





RESEARCH ARTICLE

REVISED **Bias from questionnaire invitation and response in COVID-19 research: an example using ALSPAC [version 2; peer review: 2 approved]**

Alba Fernández-Sanlés ^{1,2*}, Daniel Smith ^{1,2*}, Gemma L Clayton ^{1,2},
Kate Northstone ², Alice R Carter ^{1,2}, Louise AC Millard ^{1,2},
Maria Carolina Borges^{1,2}, Nicholas John Timpson ^{1,2}, Kate Tilling^{1,2},
Gareth J Griffith ^{1,2}, Deborah A. Lawlor ¹⁻³

¹MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK

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

³Bristol National Institute of Health Research (NIHR) Biomedical Research Centre, Bristol, UK

* Equal contributors

v2 **First published:** 19 Jul 2021, 6:184
<https://doi.org/10.12688/wellcomeopenres.17041.1>

Latest published: 08 Jul 2022, 6:184
<https://doi.org/10.12688/wellcomeopenres.17041.2>

Abstract





Background: Longitudinal studies are crucial for identifying potential risk factors for infection with, and consequences of, COVID-19, but relationships can be biased if they are associated with invitation and response to data collection. We describe factors relating to questionnaire invitation and response in COVID-19 questionnaire data collection in a multigenerational birth cohort (the Avon Longitudinal Study of Parents and Children, ALSPAC).



Methods: We analysed online questionnaires completed between the beginning of the pandemic and easing of the first UK lockdown by participants with valid email addresses who had not actively disengaged from the study. We assessed associations of pre-pandemic sociodemographic, behavioural, anthropometric and health-related factors with: i) being sent a questionnaire; ii) returning a questionnaire; and iii) item response (for specific questions). Analyses were conducted in three cohorts: the index children born in the early 1990s (now young adults; 41 variables assessed), their mothers (35 variables) and the mothers' partners (27 variables).

Results: Of 14,849 young adults, 41% were sent a questionnaire, of whom 57% returned one. Item response was >95%. In this cohort, 78% of factors were associated with being sent a questionnaire, 56% with returning one, and, as an example of item response, 20% with keyworker status response. For instance, children from mothers educated to degree-level had greater odds of being sent a questionnaire (OR=5.59; 95% CI=4.87-6.41), returning one (OR=1.60;

Open Peer Review

Approval Status  

	1	2
version 2 (revision) 08 Jul 2022		 view
version 1 19 Jul 2021	 view	  view

1. **Claire Infante-Rivard** , McGill University, Montréal, Canada
2. **Dipesh P Gopal** , Barts and The London School of Medicine and Dentistry, London, UK

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

95% CI=1.31-1.95), and responding to items (e.g., keyworker status OR=1.65; 95% CI=0.88-3.04), relative to children from mothers with fewer qualifications. Invitation and response rates and associations were similar in all cohorts.

Conclusions: These results highlight the importance of considering potential biases due to non-response when using longitudinal studies in COVID-19 research and interpreting results. We recommend researchers report response rates and factors associated with invitation and response in all COVID-19 observational research studies, which can inform sensitivity analyses.

Keywords

COVID-19, COVID-19 risk factors, Selection Bias, Longitudinal Study, ALSPAC, Missing Data, Questionnaire Invitation, Questionnaire Response.



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

Corresponding authors: Alba Fernández-Sanlés (alba.fernandez-sanles@bristol.ac.uk), Daniel Smith (dan.smith@bristol.ac.uk)

Author roles: **Fernández-Sanlés A:** Conceptualization, Data Curation, Formal Analysis, Writing – Original Draft Preparation, Writing – Review & Editing; **Smith D:** Conceptualization, Data Curation, Formal Analysis, Writing – Original Draft Preparation, Writing – Review & Editing; **Clayton GL:** Conceptualization, Data Curation, Writing – Review & Editing; **Northstone K:** Conceptualization, Data Curation, Funding Acquisition, Methodology, Project Administration, Supervision, Writing – Review & Editing; **Carter AR:** Writing – Review & Editing; **Millard LA:** Writing – Review & Editing; **Borges MC:** Conceptualization, Supervision, Writing – Review & Editing; **Timpson NJ:** Conceptualization, Funding Acquisition, Supervision, Writing – Review & Editing; **Tilling K:** Conceptualization, Funding Acquisition, Supervision, Writing – Review & Editing; **Griffith GJ:** Conceptualization, Methodology, Writing – Original Draft Preparation, Writing – Review & Editing; **Lawlor DA:** Conceptualization, Funding Acquisition, Supervision, Writing – Review & Editing

Competing interests: DAL receives support from Roche Diagnostics and Medtronic Ltd for research unrelated to this paper.

Grant information: This work was supported by the Wellcome Trust's "Longitudinal Population Study Covid-19 Steering Group and Secretariat" (221574/Z/20/Z, a Strategic Support Science Grant); the Wellcome Trust and Medical Research Council (MRC) (217065/Z/19/Z); the Elizabeth Blackwell Institute for Research at the University of Bristol for the Questionnaire 'COVID1'; the Wellcome Trust and MRC (217065/Z/19/Z) and the University of Bristol Faculty Director's Discretionary Fund for the Questionnaire 'COVID2'. A comprehensive list of grant 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 Wellcome Trust and the MRC (202802/Z/16/Z and 217065/Z/19/Z, to NJT); the Bristol BHF Accelerator Award (AA/18/7/34219, to AF-S and supporting ARC and DAL); the University of Bristol and Medical Research Council (MC_UU_00011/1, MC_UU_00011/3, MC_UU_00011/6, supporting AF-S, DS, GLC, ARC, LACM, NJT, KT, MCB, GJG and DAL); the British Heart Foundation (BHF)-National Institute of Health Research (NIHR) COVIDITY flagship project, with funds from the Bristol BHF Accelerator Award (AA/18/7/34219); the University of Bristol (Vice-Chancellor Fellowship to MCB); the Economic and Social Research Council (ES/T009101/1, postdoctoral fellowship to GJG); the University of Bristol NIHR Biomedical Research Centre and the CRUK Integrative Cancer Epidemiology Programme (C18281/A29019, to NJT); the MRC (CH/F/20/90003, to DAL); the NIHR (NF-0616-10102, to DAL); 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 Fernández-Sanlés A *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: Fernández-Sanlés A, Smith D, Clayton GL *et al.* **Bias from questionnaire invitation and response in COVID-19 research: an example using ALSPAC [version 2; peer review: 2 approved]** Wellcome Open Research 2022, 6:184 <https://doi.org/10.12688/wellcomeopenres.17041.2>

First published: 19 Jul 2021, 6:184 <https://doi.org/10.12688/wellcomeopenres.17041.1>

REVISED Amendments from Version 1

The revised version of our manuscript has taken into consideration the helpful and constructive comments of the reviewers (detailed in our responses to the reviewers). The main update that we have made to this version has been to clearly define our target population and ensure that the links between the study sample and target population are more apparent throughout the manuscript.

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

Introduction

Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) has infected more than 181 million individuals worldwide and is responsible for over 3.9 million global deaths to date (1st July 2021, WHO). Due to their wealth of pre-pandemic data, prospective longitudinal studies are making important contributions to understanding the mechanisms of both infection and disease (COVID-19), and the impact of the pandemic and its management on future health¹⁻⁶. However, sociodemographic, behavioural and health-related factors may shape not only who gets infected, progression to COVID-19 and disease severity, but also health-seeking behaviour, who gets tested, and their wider response to the pandemic (e.g., mental health impact). Furthermore, these factors may be associated with, or even influence, who is assessed or sent a questionnaire, and who responds to questionnaires or items they contain. This can result in selection bias⁷⁻¹¹, which may be exacerbated by the ways in which data collection has changed during the pandemic, for example, with the increased use of online questionnaires^{12,13}. This selection bias can lead to incorrect (or biased) estimations of the effect of a risk factor on an outcome, in this case COVID-19 and its related impacts. Importantly, having an available and rich collection of pre-pandemic data in existing longitudinal cohorts can be useful in efforts to explore potential selection pressures that lead to bias¹⁴⁻¹⁷.

The aim of this study was to describe questionnaire invitation and response rates and to explore factors associated with (i) being sent a COVID-19 questionnaire (i.e., participants who were invited to complete a questionnaire because they had not withdrawn from the study, agreed to participate in questionnaires and had a valid email address); (ii) returning a COVID-19 questionnaire; and (iii) item response (for six key variables: self-reported COVID-19 status, predicted COVID-19 cases based on symptoms¹⁸, three mental health outcomes [depression, anxiety and well-being] and keyworker status), in the Avon Longitudinal Study of Parents And Children (ALSPAC), a multi-generational longitudinal study based in the South West of England established in the early 1990s. Throughout this paper we use ‘being sent a questionnaire’ and ‘questionnaire invitation’ synonymously, and include both returning a questionnaire and item completion as ‘response’. We focused on the three adult cohorts over two generations: the index participants born in the early 1990s, Generation-1 (G1); their mothers, Generation-0 (G0) mothers; and the mothers’ partners (G0 partners). For each outcome, we examined multiple candidate predictors of invitation and response encompassing a range of

sociodemographic, behaviour, anthropometric and health-related factors. We focused our analyses on the first two online COVID-19 questionnaires, the first completed between 9th April and 15th May 2020, and the second between 26th May and 5th July 2020^{19,20}.

Methods**Study design**

ALSPAC is a three-generation birth cohort that started recruiting pregnant women resident in the former county of Avon (centred around the city of Bristol, UK), with delivery dates between April 1991 and December 1992. A total of 14,541 pregnancies were initially enrolled (14,676 fetuses), resulting in 13,988 children alive at one year of age. Those women (G0 mothers), their partners (G0 partners) and their index children (G1) have been followed with regular assessments since this time. Since the oldest children were approximately 7 years of age, the study has recruited 913 additional G1 children who did not join originally, but were part of the original target population based on date and location of birth. Hereafter we will refer to the participants where the G1 index child was alive at one year of age and who did not withdraw consent for their data to be used as the “whole cohort” (14,849 G1 children; 14,282 G0 mothers; 14,275 G0 partners). The target population therefore comprises pregnancies leading to children born in the early 1990s in the former county of Avon, in addition to their mothers and the mothers’ partners. This initial enrolled sample included approximately 75% of the target population and was broadly representative of the wider Avon population (albeit somewhat biased towards mothers who were married and from higher socioeconomic backgrounds, and biased away from ethnic minorities); further details can be found in the published cohort profiles^{14,21,22}. The study website contains details of all the data available through a fully searchable [data dictionary and variable search tool](#).

Since the start of the pandemic, participants have been sent four online COVID-19 questionnaires to assess diagnoses, symptoms, and behavioural and environmental factors related to COVID-19, and the impact of the pandemic on health^{19,20,23,24}. Questionnaires were sent to all participants who had not withdrawn from the study or declined to participate in questionnaires, and had a valid email address in the ALSPAC administrative records. Some participants would have provided updated email address information or re-engaged with the study between the questionnaires, while other participants may have withdrawn from the study during this time. Therefore, the number of participants invited to these questionnaires may differ. Data for the COVID-19 questionnaires were collected and managed using REDCap electronic data capture tools hosted at the University of Bristol²⁵. In this study we focused on the first two questionnaires^{19,20}. Analyses were conducted for each COVID-19 questionnaire separately (COVIDQ1 and COVIDQ2) and combined (data from both COVIDQ1 and COVIDQ2, and data from either COVIDQ1 or COVIDQ2).

