RESEARCH ARTICLE

Characterising the index mothers in the Avon Longitudinal Study of Parents and Children (ALSPAC) who are also UKBiobank participants [version 1; peer review: awaiting peer review]

Mark Mummé, Andy Boyd, Alison Teyhan, Richard Thomas, Nicholas Timpson

ALSPAC, Bristol Medical School, University of Bristol, Bristol, BS8 1QU, UK

Abstract

Background: This is an examination of the characteristics of the women who enrolled into both the ALSPAC cohort during the initial recruitment phase in 1991-1992 and subsequently into the UKBiobank (UKBB) cohort from 2007 to 2010. There is substantial overlap in recruitment across the two cohorts with the existing Avon Longitudinal Study of Parents and Children (ALSPAC) sampling frame being included in that deployed latterly for UKBiobank.

Methods: A process was developed using anonymisation and encryption to flag individuals participating in both studies without revealing the identity of those participants. A description of participant characteristics was undertaken to compare the demographic, socio-economic, locational, consent and participatory characteristics of the women in ALSPAC who were also eligible to participate in UKBiobank by their UKBB participation status.

Results: The ALSPAC mothers enrolled in UKBB were more engaged in and contributed more data to ALSPAC than those who were eligible to join UKBB but did not enrol. The women in both studies were more likely to be married, have higher levels of education attainment and more financial stability than their counterparts who were only enrolled in ALSPAC.

Conclusion: It is possible these characteristics are reflected more generally within UKBB compared to the wider population.

Keywords
ALSPAC, UKBiobank, record linkage, cohort study, participation, duplication
Corresponding author: Mark Mummé (mark.mumme@bristol.ac.uk)

Author roles: Mummé M: Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Visualization, Writing – Original Draft Preparation; Boyd A: Conceptualization, Methodology, Writing – Review & Editing; Teyhan A: Methodology, Writing – Review & Editing; Thomas R: Data Curation, Writing – Review & Editing; Timpson N: Conceptualization, Funding Acquisition, Supervision, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by the Wellcome Trust through the Project to Enhance ALSPAC through Record Linkage (PEARL) [Ref: 086118]. The UK Medical Research Council and Wellcome (Grant ref: 217065/Z/19/Z) and the University of Bristol provide core support for ALSPAC. This publication is the work of the authors and Mark Mummé 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). This research was funded in whole, or in part, by the Wellcome Trust [Grant 217065/Z/19/Z]. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2022 Mummé M 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: Mummé M, Boyd A, Teyhan A et al. Characterising the index mothers in the Avon Longitudinal Study of Parents and Children (ALSPAC) who are also UKBiobank participants [version 1; peer review: awaiting peer review] Wellcome Open Research 2022, 7:190 https://doi.org/10.12688/wellcomeopenres.17982.1

First published: 13 Jul 2022, 7:190 https://doi.org/10.12688/wellcomeopenres.17982.1
Cohort studies can be combined in studies of rare events or weak associations, where large numbers are needed to increase statistical power. This is often done employing methods able to account for between study heterogeneity (for example meta-analyses). ALSPAC and UKBiobank are both examples of cohorts used routinely in this way within the UK. Combining cohorts assumes that the study samples are independent. However, given the sampling frames of these two studies, it was considered highly likely a subset of the ALSPAC mothers and partners are also enrolled in UKBiobank. The level of overlap between the two samples – and the consequences of this – needs to be understood in order to understand the implications for analyses, and the potential benefits of having participants who are in both studies.

It has only recently become possible to identify the individuals in each cohort using an anonymised, encrypted code based on their unique NHS number. This method allows us to identify women enrolled in both studies. The aims of the work described in this data note were to: (1) determine which women in the ALSPAC sample are also enrolled in UKBB and create a flag so they can be easily identified; (2) examine whether the women in both ALSPAC and UKBB differ from the women in ALSPAC only, in terms of demographic, socioeconomic, location, consent, and participation characteristics. Note that at the time this work was undertaken (July 2019) ALSPAC had NHS identifiers for the women in ALSPAC (known as the ‘ALSPAC mothers’), but not for the men (the partners of the women, also referred to as the ‘ALSPAC fathers’). Therefore, only women are included in this current study. It is anticipated that enrolled partners/fathers will be linked to the NHS in the future, and the encrypted data exchange protocol may be repeated to establish the number of them also enrolled in UKBiobank.

Methods
Ethics
Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees (NHS Health Research Authority [HRA] Research Ethics Committee [REC] reference 10/H1010/70). A comprehensive list of research ethics committee approval references is available to download at: http://www.bristol.ac.uk/alspac/researchers/research-ethics/. Informed consent for the use of data collected via questionnaires and clinics and record linkage was based on the recommendations of the ALSPAC Ethics and Law Committee at the time. 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/).

