Seroprevalence of SARS-CoV-2 in children and adolescents.

Dr Helen Ratcliffe

St. Cross College
University of Oxford

A thesis submitted in partial fulfilment of the requirements for the degree of
Doctor of Philosophy

Trinity 2023

Supervisors:
Professor Matthew D Snape, Department of Paediatrics, University of Oxford, Oxford, Oxfordshire.

Professor Merryn Voysey, Department of Paediatrics, University of Oxford, Oxford, Oxfordshire.

Dr Karen Tiley, Department of Paediatrics, University of Oxford, Oxford, Oxfordshire.

Professor Susanna Dunachie, Nuffield Department of Medicine, University of Oxford, Oxfordshire.

Table of Contents

Glossary 10

Chapter 1 Introduction 17

1.1 The emergence of SARS-CoV-2 17
1.1.1 Epidemiology 17
1.1.2 Investigating the emergence of SARS-CoV-2 19
1.1.3 Incidence and prevalence 19
1.1.4 Spectrum of disease 22
1.1.5 Symptoms 23
1.1.6 Transmission 24
1.1.7 Risk factors associated with SARS-CoV-2 infection and severe COVID-19 disease 26

1.2 The SARS-CoV-2 virus and the body's immune response to infection 27
1.2.1 Coronaviruses and the identification of SARS-CoV-2 27
1.2.2 Virology of SARS-CoV-2 28
1.2.3 Variants of SARS-CoV-2 28

1.3 The immune response to SARS-CoV-2 infection 31
1.3.1 Overview of innate and adaptive immune responses to SARS-CoV-2 infection 31
1.3.2 Cross-reactive immunity with seasonal coronaviruses 32
1.3.3 Differences in immune responses with age 33

1.4 Diagnostics 35
1.4.1 PCR testing 35
1.4.2 Rapid antigen testing 36
1.4.3 Serological testing 36
1.4.4 Saliva/ oral fluid 37

1.5 Development of a paediatric prospective serological study in England 37

1.6 Aims and objectives of the thesis 38

Chapter 2 Methodology used in the 'What's the STORY?' study 40

2.1 Introduction 40
2.1.1 Aims 40

2.2 Study design 40
2.2.1 Study sites 40
2.2.2 Recruitment 44
2.2.3 Sample Size 49
2.2.4 Group one 49
2.2.5 Representativeness of participants sampled 49
2.2.6 Sample size for SARS-CoV-2 incidence based on change in prevalence 51
2.2.7 Analysis populations 52
2.2.8 Group 3 52
2.2.9 Longitudinal cohort 52
2.2.10 Screening and Eligibility Assessment 53
2.2.11 Informed Consent 54
2.2.12 Collection of sample 55
2.2.13 Data capture and Questionnaire 56
2.2.14 Vaccination history 57
2.2.15 Reimbursement 57

2.3 Laboratory methods 57
2.3.1 Serological assays 57
2.3.2 Crevicular fluid 59
2.3.3 DNA storage 59

2.4 Approvals 60

Chapter 3 A comparison of the 'STORY' study sample demographics with those of children and teenagers living in England 61

3.1 Aims 61

3.2 Capturing sample demographics 61

3.2.1 Methods 62

3.3 Results 63

3.3.1 Recruitment methods 63
3.3.2 Participant characteristics 65
3.3.3 Socioeconomic representativeness of the population at a national and regional level 73
3.3.4 Country of birth 80
3.3.5 Household size and childcare in the under five years population 81

3.4 Discussion 81

Chapter 4 Community seroprevalence of SARS-CoV-2 in children and adolescents in England, 2019-2021 90

4.1 Aims 90

4.2 Introduction 90

4.3 Methods 90

4.3.1 Study design 90
4.3.2 Data collection 91
4.3.3 Measurement of serum antibodies 92

4.4 Statistical analysis 92

4.4.1 Seroprevalence 92
4.4.2 Risk factors 93
4.4.3 Symptoms 94

4.5 Results 94

4.5.1 Comparison of assays 99
4.5.2 Seroprevalence 101
4.5.3 Presence of symptoms in seropositive participants 108
4.5.4 Risk factor analysis 114