COVID-19 questionnaire invitation, return and completion

We described invitation and response rates, and examined associations with the following three outcomes (Table 1; Figure 1

Table 1. Description of the outcome variables.

Outcome	Description	Sample	Comparison	Variable name (COVID1; COVID2)	Notes
Being sent a COVID-19 questionnaire	Whether the participant was sent a COVID-19 questionnaire	Whole cohort	Sent vs Not sent	covid1_0001; covid2_0001	-
Returning a COVID-19 questionnaire	Whether the participant returned a COVID-19 questionnaire	Participants who were sent a COVID-19 questionnaire	Returned vs Not returned	covid1_0002; covid2_0002	-
Having self-reported COVID-19 data	Whether the participant answered the question on self-reported COVID-19 diagnosis	Participants who returned a COVID-19 questionnaire	Responded vs Missing	covid1_2580; covid2_1060	-
Having predicted COVID-19 data	Whether the participant has data for predicted COVID-19 case, estimated using the Menni algorithm based on responses to a checklist of COVID-19 related symptoms	Participants who returned a COVID-19 questionnaire	Responded vs Missing	covid1_2605; covid2_1030	For participants with missing data for specific symptoms, unless all questions were blank it was assumed that the participant did not experience that symptom.
Having SMFQ data	Whether the participant has SMFQ total score data (depression)	Participants who returned a COVID-19 questionnaire	Responded vs Missing	covid1_4065; covid2_4065	Total score comprised of 13 SMFQ items. Participants missing ≥ 1 SMFQ item were coded as missing.
Having GAD-7 data	Whether the participant has GAD-7 total score data (anxiety)	Participants who returned a COVID-19 questionnaire	Responded vs Missing	covid1_4080; covid2_4080	Total score comprised of 7 GAD-7 items. Participants missing ≥ 1 GAD-7 item were coded as missing.
Having WEMWBS data	Whether the participant has WEMWBS total score data (well-being)	Participants who returned a COVID-19 questionnaire	Responded vs Missing	covid1_4120; covid2_4120	Total score comprised of 14 WEMWBS items. Participants missing ≥ 1 WEMWBS item were coded as missing.
Having keyworker status data	Whether the participant answered the question about being a keyworker	Participants who returned a COVID-19 questionnaire	Responded vs Missing	covid1_5045; covid2_6030	-

SMFQ, Short Mood and Feelings Questionnaire; GAD-7, Generalised Anxiety Disorder seven-item Assessment; WEMWBS, Warwick-Eдинburgh Mental Wellbeing Scale.

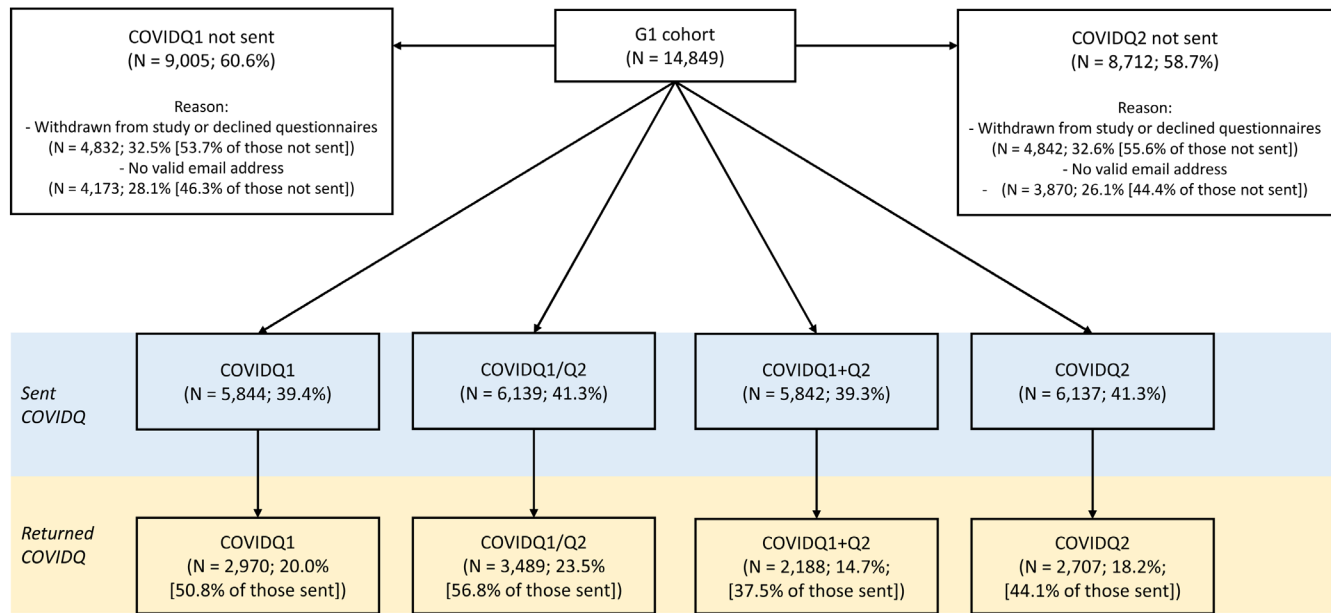


Figure 1. Flowchart of COVID-19 questionnaire invitation and response in the G1 cohort. COVIDQ1 and COVIDQ2 refer to the first and second COVID-19 questionnaires, respectively, COVIDQ1/Q2 refers to being sent/returning either COVIDQ1 or COVIDQ2, and COVIDQ1+Q2 refers to being sent/returning both COVIDQ1 and COVIDQ2. Details of reasons why participants were not sent a COVIDQ1 or COVIDQ2 questionnaire are also given.

- G1 cohort, [Figure 2](#) - G0 mothers cohort, [Extended data](#)²⁶: Figure S1 - G0 partners cohort):

1. Outcome: Being sent a COVID-19 questionnaire

Sample: All enrolled participants for which the G1 child was alive at 1 year old and who had not withdrawn consent for their data to be used.

Reference group: Participants who were not sent a COVID-19 questionnaire because they had withdrawn from the study, declined to participate in questionnaires or did not have a valid email address.

2. Outcome: Returning a COVID-19 questionnaire

Sample: Participants who were sent a COVID-19 questionnaire.

Reference group: Participants who, having been sent a COVID-19 questionnaire, did not return it.

3. Outcome: Completing items that define six key variables which may be relevant to wider COVID-19 research (item response)

Sample: Participants who returned a COVID-19 questionnaire.

Reference group: Participants who returned the questionnaire but did not provide data on the relevant questionnaire items that define the key variable.

The variables examined are listed below. The original ALSPAC variable names, along with additional details, are provided in [Table 1](#) (see the Wellcome Open Research Data notes

of COVIDQ1 and COVIDQ2 for full information on the corresponding questions and how the variables were derived^{19,20}).

- Self-reported COVID-19 diagnosis.
- COVID-19 case prediction using the Menni algorithm¹⁸, based on self-reported symptoms.
- Total score on the Short Mood and Feelings Questionnaire (SMFQ)²⁷ to assess depression.
- Total score on the Generalised Anxiety Disorder seven-item Assessment (GAD-7)²⁸ to assess anxiety.
- Total score on the Warwick-Edinburgh Mental Wellbeing Scale (WEMWBS)²⁹ to assess well-being.
- Self-reported keyworker status.

Candidate predictors of selection

We examined associations of pre-pandemic characteristics (41, 35 and 27 variables for the G1, G0 mother and G0 partner cohorts, respectively; [Extended data](#)²⁶: Tables S1 and S2) with the pre-specified outcomes defined above. Hereafter we will refer to these as “candidate predictors of selection”, acknowledging that we are not studying all possible sources of selection bias in ALSPAC and its COVID-19 data. We selected those variables *a priori* based on our knowledge of ALSPAC, COVID-19 and factors that are hypothesised to, or are known to, shape patterns of invitation and response^{14,16,21,30}. For analyses involving continuous variables, these were transformed into standard deviation (SD) units in order for all odds ratios to be interpretable on the same scale.

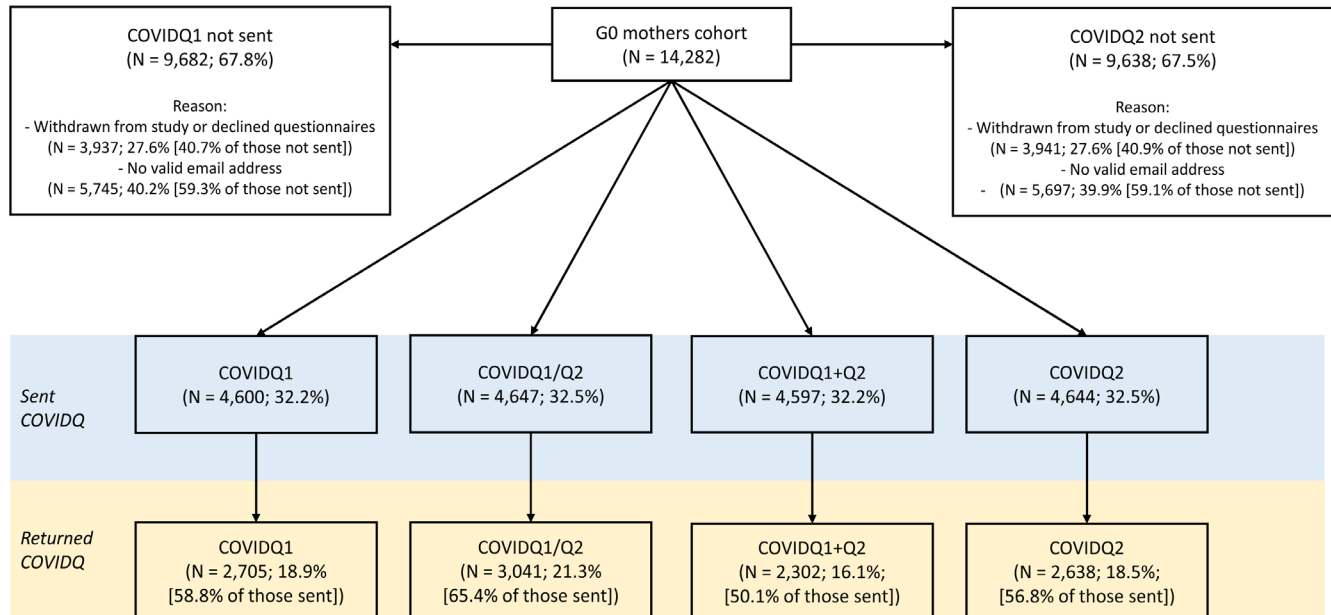


Figure 2. Flowchart of COVID-19 questionnaire invitation and response in the G0 mothers cohort. COVIDQ1 and COVIDQ2 refer to the first and second COVID-19 questionnaires, respectively, COVIDQ1/Q2 refers to being sent/returning either COVIDQ1 or COVIDQ2, and COVIDQ1+Q2 refers to being sent/returning both COVIDQ1 and COVIDQ2. Reasons why participants were not sent a COVIDQ1 or COVIDQ2 questionnaire are also given.