The ALSPAC sample
ALSPAC is a prospective population-based study. ALSPAC initially recruited 14,541 pregnant women who were resident

---

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.

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.

3. 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 Res 2019; 4:51. (https://doi.org/10.12688/wellcomeopenres.15132.1)

4. Fraser A, et al. Cohort Profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. Int J Epidemiol. 2013 Feb;42(1): 97–110. doi: 10.1093/ije/dys066. Epub 2012 Apr 16. PMID: 22507742; PMCID: PMC3600619.

5. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. (2015) UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. PLoS Med 12(3): e1001779. https://doi.org/10.1371/journal.pmed.1001779

6. Anna Fry, Thomas J Littlejohns, Cathie Sudlow, Nicola Doherty, Ligia Adamska, Tim Sprosen, Rory Collins, Naomi E Allen, Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population, American Journal of Epidemiology, Volume 186, Issue 9, 1 November 2017, Pages 1026–1034, https://doi.org/10.1093/aje/kwx246

7. S. Haworth et al., ‘Common genetic variants and health outcomes appear geographically structured in the UK Biobank sample: Old concerns returning and their implications.’ 2018
in Avon (former county covering Bristol and the surrounding areas in the Southwest UK), with expected dates of delivery between 1st April 1991 and 31st December 1992. 13,640 of these women had a study child alive at one year of age. These children and their mothers are referred to as the ‘core sample’. This core sample was boosted a few years later by the recruitment of children who met the original eligibility criteria but whose mothers had not joined the study during pregnancy. 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).

### Identifying women in both ALSPAC and UKBiobank

Both ALSPAC and UKBiobank used the ‘OpenPseudonymiser’ package (https://www.openpseudonymiser.org/) to create respective lists of encryptedhashed (AES256bit) identifiers (initially 14,875 encrypted identifiers were generated for all the non-dissenting women in ALSPAC, 502,383 for participants in UKBB, both male and female). The lists were based on the following identifiers:

- Month and year of birth
- Sex
- NHS number

The list of UKBiobank encrypted identifiers was securely transferred from UKBB to ALSPAC and merged to identify matches and generate a flag against participants enrolled in both studies. A list of matching participants was then securely transferred from ALSPAC back to UKBB. This list of 14,875 included women recruited into ALSPAC at any time. However, the descriptive analyses presented in this Data Note are restricted to women in the core sample as only they have baseline characteristic data.

A list of the basic identifiers of all the women enrolled in ALSPAC was securely compared against the identifiers of those enrolled in UKBiobank, as described above, and a marker was created to indicate women found in both groups.

### Defining the denominator of women in ALSPAC who were also eligible for UKBiobank

In order to determine the denominator of our study, it was necessary to identify which women in ALSPAC had the potential to also be a UKBiobank participant (i.e. met the eligibility criteria) (Figure 1). We first restricted our study sample to those in the core ALSPAC sample, with a study child alive at one year. These restrictions were because: (1) women who had a miscarriage, or whose child died before one year of age, did not continue participating in ALSPAC; and (2) women in the non-core sample did not complete the early questionnaires that...
contain baseline characteristics. Women who had declined consent for the use of their health records were then dropped as we have no way of identifying whether those women are in UKBB or not. Finally, as eligibility for participation in UKBB was constrained to those aged 40 to 69 years old, we only included women in ALSPAC who were in that age range during UKBB’s recruitment phase (UKBB recruitment began in April 2007 and ended in August 2010 when the target of 500,000 was reached). It was deduced that those enrolled in UKBB aged 69 in April 2007 at the start of recruitment for UKBB would have been born in 1938 and those aged 40 in August 2010 at the close of recruitment would have been born in 1970. Restricting our study sample to those born between 1938 and 1970 resulted in a final study sample of 12,127 women who are in the core ALSPAC sample, have not opted out of health record linkage, and who were eligible to enrol in UKBB. Our study sample was not restricted on geographic criteria due to uncertainties in the full residential history of all ALSPAC participants and the difficulties in obtaining the precise boundaries used during recruitment into UKBB.

**ALSPAC measures**

A number of measures, detailed below, were used to compare the women in the study sample who were in both studies compared to those in only ALSPAC, based on demographic, partner, socio-economic, locational, consent and participatory data.

Participation: a participation score was generated by counting how many of the 56 possible questionnaires each woman had returned. This was put in to ten categories (category 10 – low participation, <10% of questionnaires returned; category one – high participation, >90% questionnaires returned). A binary variable (yes, no) was derived which specifies whether the mother has a partner who is enrolled in the ALSPAC cohort.

Age: The women’s age was known directly from the study child’s birth records. The age of the women at the time of delivery was categorised into five-year brackets. A binary variable was created to indicate whether, within each couple, the partner was older than the woman based on their respective ages at the birth of the index child (yes, no).

