DATA NOTE

Data note for linking the Avon Longitudinal Study of Parents and Children (ALSPAC) with the Public Health England (PHE) COVID-19 dataset [version 1; peer review: awaiting peer review]

Mark Mummé, Andrew Boyd, John Macleod

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

Abstract
This data note describes the test results for infection by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2 or COVID-19) of the index participants of the Avon Longitudinal Study of Parents And Children birth cohort study (ALSPAC, also known as the Children of the 90s). The records were collated, processed, record-linked and then extracted by Public Health England (PHE) to ALSPAC. ALSPAC provided PHE with the NHS numbers of 12,774 of the index cohort ‘children’, who were aged between 27 and 30 years during this period, and 1,033 of the index children’s ‘parents’ (only data on a small subset of the mothers was permissible for this dataset during this period). PHE conducted the linkage using deterministic methods and returned periodic data extracts of all the COVID-19 test results they could match to those participants. ALSPAC obtained both the positive and negative COVID-19 test results from PHE from the time when testing was first available in February 2020 until late August 2021, just before PHE was dissolved. ALSPAC is uniquely placed to provide a longitudinal dataset of the health, education and other factors on the cohort participants before and during the COVID-19 pandemic. This provides the opportunity to place context to the COVID-19 test data provided by PHE (the timings, patterns, and results). The result is to provide a resource for current and future research into the COVID-19 pandemic.

Keywords
Public Health England, COVID-19, Coronavirus, record linkage, longitudinal study, cohort study, ALSPAC
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: Mark Mummé (mark.mumme@bristol.ac.uk)

Author roles: Mummé M: Data Curation, Formal Analysis, Project Administration, Writing – Original Draft Preparation, Writing – Review & Editing; Boyd A: Funding Acquisition, Resources, Writing – Review & Editing; Macleod J: Funding Acquisition, Resources, Supervision, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: The UK Medical Research Council and Wellcome [217065, https://doi.org/10.35802/217065] and the University of Bristol provide core support for ALSPAC. This publication is the work of the authors and MM will serve as guarantor for the contents of this paper. 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 and Macleod J. Data note for linking the Avon Longitudinal Study of Parents and Children (ALSPAC) with the Public Health England (PHE) COVID-19 dataset [version 1; peer review: awaiting peer review] Wellcome Open Research 2022, 7:268 https://doi.org/10.12688/wellcomeopenres.18091.1

First published: 26 Oct 2022, 7:268 https://doi.org/10.12688/wellcomeopenres.18091.1
Introduction
The Avon Longitudinal Study of Parents and Children (ALSPAC) is a multigenerational birth cohort study established with the stated aim of compiling information on participants’ health and social exposures and subsequent outcomes across the course of their life. Within this there is a need to characterise disease exposures, particularly relevant during the coronavirus disease 2019 (COVID-19) pandemic. Information about the cohort participants COVID-19 experiences was collected prospectively, at key timepoints, using self-reported questionnaires and efforts were made to collect samples. To complement the self-report data, ALSPAC was permitted access to extracts of centrally collected data from Public Health England. This Data Note describes data provided via record linkage from the PHE pillar 1 and pillar 2 COVID-19 test records. These data were available from February 2020 until August 2021 when PHE was disbanded.

Materials and methods
The ALSPAC Sample
The Avon Longitudinal Study of Parents and Children (ALSPAC) is a prospective, population-based study. The initial recruitment took place from September 1990 to December 1992 inclusive, and has been described in detail in previous papers.

Figure 1. The ALSPAC Eligible Study Area within the UK: illustrating the NHS District Health Authorities (DHAs) used to define: the ALSPAC catchment area; the historical county of Avon; and the four authorities formed following the breakup of Avon. Contains Ordnance Survey, Office of National Statistics and National Records Scotland data © Crown Copyright/database right 2014.

\[1\] Northstone K, Smith D, Bowring C et al. The Avon Longitudinal Study of Parents and Children - A resource for COVID-19 research: Questionnaire data capture May-July 2020 [version 2; peer review: 2 approved]. *Wellcome Open Res* 2020, 5:210 (https://doi.org/10.12688/wellcomeopenres.16225.2)

\[2\] Major-Smith D, Matthews S, Breeze T et al. The Avon Longitudinal Study of Parents and Children - A resource for COVID-19 research: Antibody testing results, April - June 2021 [version 2; peer review: 2 approved]. *Wellcome Open Res* 2022, 6:283 (https://doi.org/10.12688/wellcomeopenres.17294.2)

\[3\] 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.