4.6 Discussion 119

Chapter 5 Persistence of antibody concentrations against SARS-CoV-2 over time..... 123

5.1 Introduction 123

5.2 Methods 123

5.3 Results 126

5.3.1 Children are predicted to maintain a SARS-CoV-2 stable immune response to spike protein receptor binding domain. 126
5.3.2 Oral fluid concentrations of IgG wane more quickly than serum antibodies within the paediatric population 130

5.4 Discussion 132

Chapter 6 Serum HCoV-spike specific antibodies do not protect against subsequent SARS-CoV-2 infection in children and adolescents. 136

6.1 Introduction 136

6.2 Methods 137

6.2.1 Unmatched analysis 138
6.2.2 Matched case control analysis 138

6.3 Results 139
6.3.1 Assay concordance 139
6.3.2 SARS-CoV-2 antibodies in children cross-react with Beta HCoV. 140
6.3.3 HCoV spike protein antibodies do not confer protection against SARS-CoV-2 infection 142

6.4 Discussion 145

Chapter 7 Discussion 149

7.1 Establishing a nationally representative paediatric serobank 149
7.1.1 Insights into COVID-19 epidemiology gained from STORY 151
7.1.2 Insights into COVID-19 immunology gained from STORY 152

7.2 Challenges in the 'STORY' study 153
7.2.1 Running a multi-site seroepidemiological study 153
7.2.2 Processing of serological samples during a pandemic 154
7.2.3 Representation from populations of interest 155
7.2.4 The role of translations in aiding recruitment in minority ethnic groups 157
7.2.5 Building ties to the community in groups who are not well represented in the population – both adults and paediatrics 158

7.3 Further work 162
7.3.1 Applications for samples 162

List of figures

Figure 1 Timeline of COVID-19 cases in the United Kingdom from John Hopkins University CSSE COVID-19 Data adapted to show the timing of measures taken to prevent the spread of disease[3,4] 18
Figure 2 Timeline of SARS-CoV-2 variants circulating within the UK adapted to show recruitment periods for 'STORY' and 'COVID Warrior' studies. Data from John Hopkins University CSSE COVID-19 Data[3] 30
Figure 3 A map to illustrate the seven NHS regions in England 41
Figure 4 shows sites recruiting for the 'STORY' study. Postcode districts around each site were divided into – postcode districts selected by UKHSA for targeted recruitment (red), postcode districts within the site's catchment area but not selected (yellow) or out of range of the study site to recruit participants (blue). Plymouth and Nottingham's regions (grey) were not divided into postcode districts for recruitment. 43
Figure 5 shows each site using postcode districts in targeting recruitment. Each site was divided into – postcode districts selected by UKHSA for targeted recruitment (red), postcode districts within the site's catchment area but not selected (yellow) or out of range of the study site to recruit participants (blue). 47
Figure 6 Number of failed enrolments from all sites by age (years) 70
Figure 7 Number of participants per age band recruited by site 71
Figure 8 Number of participants recruited per age band by recruitment strategy 71
Figure 9 Distribution of English postcodes across(a) IMD and (b) IDACI quintiles in all NHS regions combined (i.e. whole of England), 'STORY' study sites, selected postcode districts for 'STORY' recruitment, and participants recruited to 'STORY' according to recruitment methods as described in section 2.2.2 (Group one and Group two). Error bars show 95% confidence intervals (CI) 74