Ethnicity: self-reported by the women (White; non-White – no further disaggregation possible due to small numbers). The child’s ethnicity was derived from the ethnicity of the mother and the father, when this was known. It was designated as ‘White’ if both parents were ‘White’ and ‘non-White’ if at least one parent was known to be ‘non-White’.

Partner status: woman lived with partner during pregnancy with study child (yes, no).

Pregnancy/child related: length of gestation (weeks) and birthweight (kg) of study child; parity (0; 1+); multiple birth (yes, no); breastfeeding of study child (yes, no). Whether the mum’s previous child(ren) had been breastfed (this was only relevant if the mum had had a previous child and so first it was determined whether the mum was primiparous or not). The feeding method used for previous child(ren) was then reduced to a binary (yes, no) breastfeeding variable.

Smoking: two binary (yes, no) variables were derived: (i) ever smoked and (ii) smoking during pregnancy. These were based on self-reported smoking behaviours reported at three time points: during the pregnancy (where the women were asked whether they had ever been a smoker and their current smoking); when the study child was aged two months (asked whether mother smoked in labour); and when the study child reached the age of eight years (asked about whether they have ever smoked and current smoking).

Alcohol: detailed responses about the number of glasses\(^\text{**110**}\) of alcohol consumed, the type of alcoholic beverage and the pattern of drinking through the days of the week were sought through questionnaires during the pregnancy and again in the period two months to 21 months after. The categories ‘one or two glasses per day’ and ‘three to nine glasses per day’ were combined into a single category due to the presence of small numbers.

Alcoholism: a binary variable, ever suffered from alcoholism (yes, no), was derived based on women’s self-reports during pregnancy. The ‘yes’ group includes those who answered ‘recently’ and ‘in the past’. When the study child was aged eight and 21 months, the women reported whether their partner had had alcoholism since the baby was born and a binary variable was derived (no, yes). The partner was asked directly during the pregnancy if they had suffered alcoholism and a binary variable was derived (yes, no).

The mother was also asked during her pregnancy whether either of their natural parents, or alternatively a parental figure, had an alcohol problem. The partner was asked the same question of their own parents at the same time. Any response indicating an index child’s grandparent or grandparent figure had an alcohol problem was consolidated into a ‘yes’. A ‘no’ was only recorded if it was definitely recorded that no grandparent or grandparent figure of the index child had an alcohol problem, otherwise the value was left blank (missing). These responses were consolidated into a single binary (yes, no) variable.

Substance misuse: The women were asked about substance misuse during their pregnancy, again when their child was aged about eight years and later when their child was aged about 11 years. Their partners were asked the same questions at the same time points. Any answers of ‘yes’ were aggregated together over time. Any answer of ‘no’ was included if not contradicted. The remaining data was recorded as missing.

Health and Wellbeing: The women and their partners were asked to self-assess their state of general health at several time points before and after the birth of their child. While

\(^{**110**}\) The questionnaires stipulated ‘By glass we mean a pub measure of spirits, half a pint of lager or cider, a wine glass of wine, etc.’
still pregnant the women were asked to assess their own and their partner’s health. The women and partners were asked to re-assess their own health eight weeks after the birth of their child and again at 21 months after the birth of their child.

Financial difficulties: The women were asked whether they had experienced financial difficulty when their child was aged about eight months, during the second year at about one year nine months and finally at about two years nine months. A binary flag was generated with any period of financial difficulty indicating as ‘yes’, however a ‘no’ required all periods to be recorded as being without financial difficulty. Any participants without a complete response record were excluded as ‘missing’.

Place of Residence: The women were asked how many times they had moved in the five years prior to their pregnancy. They were also asked for the status of their residential circumstances at the time of their pregnancy. The historic address records available to ALSPAC of those registered in the study were utilised to examine the residential stability of the participants. The residential address postcodes were linked to the Index of Multiple Deprivation (IMD) by decile using the year 2000 indices for the period during recruitment into ALSPAC and the 2007 indices for the period of recruitment into UKBB.

Socio-Economic Class and Education: during the pregnancy, the women and their partners were asked about their highest educational qualification. The women and their partners were also asked about their employment at the later stages of the pregnancy. The responses were provided as free text and were manually coded against the OPCS occupation codes. From the responses, the maternal and paternal social classes were derived as at the time just prior to the birth of the study child using the Standard Occupation Classification 1990.

Results
Women in both ALSPAC and UKBiobank
When the 14,875 encrypted identifiers generated for the women in ALSPAC were compared to the 502,383 for participants in UKBB, there were initially 938 matches in total, however only 866 were in the study sample as defined in Figure 1. Therefore, of the 12,127 women in ALSPAC who were also eligible to enrol in UKBiobank, 866 (7.1%) were confirmed to be enrolled in both studies. This is higher than the general rate of enrolment in UKBB (5.5%). It was observed that 88.9% of the original enrolled women were also eligible to be enrolled into UKBB, demonstrating as expected the over-lap in the sampling frame of the two projects.

