Despite rigorous piloting, the early questionnaires were found to have occasional errors that were corrected on future versions. These were identified by date of printing. In addition, there were two groups of women who had certain changes to their questionnaires, as described below.

Second pregnancies

The women undergoing second pregnancies within the study recruitment period had a version of the questionnaire that omitted questions that would not have changed in a second pregnancy (e.g. ethnic origin). The editing process therefore copied over the answers to the first questionnaire sent to these women. The numbers answering the shortened A, C and D files were 100, 105 and 117 respectively.

Women <16 years

The questionnaire 'Your Pregnancy' included intimate questions on sexual experiences. It was decided by the Ethics Committee that these were inappropriate for a girl under the legal age of consent, and so the questionnaire was adapted to omit these questions for the 30 girls concerned.

Dataset

The released data files

The data from the six questionnaires sent to the woman as described above were combined into four research data files as follows:

A File is predominantly the questions in 'Your Environment', but also includes a few relevant questions from 'Your

Health and Lifestyle'. The version of the questionnaire used is described in variable *a001*, and the gestation at which it was completed (based on the date of LMP and date of completion in completed weeks) in *a902*.

B File comprises questions from 'Having a Baby' and 'Your Health and Lifestyle'. The version of the questionnaire used is described in variable *b001*, and the gestation at which it was completed (based on the date of LMP and date of completion in completed weeks) in *b924*.

C File consists of data from the questionnaires 'Your Pregnancy' and 'Filling the Gaps'. The version of the questionnaire used is described in variable *c001*, and the gestation at which it was completed (based on the date of LMP and date of completion in completed weeks) in *c991*.

D File provides data from the questionnaire 'About Yourself'. The version of the questionnaire used is described in variable

d001, and the gestation (or time after delivery) at which it was completed (based on the date of LMP and date of completion in completed weeks) in *d990* and *d991*.

The actual gestations at completion

In [Table 2](#), we show the numbers of questionnaires completed by week since the LMP. These indicate the following peak times at completion:

A files: 45% at <15 weeks; 32% at 22–25 weeks

B files: 79% at 18–21 weeks

C files: 85% at 31–34 weeks

D files: 89% during pregnancy; 11% post-delivery

Thus, the aim of obtaining data at around 18 and 32 weeks was fairly successful, with almost 80% of B files and 85% of C files completed during a four-week period. The aim to get the

Table 2. Frequency of completion at each completed week of gestational age for each questionnaire.

Gestation at receipt (in weeks)	A	B	C	D
<7	152	5	0	5
7	195	0	0	2
8	319	0	0	0
9	559	6	1	5
10	774	4	0	6
11	889	1	1	1
12	1071	1	0	5
13	1049	14	1	9
14	1005	30	1	1203
15	577	16	0	411
16	201	28	2	253
17	130	46	1	136
18	85	5351	1	81
19	51	2407	4	62
20	31	1589	0	49
21	33	991	2	31
22	1935	591	6	42
23	631	369	11	2162
24	1253	244	16	757
25	444	167	26	364
26	256	141	32	236
27	199	126	65	1382

Gestation at receipt (in weeks)	A	B	C	D
28	297	92	155	669
29	145	90	155	888
30	105	66	161	489
31	76	53	815	222
32	50	34	5854	160
33	29	13	2454	79
34	32	25	876	44
35	44	45	442	34
36	51	47	256	579
37	66	65	197	195
38	44	39	98	61
39	47	38	57	24
40	35	31	33	12
41	31	30	18	6
42	37	33	4	1
43	41	45	3	0
Post-delivery	288	279	562	1417
Total returned with known dates	13,257	13,274	12,310	13,888
Mode	22 weeks	18 weeks	32 weeks	23 weeks
Median	15 weeks	18 weeks	32 weeks	24 weeks

A File: 'Your Environment' plus some relevant questions from 'Your Health and Lifestyle'; **B File:** 'Having a Baby' and 'Your Health and Lifestyle'; **C File:** 'Your Pregnancy' and 'Filling the Gaps'; **D File:** 'About Yourself'.

A file data within the first trimester, however, was successful for less than half the pregnancies. As planned, the D files were completed at various gestations throughout pregnancy and the postnatal period.

How to refer to the gestations at which the data were obtained

Given the above distributions, we suggest that the following descriptions are used for each data set:

A files – 45% 1st trimester

B files – 18–21 weeks

C files – 31–34 weeks

D files – 89% during pregnancy

It should be noted that although labelled ‘gestation’, the variables *a902*, *b924*, *c991*, *d990*, *pb900* and *pa900* denote the number of weeks after the woman’s stated LMP. No correction has been made for the final estimate of gestation as obtained from the clinical records. In addition, these variables include women who completed questionnaires after the baby was delivered but different variables describe the age of the child at completion in these cases.

