Prevalence of COVID-19, symptoms and testing in the U.K.

Initial findings from the COVID-19 Survey in Five National Longitudinal Studies

By Dylan M. Williams, Gabriella Conti, Nishi Chaturvedi, Alun Hughes, George B. Ploubidis and Richard J. Silverwood

CENTRE FOR LONGITUDINAL STUDIES

MRC Unit for Lifelong Health and Ageing

UKRI Economic and Social Research Council

Access the survey data

The COVID-19 survey data analysed in this briefing have been de-identified and are available for researchers. To download the data from four of the five studies coordinated by CLS (SN: 8658), visit the UK Data Service website (ukdataservice.ac.uk). Application procedures for NSHD data can be found at skylark.ucl.ac.uk.

Find out more

Email: clsfeedback@ucl.ac.uk
Visit: cls.ucl.ac.uk
Follow: @CLScohorts @MRCLHA
This briefing is one of a series produced by the Centre for Longitudinal Studies in collaboration with the MRC Unit for Lifelong Health and Ageing (LHA) using data from the COVID-19 Survey in Five National Longitudinal Studies.

Acknowledgements

This work was supported by the Economic and Social Research Council under the Centre for Longitudinal Studies, Resource Centre 2015-20, grant number ES/M001660/1, and by the Medical Research Council, grant MC_UU_00019/1.

We are extremely grateful to all the members of our studies for their contribution to this special COVID-19 survey and for their ongoing participation in our studies.

Citation

Williams, D.M., Conti, G., Chaturvedi, N., Hughes, A., Ploubidis, G.B., and Silverwood, R.J. (2020) *Prevalence of COVID-19, symptoms and testing: Initial findings from the COVID-19 Survey in Five National Longitudinal Studies*. London: UCL Centre for Longitudinal Studies.

Published by the UCL Centre for Longitudinal Studies, December 2020

Contents

About the survey

This briefing is based on data from a web survey of over 18,000 people, collected between 4 and 30 May 2020. The survey was completed by participants of five nationally representative cohort studies, who have been providing information about their lives since childhood.

These were:

- The Millennium Cohort Study (MCS), born in 2000-2002, part of ‘Generation Z’. They have been followed since birth and were age 19 at the time of the survey;
- Next Steps, who were born in 1989-1990, so-called ‘Millennials’. They have been followed since adolescence and are now age 30;
- 1970 British Cohort Study (BCS70) who were born in 1970, part of ‘Generation X’. They have been followed since birth and are now age 50;
- National Child Development Study (NCDS) who were born in 1958, into the later part of the ‘baby boomers’ generation. They have been followed since birth and are now age 62;
- National Study of Health and Development Study (NSHD) who were born in 1946, at the start of the ‘baby boomers’ generation. They have been followed since birth and are now age 74.

The survey was designed to help researchers understand the economic, health and social consequences of the novel coronavirus (SARS-CoV-2) outbreak, to give a unique insight into how people′s experiences during the pandemic vary depending on their earlier lives, and to be able to track the impact into the future. As part of the survey, response weights were created, and all the results in this briefing have been weighted, so that the results are as representative as possible of the full cohort of that age (for further information on weights and survey response data, see the survey User Guide). A number of further research briefings, using the data from the first wave of the COVID-19 survey, are available on the CLS website.

Introduction

The prevalence of COVID-19 in the community following the onset of the UK epidemic is unknown, and there are likely to be many predisposing factors which affect exposure to, or severity of, the disease. Clarity on these issues will help to inform public health strategies directed at virus suppression or elimination, and/or risk stratification measures tailored for different members of society. Here, we provide self-reported cohort-specific estimates of COVID-19 prevalence, symptoms and testing, along with estimates stratified by a range of traits. These estimates benefit from weighting for non-response using information from past data collections.

Key findings:

- The period prevalence of test-confirmed plus self-reported COVID-19 ranged from 2.1% to 10.8% across cohorts, with a clear indication of lower prevalence among the oldest study members.
- COVID-19 reporting was slightly higher among women than men, and in households where the respondents and/or their partners are key workers.
- The pattern of COVID-19 reporting across cohorts was largely concordant with serologically-confirmed prevalence of SARS-CoV-2 infections in England from the REACT-2 study.
- Testing for SARS-CoV-2 infections among participants was scarce, but followed the same distribution as the proportions that self-reported COVID-19 across cohorts.

