



DATA NOTE

REVISED **The Avon Longitudinal Study of Parents and Children - A resource for COVID-19 research: Questionnaire data capture April-May 2020 [version 2; peer review: 2 approved]**

Kate Northstone ¹, Simown Howarth ², Daniel Smith ¹, Claire Bowring ¹, Nicholas Wells ¹, Nicholas John Timpson ^{1,3}

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

²Bristol Dental School, University of Bristol, Bristol, BS1 2LY, UK

³MRC Integrative Epidemiology Unit, Department of Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, BS8 2BN, UK

v2 **First published:** 10 Jun 2020, 5:127
<https://doi.org/10.12688/wellcomeopenres.16020.1>
Latest published: 20 Nov 2020, 5:127
<https://doi.org/10.12688/wellcomeopenres.16020.2>

Abstract

The Avon Longitudinal Study of Parents and Children (ALSPAC) is a prospective population-based cohort study which recruited pregnant women in 1990-1992. The resource provides an informative and efficient setting for collecting data on the current coronavirus 2019 (COVID-19) pandemic. In early March 2020, a questionnaire was developed in collaboration with other longitudinal population studies to ensure cross-cohort comparability. It targeted retrospective and current COVID-19 infection information (exposure assessment, symptom tracking and reported clinical outcomes) and the impact of both disease and mitigating measures implemented to manage the COVID-19 crisis more broadly. Data were collected on symptoms of COVID-19 and seasonal flu, travel prior to the pandemic, mental health and social, behavioural and lifestyle factors.

The online questionnaire was deployed across parent (G0) and offspring (G1) generations between 9th April and 15th May 2020. 6807 participants completed the questionnaire (2706 original mothers, 1014 original fathers/partners, 2973 offspring (mean age ~28 years) and 114 offspring partners). Eight (0.01%) participants (4 G0 and 4 G1) reported a positive test for COVID-19, 77 (1.13%; 28 G0 and 49 G1) reported that they had been told by a doctor they likely had COVID-19 and 865 (12.7%; 426 G0 and 439 G1) suspected that they have had COVID-19. Using algorithmically defined cases, we estimate that the predicted proportion of COVID-19 cases ranged from 1.03% - 4.19% depending on timing during the period of reporting (October 2019-March 2020).

Data from this first questionnaire will be complemented with at least two more follow-up questionnaires, linkage to health records and

Open Peer Review

Reviewer Status  

Invited Reviewers

1 **2**

version 2

(revision)
20 Nov 2020



report



report




version 1


10 Jun 2020



report

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

Janis Baird, University of Southampton, Southampton, UK

2. **Anteo Di Napoli** , National Institute for Health Migration and Poverty (INMP), Rome, Italy

Martina Ventura, National Institute for Health Migration and Poverty (INMP), Rome, Italy

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

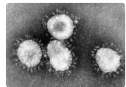
results of biological testing as they become available. Data has been released as: 1) a standard dataset containing *all* participant responses with key sociodemographic factors and 2) as a composite release coordinating data from the existing resource, thus enabling bespoke research across all areas supported by the study.

Keywords

ALSPAC, Children of the 90s, birth cohort study, COVID-19, coronavirus, online questionnaire, mental health



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



This article is included in the [Coronavirus \(COVID-19\)](#) collection.

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

Author roles: **Northstone K:** Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Methodology, Project Administration, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing; **Howarth S:** Investigation, Methodology, Writing – Review & Editing; **Smith D:** Data Curation, Formal Analysis, Writing – Review & Editing; **Bowring C:** Methodology, Project Administration, Supervision, Writing – Review & Editing; **Wells N:** Methodology, Project Administration, Software, Writing – Review & Editing; **Timpson NJ:** Conceptualization, Funding Acquisition, Investigation, Supervision, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: The work was supported by the Wellcome Trust through core support to ALSPAC [102215]. The UK Medical Research Council, the Wellcome Trust and the University of Bristol provide core support for ALSPAC. 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>). This work was also supported by the Elizabeth Blackwell Institute for Health Research, University of Bristol.

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

Copyright: © 2020 Northstone K *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: Northstone K, Howarth S, Smith D *et al.* **The Avon Longitudinal Study of Parents and Children - A resource for COVID-19 research: Questionnaire data capture April-May 2020 [version 2; peer review: 2 approved]** Wellcome Open Research 2020, 5:127 <https://doi.org/10.12688/wellcomeopenres.16020.2>

First published: 10 Jun 2020, 5:127 <https://doi.org/10.12688/wellcomeopenres.16020.1>

REVISED Amendments from Version 1

The second version of our manuscript takes into account the constructive comments made by the reviewers. In particular, we have provided detail in the introduction explaining what the UK lockdown entailed and we have emphasised the issues with applying the Menni algorithm to symptom data collected in 2019 in both the results and discussion.

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

Introduction

The coronavirus disease 2019 (COVID-19) pandemic is a rapidly developing global health challenge. There is marked heterogeneity in disease prevalence, severity and outcome both within and across populations. In part, this may be driven by the interplay between environmental, social and host factors such as age and pre-existing comorbidities which predispose or protect against infection or modify disease outcomes. Understanding this interplay requires studies with detailed environmental, health, lifestyle, and biological data – ideally measured within the context of longitudinal data and with prospective collection opportunities.