Conclusions

A note for users of the data

The Avon Longitudinal Study of Parents and Children (ALSPAC) was designed with the aim of determining features of the environment that may influence the health and development of the baby through childhood and into adulthood⁵. Knowing that the embryo is susceptible to early insults (whether from drugs, infections or other impacts), the aim was to start the study as early in pregnancy as possible, and to collect as much relevant information as possible throughout pregnancy. As can be seen above, we were not very successful in obtaining the environmental data in the first trimester. This needs to be taken into account when using data from the A files and where necessary data returned at appropriate gestations can be selected.

Ethical approval and 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/>).

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.

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. Stephens TD: **Proposed mechanisms of action in thalidomide embryopathy.** *Teratol.* 1988; **38**(3): 229–239. [PubMed Abstract](#) | [Publisher Full Text](#)
2. Boston University. Accessed 24.11.19. [Reference Source](#)
3. Sheridan MD: **Final Report of a Prospective Study of Children whose Mothers had Rubella in Early Pregnancy.** *Br Med J.* 1964; **2**(5408): 536–539. [PubMed Abstract](#) | [Free Full Text](#)
4. Jamkhande PG, Chintawar KD, Chandak PG: **Teratogenicity: a mechanism based short review on common teratogenic agents.** *Asian Pacific J Trop Dis.* 2014; **4**(6): 421–432. [Publisher Full Text](#)
5. Davis RL: **Neurodevelopment: Inflammation matters.** In: *Linking Environmental Exposure to Neurodevelopmental Disorders Volume 2.* Editors: M Aschner and L Costa. Elsevier, 2018. [Publisher Full Text](#)
6. Golding J; ALSPAC Study Team: **The Avon Longitudinal Study of Parents and Children (ALSPAC)—study design and collaborative opportunities.** *Eur J Endocrinol.* 2004; **151**(Suppl 3): U119–U123. [PubMed Abstract](#) | [Publisher Full Text](#)
7. 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](#)
8. 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](#)
9. Birmingham K, Golding J: **Pioneering ethics in a longitudinal study: The early development of the ALSPAC Ethics and Law Committee.** Bristol: Public Policy Press, 2018. [Publisher Full Text](#)

Open Peer Review

Current Peer Review Status:  

Version 2

Reviewer Report 16 October 2020

<https://doi.org/10.21956/wellcomeopenres.17982.r40751>

© 2020 Eick S. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Stephanie M. Eick 

Program on Reproductive Health and the Environment, Department of Obstetrics, Gynecology and Reproductive Sciences, University of California, San Francisco, San Francisco, CA, USA

The authors have fully addressed my concerns.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: epidemiology, environmental health, reproductive health

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

Version 1

Reviewer Report 09 September 2020

<https://doi.org/10.21956/wellcomeopenres.17521.r40219>

© 2020 Eick S. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Stephanie M. Eick 

Program on Reproductive Health and the Environment, Department of Obstetrics, Gynecology and Reproductive Sciences, University of California, San Francisco, San Francisco, CA, USA

In this manuscript, the authors provide a detailed overview of the ALSPAC study. There is detailed information about study recruitment and the timing of when certain exposures were measured.

This manuscript is well-written and would be relevant for researchers who are interested in using this data. I have several specific comments outlined below that I think would improve the clarity of this manuscript.

- Abstract: The sentence “The Avon Longitudinal Study of Parents and Children (ALSPAC) is a prime example of such a cohort” – what type of cohort, a pregnancy cohort? A well-constructed cohort? I think this should be clarified.
- First sentence in the second paragraph of the introduction: the word “caused” seems too strong here, I recommend softening it. Very few things have a single definitive “cause” in epidemiology and nearly everything is multi factorial.
- Second paragraph in the introduction needs citations. There are a lot of claims made with no reference to support them.
- Study population: What gestational ages were women targeted for recruitment? The authors state that women were recruited throughout pregnancy but was one window preferred and what was the earliest time point that women were eligible for the study?
- Recruitment: It would be helpful to specify a little more what was a “low risk” pregnancy and define more specifically the inclusion and exclusion criteria. Was there an age range for mothers, language considerations, anything beyond specific geographic region, etc.?
- Table 1: This table isn’t very clear to me as currently presented. The columns “ENV” to “YHL”, are these the different gestational age ranges that the surveys were sent out at? The heading should be more detailed here or maybe this information would be better represented as a figure.
- Table 2: To me, this table is really portraying much information as currently presented – I think it would be helpful if this information was presented as N (%) instead of just N. For example, what is the total? The raw numbers aren’t very intuitive to me without some form of reference. I think it would also be helpful if this information was grouped by gestational age as opposed to separating out each individual week gestation. My instinct is that if they were grouped somehow (maybe by 4 weeks?) there likely wouldn’t be that much difference between the individual weeks – although this is hard to tell because as presented the raw N isn’t very informative. It would also be helpful to clarify if the gestational ages in this table indicate that the questionnaire was sent to the ppt or the gestational age that the ppt actually filled it out – I imagine that could lead to some misclassification if women didn’t fill out the survey right away. I’m not sure reporting the mode is necessary, my thought would be to report the mean for consistency with other studies. Lastly, what % of ppts were lost to follow up at each time point? There is a row for “total returned” but I think it would be more informative to know what % of the total possible sample that was.
- What % of women had a second pregnancy during the study time frame and were enrolled in the study with that second pregnancy?
- For readers who are interested in using this data and are reading this manuscript as a general description of the study, I think it would be helpful to provide some general demographics as a descriptor of the overall cohort.