Statistical analysis

We used unadjusted logistic regression to quantify associations between each candidate predictor of selection and the outcomes detailed above. As we were interested in raw associations between variables, rather than estimating potential causal relationships, we did not perform multivariable analyses to adjust for potential confounders. We did not analyse any outcome with 10 or fewer participants in the reference or sent/returned/item response group (depending on the outcome being assessed). Note that, to aid interpretability of the figures displaying these results, we present a range of key candidate predictors of selection in the main text, with additional predictors displayed in the extended data. All analyses were conducted using R version 4.0.3³¹.

To describe the results in the main text, we established an arbitrary criterion based on p -values of the associations using a threshold of 0.05. This threshold has been used in similar studies to summarise large numbers of associations¹⁵. Weaker, but still potentially relevant, associations may be overlooked using this criterion, so we further described associations where the absolute z -value (log point estimate divided by the log standard error) was greater than 1 (equivalent to a p -value < 0.32). As these thresholds are arbitrary and p -values (or z -values) do not inform about the magnitude of the association³², we recommend readers consider the magnitude, direction and uncertainty of each association when interpreting these results and when undertaking COVID-19 research using ALSPAC data.

Ethical considerations

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](#).

Results

The results of the combined questionnaires are presented in the main text, while the separate COVIDQ1 and COVIDQ2 results are presented in the *Extended data*²⁶. Results for the G0 partners are also presented in the *Extended data*²⁶, as partners of the G0 mothers may change over time, the amount of data for G0 partners is lower than for G1 and G0 mothers (meaning that estimation in this cohort will be less precise), and the G0 partners' data are used less frequently than the G1 and G0 mothers' data. Figure S2 (*Extended data*²⁶) shows the overlap among those who replied to either or both questionnaires, for each cohort.

Candidate predictors of selection and invitation/response outcomes in the G1 cohort – Combined COVID-19 questionnaires

Amongst the G1 participants, 41% were sent either COVID-19 questionnaire (with 39% sent both), of whom 57% returned at

least one (37% returned both). Of those not invited to complete a questionnaire, ~55% had previously withdrawn from the study or declined to receive questionnaires, while ~45% had not actively disengaged but did not have a valid email address. Key variable response was >95% of those who returned either questionnaire for all of the six key variables, with fewer than five participants missing data on self-reported COVID-19 status and predicted COVID-19 status based on symptoms. G1 participant numbers for questionnaire invitation, return and completion are shown in [Table 2](#) and [Figure 1](#).

In the whole G1 cohort, the proportion of missing data for the candidate predictors of selection ranged from no missingness (age and sex) to three-quarters of data missing for some of the more recently collected data (e.g., education). In the G1 sample who were sent either questionnaire, completeness was generally higher, with most variables having less than 25% missingness. In the G1 sample that returned either questionnaire, completeness

was higher still, with most individual variables missing less than 17% data. [Figure 3](#) shows the proportion of missing data for each candidate predictor of selection for the three G1 samples. However, when multiple candidate predictors of selection are considered jointly, the sample size will naturally reduce further (e.g., in the whole G1 cohort, 66% of participants would be excluded from models adjusting for recent body mass index (BMI), recent smoking status, maternal education, maternal age and maternal parity; while in the ‘sent either questionnaire’ and ‘returned either questionnaire’ samples the percentage of excluded participants would be 31% and 24%, respectively).

Differences between the associations of the 41 candidate predictors with being sent or returning either questionnaire (i.e., COVIDQ1 *or* COVIDQ2) and those with being sent or returning both questionnaires (i.e., COVIDQ1 *and* COVIDQ2) were minimal ([Figure 4](#); *Extended data*²⁶: [Figure S3a](#)), so we focus on being sent and returning either questionnaire here. For

Table 2. Descriptive statistics of the outcomes from the combined COVID-19 questionnaires for the G1 cohort.

Outcome	n	Yes (%)	No (%)
Sent either questionnaire ^a	14,849	6,139 (41.34%)	8,710 (58.66%)
Sent both questionnaires ^a	14,849	5,842 (39.34%)	9,007 (60.66%)
Returned either questionnaire	6,139	3,489 (56.83%)	2,650 (43.17%)
Returned both questionnaires	5,842	2,188 (37.45%)	3,654 (62.55%)
Having self-reported COVID-19 status data ^b	3,489	>3,484 (>99.86%)	<5 ^c (<0.14%)
Having predicted COVID-19 status (from symptoms; Menni algorithm) data ^b	3,489	>3,484 (>99.86%)	<5 ^c (<0.14%)
Having SMFQ total score (depression) data ^b	3,489	3,332 (95.5%)	157 (4.5%)
Having GAD-7 total score (anxiety) data ^b	3,489	3,349 (95.99%)	140 (4.01%)
Having WEMWBS total score (well-being) data ^b	3,489	3,330 (95.44%)	159 (4.56%)
Having keyworker status data ^b	3,489	3,343 (95.82%)	146 (4.18%)

SMFQ, Short Mood and Feelings Questionnaire; GAD-7, Generalised Anxiety Disorder seven-item Assessment; WEMWBS, Warwick-Edinburgh Mental Wellbeing Scale.

^a Sample based on all enrolled G1 participants where G1 child was alive at one year of age and had not withdrawn consent for their data to be used.

^b These item/variable response outcomes are based on whether the participant returned either COVID-19 questionnaire (i.e., having data in either questionnaire).

^c Actual numbers withheld due to small cell counts (<5).

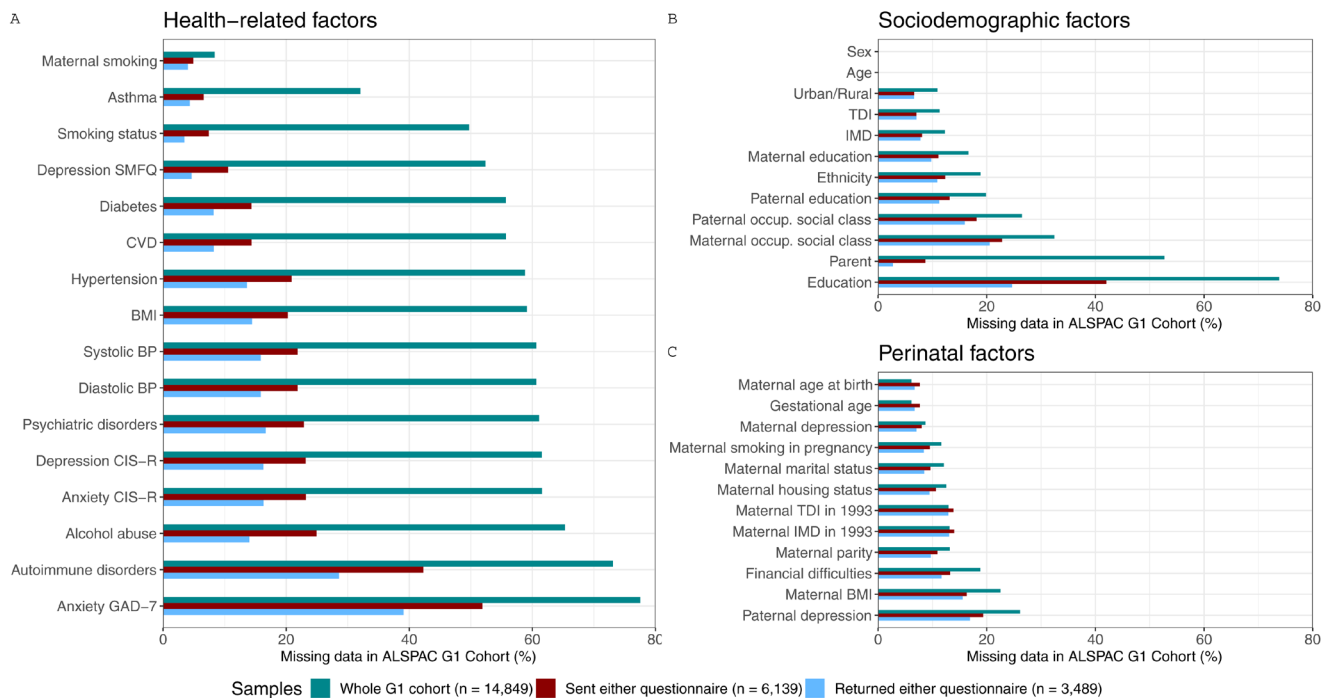


Figure 3. Proportion of missing data in each candidate predictor of selection for the three samples of the G1 cohort. These samples are the whole ALSPAC G1 cohort (green), the subsample who was sent either COVID-19 questionnaire (red), and the subsample who returned either COVID-19 questionnaire (blue). To facilitate the interpretation of the results, variables of related categories are presented in the same panel (**a.**- health-related variables, which include comorbidities, and behavioural and anthropometric factors; **b.**- sociodemographic factors; and **c.**- perinatal factors) and are ordered by the total amount of missing data in the whole sample. SMFQ, Short Mood and Feelings Questionnaire; CVD, Cardiovascular Disease; BMI, Body Mass Index; BP, Blood Pressure; CIS-R, Clinical Interview Schedule – Revised; GAD-7; Generalised Anxiety Disorder seven-item assessment; IMD, Index of Multiple Deprivation; TDI, Townsend Deprivation Index; occup., occupational.

the purpose of description and at a p -value threshold of <0.05 , 32 (78%) and 23 (56%) of the 41 candidate predictors were associated with being sent and returning a questionnaire, respectively. Equivalent results using the criteria of an absolute z -value ≥ 1 were 39 (95%) and 32 (78%). Being female, having a higher socioeconomic position, greater BMI, and older maternal age at birth were associated with higher odds of being sent a questionnaire, while ethnicity other than white, higher maternal parity, maternal perinatal depression and mother smoking during pregnancy were associated with lower odds. Similar patterns were seen for associations with returning a questionnaire, but overall effect sizes were smaller.

Figures S3b and S3c (*Extended data*²⁶) show associations between the 41 candidate predictors of selection with the outcome variables defined by completion of key questionnaire items. We did not analyse those with COVID-19 outcomes (self-reported COVID-19 status and predicted COVID-19 status based on symptoms) as most participants responded to those questions (<5 participants with missing data for these outcomes). Compared to the analyses of being sent and returning a questionnaire, few factors were associated with any of the remaining four outcomes, and results for all four outcomes were similar. For example, of 41 candidate predictors of selection, only

8 (20%) were associated with keyworker item response (at a p -value <0.05 ; 32 [56%] had an absolute z -value ≥ 1). Higher educational attainment and lower maternal age at birth were associated with higher odds of variable response, while current smoking, maternal perinatal depression and being a parent were associated with lower odds.

Candidate predictors of selection and invitation/response outcomes in the G0 mothers cohort – Combined COVID-19 questionnaire

In total, 33% of G0 mothers were sent at least one COVID-19 questionnaire (32% were sent both), of whom 65% returned at least one (50% returned both; [Table 3](#) and [Figure 2](#)). Of those not invited to complete a questionnaire, ~40% had previously withdrawn from the study or declined to receive questionnaires, while ~60% had not actively disengaged but did not have a valid email address. Key variable response was $>95\%$ of those who returned either questionnaire for all six key variables, with fewer than five participants missing data on self-reported COVID-19 status and predicted COVID-19 status based on symptoms.