Figure 10 Distribution of population over (a) IMD and (b) IDACI quintiles for the East of England according to postcode for NHS region, site, UKHSA selected postcode districts and participants recruited by group one and group two recruitment methods. Error bars show 95% confidence intervals (CI) 75
Figure 11 Distribution of population over (a) IMD and (b) IDACI quintiles for London according to postcode for NHS region, site, UKHSA selected postcode districts and participants recruited by group one and group two recruitment methods. Error bars show 95% confidence intervals (CI) 75
Figure 12 Distribution of population over (a) IMD and (b) IDACI quintiles for the Midlands according to postcode for NHS region, site, UKHSA selected postcode districts and participants recruited by group one and group two recruitment methods. Error bars show 95% confidence intervals (CI) 75
Figure 13 Distribution of population over (a) IMD and (b) IDACI quintiles for the North-East and Yorkshire & Humber region according to postcode for NHS region, site, UKHSA selected postcode districts and participants recruited by group one and group two recruitment methods. Error bars show 95% confidence intervals (CI) 76
Figure 14 Distribution of population over (a) IMD and (b) IDACI quintiles for the North-West according to postcode for NHS region, site, UKHSA selected postcode districts and participants recruited by group one and group two recruitment methods. Error bars show 95% confidence intervals (CI) 76
Figure 15 Distribution of population over (a) IMD and (b) IDACI quintiles for the South-East according to postcode for NHS region, site, UKHSA selected postcode districts and participants recruited by group one and group two recruitment methods. Error bars show 95% confidence intervals (CI) 76
Figure 16 Distribution of population over IMD quintiles for the South-West according to postcode for NHS region, site, UKHSA selected postcode districts and participants recruited by group one and group two recruitment methods. Error bars show 95% confidence intervals (CI) 77
Figure 17 Timeline for COVID-19 cases in England and recruited sample by week for 'STORY' during the same time-period. The numbered arrows denote the time periods used for analysis. Reference lines refer to actions taken by the government in England 93
Figure 18 A comparison of results by the RocheN-Anti-SARS-CoV-2 IgG nucleocapsid antibodies assay and RocheS-Anti-SARS-CoV-2 IgG spike protein receptor binding domain antibodies assay by age 100
Figure 19 Overall SARS-CoV-2 seroprevalence (RocheS-Anti-SARS-CoV-2 IgG spike protein receptor binding domain antibodies and RocheN-Anti-SARS-CoV-2 IgG nucleocapsid antibodies) by time period October 2019 to June 2021 in England, adjusted for age, NHS region and ethnicity. Error bars indicate 95% confidence intervals 102
Figure 20 SARS-CoV-2 seroprevalence (RocheS-Anti-SARS-CoV-2 IgG spike protein receptor binding domain antibodies and RocheN-Anti-SARS-CoV-2 IgG nucleocapsid antibodies) by age group and time period October 2019 to June 2021 in England, adjusted for NHS region and ethnicity. Error bars indicate 95% confidence intervals 104