Comparison by UKBB enrolment status
The 866 women in both UKBB and ALSPAC (also referred to as ‘in’ UKBB or Group B – B for Both ALSPAC and UK Biobank) were compared to the 11,261 only in ALSPAC (also referred to as ‘not in’ UKBB or ‘out’ of UKBB or Group A – A for ALSPAC only) across a range of measures. Note that the denominator may vary by variable due to missing data.

The women in both cohorts had higher levels of engagement in ALSPAC than women only in ALSPAC. Table 1 shows that women who were enrolled in ALSPAC and UKBB were more likely to have a partner also enrolled in ALSPAC, and nearly half (49.5%) of women enrolled in both cohorts completed 90% or more of the surveys compared to less than a fifth (17.5%) of mothers only in ALSPAC.

Participation in ALSPAC over time by UKBB enrolment status is shown in Figure 2. This shows the percentage of women in each group who returned questionnaires, from pregnancy up until the study children were aged twenty years. Both groups show a slow but steady decline in participation over time, however the rates of participation declined slower for the women enrolled in both cohorts.

Women enrolled in both cohorts (Group B) were more likely to report co-habitation with their partners during the pregnancy than those only enrolled in ALSPAC (Group A) (Table 2). There was negligible difference between whether women enrolled in either Group had a partner older than themselves. There is a lower proportion of women in Group B in the younger age brackets than those women in Group A and a correspondingly higher proportion in the older age brackets. There was little difference between the two groups in terms of ethnicity.

Pregnancy and birth characteristics (birth weight, gestation, parity) were similar for both groups of women (Table 3). However, women enrolled in both cohorts were more likely to have breastfed than those only enrolled in ALSPAC.

Women in Group B were less likely to report that they had ever been a regular smoker and were less likely to report they had smoked during the pregnancy (Table 4). With regards alcohol consumption before pregnancy, women in Group B were more likely to report of ‘one or more glasses a week’ than those in Group A during the periods both before and after the pregnancy. Group B were also more likely to report ‘never’ consuming alcohol during the pregnancy but reported higher rates of alcohol consumption both before and after the pregnancy.

There is negligible difference between the women enrolled in Group B, both ALSPAC and UKBiobank, and those in Group A, with regards their self-assessment of both their own and their partner’s health (Table 5).

Compared to women in Group A, the women in Group B were of higher socio-economic position across a number of measures: they had fewer financial difficulties, more stable housing and higher rates of home ownership, and lived in more affluent neighbourhoods (Table 6). The women in Group B were also of higher occupational social classes, and more likely to be educated to degree level (Table 7).

---

61 IMD – Index of Multiple Deprivation. This is produced by the Ministry of Housing, Communities and Local Government. Full details are available at https://www.gov.uk/government/collections/english-indices-of-deprivation

62 OPCS – Office of Population Censuses and Surveys, now the Office of National Statistics (ONS).

---

63 Rose D, ‘Official Social Classifications in the UK’, Social Research Update, issue 9, July 1995, https://sru.soc.surrey.ac.uk/SRU9.html
Table 1. Comparison of participation rates of the women in ALSPAC and UKBB with the women only in ALSPAC.

|                  | Group B Enrolled in UKBioBank | Group A NOT in UKBioBank |
|------------------|-------------------------------|---------------------------|
|                  | [n] [ % ] missing             | [n] [ % ] missing          | p-value         |
| Partner is enrolled in ALSPAC? | 477 55.1% {0} | 2,943 26.1% {0} | 0.000 |
| Participation measured as deciles of the return of surveys | 0 | 0 |
| a 1st (highest) | 429 49.5% | 1,969 17.5% | 0.000 |
| b 2nd           | 194 22.4% | 1,828 16.2% | 0.028 |
| c 3rd           | 86 9.9%   | 1,098 9.8%   | 0.976 |
| d 4th           | 46 5.3%   | 1,031 9.2%   | 0.367 |
| e 5th           | 23 2.7%   | 924 8.2%     | 0.339 |
| f 6th           | 24 2.8%   | 803 7.1%     | 0.415 |
| g 7th           | 26 3.0%   | 999 8.9%     | 0.293 |
| h 8th           | 16 1.8%   | 821 7.3%     | 0.399 |
| i 9th           | 13 1.5%   | 1,027 9.1%   | 0.341 |
| j 10th (lowest) | 9 1.0%    | 761 6.8%     | 0.490 |

Figure 2. Participation rates over time (20 years) of women in both ALSPAC and UKBiobank (Group B) and those not also enrolled in UKBiobank (Group A) as a percentage of each group returning each questionnaire.