In the whole G0 mothers cohort, the proportion of missing data for the candidate predictors of selection ranged from

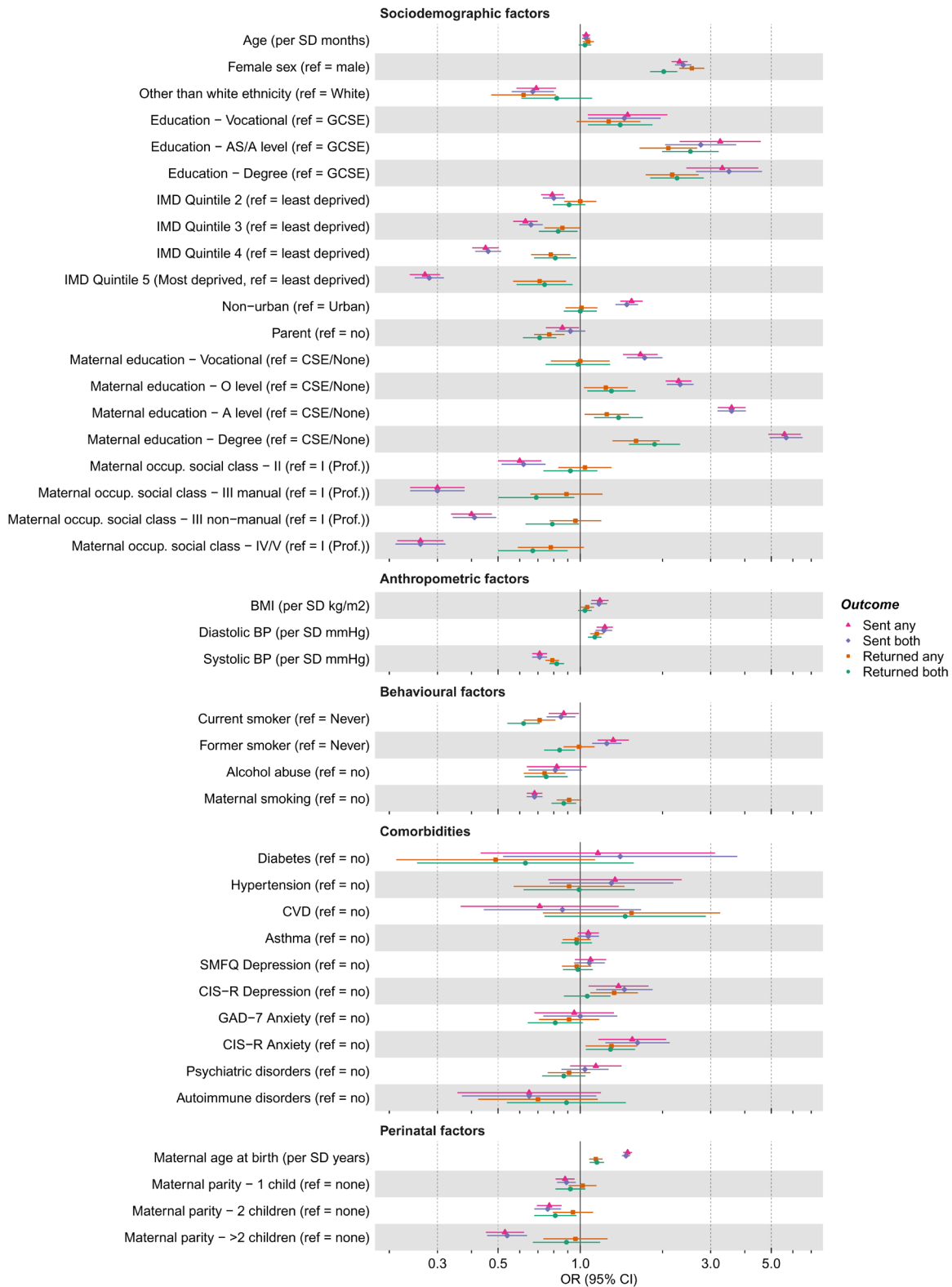


Figure 4. Associations between candidate predictors of selection and being sent and returning either or both of the COVID-19 questionnaires in the G1 cohort. For ease of presentation some candidate predictors of selection are presented in this figure, while others are displayed in Figure S3a. The x-axis is displayed on the logarithmic scale. Means and standard deviations (SD) of continuous variables: Age (months; mean = 337.0, SD = 5.89), BMI (kg/m²; mean = 24.43, SD = 5.05), SBP (mmHg; mean = 117.09, SD = 11.45) and DBP (mean = 65.99, SD = 7.75). Ref, reference; (G)CSE, (General) Certificate of Secondary Education; IMD, Index of Multiple Deprivation; Occup., occupation; Prof., professional; BMI, Body Mass Index; BP, Blood Pressure; CVD, Cardiovascular Disease; SMFQ, Short Mood and Feelings Questionnaire; CIS-R, Clinical Interview Schedule - Revised; GAD-7, Generalised Anxiety Disorder seven-item Assessment.

Table 3. Descriptive statistics of the outcomes from the combined COVID-19 questionnaires for the G0 mothers cohort.

Outcome	n	Yes (%)	No (%)
Sent either questionnaire ^a	14,282	4,647 (32.54%)	9,635 (67.46%)
Sent both questionnaires ^a	14,282	4,597 (32.19%)	9,685 (67.81%)
Returned either questionnaire	4,647	3,041 (65.44%)	1,606 (34.56%)
Returned both questionnaires	4,597	2,302 (50.08%)	2,295 (49.92%)
Having self-reported COVID-19 status data ^b	3,041	>3,036 (>99.83%)	<5 ^c (<0.17%)
Having predicted COVID-19 status (from symptoms; Menni algorithm) data ^b	3,041	>3,036 (>99.83%)	<5 ^c (<0.17%)
Having SMFQ total score (depression) data ^b	3,041	2,914 (95.82%)	127 (4.18%)
Having GAD-7 total score (anxiety) data ^b	3,041	2,944 (96.81%)	97 (3.19%)
Having WEMWBS total score (well-being) data ^a	3,041	2,913 (95.79%)	128 (4.21%)
Having keyworker status data ^b	3,041	2,951 (97.04%)	90 (2.96%)

SMFQ, Short Mood and Feelings Questionnaire; GAD-7, Generalised Anxiety Disorder seven-item Assessment; WEMWBS, Warwick-Edinburgh Mental Wellbeing Scale.

^a Sample based on all enrolled G0 mothers where G1 child was alive at one year of age and had not withdrawn consent for their data to be used.

^b These item/variable response outcomes are based on whether the participant returned either COVID-19 questionnaire (i.e., having data in either questionnaire).

^c Actual numbers withheld due to small cell counts (<5).

low/minimal (<20%, for age and most baseline covariates) to two-thirds of data missing for more recently collected data (e.g., alcohol abuse, smoking status, BMI and blood pressure). In the G0 mothers sample that were sent either questionnaire, completeness was generally higher, with all variables having less than 32% missingness. In the G0 mothers sample that returned either questionnaire, completeness was higher still, with most individual variables missing less than 21% data. Figure 5 shows the proportion of missing data in each candidate predictors of selection for the three G0 mothers samples.

As with the G1 cohort, differences between the associations of the 35 candidate predictors with being sent or returning either questionnaire and those with being sent or returning both questionnaires were minimal (Figure 6; *Extended data*²⁶: Figure S4a), so we focus on being sent and returning either questionnaire here. For the purpose of description and at a *p*-value threshold of <0.05, 27 (77%) and 22 (63%) of the 35 candidate predictors were associated with being sent and

returning a questionnaire, respectively. Equivalent results using the criteria of an absolute *z*-value ≥ 1 were 30 (86%) and 26 (74%). Results were broadly similar to the G1 cohort, with factors associated with G0 mothers being sent a questionnaire including: education, area deprivation, occupational social class, older age, White ethnicity, no history of smoking, lower BMI, lower diastolic and systolic blood pressure (DBP and SBP), and several sociodemographic variables measured at baseline (e.g., home ownership status, marital status, parity and financial difficulties). As with the G1 cohort, similar patterns were seen for the associations with returning a questionnaire, but overall effect sizes were smaller.

Figures S4b and S4c (*Extended data*²⁶) show associations between the 35 candidate predictors with the key variables defined by completion of key questionnaire items. We did not analyse those with COVID-19 outcomes as most participants responded to those questions (<5 participants with missing data for these outcomes). Results were broadly similar to the

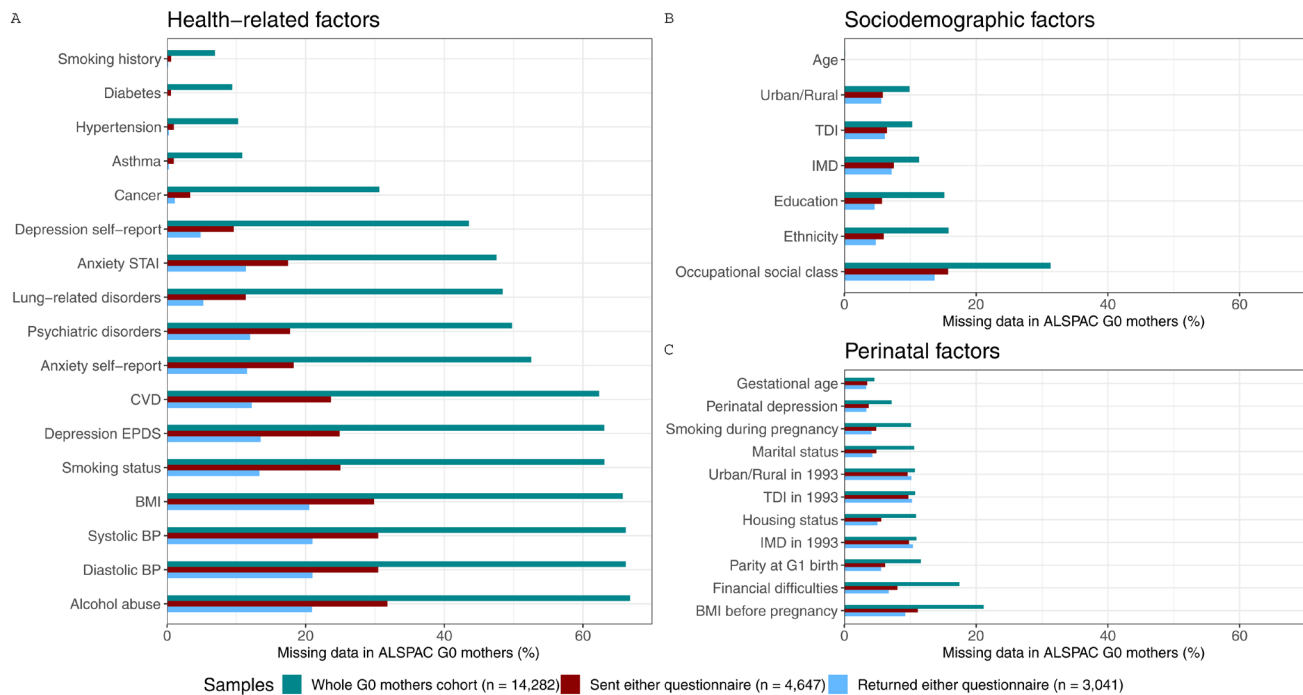


Figure 5. Proportion of missing data in each candidate predictor of selection for the three samples of G0 mothers. These samples are the whole ALSPAC cohort of G0 mothers (green), the subsample who was sent either COVID-19 questionnaire (blue). To facilitate the interpretation of the results, variables of related categories are presented in the same panel (a. health-related variables, which include comorbidities, and behavioural and anthropometric factors; b. sociodemographic factors; and c. perinatal factors) and are ordered by the total amount of missing data in the whole sample comparison. STAI, State-Trait Anxiety Inventory; CVD, Cardiovascular Disease; EPDS, Edinburgh Postnatal Depression Scale; BMI, Body Mass Index; BP, Blood Pressure; IMD, Index of Multiple Deprivation; TDI, Townsend Deprivation Index.