Measures

COVID-19 prevalence: COVID-19 reporting, as described in this report, is estimated period prevalence of SARS-CoV-2 / COVID-19 as of 30 May 2020 – i.e. the proportion of the population to have been infected from the start of the epidemic up to this date. The primary measure of interest was defined as a report of having had a positive test result for SARS-CoV-2, or reporting of a strong personal suspicion of infection or medical advice to indicate that respondents had had COVID-19. Analyses were also conducted to estimate proportions that reported having received a positive test result without combining with self-reported disease.

Symptoms: The web survey asked respondents to consider whether they had suffered from 17 symptoms of COVID-19. Fever, dry cough, cough with mucus, shortness of breath, loss of smell and loss of taste were considered key symptoms, in accordance with NHS advice for COVID-19 vigilance at the time of analysis. We analysed proportions that reported three or more key symptoms primarily, and also evaluated proportions reporting two or more key symptoms, and proportions reporting any of the full 17 symptoms.

Testing: Respondents were asked whether they had been tested for COVID-19. This question did not require respondents to specify the type of test for infection that was administered to them, but tests based on sequencing of SARS-CoV-2 genetic material were predominantly in use for clinical and public health purposes at the time of the survey (as opposed to other technologies which have since emerged, such as antibody tests).

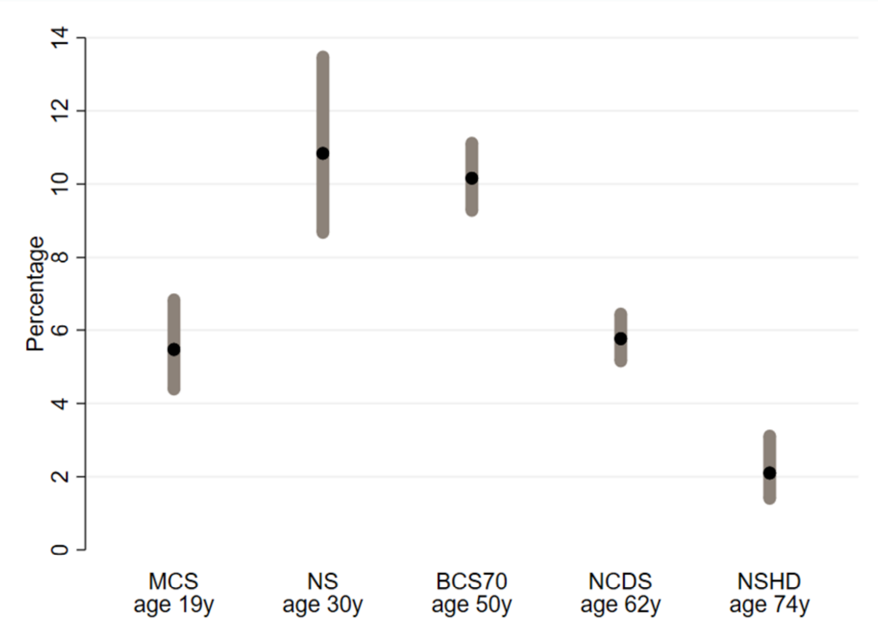
Results

Prevalence of COVID-19

Prevalence estimates for test-confirmed / self-reported COVID-19 ranged from 2.1% to 10.8% across cohorts, with the highest proportions being observed in cohorts exclusively of working age (Figure 1). This cross-cohort pattern was also observed solely for cases confirmed by positive tests, albeit with imprecise estimates due to very low proportions having been tested (Table 1). The pattern is unlikely to be wholly attributable to differences in compliance with social distancing guidelines, given the small differences in this behaviour across cohorts (with a trend for higher adherence across cohorts with older participants, where mean compliance ranged from 8.9 out of ten (95% CI: 8.8, 9.0) in MCS to 9.4 (9.3, 9.6) in NSHD). Rather, other factors that varied between cohorts might be responsible, including differences in susceptibility to and/or severity of infection (likely to be lowest in MCS participants), and factors that may affect exposure to infected individuals, such as current employment status and type of occupation, household size, and the number or type of social interactions.