Figure 21 SARS-CoV-2 seroprevalence (RocheS-Anti-SARS-CoV-2 IgG spike protein receptor binding domain antibodies and RocheN-Anti-SARS-CoV-2 IgG nucleocapsid antibodies) by region and time period October 2019 to June 2021 in England, adjusted for age and ethnicity. Error bars indicate 95% confidence intervals. 106
Figure 22 Summary of symptoms reported by participants seropositive on RocheS-Anti-SARS-CoV-2 IgG spike protein receptor binding domain antibodies assays by age group 109
Figure 23 Summary of symptoms reported by participants seropositive on RocheN-Anti-SARS-CoV-2 IgG nucleocapsid antibodies assays by age group 110
Figure 24 Univariate analysis of symptoms by age band (RocheS-Anti-SARS-CoV-2 IgG spike protein receptor binding domain 111
Figure 25 Univariate analysis of symptoms by age band (RocheN-Anti-SARS-CoV-2 IgG nucleocapsid antibodies) . 113
Figure 26 shows results of serological testing for a.Roche Elecsys® Anti-SARS-CoV-2 IgG spike receptor binding domain antibody concentration (U/ml) over time. b. Roche Elecsys® Anti-SARS-CoV-2 IgG nucleocapsid antibody concentration (COI)over time. Time-point 0 is defined as the participant's first positive result (a & b). c. Roche Elecsys® Anti-SARS-CoV-2 IgG spike receptor binding domain antibody concentration (U/ml) over time d. Roche Elecsys® Anti-SARS-CoV-2 IgG nucleocapsid antibody concentration (COI) over time. Time-point 0 is defined as the participant's highest positive result (c & d). Figures a-d are broken down into age groups 0-4 (red), 5-9 (purple),10-14 (green) and 15-24 (orange). The threshold of positivity is shown by the red line on each plot. 127
Figure 27a-b Mixed regression model predicting the adjusted means with time point 0 being the highest positive result with individual participant's serum antibody concentrations over time marked with (x) for Roche Elecsys® Anti-SARS-CoV-2 IgG spike antibodies receptor binding domain (U/ml) (a) and Roche Elecsys® Anti-SARS-CoV-2 IgG nucleocapsid antibodies (COI) (b). 95% confidence intervals are shown. The threshold of positivity is indicated by the red line on each plot. 128
Figure 28 shows matched serum and oral fluid samples for IgG concentration against a. SARS-CoV-2 viral Spike protein in spike protein receptor domain in serum (U/ml) and oral fluid IgG Spike protein (COI) b. SARS-CoV-2 viral nucleocapsid (COI) in serum and nucleopcapsid (COI) in oral fluid (NP). Red markers indicate a sample from a participant who was immunised at the time of sampling. Red dashed lines show the threshold of positivity for each assay. 130
Figure 29 a-d shows both oral fluid (red) and corresponding serum samples where available (blue). a-b demonstrate IgG concentration over time where timepoint 0 is defined as the first positive oral fluid result for oral fluid spike protein (COI) and serum spike protein receptor binding domain (Roche S Au/ml) (a) and oral fluid nucleocapsid (COI) and serum nucleocapsid (Roche N COI) (b). c-d define time-point 0 as the highest positive oral fluid result for oral fluid spike protein (COI) and serum spike protein receptor binding domain (Roche S Au/ml) (c) and oral fluid nucleocapsid (COI) and serum nucleocapsid (Roche N COI) (d). The threshold of positivity is marked by the dashed line (OF) and solid line (serum). These thresholds overlap on nucleocapsid plots. e-f shows a mixed regression model for oral fluid predicting the adjustment means with time point 0 being highest positive result with individual participants antibody concentrations over time marked with (x) for anti-spike protein (COI) € and anti-nucleocapsid antibodies (COI) (f). 95% confidence intervals shown. The threshold of positivity indicated by the red line on each plot 131
Figure 30 Age matching within the 'STORY' dataset 139
Figure 31 A comparison of assays by a. Comparing RocheS (Spike protein receptor binding domain assay) and MSD Cov2S (spike assay) b. Comparing Diasorin (spike assay) and MSD Cov2S (spike assay) c. Comparing RocheN (nucleocapsid assay) and MSD Cov2N (nucleocapsid assay). D. in-house PHE assay (Receptor binding domain assay) and MSDCov2RBD(Receptor binding domain assay). Red markers are a sample from participants who were immunised at the time of sampling. Red dashed lines show the threshold of positivity for each assay.. 140
Figure 32 A comparison of seropositive participants for SARS-CoV-2 (red), seronegative participants for SARS-CoV-2 (blue) and pre-pandemic (black) samples (positivity defined by original study assays) tested using the MSD platform. The MSD platform included IgG concentration (Au/ml) for IgG anti-spike, anti-nucleocapsid and anti-receptor binding domain (RBD) are shown for SARS-CoV-2. Anti-spike protein IgG concentration (Au/ml) is shown on for alphacoronaviruses (229E andNL63) and betacoronaviruses (HKU1, OC43, CoV1 and MERS). Mann Whitney U test was used to compare between groups. Threshold of positivity – Solid line (adult cut off) dashed line (paediatric cut off) 141
Figure 33 A matched case-control analysis comparing seropositive SARS-CoV-2 cases (red) with SARS-CoV-2 controls (blue) tested on the MSD platform. Seropositivity and seronegativity were defined by original study assays. . The MSD platform included IgG concentration (Au/ml) for IgG anti-spike, anti-nucleocapsid and anti-receptor binding domain (RBD) are shown for SARS-CoV-2. Anti-spike protein IgG concentration (Au/ml)