G1 cohort, with few strong associations between the assessed factors and item response observed, although higher education, lower BMI and no history of smoking were associated with response.

Candidate predictors of selection and invitation/ response outcomes – G0 partners cohort, and first and second COVID-19 questionnaires separately

Compared to the G1 participants and the G0 mothers, a smaller proportion of G0 partners were sent and returned either questionnaire (13% sent either questionnaire, of which 65% returned one; *Extended data*²⁶: Table S3 and Figure S1), and the missingness of the candidate predictors of selection was more substantial (*Extended data*²⁶: Figure S5). While these results therefore contain greater uncertainty than the G1 and G0 mothers cohorts, there were several candidate predictors of selection associated with being sent and returning a questionnaire among G0 partners, with effect sizes larger for being sent than for returning a questionnaire (*Extended data*²⁶: Figure S6). Few factors were strongly associated with item response, and effect estimates were somewhat inconsistent with substantial uncertainty (*Extended data*²⁶: Figure S7). Full details for the G0 partners' data and results for their analyses using the combined COVID-19 questionnaires are given in the *Extended data*²⁶ (Table S3; Figures S1 and S5–S7).

Results for the first and second COVID-19 questionnaires separately (for G1, G0 mothers and G0 partners) are presented in the *Extended data*²⁶ (Tables S4–S9; Figures S8–S25). They were not materially different to those presented above for the combined datasets.

Discussion

Associations between participant characteristics and COVID-19 questionnaire invitation and response matter because they can bias associations of those factors with COVID-19 outcomes. A comprehensive description of the analysed data and relationships with missing data allow informed inferences from association studies; this is possible in longitudinal studies in which missing data can be characterised using previously collected data. In a UK-based multigenerational birth cohort, we have demonstrated that many sociodemographic, behavioural and health-related factors were associated with (i) being sent and (ii) returning COVID-19 questionnaires, and (iii) completion of questions within these questionnaires related to six key variables for COVID-19 research. Generally, the magnitude of associations for being sent a questionnaire were larger than for returning a questionnaire. Of those who returned a questionnaire, there were fewer predictors of item response, but, due to the smaller sample size and generally high response rates, effect sizes were estimated with greater uncertainty



Figure 6. Associations between candidate predictors of selection and being sent and returning either or both of the COVID-19 questionnaires (Q) in the G0 mothers cohort. For ease of presentation some candidate predictors of selection are presented in this figure, while others are displayed in Figure S4a. The x-axis is displayed on the logarithmic scale. Means and standard deviations (SD) of continuous variables: Age (years; mean = 56.14, SD = 5.01), BMI (kg/m²; mean = 26.80, SD = 5.47), SBP (mmHg; mean = 119.93, SD = 14.28) and DBP (mean = 71.36, SD = 9.56). Ref, reference; CSE, Certificate of Secondary Education; IMD, Index of Multiple Deprivation; Occup., occupational; Prof., professional; BMI, Body Mass Index; BP, Blood Pressure; CVD, Cardiovascular Disease; EPDS, Edinburgh Postnatal Depression Scale; STAI, State-Trait Anxiety Inventory.

than for being sent or returning a questionnaire. Results were similar in all three ALSPAC cohorts (G1, G0 mothers and G0 partners).

Our findings show that – on average – participants who were invited and responded to these COVID-19 questionnaires are different in several key characteristics from those who were not invited, did not return a questionnaire and did not complete all its items. In all cohorts, fewer than half of participants were sent a questionnaire (41% in G1, 33% in G0 mothers and 13% in G0 partners). While questionnaire return rates were relatively high (57% of G1 participants, and 65% of participants in both G0 cohorts, returned either questionnaire) with mostly complete item/variable response (all >95%), together this results in considerable levels of missing data (e.g., for G1 and G0 mothers only ~20% of the whole cohort have COVID-19 questionnaire data; <10% for G0 partners). Recruitment was intended to be representative of the target population; however, due to loss to follow-up, the analytic samples with COVID-19 questionnaire data are no longer representative of the original ALSPAC study sample (and hence the target population, of which the original ALSPAC sample is broadly representative). Selection bias may therefore be a potential risk when using ALSPAC COVID-19 questionnaire data and we encourage researchers to carefully consider the results presented here to inform their work.

Two key points are worth highlighting when interpreting these findings. First, as we were interested in the raw associations between variables, we did not account for any confounding between the candidate predictors of selection, and our findings should not be interpreted as evidence for a certain factor to be independently associated with the outcomes assessed. Second, the impact of missing data needs to be considered carefully as variables from more recent data collections are biased towards certain groups (e.g., higher socioeconomic position [SEP], older G0 mothers, female G1 participants), which may result in selection bias in the observed associations. We also note that there may be selection bias when examining candidate predictors of returning a questionnaire (as this is conditional on being sent a questionnaire) and item response (as this is conditional on being sent and returning a questionnaire) due to unmeasured confounders of questionnaire invitation and response.

To illustrate these two key points, we selected a history of cancer among G0 mothers, which was associated with both being sent and returning a questionnaire. These associations may be due to confounding, as age and SEP both predict questionnaire invitation and response and may also predict cancer diagnosis (since older people are more likely to suffer from cancer and people from higher SEP backgrounds may also be more likely to engage in health-seeking behaviours³³ and, therefore, potentially be diagnosed with cancer). When we adjusted for age and education (a proxy for SEP), these associations were somewhat – although not completely – attenuated (*Extended data*²⁶: Table S10).

Additionally, data on cancer is missing for ~30% of G0 mothers, and it is likely that missing data is associated with characteristics such as age and SEP (*Extended data*²⁶: Figure S26),

which may result in biased associations. We present an example to illustrate this. When the associations of age (a nearly fully-observed predictor) with being sent or receiving either questionnaire were analysed in all participants regardless of having cancer data (an unbiased estimate), or only in those with recorded cancer data, the odds ratio estimates for being sent a questionnaire differed, indicating potential bias (*Extended data*²⁶: Table S11). Candidate predictors of selection with missing data may therefore result in biased associations with questionnaire invitation and response due to selection bias.

We have not aimed to mitigate selection bias in this paper, but rather to illustrate how it can be identified. Methods such as multiple imputation, inverse probability weighting, *g*-formula approaches, simulations, and bounds and parameter searches^{10,11,30,34–38} can be used to help explore and overcome potential selection bias. For instance, in the case of inverse probability weighting, weights should be derived according to the target population of the study (or using the total enrolled ALSPAC sample as a proxy, if weights derived from the target population cannot be constructed). For some examples of such methods applied in ALSPAC to mitigate potential selection bias, see 16,39–42. Researchers need to assess the assumptions when using these approaches as it is not possible to dictate a ‘one-size-fits-all’ approach when working with cohort data such as ALSPAC, as different research questions will be addressed using different variables and methods. However, the number of potential predictors of selection and magnitude of their associations with being sent a questionnaire were larger than for questionnaire and item response, suggesting that much – although by no means all – of the potential selection bias reported here could be minimised by using these variables as weights (if using inverse probability weighting), or auxiliary variables (if using multiple imputation) when analysing these data. In addition, as questionnaire invitation could not have been affected by the COVID-19 pandemic – since disengagement from the study and having a valid email address largely occurred prior to the pandemic – selection due to being sent a questionnaire cannot be caused by COVID-19 outcomes. That said, promotion of ALSPAC’s COVID-19 data collections could have prompted some participants to re-engage and provide the study with an up-to-date email address. In the period of interest for this study (between 9th April 2020 and 5th July 2020), we estimate that this may have been the case for a few hundred participants.

This descriptive study also demonstrates that longitudinal studies allow researchers to utilise their rich detailed pre-pandemic data as potential predictors of selection in COVID-19 studies that inform their research. While some characteristics may have a similar impact on selection across various studies (such as SEP predicting selection, as found in ALSPAC and the 1958 British birth cohort¹⁷), others are likely to be study-specific. For instance, in the ALSPAC G1 cohort, participants enrolled during their mother’s pregnancy or as children, while in other cohorts participants may have enrolled as adults (e.g. UK Biobank). We therefore cannot assume that the results described here will apply to other studies with different demographic profiles, enrolment strategies and data collection mechanisms. Additionally, given the target population (defined above), ALSPAC is not necessarily representative of the general

UK population or non-UK populations, as: i) at time of recruitment, the Bristol area comprised mainly White Europeans; and ii) the cohort that includes both females and males is a young population (G1, aged ~28 years old), while the sex-specific G0 cohorts are older populations (mean G0 mothers' age ~56 years [range: 41–75]; mean G0 partners' age ~61 years [range: 41–89]). Furthermore, the reported associations in the ALSPAC COVID-19 questionnaire data are likely to be specific to the data collection process (i.e. voluntary participation in a long-running birth cohort study), and may not be generalised to studies that acquired COVID-19 data from other sources (e.g. medical records linkage). However, birth cohorts do tend to be more representative of their target population than other study designs, which may minimise potential biases due to selection relative to these other studies (for instance, initial recruitment into ALSPAC included ~75% of the target population¹⁴, while UK Biobank only achieved a 5% recruitment rate⁷). Although we cannot extrapolate findings across cohorts, confidence in any conclusions would be amplified if we found similar results using comparable data from multiple studies/cohorts with different demographics, enrolment processes and data collection strategies (see, for example, work on the impact of COVID-19 on mental health in both ALSPAC and Generation Scotland cohorts¹). We therefore encourage other longitudinal studies to perform similar analyses to these, to help researchers plan analyses and interpret their findings.

Nonetheless, we consider that the issues related to selection bias in COVID-19 research described throughout this paper are relevant to all COVID-19 observations studies globally. This is because being infected by SARS-CoV-2, experiencing COVID-19 and the measures that have been used to limit the spread of infection (e.g., lockdowns) will influence who responds to invitations to participate in COVID-19 research. At the same time, who got tested and diagnosed would have been influenced by characteristics such as age, sex, occupation, socio-economic position and existing co-morbidities, among other factors. However, as the distribution of these characteristics, and the management of the pandemic (e.g., whether universal or selected testing is used) will vary widely between countries, we must be careful in making inferences from our specific results to other populations. Indeed, even within the UK we have shown that many of the characteristics related to response in ALSPAC also do so in UK Biobank, but for several factors the magnitude, and sometimes even the direction, of the associations differed⁴². The aim of this paper was to highlight key sources of selection bias in COVID-19 research; we feel that most observational studies (including genome-wide association studies and those using genetic data for causal inference, such as Mendelian randomization studies), exploring causes or consequences of COVID-19 are likely to have such biases and thus our conclusions and suggestion that this should be explored in other studies are widely relevant.