Conclusions
It had long been suspected that many of the women and men in ALSPAC were also enrolled into UKBiobank. ALSPAC and UKBiobank are both large cohorts that are often used together in studies, and so the level of independence between the two data sources therefore needs to be understood.
### Table 2. Comparison of ethnicity, co-habitation rates and age groups of the women in ALSPAC and UKBB with the women only in ALSPAC.

|                          | Group B Enrolled in UKBioBank | Group A NOT in UKBioBank |
|--------------------------|-------------------------------|--------------------------|
|                          | n=866                         | n=11,261                 |
|                          | [ n ] [ % ] missing           | [ n ] [ % ] missing       |
| Women's ethnicity is White | 820 97.7% [ 27 ]             | 9,811 97.4% [ 1188 ]    | 0.598 |
| Study child's ethnicity is White | 802 96.3% [ 33 ]             | 9,431 95.1% [ 1340 ]    | 0.120 |
| Woman and partner co-habited | 809 96.0% [ 23 ]             | 9,904 93.4% [ 657 ]     | 0.003 |
| Partner older than woman | 431 83.7% [ 351 ]             | 3,860 82.4% [ 6575 ]     | 0.461 |
| Women's age (in years)   | [ 21 ]                        | [ 1672 ]                 |
| a 20 – 24                | 52 6.2%                      | 1,483 15.5%               | 0.066 |
| b 25 – 29                | 295 34.9%                    | 4,137 43.1%               | 0.006 |
| c 30 – 34                | 323 38.2%                    | 2,968 31.0%               | 0.008 |
| d 35 – 39                | 151 17.9%                    | 884 9.2%                  | 0.001 |
| e 40 – 44                | 24 2.8%                      | 117 1.2%                  | 0.553 |

### Table 3. Comparison of parity and breastfeeding rates of the women in ALSPAC and UKBB with the women only in ALSPAC.

|                                      | Group B Enrolled in UKBioBank | Group A NOT in UKBioBank |
|--------------------------------------|-------------------------------|--------------------------|
|                                      | n=866                         | n=11,261                 |
|                                      | [ n ] [ % ] missing           | [ n ] [ % ] missing       |
| First pregnancy                      | 268 31.3% [ 10 ]              | 3,324 31.2% [ 598 ]      | 0.952 |
| Singleton pregnancy                  | 853 98.5% [ 0 ]               | 11,104 98.6% [ 2 ]       | 0.810 |
| Woman was primiparous                | 338 40.3% [ 27 ]              | 4,250 41.0% [ 901 ]      | 0.692 |
| Last child was breastfed             | 428 86.6% [ 7 ]               | 4,271 70.7% [ 69 ]       | 0.000 |
| Woman intended to breastfeed index child | 742 91.5% [ 55 ]             | 7,659 80.5% [ 1748 ]     | 0.000 |
| Partner intended child to breastfeed | 499 95.8% [ 345 ]             | 4,717 85.7% [ 5756 ]     | 0.000 |
| The study child ever breastfed       | 722 91.0% [ 18 ]              | 8,154 79.0% [ 945 ]      | 0.000 |

### Table 4. Comparison of smoking and alcohol consumption rates of the women in Group B (ALSPAC and UKBB) with the women only in Group A (ALSPAC only).

|                                      | Group B Enrolled in UKBioBank | Group A NOT in UKBioBank |
|--------------------------------------|-------------------------------|--------------------------|
|                                      | n=866                         | n=11,261                 |
|                                      | [ n ] [ % ] missing           | [ n ] [ % ] missing       |
| Has ever smoked?                     | 333 38.9% [ 10 ]              | 5,681 52.6% [ 458 ]      | 0.000 |
| Smoked during this pregnancy?        | 93 11.2% [ 34 ]               | 2,577 25.4% [ 1120 ]     | 0.000 |
| Alcohol consumption before pregnancy | [ 13 ]                        | [ 689 ]                  |
| a Never                              | 40 4.7%                       | 863 8.2%                  | 0.426 |
| b <1 glass/week                      | 305 35.8%                     | 4,011 37.9%               | 0.466 |
| c 1+ glasses/week                    | 403 47.2%                     | 4,532 42.8%               | 0.087 |
| d 1-9 glasses/day                    | 105 12.3%                     | 1,166 11.0%               | 0.685 |
Table 5. Comparison of the self-reported health of the women and their partners in Group B (ALSPAC and UKBB) with the women only in Group A (ALSPAC only) from before the pregnancy to 21 months after.