COVID-19 was reported slightly more frequently in females than males: in a meta-analysis of absolute differences by sex across cohorts, COVID-19 was 1.2 percentage points higher in females than males (95% confidence interval (95% CI): 0.0, 2.3; for methods and full data, see the appendix). Across the four cohorts with working age participants (MCS, NS, BCS70 and NCDS), COVID-19 reporting was also 2.1 percentage points higher on average where a respondent and/or partner were key workers, relative to households without a key worker (95% CI: 0.5, 4.6; see appendix).

Figure 1 – Prevalences of confirmed or self-reported COVID-19 by cohort



Ranges show point estimates with 95% CI

Table 1 – Percentages of respondents that had tested positive for SARS-CoV-2 by cohort

Cohort	n	N	Percent	95% CI
MCS	8	2609	0.3	0.1, 0.8
NS	10	1876	0.6	0.3, 1.1
BCS70	28	4132	0.7	0.5, 1.0
NCDS	15	5119	0.3	0.2, 0.5
NSHD	1	1170	0.1	0.0, 0.5

n – number reporting a positive test; N – total sample size

There were no detectable differences in prevalence by pre-pandemic health status, educational attainment, current income, or ethnicity (the latter examined in MCS and NS only, the two cohorts that included sufficient numbers of participants of non-European ethnic ancestry) (results not shown). However, we emphasise that in several instances, we may have lacked power to establish differences in outcomes by these characteristics where sample sizes for sub-group analyses were limited – particularly for ethnicity.

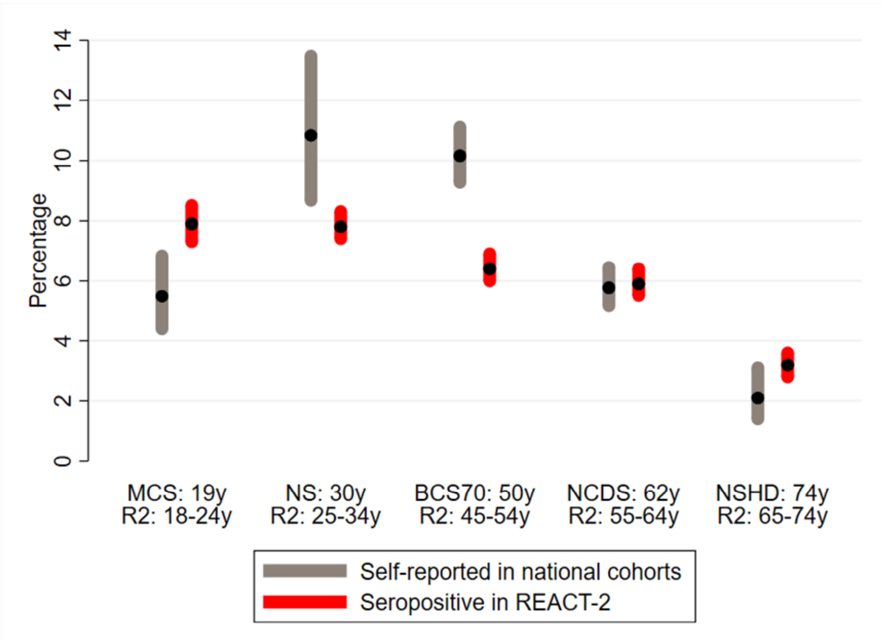
In Figure 2, test-confirmed / self-reported COVID-19 prevalence from the web survey are presented alongside seroprevalence estimates for SARS-CoV-2 (based on testing for blood-based antibodies against the virus) among 99,908 residents of England from the REACT-2 serology study.[1] Proportions reporting COVID-19 in the two oldest cohorts (NCDS, and NSHD) were concordant with serology results in corresponding age groups, suggesting that symptomatic disease reporting might have been approximately equivalent to infection status in these cohort participants. Alternatively, biases affecting each study differently might have produced convergent findings. In MCS, COVID-19 reporting was slightly lower than serostatus in the youngest REACT-2 age group. Positive serostatus and self-reported COVID-19 may both be underestimated among individuals with asymptomatic infections or mild disease. The disparity between findings from MCS and REACT-2 might have arisen if self-reporting of COVID-19 leads to relatively more under-reporting, but could also be due to other sample differences or chance. In cohorts exclusively of working age (NS and BCS70), self-reported COVID-19 prevalence was noticeably higher than seropositive proportions observed among similarly aged REACT-2 participants. The reasons for marked differences in these two cohorts are unclear. Different times of sampling (the first REACT-2 sampling finished later in the epidemic) and geographical distributions of participants (UK-wide in BCS70 versus England only in NS and REACT-2) are unlikely to have yielded over-reported prevalence specifically in these two cohorts and not others, given that total SARS-CoV-19 exposure increased over time and was higher in England than in other regions. With high