Alongside the health implications of the virus itself, the response to the pandemic is likely to affect health and wellbeing. Mitigation measures have resulted in far-reaching changes to daily activity which are likely to have an impact on nearly all aspects of work, family life, recreation and, potentially, health. The nationwide ‘lockdown’ strategy implemented on 23rd March 2020, for example, meant that UK residents were only allowed to leave the house to: 1) buy basic necessities, 2) exercise once a day, 3) attend to a medical need or care for a vulnerable person 4) travel to and from work that could not be done at home¹. These strictest measures were in place for several weeks, with easing beginning in early May, schools returning in June and many non-essential shops opening their doors again in mid-June. Any adverse effects of these mitigation strategies may themselves be heterogeneous, with certain groups at higher risk of adverse effects. Understanding the effect of mitigation strategies on health and identifying the social, environmental, or other factors which help reduce their impact will be an important part of planning both ongoing and future COVID-19 mitigation strategies as the pandemic develops.

The Avon Longitudinal Study of Parents and Children (ALSPAC) is a unique 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 has a wealth of biological, genetic and phenotypic data across these generations²⁻⁵. ALSPAC has an opportunity to capture information across key parts of the population in light of the COVID-19 pandemic – in particular the contrast between those in higher risk (G0 mean age: ~58years) and lower risk (G1 mean age: ~28yrs) groups. We were well placed to collect data quickly using our existing infrastructure for online data collection.

The wider COVID-19 data collection in ALSPAC will include data from three main sources: self-reported data from questionnaires, data from clinical services based on linkage to medical and other records and though biological samples. The data from these sources are intended to be complementary and help address different potential research questions around COVID-19. This data note describes the data collected via our first online questionnaire between 9th April and 14th May 2020 and provides a summary of the participants who responded. Updates on subsequent questionnaires and other sources of data will be released as these additional datasets are made available.

Methods**Setting**

ALSPAC is an intergenerational longitudinal cohort that recruited pregnant women residing in Avon, UK with expected dates of delivery 1st April 1991 to 31st December 1992^{2,3}. The initial cohort consisted of 14,541 pregnancies resulting in 14,062 live births and 13,988 children who were alive at 1 year of age. From the age of seven onwards, the initial sample was bolstered with eligible cases who had originally failed to join the study and there were subsequently 14,701 children alive at 1 year of age following this further recruitment⁴. Please note, the study website contains details of all the data that is available through a fully [searchable data dictionary and variable search tool](#).

In response to the COVID-19 it was necessary to develop a data collection strategy which was practical, would yield data quickly and could be updated and repeated. For these reasons, we chose to use an online only data collection approach for this, restricting our invites to those participants with a valid email address (and coordinated with a systematic communications/outreach campaign). The questionnaire was developed and deployed using REDCap (Research Electronic Data Capture tools⁶); a secure web application for building and managing online data collection exercises, hosted at the University of Bristol.

Content design

Content was initially developed internally, we then consulted with a network of 16 UK and international longitudinal population studies and partners through a process facilitated by Wellcome (see acknowledgements). This resulted in a core set and a recommended set of questions about health, behaviour, social, economic and environmental impact of COVID-19. Whilst we were able to align many features of the ALSPAC questionnaire and the Wellcome core questionnaire, there are some areas of divergence. This is because the ALSPAC questionnaire was deployed while the Wellcome core questionnaire was still being finalized, and because we chose to use mental health measures that we have used previously to facilitate longitudinal analyses rather than the Wellcome recommended measures. It is worth noting that ALSPAC is the contact point for using the Wellcome questionnaire and we can provide data dictionaries in either REDCap or Qualtrics (via Generation Scotland) on request. This work is therefore part of a coordinated effort to generate and promote a Wellcome Trust supported core questionnaire. This is now complete and available and access to this

can be organised through ALSPAC. The questionnaire has not been formally validated, however, extensive testing by our ethics committee, participant advisory group and ALSPAC staff led to clarity of wording and ensuring REDCap functionality worked as expected.

The questionnaire included 4 sections, and captured information on the following:

- A. General health, recent travel and seasonal symptoms
 - Conditions making people high risk
 - Frailty assessed using PRISMA -7⁷
 - Regular medications (prescription and over the counter)
 - Home country and travel outside that country since October 2019
 - Symptoms of COVID-19 and negative control symptoms since October 2019
 - Tested/diagnosed with COVID-19
- B. Behaviour as a result of COVID-19
 - Self-isolation and reasons why
 - Behaviour changes prior to lockdown
 - Lifestyle changes since lockdown
 - Social contacts and methods of communication
- C. Impact of the pandemic
 - Worries during the pandemic
 - Depression assessed using the Short Moods and Feelings questionnaire (SMFQ; 8)
 - Anxiety assessed using the General Anxiety Disorder-7 questionnaire (GAD7; 9)
 - Well-being assessed using the Warwick-Edinburgh Mental Wellbeing Scales (WEMWBS; 10)
- D. About you during the pandemic
 - Understanding of official guidance on COVID-19
 - Time spent talking/reading about or listening to information about COVID-19
 - Living arrangements
 - Healthcare worker and keyworker status
 - Effect of pandemic on plans to have children (G1 only)
 - Free text inviting participants to provide details of other ways they have been affected by the pandemic

The final questionnaire (REDCap PDF) used is available with the associated data dictionary (which includes frequencies of all variables that are available) and both are available as extended data¹¹.

