Is the presence of children in a household protective against the risk of COVID-19 infection and severe outcomes?

Study Protocol

Version: v1.0

Date: 10/07/2020

This is a collaboration between the following institutions as part of OpenSAFELY.org:

- The DataLab, Nuffield Department of Primary Care Health Sciences, University of Oxford, OX26GG
- Electronic Health Records Research Group, Department of Non-communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine

Table of Contents

Background

Susceptibility to COVID-19 disease may be influenced by prior infection with other coronaviruses. Four human coronaviruses (hCoVs) are endemic globally and are estimated to cause 10 to 30% of upper respiratory tract infections in adults.[1] These common hCoVs are hypothesised to provide some immunity against infection with SARS-CoV-2 or severe outcomes of COVID-19 disease. Neutralizing antibodies induced by previous hCoV infections essentially block future infections; these antibodies appear to have some cross-neutralizing activity against SARS-CoV-2.[2,3] Protective immunity from these hCoVs may only be sustained for a short period of time, as the duration of immunity to hCoVs in some individuals is as little as 6 to 12 months.[4] However, even small amounts of protective coronavirus immunity within populations may have important impacts on the dynamics of COVID-19 epidemiology.

We can assume that adults in close contact with children have a higher frequency of viral upper respiratory tract infections, with more frequent infections among adults exposed to younger children and among women.[5] This is reflected in parents having greater resistance to common cold viruses, such as rhinovirus and influenza virus, than adults without children.[6] As such, parents may have some protective immunity against SARS-CoV-2, due to previous exposure to hCoVs. Children may, by contrast, increase the risk of infection with SARS-CoV-2 within a household, due to increased risk of SARS-CoV-2 exposure in school environments, negating any protective effects.[7] However, emerging evidence suggests that children may play a minor role in the transmission of SARS-CoV-2; primary care and household study data from the Netherlands suggests SARS-CoV-2 is mainly spread between adults and from adult members of a household to children.[8]

With many countries determining policy related to school opening, understanding the net risk of SARS-CoV-2 infection and COVID-19 disease to adults with and without children in a household is critical. We therefore carried a cohort study using UK data from the OpenSafely platform.

Objectives

Primary Objectives

To determine whether the risk of being admitted into ICU or dying from COVID-19 differs between adults living in households with and without school age children.

Secondary Objectives

To determine whether the risk of developing clinical infection with SARS-CoV-2 differs between adults living in households with and without school age children.

Exploratory Objectives

To confirm that our household identifier correctly identifies parents/care-givers and children we will conduct a similar analysis over an earlier time period in 2018-2019 where the outcome is consultation with threadworm infection (a condition for which we would anticipate transmission from children in nurseries and early primary school years to adults care-givers).

Methods

Database Description

The analyses will be conducted using the OpenSAFELY platform, which has been previously described in detail (Williamson et al, 2020). In addition OpenSAFELY will include a linkage to the Intensive Care National Audit & Research Centre (ICNARC) database.[9]

Study Design and Population

Adults (males and females 18 years and above) currently registered as active patients in a TPP general practice in England on 1st February 2020.
For the primary analysis, we will follow participants from 1st of February 2020 until the earliest of:

1. Admission to ITU with COVID-19 or death from COVID-19
2. Deregistration from GP practice (if available)
3. Death from other cause
4. Latest TPP linkage

We will censor datasets prior to the earliest of the linkages to make the datasets comparable.

For the secondary analysis, we will follow participants from 1st of February 2020 until the earliest of:

1. Diagnosis of COVID-19 in primary care
2. Deregistration from GP practice (if available)
3. Death
4. Latest TPP linkage

Currently, the date of deregistering from a GP practice is not available in TPP. Work is ongoing to include this variable in study definitions, and, should it become available during the study conduct this will additionally be used to censor individuals. Early on during the pandemic we expect this to have made little difference however, as few individuals are expected to have swapped GP practices during lockdown due to the restrictions on movement.

Inclusion Criteria

At least 3 months of prior follow-up in the GP practice (allowing some time for GP records to be updated following a patient changing practices, while minimising loss of households with young children who may move more frequently).
Households where any individuals have missing recorded sex, age, and measure of deprivation status.

Exclusion Criteria

Household sizes greater than 10 individuals

Study Measures

Exposure

Primary Exposure

An ordered categorical variable with three levels:

1. No children under 18 years in the household
2. Any children under 12 years of age in the household
3. No children under 12 years of age but one or more children aged 12-18 years in the household

Outcome

Primary outcome:
A composite outcome of ITU admission or death with COVID-19 ascertained from ICNARC data on ITU admission with COVID-19 and ONS death certificate data.