---

[1] Ward H. *et al* (2020) Antibody prevalence for SARS-CoV-2 following the peak of the pandemic in England: REACT2 study in 100,000 adults. *medRxiv*. 2020.08.12.20173690; doi: https://doi.org/10.1101/2020.08.12.20173690

proportions in the workforce, there might have been more anxiety or perceived risk of catching the virus by participants of NS and BCS70 than in the other three cohorts, which might have led to over-reporting. Differences in approaches used to weight for non-response or chance might also be responsible to some extent.

Figure 2 – Confirmed or self-reported COVID-19 prevalence in the national birth cohorts alongside seroprevalence estimates from the nearest corresponding age groups in the REACT-2 serology study



In the X-axis labels, the top line refers to the ages of each national cohort, and the second line refers to the age groups of REACT-2 participants

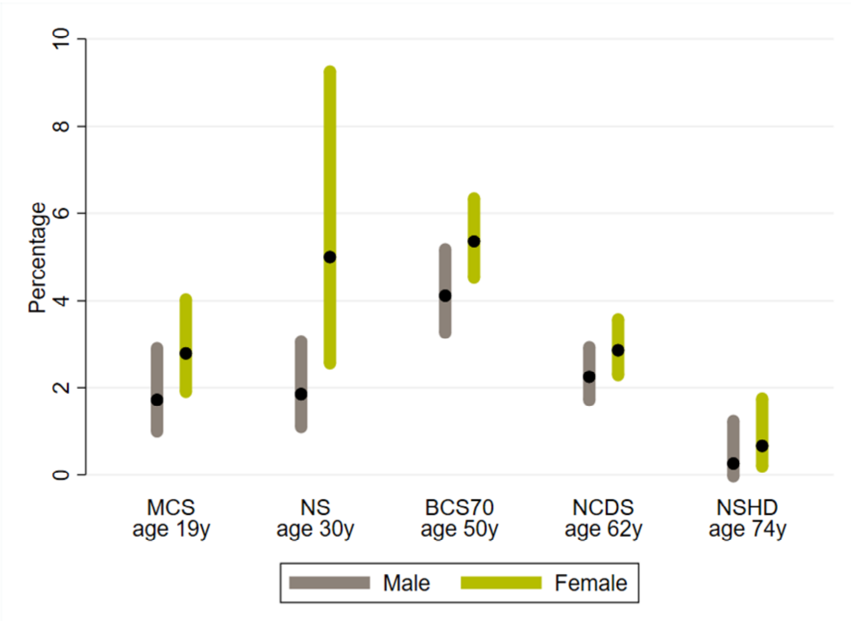
Seroprevalence estimates from REACT-2 were adjusted for test performance and re-weighted towards characteristics of the English population

COVID-19 symptoms

The pattern of respondents reporting three or more key symptoms by cohort was broadly consistent with the pattern of reporting of COVID-19 (Figure 3), though proportions reporting multiple symptoms were two to three times lower than the corresponding proportions for disease status – perhaps reflecting the diversity of COVID-19 symptoms that individuals experience beyond the key symptoms selected. Reporting of three or more symptoms was modestly higher in females, with a combined absolute difference of 0.6 percentage points (95% CI: 0.1, 1.2) from meta-analysis (data shown in the appendix). A much larger proportion of females reported having had any of the 17 symptoms; the combined difference being 9.3 percentage points higher in females than males (95% CI: 6.4, 12.1).