Invitation and reminder strategy

Between the 9th and 15th April 2020, all participants (G0, G1 and G1 partners enrolled as part of G2 (children of the Children of the 90s⁵) for whom we had an active email

address were sent an invitation to complete the questionnaire. Participants were not contacted if our administrative database record indicated that they were deceased, had withdrawn from the study, had declined further contact or had declined questionnaires. The questionnaire survey was live on the online platform for just over one month. After 2 weeks, any non-responders were sent a reminder email to complete the questionnaire. In addition, traditional (print, radio, tv) & social media (Facebook, Instagram and Twitter) were used to inform participants that the questionnaire was live, asking them to contact us if they had not received it and to encourage completion. These communication channels were also used to encourage re-engagement of friends and family back into the study. Unlike our standard questionnaires (usually completed annually) we did not provide any incentive for completion; however, we did offer a prize draw (three prizes of £100) for those who completed their questionnaire by 11th May.

Response rate

A total of 12,520 invitations were sent out and responses were received from 6811 participants (overall response rate of 54%). Over 4,000 participants completed the questionnaire within the first week of the invitation (see [Figure 1](#)), this represents a snapshot of participant experience around 3 weeks after the start of 'lockdown'. The questionnaire was closed on 15th May 2020.

Overall, female participants were more likely to respond ([Table 1](#)), and this was particularly true of the younger generation. A potential explanation for the difference in male response rates between younger and older participants we note that G0 fathers included in this data collection exercise are already relatively engaged with the study as they had to enrol in their own right in 2011 as part of our 'Focus on Fathers' data collection (previously they were only invited to participate via the mother). The response from G1 males was disappointing and we aim to re-engage this group using targeted social media and other activities in the future. [Table 1](#) summarises the response rate within each group organised by cohort structure.

Characteristics of responders according to key variables that will be released with the complete dataset can be seen in [Table 2](#).

Key results

Participants were asked whether they thought they have had COVID-19. Options were: 'Yes, confirmed by a positive test', 'Yes, suspected by a doctor but not tested', 'Yes, my own suspicions' or 'No'. Overall 8 respondents reported that they had tested positive to COVID-19 (when combined with participant group this information is potentially disclosive it will therefore be combined with 'yes, suspected by a doctor' to create a new category in the released datasets 'Yes, tested positive or suspected by a doctor'). [Table 3](#) summarises the responses to this question by cohort structure.

Menni and Colleagues¹² recently published a model combining symptoms to predict 'probable infection' using data collected from an app-based symptom tracker¹³. This algorithm uses