Secondary outcome

Evidence of SARS-CoV-2 infection in primary care defined as primary care codes that are judged to best define COVID-19 cases.

The primary care coding for COVID-19 is being explored in a parallel project (https://github.com/opensafely/covid-positive-infection-research). The specific codes used to define COVID-19 cases will depend on the conclusions from this project, and it is therefore not possible to prespecify the codes that will be used at this stage. However, the codes will all be shared as part of opencodelist and available for inspection.

Covariates

The presence of children in the house will be defined using the household identifier, developed by TPP, which identifies people living in the same property. Additional covariates include:

- Demographics: Age, sex, body mass index (BMI; kg/m2), smoking status, Index of Multiple Deprivation, geographical variation (STP), the total number of people in the house categorised as 1, 2, 3, 4, 5 or more (if this leads to problems with model fitting we will consider revising the boundaries after seeing the range and frequency of numbers).
- Chronic comorbidities (all defined as ever diagnosis prior to index date unless specified otherwise): asthma, chronic respiratory disease, chronic heart disease, diabetes mellitus, chronic liver disease, stroke/dementia, other chronic neurological diseases, common autoimmune diseases (Rheumatoid Arthritis (RA), Systemic Lupus Erythematosus (SLE) or psoriasis), solid organ transplant, asplenia, immunosuppression (including permanent immunodeficiency such as HIV, sickle cell disease, other immunosuppressive conditions and temporary immunodeficiency), cancer (haematological or non haematological diagnosed < 1 year ago, 1-4.9 years ago or ≥5 years ago), estimated GFR (eGFR), renal replacement therapy, high blood pressure or diagnosis of hypertension.
- Ethnicity: categorised as 5-level variable (White, Mixed, South Asian, Black or Other)
- Probable shielding behaviour, including organ transplant recipients, renal replacement therapy, haematological cancers, non-haematological cancers, immunodeficiencies/asplenia and severe respiratory conditions.

Due to anticipated missing data we will adjust for ethnicity only in a complete record sensitivity analysis.

Statistical Analysis

Primary Model

We will describe the proportion of individuals developing each outcome, by the covariates. In initial descriptive analyses, we will describe the risk of primary and secondary outcomes according to the presence of children in the household, the number of children and the age of the oldest and youngest child in the household.

We will use Cox proportional hazards modelling to determine hazard ratios for the primary and secondary outcomes using robust standard errors to account for the clustering by household identifier, stratifying by geographic area (STP) to allow for regional variation in infection rates. In sequential models we will adjust for sex and age using restricted cubic splines, demographic covariates (+ additional people in household, IMD, BMI and smoking) and fully adjusted models (+additional clinical comorbidities). We will report Hazard ratios with 95% confidence intervals. We will check for collinearity of variables, particularly comorbidities.

Identifying true confounders of the relationship between having children and risk of severe outcomes requires identifying severity of conditions prior to conception, and around childbearing age of people without children. We felt this would be difficult to define accurately using primary care data, particularly in young adults who may change GP practices regularly and therefore have data re-entered. Therefore we will examine the impact of chronic conditions that may differ between people with and without children by comparing HRs in sequentially adjusted models. Violations of the proportional hazards assumption will be explored by testing for a zero slope in the scaled Schoenfeld residuals and by visual inspection of plots of the Schoenfeld residuals.

In the primary analysis, we will assume those with missing BMI to be non-obese and those with missing smoking information to be non-smokers on the assumption that both obesity and smoking would be likely to be recorded if present.

Secondary models

- We will examine for an interaction with age of the household resident between adults: 18-66 (primary analysis: adults of working age) and older adults: >66 years
- We will examine for an interaction between men and women
- We will examine for an interaction between time periods before and after 3rd April 2020 (2 weeks after school closure to allow for infections related to school transmission to become clinically evident, and after lockdown).
- We will examine for an interaction by probable shielding behaviour comparing the risk to households where adults were within a risk group who were likely to adopt early shielding behaviour .
- We will examine for a ‘dose response’ effect of exposure to previous infections by adding the number of children recategorised as:
    - none
    - only ≥12 years
    - 1 child <12
    - 2 children <12
    - 3+ children <12

Sensitivity analyses

1. We will repeat the main analysis, including a minimum 12 months follow-up in general practice, to allow time to capture pre-existing comorbidities. Within TPP, transfer of electronic medical records from other GP practices is very quite good, however there will be some patients without a pre-existing medical record (such as immigrants) whose pre-existing comorbidities will be missing.
2. We will repeat the main fully adjusted model, adding ethnicity (5 categories) as an additional covariate
3. We will repeat the main fully adjusted model, adding ethnicity and also including BMI and smoking with records with missing data excluded.
4. We will repeat the main fully adjusted model including BMI and smoking with records with missing data excluded.
5. Additional definition of COVID-19 in primary care to include codes for suspected cases.