| Variables                                    | Group B Enrolled in UKBioBank | Group A NOT in UKBioBank |
|----------------------------------------------|-------------------------------|--------------------------|
|                                              | n=866                         | n=11,261                 |
|                                              | [ n ] [%] missing              | [ n ] [%] missing        |
| Alcohol consumption during pregnancy         |                               |                          |
| a Never                                       | 404 52.5%                     | 5,619 58.3%              |
| b <1 glass/week                               | 134 17.4%                     | 1,257 13.0%              |
| c 1+ glasses/week                             | 183 23.8%                     | 2,113 21.9%              |
| d 1-9 glasses/day                             | 49 6.3%                       | 650 6.7%                 |
| Alcohol consumption after pregnancy          |                               |                          |
| a Never                                       | 138 16.7%                     | 2,130 22.2%              |
| b <1 glass/week                               | 394 47.7%                     | 4,813 50.1%              |
| c 1+ glasses/week                             | 244 29.5%                     | 2,249 23.4%              |
| d 1-9 glasses/day                             | 50 6.1%                       | 421 4.4%                 |

| Women's health before pregnancy              | [ 55 ]                        | [ 1472 ]                 |
| a Always Well                                 | 297 36.6%                     | 3,153 32.2%              |
| b Mostly Well                                 | 478 58.9%                     | 5,899 60.3%              |
| c Sometimes/ Often/ Always Unwell             | 36 4.4%                       | 737 7.5%                 |

| Women's health 8 weeks after child's birth    | [ 50 ]                        | [ 1720 ]                 |
| a Always Well                                 | 309 37.9%                     | 3,333 34.9%              |
| b Mostly Well                                 | 488 59.8%                     | 5,978 62.7%              |
| c Sometimes/ Often/ Always Unwell             | 19 2.3%                       | 230 2.4%                 |

| Women's health 21 months after child's birth  | [ 78 ]                        | [ 2722 ]                 |
| a Always Well                                 | 450 57.1%                     | 4,704 55.1%              |
| b Mostly Well                                 | 308 39.1%                     | 3,443 40.3%              |
| c Sometimes/ Often/ Always Unwell             | 30 3.8%                       | 392 4.6%                 |

| Partner / dad's health during early pregnancy | [ 29 ]                        | [ 810 ]                  |
| a Always Well                                 | 388 46.4%                     | 4,869 46.6%              |
| b Mostly Well                                 | 418 49.9%                     | 5,137 49.2%              |
| c Sometimes/ Often/ Always Unwell             | 31 3.7%                       | 445 4.3%                 |

| Partner / dad's health 8 weeks after child's birth | [ 210 ]                        | [ 4374 ]                 |
| a Always Well                                 | 352 53.7%                     | 3,690 53.6%              |
| b Mostly Well                                 | 289 44.1%                     | 3,056 44.4%              |
| c Sometimes/ Often/ Always Unwell             | 15 2.3%                       | 141 2.0%                 |

| Partner / dad's health 21 months after child's birth | [ 305 ]                        | [ 6130 ]                 |
| a Always Well                                 | 349 62.2%                     | 3,175 61.9%              |
| b Mostly Well                                 | 193 34.4%                     | 1,830 35.7%              |
| c Sometimes/ Often/ Always Unwell             | 19 3.4%                       | 126 2.4%                 |
Table 6. Comparison of early financial and residential circumstances.