\[4\] 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.
together with the current county boundaries for context. Within ALSPAC the original pregnant women and their partners are referred to as Generation Zero (G0) and the index children as Generation One (G1). This provides a baseline sample of 14,901 G1 participants who were alive at 1 year of age\(^7\) and many of their parents and carers also still involved in the study. Starting in 2011, the participants were sent ‘fair processing’ materials describing ALSPAC’s intended use of their health and administrative records and were given clear means to consent or object via a written form. Data were not extracted for participants who have objected, or who were not sent fair processing materials.

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.

Public Health England and COVID-19 data
Public Health England (PHE) was an executive agency of the Department of Health and Social Care (DHSC) in England which began operating on 1 April 2013 to protect and improve health and wellbeing and reduce health inequalities. Its formation came as a result of the reorganisation of the National Health Service (NHS) in England outlined in the Health and Social Care Act 2012. It took on the role of the Health Protection Agency, the National Treatment Agency for Substance Misuse and a number of other health bodies. It was an executive agency of the Department of Health and Social Care, and a distinct delivery organisation with operational autonomy.

The DHSC set PHE’s priorities annually, and for 2019/20 this included an “integrated surveillance system” and “investigation and management of outbreaks of infectious diseases”. PHE carried out contact tracing in the early stages of the COVID-19 pandemic and began providing COVID-19 epidemiology surveillance information from late April 2020, combining community, primary care, secondary care, virology and mortality surveillance data to support national and regional planning in relation to the pandemic.

Tests in the UK are carried out through a number of different routes\(^6\):

- pillar 1: swab testing in Public Health England (PHE) labs and NHS hospitals for those referred with a clinical need, and for health and care workers
- pillar 2: swab testing for the wider population, as set out in government guidance
- pillar 3: serology testing to show if people have antibodies from having had COVID-19
- pillar 4: blood and swab testing for national surveillance supported by PHE, the Office for National Statistics (ONS), and research, academic, and scientific partners to learn more about the prevalence and spread of the virus and for other testing research purposes, such as the accuracy and ease of use of home testing

PHE provided ALSPAC with data from Pillar 1 and Pillar 2 testing.

On 29 March 2021, the UK Government announced that PHE would be disbanded and that its public health functions would be transferred, in proposals to reforms public health structures. From 1 October 2021, PHE’s health protection functions were formally transferred into the UK Health Security Agency (UKHSA), while its health improvement functions were transferred to the Office for Health Improvement and Disparities (as part of DHSC), NHS England and NHS Digital.

The ALSPAC-PHE COVID-19 dataset
The whole ALSPAC cohort was initially run through a filter to remove any participants of the index children’s generation (G1) who had either opted out of sharing any health data or for whom we did not have permission under section 251\(^8\) to request any health data and further to remove any participating parents or carers (G0) who had not given explicit permission to use their health data. This was initially performed by ALSPAC and then PHE additionally applied the National Opt-Out to those participants supported under section 251. ALSPAC did not have any section 251 support for the parents or carers at the time that this dataset was generated. Figure 2 shows how the sub-cohort of 13,807 ALSPAC participants was constructed to be sent to PHE for data matching.

Record linkage methodology
A unique Link ID was generated, and the identifiers were sent to PHE through a secure file transfer, after being encrypted to AES-256 standard, and a 20-character password was provided through an alternative channel. Two files were sent to PHE containing the NHS numbers and dates of birth of the ALSPAC cohort for whom data linkage was allowed. The first file contained the NHS numbers of the consented participants and the second file contained details of the participants supported under section 251. This allowed PHE to apply the National Opt-Out\(^9\) only to the second (section 251) file, and to allow the full flow of data for the consented file. The record linkage was performed using the NHS numbers, with the dates of birth as used as a confirmation cross-check.