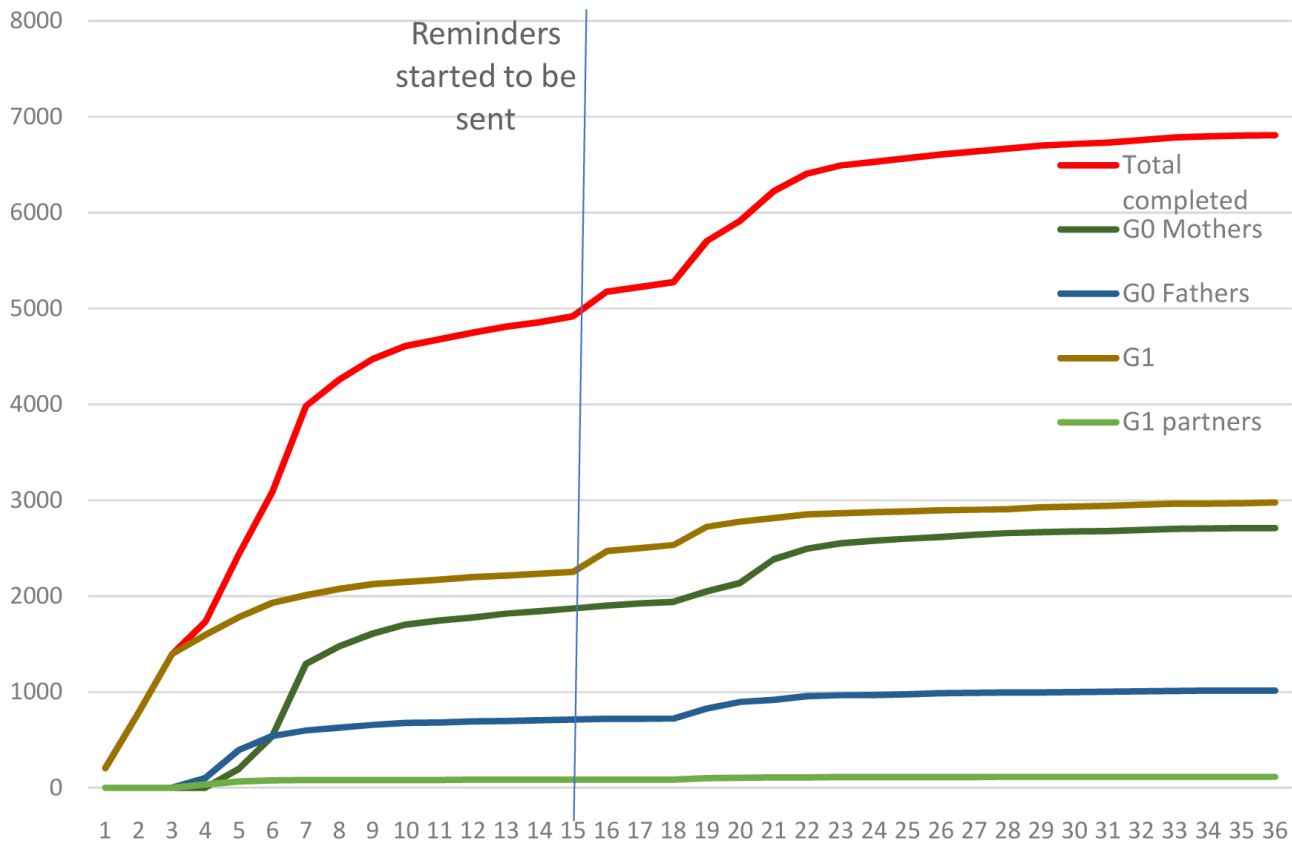


Figure 1. Completion rate by number of days questionnaire was live. Invitations were sent in batches over several days - G0 invitations did not start going out until 3 days after the questionnaire was live.

Table 1. Number of participants who were eligible and who responded to the first COVID-19 questionnaire.

Cohort Group	Eligible ¹	Responded ²
G0 Mothers	4590	2706 (59%)
G0 Fathers/partners	1803	1014 (56%)
G1 Offspring daughters	3617	2126 (59%)
G1 Offspring sons	2225	847 (38%)
G1 Offspring partners (female)	103	63 (61%)
G1 Offspring partners (male)	182	51 (28%)
TOTAL	12520	6807 (54%)

¹valid email address, marked as contactable for questionnaires

²Proportions of those invited (i.e. eligible)

four symptoms: loss of smell and taste, severe or significant persistent cough, severe fatigue and skipped meals (coded as 1 if present and 0 otherwise), together with age and sex (1 male; 0 female). We had slight difference in wording and

thus the algorithm (using the same weightings) applied was as follows:

$$\begin{aligned}
 & -1.32 - (0.01 \times \text{age}) + (0.44 \times \text{sex}) + (1.75 \times \text{loss of loss of smell or taste}) \\
 & + (0.31 \times \text{new persistent cough}) + (0.49 \times \text{severe fatigue}) \\
 & + (0.39 \times \text{decreased appetite}).
 \end{aligned}$$

Probable COVID-19 cases were obtained by applying an $\exp(x)/[1+(\exp(x))]$ transformation and coding values >0.5 as probable cases. We applied this algorithm to our monthly symptom data and the number of predicted 'cases' each month are shown in Figure 2. We used Stata v.15.0 and our Stata do file is available as extended data¹¹. Peak predicted cases were seen in March (4.19%) with similar proportions seen in all other months (2.10% - 2.38%), except November which was substantially lower (1.03%). Our figures were consistently lower than the 5.36% of responders to the app¹² who were reported as likely being infected by the virus. It should be noted that Menni *et al.* developed their algorithm based on data collected in the spring of 2020 when COVID-19 levels were at their highest. We have applied the algorithm to data collected

Table 2. Summary of key characteristics for those who responded; n (%) for categorical variables, mean (sd) for BP and median (Inter-quartile range) for BMI.

	Mothers	Fathers/ partners	Offspring	Offspring partners
Age (years)	57.9 (4.44)	60.8 (5.17)	27.6 (0.54)	29.9 (4.37)
Latest BMI ¹	25.4 (22.9, 28.7)	26.9 (24.9, 29.1)	23.4 (21.1, 26.9)	26.5 (23.3, 30.4)
Latest Systolic BP ¹	119.4 (14.10)	132.9 (13.77)	115.2 (10.83)	116.3 (12.44)
Latest Diastolic BP ¹	70.6 (9.33)	77.2 (9.03)	66.7 (7.77)	65.9 (9.97)
Education level ² ≥A level	1385 (53.7%)	674 (70.4%)	1793 (77.8%)	26 (60.5%)
Ethnicity ³ White	2533 (98.4%)	948 (99.2%)	2559 (96.6%)	Not available

¹Data taken from the most recent clinic that individual attended where available

²Data taken from pregnancy questionnaires for G0 and from most recent questionnaire for G1 where available

³Data taken from pregnancy questionnaires for all

Table 3. Participant response to whether they have had COVID-19.

	G0 - parents	G1 - offspring (+partners)	Total
Yes, positive test	4 (0.01%)	4 (0.01%)	8 (0.01%)
Yes, doctor suspected, no test	28 (0.76%)	48 (1.56%)	76 (1.12%)
Yes, own suspicions	426 (11.5%)	439 (14.2%)	865 (12.8%)
No	3229 (87.4%)	2584 (83.8%)	5734 (84.6%)