IS SHOWN ON FOR ALPHACORONAVIRUSES (229E ANDNL63) AND BETACORONAVIRUSES (HKU1, OC43, COV1 AND MERS)MANN WHITNEY U TEST WAS USED TO COMPARE BETWEEN BAESLINE(V1) VISITS AND SEROCONVERSION (V2) VISITS. .THRESHOLD OF POSITIVITY SOLID LINE (ADULT CUT OFF) DASHED LINE (PAEDIATRIC CUT OFF) 144

List of tables

TABLE 1 SARS-COV-2 STRAINS CIRCULATING AT THE TIME OF DATA COLLECTION FOR THE 'STORY' AND 'COVID WARRIOR' STUDIES. TABLE RECREATED FROM THE WHO TRACKING OF SARS-COV-2 VARIANTS[65] 30
TABLE 2 THE LANGUAGES IN WHICH STUDY INFORMATION WAS TRANSLATED 44
TABLE 3 UKHSA SELECTED POSTCODES BY SITE 46
TABLE 4 SUMMARY OF RECRUITMENT STRATEGIES BY GROUP 48
TABLE 5 PRECISION OF SEROPREVALENCE ESTIMATES 49
TABLE 6 ASSESSMENT OF REPRESENTATIVENESS, PRECISION AND DETECTABLE DIFFERENCES 51
TABLE 7 PRECISION OF ESTIMATES OF CHANGE IN PREVALENCE WITH 100 SAMPLES AT TWO TIME POINTS. 51
TABLE 8 MAXIMUM (TARGET) BLOOD VOLUMES BY AGE 56
TABLE 9 SEROLOGICAL ASSAYS 59
TABLE 10 THE NUMBER OF PARTICIPANTS RECRUITED BY EACH RECRUITMENT METHOD AS RECORDED ON THE PARTICIPANTS CASE REPORT FORM (CRF) 64
TABLE 11 SUMMARISES THE NUMBER OF MAIL-OUT INVITATIONS SENT AND RESPONSES RECEIVED PER SITE. 65
TABLE 12 SHOWS PARTICIPANT CHARACTERISTICS BY SITE AND REFERENCE REGIONAL AND NATIONAL ESTIMATES. 67
TABLE 13 NUMBERS OF ENROLMENT AND FAILED ENROLMENT PER SITE 69
TABLE 14 RECRUITMENT BY NHS REGION AND AGE BAND 72
TABLE 15 SUMMARY OF MEASURES OF SOCIOECONOMIC DEPRIVATION IN CHILDREN AND YOUNG ADULTS AGED 0-24 YEARS. IMD AND IDACI DERIVED FROM POSTCODE OF PARTICIPANT. FASIII AND CENSUS 2011 HOME OWNERSHIP AND HOUSE SIZE DATA WERE OBTAINED BY QUESTIONNAIRE. QUINTILE 1 WAS MOST DEPRIVED AND QUINTILE 5 WAS MOST AFFLUENT. 78
TABLE 16 A COMPARISON OF THE 'STORY' SAMPLE AND ONS DATA OF MIGRANT POPULATIONS ACROSS THE NHS REGIONS 81
TABLE 17 DISTRIBUTION OF HOURS SPENT IN CHILDCARE IN PARTICIPANTS AGED LESS THAN FIVE YEARS WHO ATTENDED CHILDCARE 81
TABLE 18 RECRUITING SITES NHS REGION(S) AND POSTCODE DISTRICTS WHERE APPLICABLE 91
TABLE 19 SENSITIVITY, SPECIFICITY AND THRESHOLDS OF POSITIVITY FOR ANTI-SARS-COV-2 IGG SPIKE PROTEIN RECEPTOR BINDING DOMAIN ANTIBODIES (ROCHES) AND ANTI-SARS-COV-2 IGG NUCLEOCAPSID ANTIBODIES (ROCHEN) [129,132] 92
TABLE 20 SUMMARY OF PARTICIPANTS AGED 0-18 YEARS TESTED FOR COVID-19 WITH ROCHES-ANTI-SARS-COV-2 IGG SPIKE PROTEIN RECEPTOR BINDING DOMAIN ANTIBODIES AND ROCHEN-ANTI-SARS-COV-2 IGG NUCLEOCAPSID ANTIBODIES ASSAYS FOR THE SEROPREVALENCE ANALYSIS, SYMPTOMS ANALYSIS (A FULL SYMPTOM HISTORY AVAILABLE) AND RISK FACTOR ANALYSIS (DEMOGRAPHICS KNOWN) 95