| Variables                                      | Group B Enrolled in UKBioBank | Group A NOT in UKBioBank |
|------------------------------------------------|-------------------------------|--------------------------|
|                                                 | n=866                         | n=11,261                 |
|                                                 | [ n ]  | [%]  | missing | [ n ]  | [%]  | missing | p-value |
| Had any financial difficulties at any stage?    | 125   | 16.7%| { 117 } | 2,050 | 27.1%| { 3689 }| 0.000   |
| Number of home moves during five years prior to pregnancy | [ 14 ] |  | { 373 } |  |  |  |
| a 0                                             | 232   | 27.5%| 2,475 | 23.6%| 373 | 3.6% | 0.183   |
| b 1                                             | 313   | 37.0%| 3,361 | 32.1%| 23  | 2.7% | 0.077   |
| c 2                                             | 165   | 19.5%| 2,316 | 22.1%| 511 | 4.9% | 0.629   |
| d 3                                             | 78    | 9.2% | 1,188 | 11.3%| 256 | 2.4% | 1.000   |
| e 4                                             | 23    | 2.7% | 511   | 4.9% | 373 | 3.6% | 0.705   |
| f 5                                             | 20    | 2.4% | 256   | 2.4% | 373 | 3.6% | 0.705   |
| g 6+                                            | 14    | 1.7% | 373   | 3.6% | 373 | 3.6% | 0.705   |
| Number of addresses known to ALSPAC, since enrolling to 2019 | [ 5 ] |  | [ 112 ] |  |  |  |
| a 1                                             | 328   | 38.1%| 3,379 | 30.3%| 0.004 |  |
| b 2                                             | 312   | 36.2%| 3,700 | 33.2%| 0.281 |  |
| c 3                                             | 122   | 14.2%| 2,003 | 18.0%| 0.287 |  |
| d 4                                             | 58    | 6.7% | 1,072 | 9.6% | 0.462 |  |
| e 5                                             | 25    | 2.9% | 504   | 4.5% | 0.704 |  |
| f 6+                                            | 16    | 1.9% | 491   | 4.4% | 0.628 |  |
| Status of women's residence at time of pregnancy | [ 11 ] |  | [ 553 ] |  |  |  |
| a homeowner                                     | 782   | 91.5%| 8,115 | 75.8%| 0.000 |  |
| b private rental                                 | 19    | 2.2% | 712   | 6.6% | 0.442 |  |
| c public housing                                 | 39    | 4.6% | 1,572 | 14.7%| 0.076 |  |
| d other                                         | 15    | 1.8% | 309   | 2.9% | 0.803 |  |
| IMD decile of women's residential address at time of pregnancy | [ 74 ] |  | [ 861 ] |  |  |  |
| a 1<sup>st</sup> (highest)                      | 133   | 16.8%| 1,390 | 13.4%| 0.276 |  |
| b 2<sup>nd</sup>                                 | 149   | 18.8%| 1,283 | 12.3%| 0.025 |  |
| c 3<sup>rd</sup>                                 | 99    | 12.5%| 1,210 | 11.6%| 0.789 |  |
| d 4<sup>th</sup>                                 | 41    | 5.2% | 557   | 5.4% | 0.956 |  |
| e 5<sup>th</sup>                                 | 85    | 10.7%| 992   | 9.5% | 0.718 |  |
| f 6<sup>th</sup>                                 | 51    | 6.4% | 818   | 7.9% | 0.699 |  |
| g 7<sup>th</sup>                                 | 80    | 10.1%| 1,017 | 9.8% | 0.931 |  |
| h 8<sup>th</sup>                                 | 50    | 6.3% | 768   | 7.4% | 0.772 |  |
| i 9<sup>th</sup>                                 | 60    | 7.6% | 1,418 | 13.6%| 0.181 |  |
| j 10<sup>th</sup> (lowest)                      | 44    | 5.6% | 947   | 9.1% | 0.426 |  |
### IMD decile of women’s residence, at start of UKBiobank (January 2008)

| IMD Decile | Group B Enrolled in UKBioBank | Group A NOT in UKBioBank |
|------------|-------------------------------|--------------------------|
|            | n=866                         | n=11,261                 |
|            | [ n ] [ % ]                   | [ n ] [ % ]              |
| a 1<sup>st</sup> (highest) | 188 22.1% | 1,962 18.7% | 0.256 |
| b 2<sup>nd</sup> | 112 13.2% | 1,225 11.6% | 0.615 |
| c 3<sup>rd</sup> | 138 16.2% | 1,180 11.2% | 0.084 |
| d 4<sup>th</sup> | 122 14.4% | 1,390 13.2% | 0.708 |
| e 5<sup>th</sup> | 105 12.4% | 1,066 10.1% | 0.460 |
| f 6<sup>th</sup> | 47 5.5% | 786 7.5% | 0.611 |
| g 7<sup>th</sup> | 50 5.9% | 722 6.9% | 0.786 |
| h 8<sup>th</sup> | 50 5.9% | 758 7.2% | 0.729 |
| i 9<sup>th</sup> | 29 3.4% | 689 6.6% | 0.493 |
| j 10<sup>th</sup> (lowest) | 9 1.1% | 740 7.0% | 0.488 |

### Table 7. Comparison of early social class and education achievement.

|                        | Group B Enrolled in UKBioBank | Group A NOT in UKBioBank |
|------------------------|-------------------------------|--------------------------|
|                        | n=866                         | n=11,261                 |
|                        | [ n ] [ % ]                   | [ n ] [ % ]              |
|                        | missing                       | missing                  |
| p-value                |                               |                          |

#### Women’s highest education level

| Level         | Group B Enrolled in UKBioBank | Group A NOT in UKBioBank |
|---------------|-------------------------------|--------------------------|
| a CSE/none    | 56 6.7% | 1,619 17.6% | 0.034 |
| b Vocational  | 55 6.6% | 853 9.3% | 0.501 |
| c O Level     | 274 33.0% | 3,270 35.5% | 0.406 |
| d A Level     | 241 29.0% | 2,216 24.1% | 0.093 |
| e Degree      | 204 24.6% | 1,249 13.6% | 0.000 |

#### Partner / father’s highest education level

| Level         | Group B Enrolled in UKBioBank | Group A NOT in UKBioBank |
|---------------|-------------------------------|--------------------------|
| a CSE/none    | 112 13.5% | 2,461 25.1% | 0.005 |
| b Vocational  | 68 8.2% | 814 8.3% | 0.977 |
| c O Level     | 143 17.2% | 2,104 21.5% | 0.224 |
| d A Level     | 248 29.8% | 2,622 26.7% | 0.293 |
| e Degree      | 261 31.4% | 1,803 18.4% | 0.000 |