It is also possible that rates of response to future COVID-19 questionnaires and associations of candidate predictors may change over time within a cohort like ALSPAC (for instance, because of increased understanding in the importance of COVID-19 research and changes in the restrictions used to manage

COVID-19). In our supplementary analyses, we compared candidate predictors of questionnaire invitation and response in the two COVID-19 questionnaires separately. Overall, they appeared similar, but more subtle differences are important to consider when using repeated data across multiple waves of COVID-19 data collection. For instance, among ALSPAC participants who returned the first COVID-19 questionnaire, those who returned the second COVID-19 questionnaire were more likely to be older (i.e., G0 participants), had higher educational qualifications, and had fewer recent financial worries in the first COVID-19 questionnaire²⁰. As we only focused on questionnaires completed early in the pandemic, repeating these analyses with subsequent rounds of COVID-19 questionnaires will be important for researching impacts of the long-term effects of COVID-19 and its management, new variants and on long-COVID.

We also note some specific caveats when interpreting these results and working with these ALSPAC data. First, we could not investigate whether those who reported having had COVID-19 were more (or less) likely to respond to these questionnaires, which may result in selection bias when using this as an exposure or outcome. This risk of bias may be especially concerning as we only have COVID-19 questionnaire data for ~20% of the G1 and G0 mother cohorts (~10% for G0 partners), meaning that COVID-19 status data is missing for ~80% of the cohorts. Additional linkage data – such as from Public Health England COVID-19 testing data – or data from other sources is required to answer such questions³⁹. For instance, recent research using ALSPAC COVID-19 serology test data demonstrates that individuals consenting to have a COVID-19 antibody test were more likely to report having had COVID-19²³. Questionnaire completion may have been similarly biased, thus increasing the risk of selection bias in studies where COVID-19-related outcomes are the outcome of interest.

Second, as ALSPAC data was collected repeatedly using the same questions/instruments, appropriate methods to model this repeated data should be used. Here, for repeated measures we chose a simple method consisting of using the most recent observation, and, if missing, back-filling with previous data. However, if the time-points are not comparable, this may result in bias (e.g., different rates of depression or anxiety at different ages, or different 'smoking status' at age 18 vs age 24). Nonetheless, using G1 depression and asthma as examples, we compared different approaches to define these variables (such as using single time-points, forward-filling data and averaging), and found that they were broadly consistent (*Extended data*²⁶: Table S12; although for depression using data from single time-points gave slightly lower estimates relative to collating over repeated measures). However, we noted that using 'any history of X' to derive these variables biased cases to those with more data, effectively turning these variables into measures of repeated participation, so we recommend researchers *not* to use this approach. As such, we suggest that researchers apply appropriate statistical techniques when working with these (possibly missing) longitudinal data, such as longitudinal multiple imputation⁴³, latent variable/structural equation modelling (e.g., 44), and/or sensitivity analyses using different variable derivations.

Finally, as 913 G1 children (6% of the whole G1 cohort) were enrolled after the age of seven years, candidate predictors of selection measured during pregnancy and in early childhood will be missing for these children and their G0 parents. This includes potentially important factors associated with selection such as parental education, parental occupational social class, gestational age, maternal parity, maternal age at birth, maternal smoking in pregnancy and parental perinatal depression. Therefore these variables cannot be used to derive weights for these participants if using inverse probability weighting; while if there are few observed auxiliary variables associated with these factors, then multiple imputation will have little information to draw upon when predicting these variables, resulting in greater uncertainty in parameter estimates.

Conclusion

Missing data from not being sent an invitation or questionnaire and not responding (as well as other sources of missing data) can lead to spurious inferences and counterintuitive results due to selection bias, which may result in incorrect policy recommendations. This is particularly important in the fast-moving area of COVID-19 research. We found several factors associated with selection due to questionnaire invitation and response that may bias findings in COVID-19 research in ALSPAC. This work can be used as a basis for future research using ALSPAC COVID-19 data and highlights the importance of using longitudinal pre-pandemic data to assess potential selection pressures in observational COVID-19 research, and make informed inferences.

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 study and all other ALSPAC data. The datasets presented in this article are linked to ALSPAC project number B3543, please quote this project number during your application. The ALSPAC variable codes highlighted in the dataset descriptions can be used to specify required variables.

1. 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.

2. 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.

3. 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.

If you have any questions about accessing data, please email alspac-data@bristol.ac.uk.

Please note that a standard COVID-19 dataset will be made available at no charge (see^{19,20}); however, costs for required paperwork and any bespoke datasets required additional variables will apply.

Extended data

Open Science Framework: Questionnaire Invitation/Response and Selection Bias, <https://doi.org/10.17605/OSF.IO/TP45V>²⁶.

This project contains the following extended data:

- ALSPACSelectionBias_SuppInfo.pdf (supplementary information file: Tables S1–12; Figures S1–26)

Reporting guidelines

Open Science Framework: STROBE checklist for ‘Bias from questionnaire invitation and response in COVID-19 research: an example using ALSPAC’, <https://doi.org/10.17605/OSF.IO/TP45V>²⁶.

Data are available under the terms of the [Creative Commons Attribution 4.0 International license](#) (CC-BY 4.0).

Acknowledgements

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

References

1. Kwong ASF, Pearson RM, Adams MJ, *et al.*: **Mental health before and during the COVID-19 pandemic in two longitudinal UK population cohorts.** *Br J Psychiatry.* 2020; 1–10. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
2. Di Gessa G, Maddock J, Green MJ, *et al.*: **Mental health inequalities in healthcare, economic, and housing disruption during COVID-19: an investigation in 12 longitudinal studies.** *medRxiv.* 2021. [Publisher Full Text](#)
3. Kwong ASF, Pearson RM, Smith D, *et al.*: **Longitudinal evidence for persistent anxiety in young adults through COVID-19 restrictions [version 1; peer review: 2 approved with reservations].** *Wellcome Open Res.* 2020; 5: 195. [Publisher Full Text](#)
4. Paul E, Kounali D, Kwong ASF, *et al.*: **Trajectories of child emotional and behavioural difficulties before and during the COVID-19 pandemic in a longitudinal UK cohort.** *medRxiv.* 2021. [Publisher Full Text](#)
5. Wielgoszewska B, Maddock J, Green MJ, *et al.*: **The UK Coronavirus Job Retention Scheme and changes in diet, physical activity and sleep during the COVID-19 pandemic: Evidence from eight longitudinal studies.** *medRxiv.* 2021. [Publisher Full Text](#)
6. Thompson EJ, Williams DM, Walker AJ, *et al.*: **Risk factors for long COVID: analyses of 10 longitudinal studies and electronic health records in the UK.** *medRxiv.* 2021. [Publisher Full Text](#)
7. Munafò MR, Tilling K, Taylor AE, *et al.*: **Collider scope: When selection bias can**

- substantially influence observed associations. *Int J Epidemiol.* 2018; **47**(1): 226–35.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
8. Greenland S, Pearl J, Robins JM: **Causal diagrams for epidemiologic research.** *Epidemiology.* 1999; **10**(1): 37–48.
[PubMed Abstract](#)
 9. Greenland S: **Quantifying biases in causal models: classical confounding vs collider-stratification bias.** *Epidemiology.* 2003; **14**(3): 300–6.
[PubMed Abstract](#) | [Publisher Full Text](#)
 10. Hernán M, Robins J: **Causal Inference: What If.** Boca Raton: Chapman & Hall/CRC. 2020.
[Reference Source](#)
 11. Griffith GJ, Morris TT, Tudball M, et al.: **Collider bias undermines our understanding of COVID-19 disease risk and severity.** *Nat Commun.* 2020; **11**(1): 5749.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 12. Blom AG, Herzing JME, Cornesse C, et al.: **Does the Recruitment of Offline Households Increase the Sample Representativeness of Probability-Based Online Panels? Evidence From the German Internet Panel.** *Soc Sci Comput Rev.* 2017; **35**(4): 498–520.
[Publisher Full Text](#)
 13. Antoun C, Zhang C, Conrad FG, et al.: **Comparisons of Online Recruitment Strategies for Convenience Samples: Craigslist, Google AdWords, Facebook, and Amazon Mechanical Turk.** *Field methods.* 2016; **28**(3): 231–46.
[Publisher Full Text](#)
 14. 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): 111–27.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 15. Taylor AE, Jones HJ, Sallis H, et al.: **Exploring the association of genetic factors with participation in the Avon Longitudinal Study of Parents and Children.** *Int J Epidemiol.* 2018; **47**(4): 1207–16.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 16. Cornish RP, Macleod J, Boyd A, et al.: **Factors associated with participation over time in the Avon Longitudinal Study of Parents and Children: a study using linked education and primary care data.** *Int J Epidemiol.* 2021; **50**(1): 293–302.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 17. Mostafa T, Narayanan M, Pongiglione B, et al.: **Improving the plausibility of the missing at random assumption in the 1958 British birth cohort: A pragmatic data driven approach.** CLS working paper 2020/6. London: UCL Centre for Longitudinal Studies. 2020.
[Reference Source](#)
 18. Menni C, Valdes AM, Freidin MB, et al.: **Real-time tracking of self-reported symptoms to predict potential COVID-19.** *Nat Med.* 2020; **26**(7): 1037–40.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 19. Northstone K, Haworth 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 Res.* 2020; **5**: 127.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 20. 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.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 21. 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](#)
 22. 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.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 23. Northstone K, Smith D, Bowring C, et al.: **The Avon Longitudinal Study of Parents and Children - A resource for COVID-19 research: Home-based antibody testing results, October 2020 [version 1; peer review: awaiting peer review].** *Wellcome Open Res.* 2021; **6**: 34.
[Publisher Full Text](#)
 24. Smith D, Bowring C, Wells N, et al.: **The Avon Longitudinal Study of Parents and Children - A resource for COVID-19 research: Questionnaire data capture November 2020 - March 2021 [version 1; peer review: awaiting peer review].** *Wellcome Open Res.* 2021; **6**: 155.
[Publisher Full Text](#)
 25. Harris PA, Taylor R, Thielke R, et al.: **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](#)
 26. Fernández-Sanlés A, Smith D, Clayton G, et al.: **ALSPAC COVID-19 - Questionnaire Invitation/Response and Selection Bias.** 2021.
<http://www.doi.org/10.17605/OSF.IO/TP45V>
 27. 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): 237–49.
[Reference Source](#)
 28. 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): 1092–7.
[PubMed Abstract](#) | [Publisher Full Text](#)
 29. 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](#)
 30. Howe LD, Tilling K, Galobardes B, et al.: **Loss to follow-up in cohort studies: Bias in estimates of socioeconomic inequalities.** *Epidemiology.* 2013; **24**(1): 1–9.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 31. R Development Core Team: **R: A language and environment for statistical computing.** 2021.
[Reference Source](#)
 32. Sterne JA, Davey Smith G: **Sifting the evidence—what’s wrong with significance tests?** *BMJ.* 2001; **322**(7280): 226–31.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 33. Adler NE, Stewart J: **Health disparities across the lifespan: Meaning, methods, and mechanisms.** *Ann N Y Acad Sci.* 2010; **1186**: 5–23.
[PubMed Abstract](#) | [Publisher Full Text](#)
 34. van Buuren S: **Flexible Imputation of Missing Data.** Boca Raton: Chapman and Hall/CRC Interdisciplinary Statistics Series. 2nd Edition. ISBN: 978-1-138-58831-8. 2018.
[Publisher Full Text](#)
 35. Smith LH, Vanderweele TJ: **Bounding Bias Due to Selection.** *Epidemiology.* 2019; **30**(4): 509–16.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 36. Hughes RA, Heron J, Sterne JAC, et al.: **Accounting for missing data in statistical analyses: Multiple imputation is not always the answer.** *Int J Epidemiol.* 2019; **48**(4): 1294–304.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 37. Seaman SR, White IR: **Review of inverse probability weighting for dealing with missing data.** *Stat Methods Med Res.* 2013; **22**(3): 278–95.
[PubMed Abstract](#) | [Publisher Full Text](#)
 38. Wolke D, Waylen A, Samara M, et al.: **Selective drop-out in longitudinal studies and non-biased prediction of behaviour disorders.** *Br J Psychiatry.* 2009; **195**(3): 249–56.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 39. Cornish RP, Tilling K, Boyd A, et al.: **Using linked educational attainment data to reduce bias due to missing outcome data in estimates of the association between the duration of breastfeeding and IQ at 15 years.** *Int J Epidemiol.* 2015; **44**(3): 937–45.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 40. Lee KJ, Tilling KM, Cornish RP, et al.: **Framework for the treatment and reporting of missing data in observational studies: The Treatment And Reporting of Missing data in Observational Studies framework.** *J Clin Epidemiol.* 2021; **134**: 79–88.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 41. Cadman T, Kwong AS, Moran P, et al.: **Joint associations of parental personality traits and socio-economic position with trajectories of offspring depression: Findings from up to 6925 families in a UK birth cohort.** *JCPP advances.* 2021; **1**(3): e12028.
[Publisher Full Text](#)
 42. Millard LAC, Fernández-Sanlés A, Carter AR, et al.: **Exploring selection bias in COVID-19 research: Simulations and prospective analyses of two UK cohort studies.** *MedRxiv.* 2021.
[Publisher Full Text](#)
 43. Huque MH, Carlin JB, Simpson JA, et al.: **A comparison of multiple imputation methods for missing data in longitudinal studies.** *BMC Med Res Methodol.* 2018; **18**(1): 168.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 44. Thapar A, Heron J, Jones RB, et al.: **Trajectories of change in self-reported psychotic-like experiences in childhood and adolescence.** *Schizophr Res.* 2012; **140**(1–3): 104–9.
[PubMed Abstract](#) | [Publisher Full Text](#)