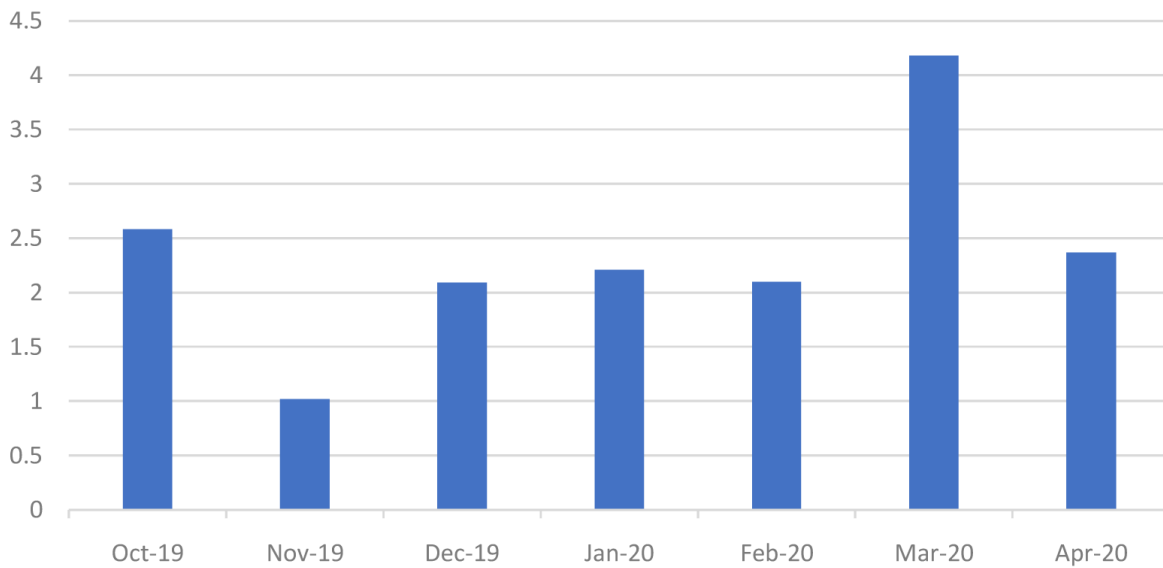


Figure 2. Predicted cases (% of population) of COVID-19 per month according to symptoms reported for those months using Menni et al algorithm¹².

prior to this and demonstrate a predicted prevalence of between 1 and 2.5% in 2019. Whilst, COVID-19 may have been present in the UK population during this time, it is more likely that symptoms reported by participants during this period were as a result of other respiratory viruses and these predicted cases should therefore be interpreted with caution. Table 4 summarises the predicted cases of COVID-19 each month according to self-reported infection status. There is a clear temporal trend such that the proportion of predicted cases increases over time in those who self-reported that they had COVID-19. As noted above, the results from 2019 should be interpreted with caution. Indeed the trends of self-reported infection status are much lower in those months, reflecting the likelihood that the algorithm was picking up a different respiratory infection.

Whilst we have presented these results as a mark of the type of analysis one can undertake using these data, we note that these predictions are subject to important assumptions. First, the *baseline risk of having COVID-19* (intercept term in the model) is assumed to be the same in the ALSPAC study population as the Menni study population. This assumption may be invalid as there are fewer reported COVID-19 cases in South West England compared to other regions of the UK¹⁴, potentially over-estimating prevalence in ALSPAC. Secondly, the symptoms we termed ‘new persistent cough’, ‘severe fatigue’ and ‘decreased appetite’ are assumed to capture the same information as the symptoms used in the Menni study but may in truth have subtly different meaning, leading to either over or under-estimation of prevalence in the ALSPAC study. Thirdly, the *association of these symptoms with COVID-19* (fixed effects in the model) is assumed to be the same in the ALSPAC study population as the Menni study population. This assumption might be violated if the pattern of symptom presentation varies in different

groups of people, leading to over or under-estimation of prevalence in the ALSPAC study.

Strengths and limitations of the data

The primary strengths of this data are the timelines within which the collection occurred, the retrospectively available and prospectively continuing longitudinal collection forming the context for these new data and the potential for cross-cohort comparisons with a set of measures aligned to other UK studies. We believe the timeframe is important, as the data collected here reflects the feelings of the cohort early on in mitigation during a period of stringent lockdown measures. In addition to this, given that the questionnaire was active for over a month it will be important to take date of completion into account for certain analyses, as responses may differ between early- versus late- completers (e.g. late completers experienced an additional month of lockdown measures and have adapted more to the new circumstances). Our second COVID-19 data collection, funded and planned to start at the end of May will provide invaluable comparisons as (at the time of writing) some lockdown features are starting to be relaxed (certain work sectors being encouraged to return to work etc). Through the Wellcome coordinated group developing a core set of questions we will be able to make valuable comparisons with other cohort groups of different ages, backgrounds and from different countries and cultures.

We had a good response rate given the pandemic; however, it should be noted that invitations were only sent to those participants for whom we had a valid email address and we therefore have limitations to coverage. Our usual questionnaire strategy is to first send emails to encourage online completion. We then follow up as part of the reminder process and send

Table 4. Predicted cases of COVID-19 each month using Menni *et al.* algorithm¹³ according to self-reported infection status (note that valid self-report infection status data available for n=6664 which is why the total column is higher than the sum of the other columns).

	Tested positive (0.01%)	Doctor suspected (1.13%)	Participant suspected (12.7%)	Not had (84.6%)	TOTAL
Oct 2019 n=166	0	1 (0.6%)	23 (13.9%)	142 (85.5%)	176 (2.58%)
Nov 2019 n=64	0	1 (1.6%)	15 (23.4%)	48 (75.0%)	70 (1.03%)
Dec 2019 n=136	0	1 (0.7%)	40 (29.4%)	95 (69.9%)	143 (2.10%)
Jan 2020 n=145	0	0	51 (35.2%)	94 (64.8%)	151 (2.22%)
Feb 2020 n=137	0	3 (2.2%)	56 (40.9%)	78 (56.9%)	144 (2.11%)
March 2020 n=271	0	37 (13.7%)	139 (51.3%)	95 (35.1%)	285 (4.19%)
April 2020 n=153	0	29 (18.9%)	70 (45.6%)	54 (35.3%)	162 (2.38%)

paper questionnaires through the post. This online only strategy will have particularly affected G0 mothers who historically have tended to use paper questionnaires more than other sub-groups and for whom we are least likely to hold a current email address. This online only strategy will no doubt result in selection bias, which must be taken into account when using this data, as the proportion of original participants in the study taking part is small. Work is ongoing to describe this selection bias and will be published in due course. However, the pandemic has led to a number of participants reaching out and getting in touch to provide these details, and indeed to re-engage with the study having dropped out previously.