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?

No

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

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: epidemiology, environmental health, reproductive health

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 29 Sep 2020

Jean Golding, University of Bristol, Oakfield House, Oakfield Grove, Bristol, UK

Many thanks for your useful comments. I have answered them below, and changed the paper where indicated.

1. The Abstract: the words 'of such a cohort' have been changed to: 'of a cohort where certain data were collected at specific time points and others at variable times depending on the gestation at enrolment'.
2. The word 'caused' has been changed to 'initiated'.
3. The statement in this paragraph is now accompanied by a reference to a comprehensive review (new ref 4).
4. The following sentence has been added to clarify. 'There were no exclusion criteria, and women were encouraged to enrol as early in pregnancy as possible.'
5. As stated in the sentence added above, there were no exclusions, and all pregnant women were targeted. 'Low risk' was mentioned as a description of where such pregnant women were likely to receive their obstetric care – but did not change the way in which the pregnant woman was approached.
6. Apologies for omitting the meaning of the questionnaires PQ and YYE. These were questions sent to the partner of the pregnant woman, and have now been omitted, as they

will be the topic of a separate Data Note.

7. (a) Table 2 presents the numbers of women responding to the various questionnaires, with the actual gestation at completion. This table is presenting the details so that a researcher can determine the numbers that would be available for any particular analysis based on the gestational window of interest; for this reason we feel it is important to display the data by actual week of gestation. The interested reader can calculate percentages from the totals at the foot of each column.

(b) We have quoted the median and mode gestational ages. We have not quoted the mean as the distributions are so skewed for two of the variables that the mean would be inappropriate.

(c) The referee asks for % lost to follow-up. This is not as obvious a question as it appears. As the referee will be aware, a survey of pregnancy is a fluid event, with people enrolling, gestations changing as more information becomes available, pregnancies miscarrying or being terminated, and delivering preterm. Consequently the numbers eligible will vary by week of gestation as well as by week of enrolment. Calculation of this statistic is too detailed for this simple data note.

8. The number of second pregnancies completing the AY questionnaire was 117. This information has been added to the appropriate section, as has the number of women who had shorter questionnaires because they were < 16 years of age.

9. The reviewer requests overall demographics of the study. We have not included this as it is adequately covered in the references quoted [refs 6-8].

Competing Interests: None

Reviewer Report 05 June 2020

<https://doi.org/10.21956/wellcomeopenres.17521.r38815>

© 2020 Pembrey L. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Lucy Pembrey 

Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, UK

This Data Note describes how data were collected with questionnaires at several points during pregnancy from mothers enrolled in the Avon Longitudinal Study of Parents and Children (ALSPAC) from April 1991 to December 1992. Details of the recruitment procedures are provided with a detailed description of the timing of administration of the six different questionnaires according to gestational age at the time of enrolment. The four available datasets are described,

and the number of questionnaires collected at each gestational age for each dataset are shown.

Given the need to enrol women at different stages of pregnancy and the importance of taking into account the gestational age at the time questions were asked for certain environmental exposures, this Data Note provides a useful summary of the complexity of the ALSPAC pregnancy questionnaires. ALSPAC data are now widely used by researchers in the UK and internationally, accessed via application to the ALSPAC Executive Committee. Researchers using variables from the pregnancy questionnaires can be referred to this publication and it can be referenced in papers of their results.

I have a few minor comments as follows:

1. The last sentence of the Introduction could be re-phrased for clarity, from 'Then to describe the problems arising in order to inform...' to 'We then aim to describe the problems which arose in order to inform...'
2. In Table 1, please indicate in the footnote what PQ in column 3 and YYE in column 5 refer to. I can't see where these are mentioned in the text. If these are additional questionnaires, further explanation in the text on page 4 may be needed.
3. The first sentence on page 5 could be re-phrased for clarity, from 'Nevertheless, for those women who had enrolled late, there was a certain amount of information concerning the environment and lifestyle that could and should validly be collected.' to 'Nevertheless, for those women who had enrolled late, there was a certain amount of valid information concerning the environment and lifestyle that could be collected.'

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: epidemiology, birth cohort studies

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

Author Response 29 Sep 2020

Jean Golding, University of Bristol, Oakfield House, Oakfield Grove, Bristol, UK

Thank you so much for your comments. We have changed the text in the way you have suggested in points 1 and 3. In regard to point 2 we have omitted the labels in the Table. They referred to questionnaires to the partners which are not the topic of this questionnaire.

Competing Interests: None