In contrast, there was no clear evidence for overall differences or consistent trends in key symptoms reporting across cohorts when stratified by pre-pandemic health status, key worker(s) in the household, educational attainment, income, or ethnicity (results not shown). As for differences in prevalence results, no notable evidence of differences may be due to a lack of power to detect differences.

Figure 3 – Reporting of three or more key COVID-19 symptoms by cohort and sex



Overall percentages and 95% CI in the unstratified samples were 2.3% (1.7, 3.1) in MCS, 3.6% (2.1, 6.0) in NS, 4.7% (4.1, 5.4) in BCS70, 2.5% (2.1, 3.0) in NCDS, and 0.5% (0.2, 1.1) in NSHD

Testing for SARS-CoV-2

As reported in section on COVID-19 prevalence, the proportions of respondents that had been tested for SARS-CoV-2 were very low, but a clear pattern across cohorts was present – consistent with the proportions in cohorts that had, or suspected having had, COVID-19 (Table 2).

Table 2 – Percentages of respondents tested for SARS-CoV-2 by cohort

Cohort	n	N	Percent	95% CI
MCS	58	2607	2.2	1.4, 3.6
NS	79	1872	4.2	2.8, 6.2
BCS70	148	4119	3.6	3.1, 4.2
NCDS	129	5086	2.5	2.1, 3.0
NSHD	16	1155	1.4	0.8, 2.2

n – number reporting testing; N – total sample size

In a combined analysis of the four cohorts in which we examined differences by key worker status (i.e. excluding NSHD), testing was 2.2 percentage points higher in households where the survey respondent and/or partner were key workers (95% CI: 0.1, 4.3; data shown in the appendix). There was no clear evidence for differences in testing proportions by sex, educational attainment, income, pre-pandemic health status, or ethnicity, but again, this may have been due to limited power (results not shown).

Conclusions

Across five national cohorts, test-confirmed or self-reported COVID-19 period prevalence as of 30 May 2020 ranged from 2.1% to 10.8%, with the lowest estimate from the oldest cohort.

These proportions were in keeping with other UK sources, including concurrent serological data on the prevalence of infections. Combining this survey information with lifelong data collections within the national cohorts will provide a unique opportunity to explore life course determinants of COVID-19 susceptibility and severity in future research, including the impact of traits and exposures that occurred in early life and may have predisposed individuals to COVID-19 decades later. The prospects for these research avenues can be enhanced with follow-up web surveys, detailed clinical studies on participants with and without evidence of infection, and more accurate establishment of COVID-19 exposure, which should include serological ascertainment.

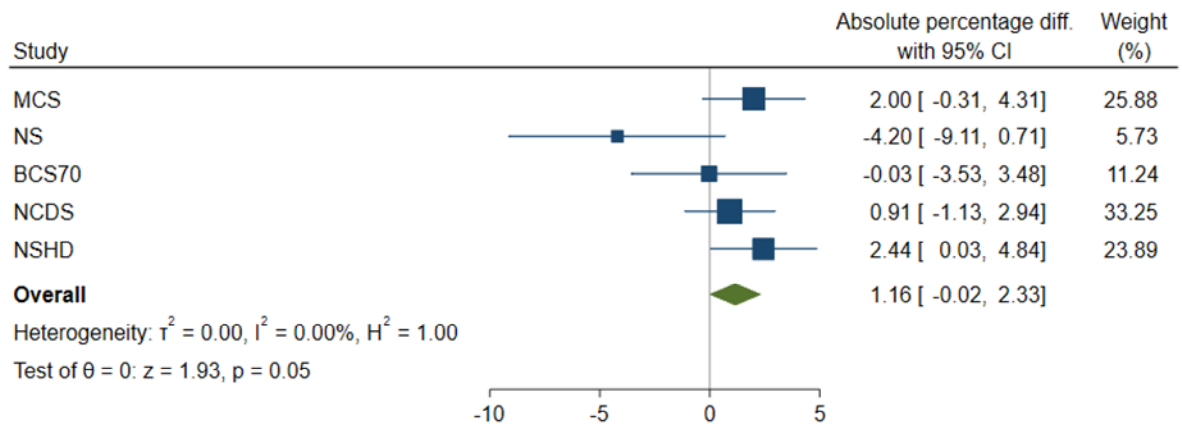
Appendix

Statistical methods used to produce the findings in this report

Proportions and mean compliance were estimated per cohort (i.e. not pooled) to allow for the application of cohort-specific weighting for non-response, and adjustment for survey sampling in estimates for MCS and NS. Confidence intervals for proportions were calculated using the Agresti-Coull method.[2]