In some cases, the data recorded is potentially identifiable. We went through each variable one by one and made decisions about whether to combine categories. We have combined categories where we feel the data is at high risk of potential disclosure.

A further key limitation of this data is the reduced response rate in our male G1 participants (28% of all G1 responders were male), even compared to previous data collection exercises in this group where around a third of G1 responders have been male.

Finally, the collection was of course subject to the frequency of COVID-19 infection in the population and sample. Very few participants reported a positive test to COVID-19 and the frequency of algorithmically assigned case status suggests a frequency of symptomatic presentation which is consistent with regional estimates¹⁴. This is partly due to UK policy not to test widely at the time that data was being collected but also the fact that the majority of our participants remain in the South-West of England which has had the lowest COVID-19 death rate in the country (and potentially the lowest infection rate¹⁵). All symptoms were self-reported and the use of the algorithm to predict cases based on these symptoms is subject to a number of limitations as discussed in the results section. We aim to address these limitations by linking to health records (including Public Health England testing results) and direct serological testing of our participants in order to assess the true symptoms and prevalence of COVID-19 in this population.

Data availability

Underlying data

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](#)¹⁶ 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](#)¹⁷, which lists all research projects that have been approved since April 2011.
3. Please [submit your research proposal](#)¹⁸ 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.

Please note that a standard COVID-19 dataset will be made available at no charge (see description below); however, costs for required paperwork and any bespoke datasets required additional variables will apply.

COVID-19 Questionnaire 1 Data File

Data from the first ALSPAC COVID-19 questionnaire is available in two ways.

1. A freely available standard set of data containing *all* participants together with key sociodemographic variables (where available) is available on request (see data availability section). Subject to the relevant paperwork being completed (costs may apply to cover administration) this dataset will be made freely available to any bona fide researcher requesting it. Variable names will follow the format *covid1_xxxx* where *xxxx* is a four-digit number. A full list of variables released is available as extended data¹¹. Frequencies of variable and details of any coding/editing decisions and derived variables are also available in the data dictionary (see extended data¹¹).
2. Formal release files have been created for G0 mothers, G0 fathers and G1 participants in the usual way and now form part of the ALSPAC resource (Due to the small number of G1 partners contributing we will not be formally releasing this data, however, it may be available on request for specific G2 projects). These datasets (or sections therein) can be requested in the usual way. Variable names will replicate those in 1) above but as each variable in ALSPAC is uniquely defined we have added letters to denote the source of the variable. For example, in dataset 1, the age of the participant at completion (in years) is denoted by *covid1_9650*. In the mother's dataset this will be denoted by *covid1m_9650*, for fathers/partner this will be *covid1p_9650* and for the G1 generation it will be *covid1yp_9650*. Frequencies for all variables for each participant group are available in the data dictionary in the usual way¹⁹.

Text data and other potentially disclosive information will not be released until they have been coded appropriately. [Table 5](#) describes the data that is withheld at the time of first release. Text from Section C, question 1 and Section D, other are being thematically coded by qualitative researchers, the remainder will be coded as it is required/requested. Data will be incorporated back into both file sets as they become available.

Extended data

Open Science Framework: ALSPAC COVID-19 Q1. <https://doi.org/10.17605/OSF.IO/ZU8WY>¹¹

This project contains the following extended data:

- ALSPAC COVID Q1 FINAL.pdf (The final questionnaire; REDCap PDF)

Table 5. Data from questions that will not be released until coded.

Question number	Question text
Section A	
1	Please tell us the type of: <ul style="list-style-type: none"> • Organ transplant • Diabetes • Heart disease or heart problems • Other lung condition • Cancer • Condition affecting the brain and nerves • Psychiatric disorder Please can you tell us why your immune system is weakened?
3	For each medication: <ul style="list-style-type: none"> • Name of medication • Amount • How often • Reason for taking
5	Which country do you live in? If travelled outside home country: <ul style="list-style-type: none"> • Country and region/city/resort • Date arrived/left • Purpose of trip
6	What kind of other medical attention did you access? What other medication did you take?
9	Date first told had COVID-19
Section B	
1	When did you start self-isolating? Other reason for self-isolating
2	Other reason for changing normal day to day behaviour
Section C	
1	What other reason causing worry
Section D	
7	Other type of accommodation lived in
Other	Is there anything else you would like to tell us about how the pandemic has affected you?

- ALSPAC_COVID1_varlist.pdf (List of variable names and labels)
- ALSPAC_CovidQ1_data dictionary.pdf (Associated data dictionary including frequencies of all variables that are available)
- ALSPAC_symptom_algorithm_14052020.do (Stata script for obtaining probable COVID-19 cases using the Menni algorithm)

Data are available under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/) (CC-BY 4.0).