TABLE 21 SHOWS A SUMMARY OF HOUSEHOLD SIZES FOR PARTICIPANTS BY ETHNICITY. 97
TABLE 22 COMPARING CHILDREN AGED 0-15 YEARS ON IDACI AND IMD BY REGION AS A MEASURE OF SOCIOECONOMIC DEPRIVATION IN CHILDREN IMD AND IDACI DERIVED FROM POSTCODE OF PARTICIPANT, FASIII AND CENSUS 2011 HOME OWNERSHIP AND HOUSE SIZE QUESTIONS DATA WERE COLLECTED ON THE QUESTIONNAIRE. QUINTILE 1 WAS MOST DEPRIVED AND QUINTILE 5 WAS MOST AFFLUENT. 97
TABLE 23 SARS-COV-2 SEROPREVALENCE (ROCHES-ANTI-SARS-COV-2 IGG SPIKE PROTEIN RECEPTOR BINDING DOMAIN ANTIBODIES AND ROCHEN-ANTI-SARS-COV-2 IGG NUCLEOCAPSID ANTIBODIES) BY TIME PERIOD OCTOBER 2019 TO JUNE 2021 IN ENGLAND, ADJUSTED FOR AGE, REGION AND ETHNICITY 103
TABLE 24 SARS-COV-2 SEROPREVALENCE (ROCHES-ANTI-SARS-COV-2 IGG SPIKE PROTEIN RECEPTOR BINDING DOMAIN ANTIBODIES AND ROCHEN-ANTI-SARS-COV-2 IGG NUCLEOCAPSID ANTIBODIES) BY AGE GROUP AND TIME PERIOD OCTOBER 2019 TO JUNE 2021 IN ENGLAND, ADJUSTED FOR NHS REGION AND ETHNICITY 105
TABLE 25 SARS-COV-2 SEROPREVALENCE (ROCHES-ANTI-SARS-COV-2 IGG SPIKE PROTEIN RECEPTOR BINDING DOMAIN ANTIBODIES AND ROCHEN-ANTI-SARS-COV-2 IGG NUCLEOCAPSID ANTIBODIES) BY REGION AND TIME PERIOD OCTOBER 2019 TO JUNE 2021 IN ENGLAND, ADJUSTED FOR AGE AND ETHNICITY· 107
TABLE 26 LOGISTIC REGRESSION MODELS TO DETERMINE SYMPTOMS ASSOCIATED WITH POSITIVITY THE ROCHES-ANTI-SARS-COV-2 IGG SPIKE PROTEIN RECEPTOR BINDING DOMAIN ANTIBODIES ASSAY RESULTS. SYMPTOMS WITH THE HIGHEST P-VALUE WERE SEQUENTIALLY EXCLUDED AND MODEL AKAIKE INFORMATION CRITERION (AIC) VALUES WERE COMPARED UNTIL A MODEL WITH THE LOWEST AIC VALUE HAD BEEN REACHED. SEPARATE MODELS CREATED FOR 0-9 YEARS AND 10-18 YEARS. 112
TABLE 27 LOGISTIC REGRESSION MODELS TO DETERMINE SYMPTOMS ASSOCIATED WITH POSITIVITY THE ROCHEN-ANTI-SARS-COV-2 IGG NUCLEOCAPSID ANTIBODIES ASSAY RESULT. SYMPTOMS WITH THE HIGHEST P-VALUE WERE SEQUENTIALLY EXCLUDED AND MODEL AKAIKE INFORMATION CRITERION (AIC) VALUES WERE COMPARED UNTIL A MODEL WITH THE LOWEST AIC VALUE HAD BEEN REACHED· SEPARATE MODELS CREATED FOR 0-9 YEARS AND 10-18 YEARS 114
TABLE 28 UNIVARIATE AND MULTIVARIABLE LOGISTIC REGRESSION MODELS TO ESTABLISH RISK OF SARS-COV-2 SEROPOSITIVITY ON ROCHE ELECSYS ANTI-SARS-COV-2 SEROLOGICAL ASSAYS FOR THE DETECTION OF ANTI-SARS-COV-2 IGG SPIKE PROTEIN RECEPTOR BINDING DOMAIN ANTIBODIES (ROCHES-ANTI-SARS-COV-2 IGG SPIKE PROTEIN RECEPTOR BINDING DOMAIN ANTIBODIES) IN CHILDREN AGED 0-18 YEARS. MULTIVARIABLE ANALYSIS ADJUSTS FOR AGE, SEX, NHS REGION, TIME PERIOD, ETHNICITY, SOCIOECONOMIC DEPRIVATION (USING EITHER INDEX OF MULTIPLE DEPRIVATION (IMD) OR INCOME DEPRIVATION