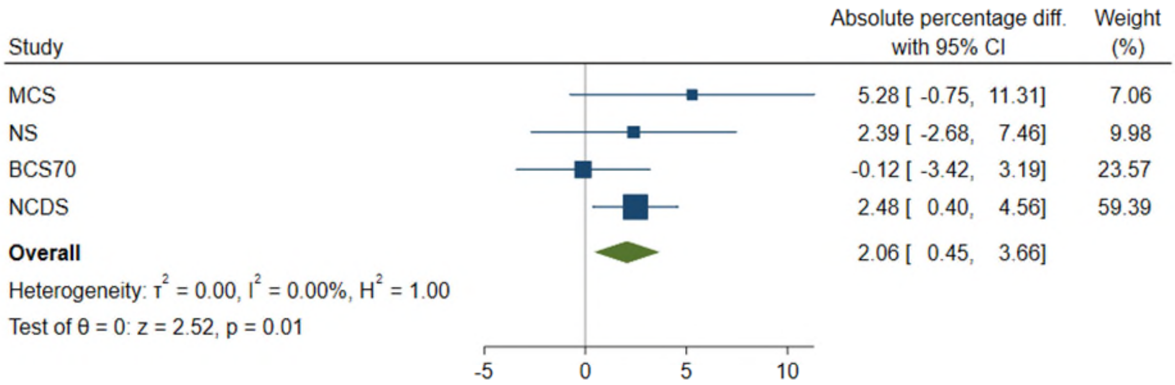
To compare overall differences in COVID-19 prevalence, symptoms and testing proportions across cohorts, we first evaluated cohort-specific differences in these three outcomes by all characteristics examined using Wald tests. We then conducted random-effects meta-analyses of point estimates and standard errors for each cohort-specific difference in the outcomes by characteristics, using restricted maximum likelihood. The consistency of differences across cohorts is quantified by heterogeneity statistics, e.g. a low I^2 statistic meaning differences were similar across cohorts. Forest plots of the analyses cited in the main document are displayed below. Data for any other null findings referred to in the text are available on request.

Supplemental figure 1: Meta-analysis of percentage differences in confirmed or self-reported COVID-19 prevalence by sex (females relative to males)



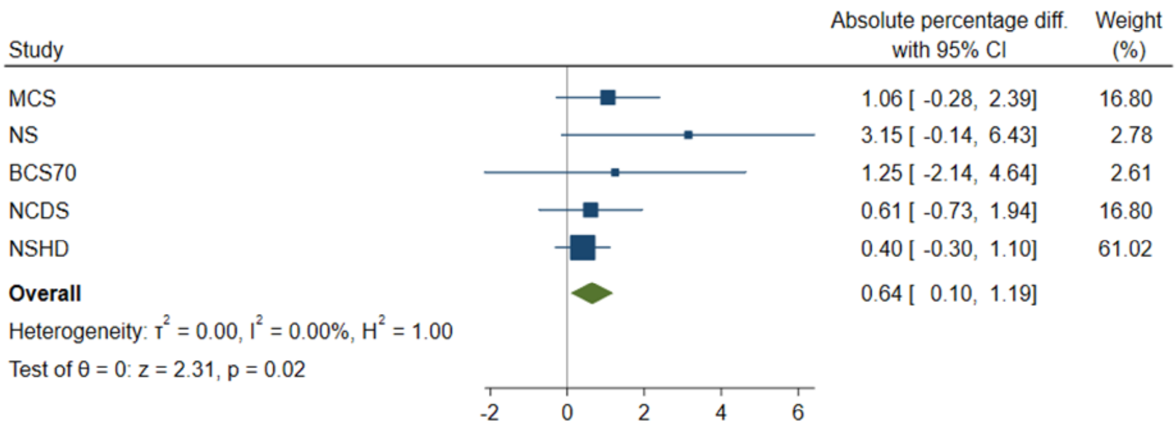
[2] Agresti A., Coull B.A. (1998) Approximate is better than "exact" for interval estimation in binomial proportions. *The American Statistician*. 52(2) 119-126.