---

\(^6\) Northstone K, Lewcock M, Groom A, Boyd A, Macleod J, Timpson NJ, Wells N. The Avon Longitudinal Study of Parents and Children (ALSPAC): an updated on the enrolled sample of index children in 2019. Wellcome Open research 2019; 4:51 (https://doi.org/10.12688/wellcomeopenres.15132.1)

\(^7\) For further details - https://digital.nhs.uk/services/national-data-opt-out

\(^8\) This refers to section 251 of the National Health Service Act 2006 and its current Regulations, the Health Service (Control of Patient Information) Regulations 2002.

\(^9\) For further details - https://www.gov.uk/government/publications/coronavirus-covid-19-testing-data-methodology/covid-19-testing-data-methodology-note
The PHE data was returned to ALSPAC through the secure file transfer system EGRESS. These files were also encrypted, and password protected. They were provided as separate excel files. There were initially six files provided although these were quickly consolidated into two:

- **COVID Positive** – initially as separate files for section 251 G1 participants, consented G1 participants and consented G0 mums.
- **COVID Negative** – initially as separate files for section 251 G1 participants, consented G1 participants and consented G0 mums.

The first dataset was provided in October 2020 and included 3,851 test results. It is important to note the effect of the National Opt-Out here. There were 2,003 test results from 1,189 of the 7,573 participants whose data provision was supported by section 251. It is known that there were 35 individuals who had at least one test record but whose data was not permitted to be shared. It is not known how many tests this represented or how many, if any, of these tests were positive.

There were duplicate samples with test records which had the same specimen date but more than one report date. These
were consolidated into a single test record and the latest report date was retained.

A refresh and update of the data was provided in November 2020 which included 5,394 test results. There were an additional 1,563 test results in the November 2020 update however it was observed that 20 test results from the October 2020 dataset were dropped from the November 2020 dataset. PHE explained that the dataset was constructed so that once an individual tested positive to COVID-19 then all their earlier negative results were dropped from the dataset. PHE advised they could not provide any information on how many negative test results were dropped. For this reason, the final dataset was constructed by consolidating all the individual datasets. This maximised the granularity of the negative tests and the positive test data was unchanged.

PHE provided a total of 43,438 test results. There were 1,451 (3.34%) positive test results and 41,987 negative tests. It is unknown how many negative test results were dropped throughout the reporting period. The further descriptions of the ALSPAC-PHE COVID-19 dataset are regarding the final combined dataset. It is important to be aware that each time a data extract was provided the National Opt-Out was applied to the group of participants supported by section 251, although the number of individuals was only provided for the first data extract. This may explain why there are 15 positive cases missing from the September 2021 data extract when compared to the combined dataset.

Figure 3, below, shows how the number of COVID-19 tests performed on the ALSPAC participants and reported by PHE changed over time. It is important to note that once a participant had a positive test then any negative tests results for that participant since the previous update sent to ALSPAC were lost. The tests reported in Figure 3 are aggregated weekly.

Figure 4, below, shows the COVID-19 cases of the ALSPAC participants over time as indicated by a positive test result provided by PHE. The graph shows the individual regions of the UK aggregated into weekly totals. The regional sub-totals are provided below in Table 2. The South-West region accounts for 84.2% of cases. The UK national cases, weekly, are shown in Figure 5, below, for reference.

Table 3, below, shows the numbers and proportions of the two generations of ALSPAC participants that had at least one test result recorded by PHE and returned to ALSPAC.

Table 4, below, shows the numbers and proportions of the ALSPAC participants by sex that had at least one test result recorded by PHE and returned to ALSPAC.

Table 5, below, shows the numbers and proportions of the ALSPAC participants by consent status that had at least one test result recorded by PHE and returned to ALSPAC. The number of participants whose data was not provided under section 251 because they had exercised their right to opt-out of their confidential patient information being used for research and planning was only provided by PHE from the first data extract, October 2020. PHE advised that 35 individuals with test records were not made available to ALSPAC under the National Opt-Out which was 3% of participants. It is unknown how many tests this corresponded to. Figures from NHS-Digital confirm the national rate of patients choosing to opt-out increased from 2.67% in September 2020 to 5.25% in August 2021.