Consent

Completion of the questionnaire was optional and choosing to complete the questionnaire is considered informed consent for the questionnaire.

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. 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](#)²⁰.

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.

We are grateful to many people who contributed to the design, development, and testing of the questionnaire. Specifically, we would like to acknowledge contributions from:

Professor Deborah Lawlor (University of Bristol)

Professor George Davey Smith (University of Bristol)

Professor Peter Vickerman (University of Bristol)

Dr Robert Aldridge (University College London)

Professor David Porteous (University of Edinburgh, Gen Scot)

Honorary Professor Rosie McEachan (University of Bradford, Born in Bradford)

Professor Goncalo Abecasis (Regeneron)

Professor Naomi Allen (University of Oxford, UKBB)

Dr Mary de Silva (Wellcome)

Dr Bruna Galobardes (Wellcome)

Dr Claire Steves (Kings College London, TwinsUK)

Professor Stephan Collishaw (University of Cardiff)

Professor Martin Tobin (University of Leicester, EXCEED and HDRUK BREATHE)

References

- <https://www.parliament.uk/business/news/2020/march/update-on-coronavirus-24-march-2020/>.
- 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](#)
- 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](#)
- 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.** *Wellcome Open Res.* 2019; **4**: 51.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Lawlor DA, Lewcock M, Rena-Jones L, *et al.*: **The second generation of The Avon Longitudinal Study of Parents and Children (ALSPAC-G2): a cohort.** *Wellcome Open Res.* 2019; **4**: 36.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Harris PA, Taylor R, Thielke R, *et al.*: **Conde, Research electronic data capture (REDCap) - A metadata-driven methodology and workflow process for providing translational research informatics support.** *J Biomed Inform.* 2009; **42**(2): 377–81.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Raïche M, Hébert R, Dubois MF, *et al.*: **User guide for the PRISMA-7 questionnaire to identify elderly people with severe loss of autonomy.** In: *Integrated service delivery to ensure persons' functional autonomy*, ed. R. Hébert, A. Tourigny, and M. Gagnon, Quebec: Edisem. 147–65.
- Angold A, Costello EJ, Messer SC, *et al.*: **Development of a short questionnaire for use in epidemiological studies of depression in children and adolescents.** *Int J Methods Psychiatr Res.* 1995; **5**(4): 237–249.
[Reference Source](#)
- Spitzer RL, Kroenke K, Williams JBW, *et al.*: **A Brief Measure for Assessing Generalized Anxiety Disorder: The GAD-7.** *Arch Intern Med.* 2006; **166**(10): 1092–1097.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Tennant R, Hiller L, Fishwick R, *et al.*: **The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): development and UK validation.** *Health Qual Life Outcomes.* 2007; **5**: 63.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Timpson NJ, Northstone K, Wells N, *et al.*: **ALSPAC COVID-19 Q1.** 2020.
<http://www.doi.org/10.17605/OSF.IO/ZU8WY>
- Menni C, Valdes AM, Freidin MB, *et al.*: **Real-time tracking of self-reported symptoms to predict potential COVID-19.** *Nat Med.* 2020.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Drew D, Nguyen LH, Steves CJ, *et al.*: **COPE Consortium. Rapid implementation of mobile technology for real-time epidemiology of COVID-19.** *Science.* 2020; **368**(6497): 1362–1367.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- <https://covid.joinzoe.com/data>.
- <https://www.england.nhs.uk/statistics/statistical-work-areas/covid-19-daily-deaths/>.
- http://www.bristol.ac.uk/media-library/sites/alspac/documents/researchers/data-access/ALSPAC_Access_Policy.pdf.
- <https://proposals.epi.bristol.ac.uk/?q=proposalSummaries>.
- <https://proposals.epi.bristol.ac.uk/>.
- http://www.bristol.ac.uk/alspac/external/documents/ALSPAC_DataDictionary.zip.
- <http://www.bristol.ac.uk/alspac/researchers/research-ethics/>.

Open Peer Review

Current Peer Review Status:  

Version 2

Reviewer Report 12 February 2021

<https://doi.org/10.21956/wellcomeopenres.18097.r42302>

© 2021 Di Napoli A et al. 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.



Anteo Di Napoli 

National Institute for Health Migration and Poverty (INMP), Rome, Italy

Martina Ventura

National Institute for Health Migration and Poverty (INMP), Rome, Italy

This data note represents a useful instrument to evaluate the relationship between lifestyle, health and biological data, environmental features and COVID-19 infection incidence. Design of Avon Longitudinal Study of Pregnancy and Childhood (ALSPAC) is suitable to follow dynamic changes associated with SARS-COV2 pandemic. In this modified version of the manuscript, the authors improved the first version, responding to all comments from reviewers.

However, it is important to argue the potential selection bias cited by reviewers, considering the age selection of the ALSPAC study population. This study collects a cohort of pregnant people, their partners and their children from the 90s. As a result, the average age of the cohort is currently about sixty years old. In addition, the use of an emailed questionnaire excluded the older population without a valid email address, which represents the population with high risk of severe form of SARS-COV2 infection. This "age population" selection bias could affect the results of studies that evaluate the association between environment, health, lifestyle, biological data and COVID-19 incidence. The authors should then discuss this issue.

Lastly, it would be interesting if authors discussed their findings about the Self-Reported Symptoms of SARS-COV2 in the south-west of England comparing them with data from institutional sources, at last for the period from March to April 2020.