Software and Reproducibility

Data management will be performed using Python 3.8 and SQL with analysis carried out using Stata 16. Code for data management and analysis as well as codelists archived online https://github.com/opensafely/school-age-children-and-covid

Sample size calculations

Assuming an alpha-level of 0.05 and inclusion of at least 8 million households with no dependants and 8 million with children, and 0.02% of patients experiencing COVID19 death or ITU admission during the study period (the rate of primary outcome in the 30-40 year age group), we would have 90% power to detect a hazard ratio of 0.89 or larger.
Using a more conservative definition of 12 million households with no children and 4 million with children we would have 97% power to detect HRs of 0.85 and above.

Table Shells

Table 1. Cohort description by presence of children or young people in the household.

	N (column %)	No children under 18 years in the household (% within stratum)	Any children under 12 years of age in the household (% within stratum)	No children under 12 years of age but one or more children aged 12-18 years in the household (% within stratum)
Total				
Age				
18-<30				
30-<40				
40-<50				
50-<60				
60-<70				
70-<80				
80+				
Sex				
Female				
Male				
BMI (kg/m2)				
<18.5				
18.5-24.9				
25-29.9				
30-34.9 (Obese class I)				
35-39.9 (Obese class II)				
≥40 (Obese class III)				
Missing				
Smoking				
Never				
Former				
Current				
Missing				

Ethnicity				
White				
Mixed				
South Asian				
Black				
Other				
Missing				
IMD quintile				
1 (least deprived)				
2				
3				
4				
5 (most deprived)				
Total number people in household				
1				
2				
3				
4				
5 or more				
Blood pressure				
Normal				
Elevated				
High Stage 1				
High Stage 2				
Missing				
High bp or diagnosed hypertension				
Comorbidities				
Chronic respiratory disease ex asthma				
Asthma				

Chronic cardiac disease				
Diabetes**				
Haematological cancer				
Diagnosed < 1 year ago				
Diagnosed 1-4.9 years ago				
Diagnosed ≥5 years ago				
Non-haematological cancer				
Diagnosed < 1 year ago				
Diagnosed 1-4.9 years ago				
Diagnosed ≥5 years ago				
Reduced kidney function				
Estimated GFR 30-60				
Estimated GFR <30				
Renal replacement therapy				
Chronic Liver disease				
Stroke/dementia				
Other neurological disease				
Solid organ transplant				
Asplenia				
Rheumatoid/Lupus/ Psoriasis				
Other immunosuppressive condition				

Table 2. Cohort description with number of SARS-CoV-2 clinical infections and COVID-19 ITU admissions or deaths by covariates

	N (column %)	Number of SARS-CoV-2 clinical infections (% within stratum)	Number of COVID-19 ITU admissions or deaths (% within stratum)
Total			
Age			
18-<30			
30-<40			
40-<50			
50-<60			
60-<70			
70-<80			
80+			
Sex			
Female			
Male			
BMI (kg/m2)			
<18.5			
18.5-24.9			
25-29.9			
30-34.9 (Obese class I)			
35-39.9 (Obese class II)			
≥40 (Obese class III)			
Missing			
Smoking			
Never			
Former			
Current			
Missing			
Ethnicity			
White			
Mixed			
South Asian			
Black			

Other			
Missing			
IMD quintile			
1 (least deprived)			
2			
3			
4			
5 (most deprived)			
Total number people in household			
1			
2			
3			
4			
5 or more			
Blood pressure			
Normal			
Elevated			
High Stage 1			
High Stage 2			
Missing			
High bp or diagnosed hypertension			
Comorbidities			
Chronic respiratory disease ex asthma			
Asthma			
Chronic cardiac disease			
Diabetes**			
Haematological cancer			
Diagnosed < 1 year ago			
Diagnosed 1-4.9 years ago			
Diagnosed ≥5 years ago			