Figure 3. All ALSPAC test results reported by PHE by week.

https://digital.nhs.uk/data-and-information/publications/statistical/national-data-opt-out/august-2021
PHE provided a total of 43,438 test results. There were 41,987 negative tests and 1,451 (3.34%) positive test results. Table 6, below, shows the source of the positive test results as

**Table 6.** Positive ALSPAC test results returned by PHE by week.

| Date           | Tests | Positive | Negative | Added | Dropped |
|----------------|-------|----------|----------|-------|---------|
| 03 October 2020| 3,851 | 67       | 3,784    | 3,851 | -       |
| 02 November 2020| 5,394 | 199      | 5,195    | 1,563 | 20      |
| 18 December 2020| 6,998 | 330      | 6,668    | 1,784 | 200     |
| 06 February 2021| 11,012| 687      | 10,325   | 4,373 | 559     |
| 07 March 2021  | 14,383| 752      | 13,631   | 3,537 | 725     |
| 15 April 2021  | 19,083| 783      | 18,300   | 4,742 | 767     |
| 28 May 2021    | 26,136| 785      | 25,531   | 7,277 | 811     |
| 01 September 2021| 41,113| 1,436    | 39,677   | 16,311| 2,325   |
| **Final Combined Dataset** | **43,438** | **1,451** | **41,987** | -     | -       |

**Table 2.** Region of the UK where the ALSPAC positive test was taken.

| UK Region            | Cases |
|----------------------|-------|
| East Midlands        | 19    |
| East of England      | 10    |
| London               | 96    |
| North-East           | 3     |
| North-West           | 19    |

**Table 1.** Summary details of test result data returned from Public Health England (PHE) at each dataset.

| Date           | Tests | Positive | Negative | Added | Dropped |
|----------------|-------|----------|----------|-------|---------|
| 03 October 2020| 3,851 | 67       | 3,784    | 3,851 | -       |
| 02 November 2020| 5,394 | 199      | 5,195    | 1,563 | 20      |
| 18 December 2020| 6,998 | 330      | 6,668    | 1,784 | 200     |
| 06 February 2021| 11,012| 687      | 10,325   | 4,373 | 559     |
| 07 March 2021  | 14,383| 752      | 13,631   | 3,537 | 725     |
| 15 April 2021  | 19,083| 783      | 18,300   | 4,742 | 767     |
| 28 May 2021    | 26,136| 785      | 25,531   | 7,277 | 811     |
| 01 September 2021| 41,113| 1,436    | 39,677   | 16,311| 2,325   |
| **Final Combined Dataset** | **43,438** | **1,451** | **41,987** | -     | -       |

**Figure 4.** Positive ALSPAC test results returned by PHE by week.
**Figure 5.** UK government figures for COVID-19 cases nationally.

**Table 3.** Test results returned by Public Health England (PHE), shown by generation.

|                      | Individuals with a test reported | Number of individuals sent to PHE | proportion with test result |
|----------------------|----------------------------------|-----------------------------------|----------------------------|
| Generation 0 (Parents and carers) | 660                              | 1,033                             | 64%                        |
| Generation 1 (Index children)      | 7,477                             | 12,774                            | 59%                        |

**Table 4.** Test results returned by Public Health England (PHE), shown by sex.

|                      | Male     | Female   | Total    |
|----------------------|----------|----------|----------|
| number of individuals in cohort | 6,451    | 7,356    | 13,807   |
| total tests reported  | 11,936   | 31,502   | 43,438   |
| individuals with a test reported | 3,256    | 4,881    | 8,137    |
| tests per person in cohort      | 1.85     | 4.28     | 3.15     |
| tests per person in sample     | 3.67     | 6.45     | 5.34     |

**Table 5.** Test results returned by Public Health England (PHE), shown by consent status within the study.

|                      | Individuals with a test reported | Number of individuals sent to PHE | proportion with test result |
|----------------------|----------------------------------|-----------------------------------|----------------------------|
| Consent              | 4,096                            | 6,234                             | 66%                        |
| Section 251          | 4,041                            | 7,573                             | 53%                        |
either Pillar 1 (swab testing in PHE labs and NHS hospitals) or Pillar 2 (swab testing for the wider population).

Table 7, below, shows the type of test used to determine a positive test result for infection with COVID-19.

The delay in days between the date given as when the specimen was taken and the date given as when the result was reported is shown in Table 8, below.