Supplemental figure 2: Meta-analysis of percentage differences in confirmed or self-reported COVID-19 prevalence by key worker status (households with key worker(s) relative to households without)

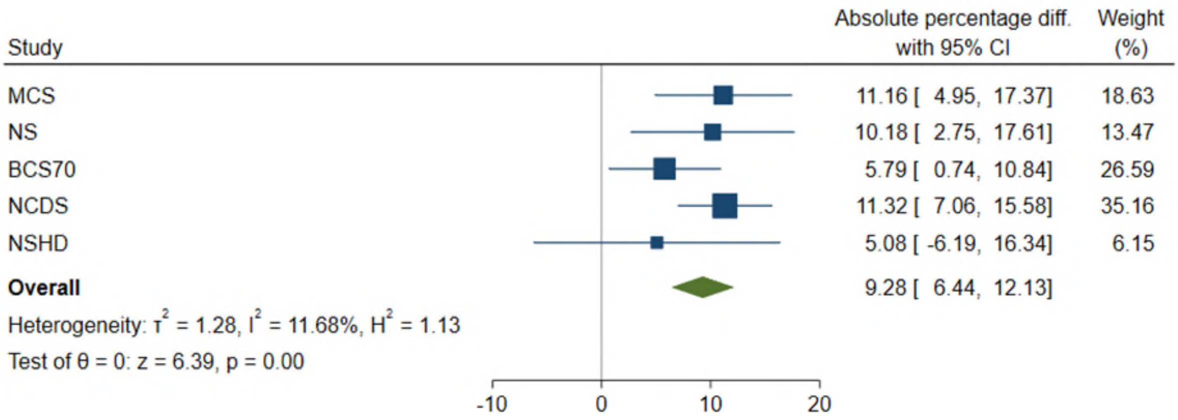


Note: NSHD data were not included in key worker analyses because very few cohort participants are in the work force

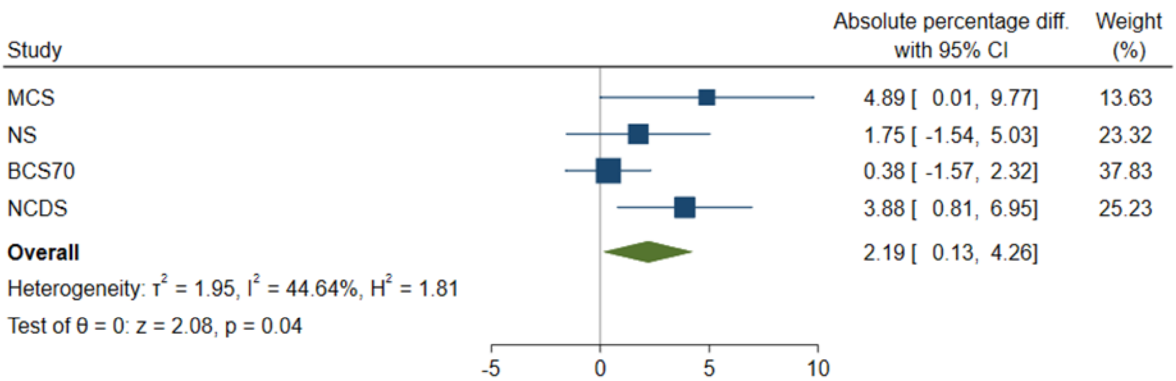
Supplemental figure 3: Meta-analysis of percentage differences in proportions reporting three or more key COVID-19 symptoms by sex (females relative to males)



Supplemental figure 4: Meta-analysis of percentage differences in proportions reporting any of the 17 COVID-19 symptoms by sex (females relative to males)



Supplemental figure 5: Meta-analysis of percentage differences in SARS-CoV-2 testing by key worker status (households with key worker(s) relative to households without)



Note: NSHD data were not included in key worker analyses because very few cohort participants are in the work force