Open Peer Review

Current Peer Review Status:  

Version 2

Reviewer Report 18 July 2022

<https://doi.org/10.21956/wellcomeopenres.19985.r51457>

© 2022 Gopal D. 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.



Dipesh P Gopal 

Centre of Primary Care, Wolfson Institute of Population Health, Barts and The London School of Medicine and Dentistry, London, UK

Thank you for responding. I have no further comments to make. Thank you. I would be great to see this work indexed.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Primary Care, Cancer Survivorship, Health Inequalities

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

Version 1

Reviewer Report 04 January 2022

<https://doi.org/10.21956/wellcomeopenres.18822.r47518>

© 2022 Gopal D. 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.



Dipesh P Gopal 

¹ Centre of Primary Care, Wolfson Institute of Population Health, Barts and The London School of Medicine and Dentistry, London, UK

² Centre of Primary Care, Wolfson Institute of Population Health, Barts and The London School of

Medicine and Dentistry, London, UK

The research is comprehensively undertaken and is a marked improvement on other work using the ALSPAC data by using validated symptom scores where appropriate.

However, it is unclear what this research adds to the literature especially since previous ALSPAC work showed that people from low socio-economic backgrounds have lower response rates (BMJ Open. 2017 Mar 14;7(3):e014238.).¹

The data presented graphically is difficult to follow and much of the data referenced in the text is in 'Extended data' section. Hugely frustrating for the reader. Could it all be organised so the relevant data is presented?

References

1. Heron J, Grzeda M, von Gontard A, Wright A, et al.: Trajectories of urinary incontinence in childhood and bladder and bowel symptoms in adolescence: prospective cohort study. *BMJ Open*. 2017; 7 (3). [Publisher Full Text](#)

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

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

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Partly

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Primary Care, Cancer Survivorship, Health Inequalities

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

Author Response 01 Jul 2022

Daniel Smith

We thank the author for their review, and hope that we have addressed any remaining concerns. We also apologise for the length of time taken to respond to the reviewer's comments; we were expecting a further review before revising the paper, but this never materialised. We have responded to the reviewer's comments in turn below (with the reviewer's original comments in italics).

However, it is unclear what this research adds to the literature especially since previous ALSPAC work showed that people from low socio-economic backgrounds have lower response rates (BMJ Open. 2017 Mar 14;7(3):e014238.).1

While socioeconomic position (SEP) is indeed a key predictor of response, we also explored other factors which could inform future work using this data. We show that the assessed factors (which are not limited to SEP) predict response to COVID-19 data collections specifically, which may be different from previous ALSPAC data collections. For example, the analysed COVID-19 data was collected through online-only questionnaires (unlike the reference provided) and during a pandemic (which is an additional selection pressure on top of those arising from different data collections and a different time point).

The data presented graphically is difficult to follow and much of the data referenced in the text is in 'Extended data' section. Hugely frustrating for the reader. Could it all be organised so the relevant data is presented?

We understand that the amount of data described may be overwhelming for the reader. Therefore, we tried to be clear giving enough information in the main text to get the overall take-home message. Additional details and further depictions are shown in the extended data for those readers interested in exploring the finer details of this work. We have added a sentence in the 'statistical analysis' section of the results (previously just in figure footnotes) to make this clearer:

"Note that, to aid interpretability of the figures displaying these results, we present a range of key candidate predictors of selection in the main text, with additional predictors displayed in the extended data."

We would appreciate more specific feedback on suggestions on how to improve the organisation of the paper; we would be happy to revise the paper accordingly if those suggestions improved clarity.

Competing Interests: None.

Reviewer Report 23 August 2021

<https://doi.org/10.21956/wellcomeopenres.18822.r45126>

© 2021 Infante-Rivard C. 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.



Claire Infante-Rivard 

¹ Department of Epidemiology, Biostatistics and Occupational Health, Faculty of Medicine, McGill University, Montréal, QC, Canada

² Department of Epidemiology, Biostatistics and Occupational Health, Faculty of Medicine, McGill University, Montréal, QC, Canada

The authors study selection bias while investigating COVID-19 issues with online questionnaires in the ALSPAC cohorts.

The results of this descriptive study are important and particularly useful; selection bias, a rather complex and misunderstood bias, is often ignored in the interpretation of results or addressed informally, while potentially present in most studies. The report is very clearly written, the study methods are impeccably presented and thus provide an intelligible model for investigators concerned with the bias but uncertain about how to go about uncovering its presence in their data.

I have a minor comment about the wording of one aspect of the report. I would strongly suggest changing the wording around the issue of bias related to “being sent a questionnaire”. This is initially interpreted as a meaning bias originating from the research team when selecting who is to be sent a questionnaire whereas it has nothing to do with that. I am sure the authors can come up with a less ambiguous formulation: for example, “subjects who were reachable” or “subjects who could be contacted”, etc.

My main comment though is about the need for a more advanced discussion with regards to the effects of selection bias; briefly, the effect can be biased in the sample (and therefore in the target population) or an unbiased in the sample but biased for the target. Without getting into the specific, and sometimes subtle, aspects of these observations, for sure, the reader would benefit from an enhanced discussion on the issue of generalizability from sample results to the target.

All studies have the goal of interpreting results from the sample to the target population of inferential interest. Few investigators are interested in biomedical results in a few hundred or even a few thousand children or mothers (i.e., the study sample). All are interested in knowing if sample results are generalizable to a well characterized population of children and/or mothers. Whereas the report recognizes that “ALSPAC is not necessarily representative of the general UK population or non-UK population” the target populations at any of the analyses time points in this report remain undefined, with the possible exception of recruitment time, assuming a sampling frame allowed the enumeration of eligible subjects. With the successive follow-ups and the marked losses at each step, defining the target at these steps is a challenge. However, this knowledge impacts on the usefulness of study results, including the ones shown in this report.

While the issue of defining the target population of inferential interest may seem theoretical, it becomes very concrete when it comes to generalizing results from a sample meeting the identification conditions. For example, if one were to use IPW or a G-formula approach (which by

the way should be added to the list of methods in the report), distribution of key selection-influencing variables from the target would be needed (let's assume no confounding for simplicity). Should the distribution of these selection-influencing factors be from the eligible at recruitment or from those remaining in the study at later follow-up times? These may well represent different target populations. There are many possible selection models: at entry, at follow-up, at death, for item response. Using IPW for example, should the weights be specific to each mechanism or is the use of a simpler average weight over the mechanisms interpretable? More concretely, from a public health standpoint, assuming generalizations from the sample could be applied, how would the target populations be defined in these applications?

Another problem arises when the authors use additional children in the G1 cohort; they are defined as eligible but non-participant children: from the ALSPAC literature, it seems impossible to define a sampling frame for these, so the target remains undefined. Generalization with methods previously mentioned could not be performed. Adding these children only makes the results less interpretable with respect to generalization.

In summary, I think the paper would gain considerable depth if it at least addressed some of the issues raised in the previous paragraphs in the discussion.

Finally, whereas the authors are prudent about the variables used as driving selection, stating that they are not studying all possible sources of selection bias in ALSPAC, I believe that providing additional insight from the ALSPAC experience could be useful to the readership. What would a reasonable strategy be to choose the selection-influencing variables for identification (and used with IPW and the G-formula for example). This can be a complex task but most likely best addressed by investigators with an intimate knowledge of the study. Apart from these variables, a reflection on confounders is also needed for identification. Non-causal paths from confounders are familiar but non-causal paths created by selection bias, probably less so. In summary, based on the authors' profound knowledge of the ALSPAC data and the present study results, what suggestions can they make to investigators with respect to measuring key variables needed for identification and eventually generalization in a cohort study such as ALSPAC.

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

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

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Yes

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Epidemiology

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

Author Response 01 Jul 2022

Daniel Smith

We are delighted with the positive review and thank the reviewer for their constructive comments. We also apologise for the length of time taken to respond to the reviewer's comments; we were expecting a further review before revising the paper, but this never materialised. We have responded to the reviewer's comments in turn below (with the reviewer's original comments in italics).

I have a minor comment about the wording of one aspect of the report. I would strongly suggest changing the wording around the issue of bias related to "being sent a questionnaire". This is initially interpreted as a meaning bias originating from the research team when selecting who is to be sent a questionnaire whereas it has nothing to do with that. I am sure the authors can come up with a less ambiguous formulation: for example, "subjects who were reachable" or "subjects who could be contacted", etc.