#### Women’s Social Class

| Social Class                        | Group B Enrolled in UKBioBank | Group A NOT in UKBioBank |
|-------------------------------------|-------------------------------|--------------------------|
| a I - Professional etc. occupations | 81 10.6% | 474 5.7% | 0.096 |
| b II - Managerial and Technical occupations | 255 33.3% | 2,708 32.8% | 0.871 |
| c III - Skilled occupations (non-manual) | 345 45.0% | 3,482 42.1% | 0.298 |
| d III - Skilled occupations (manual) | 32 4.2% | 633 7.7% | 0.464 |
| e IV - Partly skilled occupations   | 45 5.9% | 786 9.5% | 0.419 |
| f V - Unskilled occupations        | 8 1.0% | 179 2.2% | 0.819 |
For the first time, using a novel encryption method, we have identified a cohort of women who are enrolled in both ALSPAC and UKBB (7.1% of the ALSPAC women who met the UKBB age eligibility criteria). It is not clear whether it will remain possible to identify members-in-common between two different research projects if they are held within two different TREs (Trusted Research Environments).

ALSPAC can be considered as broadly representative of the population at the time due to being a birth cohort with very basic eligibility criteria and the project’s high recruitment rate. UKBB is a cohort with a significantly lower recruitment rate possibly indicating a self-selection bias in the study sample.

We compared those women who are in both cohorts, Group B, to the women in ALSPAC who were eligible but did not enrol into UKBiobank (Group A) and found that they differed on many characteristics. Notably the women in Group B participated more fully in ALSPAC and provided more data over time. Those in Group B were older and more advantaged across a range of important socio-economic measures such as housing and education, and they had some healthier behaviours such as smoking less, drinking less during pregnancy and being more likely to have breastfed their babies. However, they tended to drink more when not pregnant.

It is conjectured any differences in the characteristics of the ALSPAC mothers enrolled in UKBB are indicative of those in UKBB compared to the wider population, as indicated by the ALSPAC mothers not in in UKBB. This could be of assistance when interpreting results using UKBB data.

As there is a very large amount of data on the approximately 900 women who are in both studies, there could be the potential for studies focused only on these women. The knowledge that there are disparate datasets holding detailed, and different, information on these individuals and that potentially these are able to be linked quickly and accurately is important. This could be leveraged to facilitate rapid research into evolving real-world situations, such as the COVID pandemic, with minimal extra resource or through sharing existing resource.

### Data availability

UKBiobank details can be found at [https://www.ukbiobank.ac.uk/](https://www.ukbiobank.ac.uk/)

Further details on ALSPAC can be found at [http://www.bristol.ac.uk/alspac/](http://www.bristol.ac.uk/alspac/)

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:

i. Please read the ALSPAC access policy ([http://www.bristol.ac.uk/media-library/sites/alspac/documents/researchers/data-access/ALSPAC_Access_Policy.pdf](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.

ii. You may also find it useful to browse our fully searchable research proposals database ([https://proposals.epi.bristol.ac.uk/?q=proposalSummaries](https://proposals.epi.bristol.ac.uk/?q=proposalSummaries)), which lists all research projects that have been approved since April 2011.

iii. Please submit your research proposal ([https://proposals.epi.bristol.ac.uk/](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.

The availability of our linked participant records is dependent on our ethical approvals and contractual arrangements with the NHS. If you are interested in using these data, then please

| Partner / Paternal Social Class | Group B Enrolled in UKBiobank | Group A NOT in UKBiobank |
|-------------------------------|-------------------------------|--------------------------|
|                               | n=866                         | n=11,261                 |
|                               | [ n ] | [ % ] | missing | [ n ] | [ % ] | missing | p-value |
| a I - Professional etc. occupations | 134   | 17.0% | [77]    | 1,010 | 11.1% | [2192] | 0.047   |
| b II - Managerial and Technical occupations | 304   | 38.5% |         | 3,133 | 34.5% |         | 0.162   |
| c III - Skilled occupations (non-manual) | 103   | 13.1% |         | 981   | 10.8% |         | 0.478   |
| d III - Skilled occupations (manual) | 196   | 24.8% |         | 2,834 | 31.2% |         | 0.060   |
| e IV - Partly skilled occupations | 42    | 5.3%  |         | 875   | 9.6%  |         | 0.351   |
| f V - Unskilled occupations     | 10    | 1.3%  |         | 236   | 2.6%  |         | 0.798   |
Contact the ALSPAC Data Linkage Team (alspac-linkage@bristol.ac.uk).

Author contributions
Data note conceptualisation and funding organised by NT, who also assisted with review and editing. Methodology, review and editing by AB. RT contributed to creating the dataset of matched participants. AT assisted with review and editing.

Visualisation, investigation, curation and writing of the data note by MM.

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.