Other variables were provided by PHE for 1,451 positive COVID-19 test results, but these were either very poorly populated, completed with a value indicating the variable status was generally unknown, or is unavailable for research due to a potential to be disclosive of ALSPAC participants.

---

### Table 6. Source of positive test results returned by Public Health England (PHE).

| Source     | Count |
|------------|-------|
| Pillar 1   | 99    |
| Pillar 2   | 1352  |

### Table 7. Type of test used to determine positive COVID-19 test result. PCR=Polymerase chain reaction; LFT=lateral flow test.

| Test Type          | Count |
|--------------------|-------|
| PCR only           | 1,288 |
| LFT only           | 31    |
| LFT with PCR       | 128   |

### Table 8. Number of days between date given as date of test and date given as date of test result reported.

| Days delay | Count | Percentage |
|-----------|-------|------------|
| -ve (probable in error) | 25 | 0.1% |
| 0         | 25,240 | 58.1% |
| 1         | 10,664 | 24.5% |
| 2         | 4,915  | 11.3% |
| 3         | 1,658  | 3.8% |
| 4         | 597    | 1.4% |
| 5         | 180    | 0.4% |
| 6         | 59     | 0.1% |
| 7         | 21     | 0.0% |
| 8         | 12     | 0.0% |
| 9         | 5      | 0.0% |
| 10        | 11     | 0.0% |
| 11 - 20   | 15     | 0.0% |
| 21 +      | 36     | 0.1% |

### Table 9. Asymptomatic indicator from when an ALSPAC participant presented for a COVID-19 test which was positive.

| Asymptomatic Indicator | Count | Percentage |
|------------------------|-------|------------|
| Yes                    | 377   | 26%        |
| No                     | 953   | 66%        |
| Unknown                | 121   | 8%         |

### Table 10. Variables provided by Public Health England (PHE) in the COVID-19 dataset which are of little or no value for research due to incompleteness.

| Variable                                               | Count |
|--------------------------------------------------------|-------|
| 1 Ethnicity                                            |       |
| 2 Age in years                                         |       |
| 3 Age in months                                        |       |
| 4 Outbreak Indicator, Outbreak Indicator1              |       |
| 5 Patient Immunocompromised Indicator                  |       |
| 6 Hospital Acquired Indicator                          |       |
| 7 Travel Abroad Indicator                              |       |
| 8 Invasive Indicator                                   |       |
| 9 Meningitis Indicator                                 |       |
| 10 Bacteraemia Indicator                               |       |
| 11 Patient Death Indicator                             |       |
| 12 Reporting Lab Used As Source Lab                    |       |
| 13 Symptom Onset Date                                  |       |
| 14 Postcode Source                                     |       |
| 15 Reporting Lab                                       |       |
| 16 Source Lab                                          |       |
| 17 Reporting Lab Region                                |       |
| 18 County Description                                 |       |
| 19 UK Region Description                               |       |
| 20 Health Protection Team                              |       |
| 21 Local Authority Name                                |       |
The ALSPAC-PHE COVID-19 testing data can be linked with other ALSPAC self-reported data, assayed biological samples, abstracted clinical notes and other longitudinal health and education records. This resource can be used for COVID-19 and other research and has been used to facilitate COVID research case selection.

**Ethics**

Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees. A comprehensive list of research ethics committee approval references is available to download at: [http://www.bristol.ac.uk/alspac/researchers/research-ethics/](http://www.bristol.ac.uk/alspac/researchers/research-ethics/).

The initial approvals for the University of Bristol project were obtained as:

IRAS Number: 54370

NHS REC Reference: 10/H1010/70

**Consent**

Permissions 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 and NHS Research Ethics Committee’s 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.

**Data availability**

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

i. Please read the ALSPAC access policy 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, which lists all research projects that have been approved since April 2011.

iii. Please submit your research proposal 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 contact the ALSPAC Data Linkage Team ([alspac-linkage@bristol.ac.uk](mailto:alspac-linkage@bristol.ac.uk)).

**Acknowledgements**

We are extremely grateful to all the families who take part in the ALSPAC 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 acknowledge the work of and are grateful to all the staff at Public Health England who assisted in the provision of this data to ALSPAC.