We appreciate this suggestion and understand the confusion this unusual wording may introduce. However, 'being sent a questionnaire' seems our best choice for this outcome for a few reasons. First, it works as a nice counterpoint to the next outcome, 'returning a questionnaire'. Second, 'subjects who were reachable' is not technically accurate, as some participants were reachable but were not sent a questionnaire (if they had declined to participate in questionnaires, for example). As we are comparing participants who were sent vs "not sent" a questionnaire, we feel this is the most appropriate label for this outcome.

To try and avoid any potential confusion, in the second paragraph of the introduction where we first mention the outcome 'being sent a questionnaire', we have expanded the definition to say:

"The aim of this study was to describe questionnaire invitation and response rates and to explore factors associated with (i) being sent a COVID-19 questionnaire (**i.e., participants who were invited to complete a questionnaire because they had not withdrawn from the study, agreed to participate in questionnaires and had a valid email address**); (ii) returning a COVID-19 questionnaire..."

My main comment though is about the need for a more advanced discussion with regards to the

effects of selection bias; briefly, the effect can be biased in the sample (and therefore in the target population) or an unbiased in the sample but biased for the target. Without getting into the specific, and sometimes subtle, aspects of these observations, for sure, the reader would benefit from an enhanced discussion on the issue of generalizability from sample results to the target.

We appreciate this comment and thank the reviewer for raising this important issue. First, we added a sentence in the first paragraph of the introduction to briefly extend the implications of selection bias:

“...This can result in selection bias [7– 11], which may be exacerbated by the ways in which data collection has changed during the pandemic, for example, with the increased use of online questionnaires [12, 13]. **This selection bias can lead to incorrect (or biased) estimations of the effect of a risk factor on an outcome, in this case COVID-19 and its related impacts.** Importantly, having an available and rich collection of pre-pandemic data in existing longitudinal cohorts can be useful in efforts to explore potential selection pressures that lead to bias [14– 17].”

In the ‘study design’ section of the methods, we have also included an sentence explicitly defining the target population:

“The target population therefore comprises pregnancies leading to children born in the early 1990s in the former county of Avon, in addition to their mothers and the mothers’ partners. This initial enrolled sample included approximately 75% of the target population and was broadly representative of the wider Avon population (albeit somewhat biased towards mother who were married and from higher socioeconomic backgrounds, and biased away from ethnic minorities); further details can be found in the published cohort profiles [14, 21, 22]”

In addition to clearly defining the target population, we have also expanded our discussion with respect to representativeness. While our conclusions regarding selection bias are widely applicable globally for any studies concerned with COVID-19 research, we acknowledge that the patterns of selection are likely to vary for different populations that we would want to make inference to, and have tried to clarify this in the discussion:

“Nonetheless, we consider that the issues related to selection bias in COVID-19 research described throughout this paper are relevant to all COVID-19 observations studies globally. This is because being infected by SARS-CoV-2, experiencing COVID-19 and the measures that were used to limit the spread of infection (e.g., lockdowns) will influence who responds to invitations to participate in COVID-19 research. At the same time, who got tested and diagnosed would have been influenced by characteristics such as age, sex, occupation, socioeconomic position and existing co-morbidities, among other factors. However, as the distribution of these characteristics, and the management of the pandemic (e.g., universal or selected testing) will vary widely between countries, we must be careful in making inferences from our specific results to other populations. Indeed, even within the UK we have shown that many of the characteristics related to response in ALSPAC also do so in UK Biobank, but for several

factors the magnitude, and sometimes even the direction, of the associations differed [<https://www.medrxiv.org/content/10.1101/2021.12.10.21267363v1>]. The aim of this paper was to highlight key sources of selection bias in COVID-19 research; we feel that most observational studies (including genome-wide association studies and those using genetic data for causal inference such as Mendelian randomization studies) exploring causes or consequences of COVID-19 are likely to have such biases, and thus, our conclusions and suggestion that this should be explored in other studies are widely relevant.”

All studies have the goal of interpreting results from the sample to the target population of inferential interest. Few investigators are interested in biomedical results in a few hundred or even a few thousand children or mothers (i.e., the study sample). All are interested in knowing if sample results are generalizable to a well characterized population of children and/or mothers. Whereas the report recognizes that “ALSPAC is not necessarily representative of the general UK population or non-UK population” the target populations at any of the analyses time points in this report remain undefined, with the possible exception of recruitment time, assuming a sampling frame allowed the enumeration of eligible subjects. With the successive follow-ups and the marked losses at each step, defining the target at these steps is a challenge. However, this knowledge impacts on the usefulness of study results, including the ones shown in this report.

Thank you for raising this point. As mentioned above, we introduced a definition of the target population (children and associated mothers and their partners born in Avon in the early 1990s), which we believe helps to clarify things. We have now amended the sentence highlighted in the discussion to make it clearer that, as the ALSPAC target population is a relatively small geographically- and temporally-defined area, results may not be generalisable to wider UK and non-UK populations:

“Additionally, **given the target population (defined above)**, ALSPAC is not necessarily representative of the general UK population or non-UK populations”

In addition, in the second paragraph of the discussion we have made it clear that, due to selection at each successive stage explored here (sent a questionnaire vs returned a questionnaire vs item non-response), samples using this COVID-19 questionnaire data are not representative of the target population, and therefore at risk of selection bias:

“Recruitment was intended to be representative of the target population; however, due to loss to follow-up, the analytic samples with COVID-19 questionnaire data are no longer representative of the original ALSPAC study sample (and hence the target population, of which the original ALSPAC sample is broadly representative). Selection bias may therefore be a potential risk when using ALSPAC COVID-19 questionnaire data and we encourage researchers to carefully consider the results presented here to inform their work.”

While the issue of defining the target population of inferential interest may seem theoretical, it becomes very concrete when it comes to generalizing results from a sample meeting the

identification conditions. For example, if one were to use IPW or a G-formula approach (which by the way should be added to the list of methods in the report), distribution of key selection-influencing variables from the target would be needed (let's assume no confounding for simplicity). Should the distribution of these selection-influencing factors be from the eligible at recruitment or from those remaining in the study at later follow-up times? These may well represent different target populations. There are many possible selection models: at entry, at follow-up, at death, for item response. Using IPW for example, should the weights be specific to each mechanism or is the use of a simpler average weight over the mechanisms interpretable? More concretely, from a public health standpoint, assuming generalizations from the sample could be applied, how would the target populations be defined in these applications?

We appreciate this comment following the previous one. In addition to adding g-formula approaches to our list of potential methods, we have given an example of how one could apply inverse probability weights in this study (similar considerations would apply for multiple imputation or g-formula approaches). Given that the enrolled ALSPAC sample includes ~75% of the target population, using this 'whole sample' as a proxy for the target population is likely to be a reasonable approximation (even though there are some differences between those enrolled vs non-enrolled, as discussed above), especially given the difficulties in obtaining external information on the non-enrolled pregnancies in the target population:

"Methods such as multiple imputation, inverse probably weighting, g-formula approaches, simulations, and bounds and parameter searches [11, 30, 34– 38] can be used to help explore and overcome potential selection bias. **For instance, in the case of inverse probability weighting, weights should be derived according to the target population of the study (or using the total enrolled ALSPAC sample as a proxy, if weights derived from the target population cannot be constructed).**"

Another problem arises when the authors use additional children in the G1 cohort; they are defined as eligible but non-participant children: from the ALSPAC literature, it seems impossible to define a sampling frame for these, so the target remains undefined. Generalization with methods previously mentioned could not be performed. Adding these children only makes the results less interpretable with respect to generalization.

We have added the following in the methods section to make clear these children were part of the target population:

"Since the oldest children were approximately 7 years of age, the study has recruited 913 additional G1 children who did not join originally, but were **part of the original target population** based on date and location of birth."

In summary, I think the paper would gain considerable depth if it at least addressed some of the issues raised in the previous paragraphs in the discussion.

We have added the amendments explained above to make those important points clearer,

but we are concerned about the length of this manuscript, particularly the discussion. We hope these modifications improve those previously unclear or misleading issues.

Finally, whereas the authors are prudent about the variables used as driving selection, stating that they are not studying all possible sources of selection bias in ALSPAC, I believe that providing additional insight from the ALSPAC experience could be useful to the readership. What would a reasonable strategy be to choose the selection-influencing variables for identification (and used with IPW and the G-formula for example). This can be a complex task but most likely best addressed by investigators with an intimate knowledge of the study. Apart from these variables, a reflection on confounders is also needed for identification. Non-causal paths from confounders are familiar but non-causal paths created by selection bias, probably less so. In summary, based on the authors' profound knowledge of the ALSPAC data and the present study results, what suggestions can they make to investigators with respect to measuring key variables needed for identification and eventually generalization in a cohort study such as ALSPAC.

As we state in the discussion, although we cannot dictate how researchers analyse ALSPAC data, we agree that this might be useful to help guide some readers. In the original version we did present a few suggestions to help researchers (paragraph 6 of the discussion). However, given the length of the discussion already we are hesitant about providing much more additional detail – which could easily fill multiple additional papers! – especially given that we have referenced methods to mitigate selection bias already. Nonetheless, we have introduced a couple of sentences in the discussion to make clear that this study aimed to describe selection pressures in ALSPAC's self-reported COVID-19 data, and that it is beyond its scope to describe in detail how to mitigate the subsequent bias, but have also provided references to some ALSPAC studies which apply these methods in an effort to overcome potential selection bias:

"We have not aimed to mitigate selection bias in this paper, but rather to illustrate how it can be identified. Methods such as multiple imputation, inverse probability weighting, g-formula approaches, simulations, and bounds and parameter searches [11, 30, 34– 38] can be used to help explore and overcome potential selection bias. [...] **For some examples of such methods applied in ALSPAC to mitigate potential selection bias, see [references cited in updated manuscript]. Researchers need to assess the assumptions when using these approaches as** it is not possible to dictate a 'one-size-fits-all' approach when working with cohort data such as ALSPAC, as different research questions will be addressed using different variables and methods."

These references giving examples of how such approaches have been used in ALSPAC to overcome potential selection bias are:

Cornish RP et al. Factors associated with participation over time in the Avon Longitudinal Study of Parents and Children: a study using linked education and primary care data. *Int J Epidemiol.* 2021;50(1):293–302. 33057662 10.1093/ije/dyaa192 7938505

Cornish RP et al. Using linked educational attainment data to reduce bias due to missing outcome data in estimates of the association between the duration of breastfeeding and IQ

at 15 years. *Int J Epidemiol.* 2015;44(3):937–45. 25855709 10.1093/ije/dyv035 4521129

Lee KJ et al. Framework for the treatment and reporting of missing data in observational studies: The Treatment And Reporting of Missing data in Observational Studies framework. *Journal of clinical epidemiology.* 2021 Jun 1;134:79-88.

Cadman T et al. Joint associations of parental personality traits and socio-economic position with trajectories of offspring depression: Findings from up to 6925 families in a UK birth cohort. *JCPP advances.* 2021 Oct;1(3):e12028.

Millard LAC et al. Exploring selection bias in COVID-19 research: Simulations and prospective analyses of two UK cohort studies, *MedRxiv.* 2021.
(<https://www.medrxiv.org/content/10.1101/2021.12.10.21267363v1>)

Competing Interests: None.