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. My expertise is mainly in the field of health inequalities, migrations and nephrology.**We confirm that we have read this submission and believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Reviewer Report 23 November 2020

<https://doi.org/10.21956/wellcomeopenres.18097.r41525>

© 2020 Crozier S et al. 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.

**Sarah Crozier** 

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

Janis Baird

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

The authors have addressed our previous comments well and we have no further comments to make.

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:** Sarah Crozier: Associate Professor of Statistical Epidemiology; Janis Baird, Professor of Public Health and Epidemiology, Honorary Consultant in Public Health**We confirm that we have read this submission and believe that we have an appropriate level**

of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 26 October 2020

<https://doi.org/10.21956/wellcomeopenres.17572.r40286>

© 2020 Crozier S et al. 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.

? **Sarah Crozier** 

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

Janis Baird

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

Northstone and colleagues provide a valuable Data Note describing an ALSPAC data-collection wave completed during the COVID-19 pandemic. The ALSPAC team are to be commended for implementing a fast and effective response to a dynamic global situation, allowing prospective data collection at a time of enormous significance. The context of this data resource within a longitudinal study provides great potential for future analyses, and the use of a questionnaire aligned with other longitudinal studies maximises the potential for cross-cohort harmonisation and collaboration.

The authors have been appropriately cautious to ensure confidentiality in the resource available for release and restriction to non-disclosive variables. Online resources such as the variable list, questionnaire and provision of Stata code will be of great benefit to future researchers. The dataset is large (nearly 7,000 participants). 54% of eligible participants responded to the questionnaire, although the greater proportion of females than males responding could lead to potential biases in some analyses.

In the Introduction to the paper, the UK 'lockdown' strategy is referenced; for an international audience, it might be helpful to describe what measures were employed by the UK government in March to May 2020 in order to illustrate the restrictions being experienced by cohort members.

In Table 2 BMI should be summarised using a median and interquartile range since this variable will not be normally distributed in this population.

Figure 2 requires more comment. The Menni algorithm suggests that 1-2% of participants had probable COVID-19 infection in October, November and December 2020. Whilst it is possible that there were cases of COVID-19 in the UK at this time, it is highly unlikely that the proportion of the population infected was 1-2%. The results, therefore, suggest that other respiratory viruses can mimic the symptoms of COVID-19 and bring into question the applicability of the Menni algorithm to data collected before the pandemic in the UK. The fact that the algorithm was developed on

data collected in the spring when COVID-19 levels were high but other respiratory virus levels are likely to have been lower, may be relevant.

Similarly, further discussion is required regarding the message of Table 4. The percentages suggest that participants who were ill with symptoms resembling COVID-19 in October were less likely to report having ever had suspected COVID-19 than those ill with symptoms resembling COVID-19 in more recent months. The authors could comment further on these trends. Also, Table 4 would perhaps be more logical, and would certainly follow better from Figure 2, if October 2019 was the first line in the table and April 2020 was the last.

The term 'excellent' (regarding the response rate) on page 7 might be moderated somewhat; the authors should certainly be commended for the success of their data collection within the pandemic situation, but the response rate might not be seen as 'excellent' in a broader context. The authors might also note as a limitation that the proportion of total participants responding was not large (i.e. 2,973 G1 responses were received from the original 14,701 children alive at 1 year (20.2%)) leading to potential implications for the representativeness of the responding sample.

Minor comments:

- In the Abstract the G0 and G1 notation needs to be explained.
- The following sentence in the Abstract is potentially confusing: 'Using algorithmically defined cases, we estimate that the predicted proportion of COVID-19 cases fell between 1.03% - 4.19% depending on the timing of measurement during the period of reporting.'. The term 'fell between' potentially implies there has been a fall in cases over time which is not what is described. Also, the period of reporting could be defined fully in this sentence as October 2019 to March 2020.
- Table 1 should be referenced earlier in the paragraph beginning 'Overall, female participants...'
- Page 4, first paragraph under 'Key results': the last sentence here is unnecessary since this has already been stated.
- Page 4, third bullet point under 'C': 'Disorder' rather than "Disortder"
- Page 8, line 7: Add close bracket.
- Page 8: In the second point describing availability of data, letters (rather than digits) have been added to denote the source of the variable.

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: Sarah Crozier: Associate Professor of Statistical Epidemiology; Janis Baird, Professor of Public Health and Epidemiology, Honorary Consultant in Public Health

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

Author Response 10 Nov 2020

Kate Northstone, University of Bristol, Bristol, UK

Thank you for your positive review of our data note. We respond to your queries according to paragraph of your review.

Paragraph 3: Thankyou for this helpful suggestion. We have provided further detail as suggested on the UK lockdown.

Paragraph 4: We have replaced mean(sd) with median(IQR) for BMI as suggested

Paragraphs 5 and 6: Thankyou for these insightful comments. We agree with the suggestion that the algorithm may not be applicable for data collected pre-2020. We considered removing this data all together but believe it is of interest. However, we have added to both the results and discussion sections as suggested by the reviewers. We have also changed the order of Table 4 as recommended.

Paragraph 7: Fair point! We have amended this and also acknowledge the inevitable selection bias introduced by our online-only strategy.

Minor comments: We have taken into account all the minor comments you suggest and amended the text accordingly. Please note we had to make other minor changes to the abstract to be within the word limit.

Competing Interests: None