Non-haematological cancer			
Diagnosed < 1 year ago			
Diagnosed 1-4.9 years ago			
Diagnosed ≥5 years ago			
Reduced kidney function			
Estimated GFR 30-60			
Estimated GFR <30			
Renal replacement therapy			
Chronic Liver disease			
Stroke/dementia			
Other neurological disease			
Solid organ transplant			
Asplenia			
Rheumatoid/Lupus/ Psoriasis			
Other immunosuppressive condition			

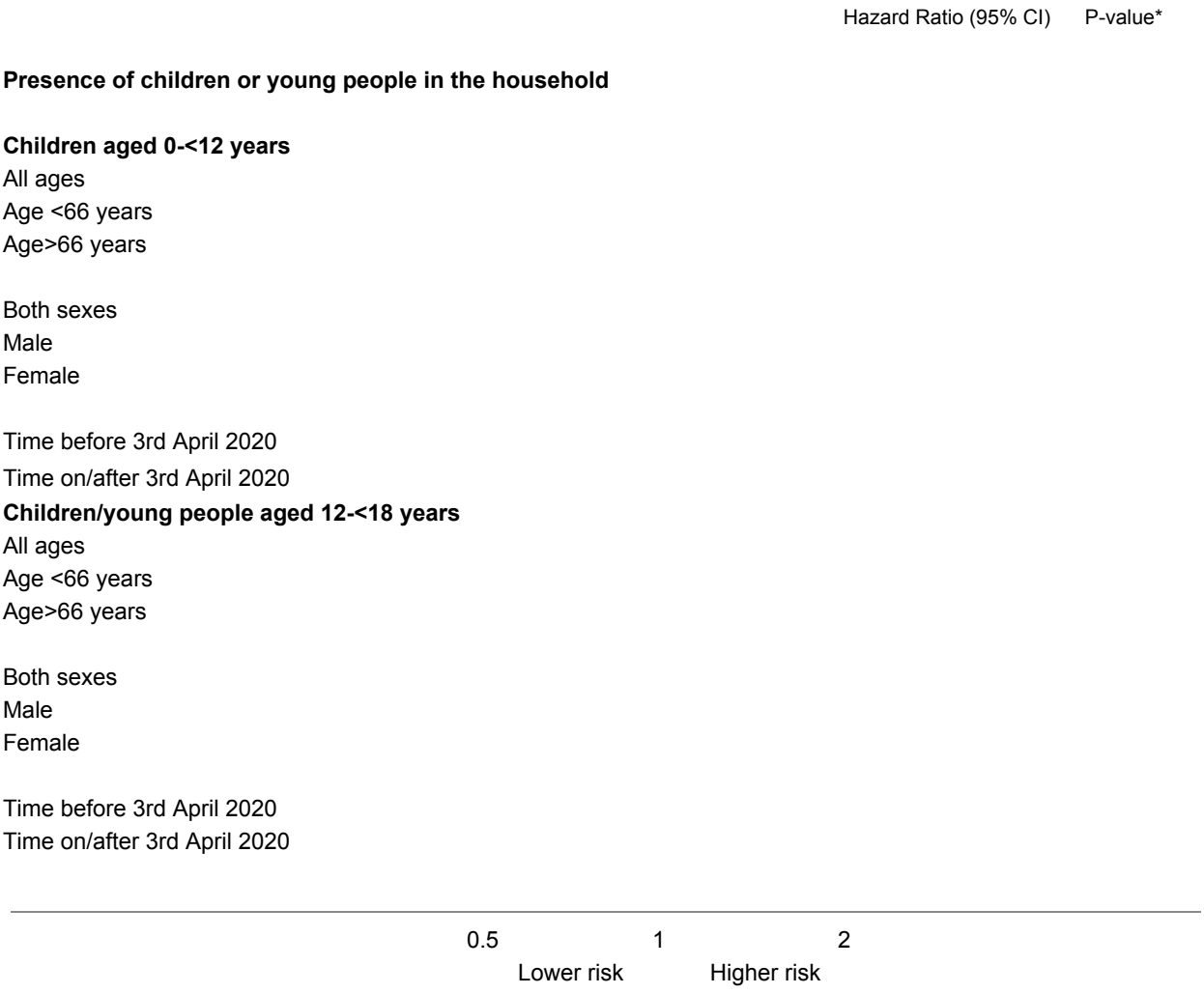
Table 3. Hazard Ratios (HRs) and 95% confidence intervals (CI) for SARS-CoV-2 clinical infections

	SARS-CoV-2 clinical infection HR (95% CI)					
	N (column %)	Person years follow-up	Rate (per 1000 person weeks)	Age-sex adj	Additionally adjusted for additional people in household IMD, BMI, smoking	Fully adj
Presence of children or young people in the household						
None					1.00 (ref)	1.00 (ref)
Children aged 0-<12 years						
Children/young people aged 12-<18 years						
Number of children aged 1-<11 years in household						
None						
1						
2						
3						
4 or more						

Table 4. Hazard Ratios (HRs) and 95% confidence intervals (CI) for COVID-19 ITU admission or death

	COVID-19 Death HR (95% CI)					
	N (column %)	**Person years follow-up**	**Rate (per 1000 person weeks)**	**Age-sex adj**	**Additionally adjusted for additional people in household IMD, BMI, smoking**	**Fully adj**
Presence of children or young people in the household						
None					1.00 (ref)	1.00 (ref)
Children aged 0-<11 years						
Children/young people aged 11-<18 years						
Number of children aged 1-<11 years in household						
None						
1						
2						
3						
4 or more						

Figure 1. Hazard Ratios (HRs) and 95% confidence intervals (CI) for each outcome, compared to having no children in the household, by age group, sex and time periods before and after 3rd April 2020



*P-Value for interaction

Limitations

- Our analysis is based on the assumption that the presence of children is a proxy measure for the frequency of colds. However, school age children may also have provided additional opportunity for infection from SARS-CoV-2 through exposure at schools and we would anticipate that these two factors have different directions of effect. Therefore our outcome can be seen in terms of the net effect of these two factors.
- We lack information about occupation, both in terms of exposure to SARS-CoV-2 (such as healthcare workers) and work with children (such as nursery workers) who may not have children but may have frequent exposure to cold viruses.
- We will misclassify the presence of children in a household for divorced or separated parents where the child is registered with a GP from the other parents address. In addition, children of divorced or separated parents were able to travel between homes during lockdown and could therefore provide additional opportunities for infection.
- We were not able to adjust for comorbidities that affected ability to have children and subsequent risk of development of severe outcomes from SARS-CoV-2 (true confounders such as Type 1 diabetes). However, we examined the impact on adjusting for serious comorbidities in sequentially adjusted models.
- We anticipate that ethnicity may be strongly associated with household numbers but due to missing data were not able to adjust for it in the main model.
- Limitations in the granularity of data may mean that we misclassify the number of people living in a household, e.g. for flats within a larger property.
- In this analysis we are assuming a constant relationship between infections between people through clustering at the household level rather than detailed modelling of how infections are transmitted within households.
- Splitting analysis time could mean that people who were most susceptible to developing infection with SARS-CoV-2 or severe outcomes from it all developed these within the early study period. However national seroprevalence data has shown that a minority of people showed evidence of infection within this time period, limiting impact of this.

References

1. Monto, A. S. Epidemiology of viral respiratory infections. *Dis. Mon.* **49**, 160 (2003).
2. Jiang, S., Hillyer, C. & Du, L. Neutralizing Antibodies against SARS-CoV-2 and Other Human Coronaviruses: (Trends in Immunology 41, 355-359; 2020). *Trends Immunol.* **41**, 545 (2020).
3. Grifoni, A. *et al.* Targets of T Cell Responses to SARS-CoV-2 Coronavirus in Humans with COVID-19 Disease and Unexposed Individuals. *Cell* (2020) doi:10.1016/j.cell.2020.05.015.
4. Edridge, A. W. D. *et al.* Coronavirus protective immunity is short-lasting. doi:10.1101/2020.05.11.20086439.
5. Monto, A. S. & Ross, H. Acute respiratory illness in the community: effect of family composition, smoking, and chronic symptoms. *Br. J. Prev. Soc. Med.* **31**, 101–108 (1977).
6. Sneed, R. S., Cohen, S., Turner, R. B. & Doyle, W. J. Parenthood and host resistance to the common cold. *Psychosom. Med.* **74**, 567–573 (2012).
7. Ball, T. M., Holberg, C. J., Aldous, M. B., Martinez, F. D. & Wright, A. L. Influence of attendance at day care on the common cold from birth through 13 years of age. *Arch. Pediatr. Adolesc. Med.* **156**, 121–126 (2002).
8. National Institute for Public Health and the Environment. Children and COVID-19. *National Institute for Public Health and the Environment* www.rivm.nl/en/novel-coronavirus-covid-19/children-and-covid-19.
9. *COVID-19 Report*. https://web.archive.org/web/20200425133758/https://www.icnarc.org/Our-Audit/Audits/Cmp/Reports (2020).