Affecting Children Index (IDACI)), urban/rural geography and presence of healthcare workers within the household. 116
Table 29 Univariate and multivariable logistic regression models to establish risk of SARS-CoV-2 seropositivity on Roche Elecsys Anti-SARS-CoV-2 serological assays for the detection of anti-SARS-CoV-2 IgG nucleocapsid antibodies (RocheN-Anti-SARS-CoV-2 IgG nucleocapsid antibodies) in children aged 0-18 years. Multivariable analysis adjusts for age, sex, NHS region, time period, ethnicity, socioeconomic deprivation (using either Index of multiple deprivation (IMD) or Income Deprivation Affecting Children Index (IDACI)), urban/rural geography and presence of healthcare workers within the household. 117
Table 30 Number of repeat samples by age group and assay deployed..... 125
Table 32 A comparison of the predicted antibody decline of anti-spike and anti-nucleocapsid IgG concentrations in serum and oral fluid..... 129
Table 33 Shows the odds being infected per 10-fold higher baseline titre (V1) results with value of P<0.01 were considered statistically significant 142

Glossary

ARDS	Acute respiratory syndrome
AIC	Akaike Information Criterion
AM	Alveolar macrophages
ACE2	Angiotensin-converting enzyme 2
BAL	Bronchoalveolar lavage
CRF	Case report form
CH	Central helix
CHIS	Child Health Information Service
cDNA	complementary DNA
CD	Connector domain
COI	Cut off index
COVID-19	Coronavirus disease previously known as "2019 novel coronavirus"
CT	Cytoplasmic domain
DC	Dendritic cell
S^B	Domain B of S1 subunit
ECLIA	Electrochemiluminescent immunoassays
E	Envelope
FP	Fusion peptide
GP	General Practitioner
HR1	Heptad repeat 1
HR2	Heptad repeat 2
HCoV	Human Coronaviruses
IDACI	Income Deprivation Affecting Children Index
IMD	Index of Multiple Deprivation
IMM	inflammatory monocyte-macrophages
IFN	Interferon
LSOA	Lower-Layer Super Output
M	Membrane
MERS	Middle East respiratory syndrome
MPIDR	Max Planck Institute for Demographic Research
NTD	N-terminal domain
NIMS	National Immunisation Management Service
NIHR	National Institute for Health Research
NET	Neutrophil extracellular trap
Nsps	Nonstructural proteins
N	Nucleocapsid
OR	Odds ratio
OF	Oral fluid
PIMS-TS	Paediatric Inflammatory Multisystem Syndrome
PRR	Pattern recognition receptors
PPE	Personal protective equipment
POC	Point of care

PCR	Polymerase chain reaction
ROS	Reactive oxygen species
RBD	Receptor binding domain
REDCap	Research Electronic Data capture
RT–PCR	Reverse transcription–polymerase chain reaction
SARS-CoV-1	Severe Acute Respiratory Syndrome Coronavirus -1
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus -2
SIS	School infection survey
S	Spike protein
Tfh	T follicular helper cells
TM	Transmembrane domain
T/N ratio	Samples analysed in duplicate. Optical density evaluated at 450 nm (OD_{450}) data by dividing average OD_{450} values for individual samples by average OD_{450} of a known calibrator with negative antibody levels (T/N ratio).
UK GDPR	UK General Data Protection Regulation
UKHSA	UK Health Security Agency
VOC	Variants of concern
VOI	Variants of interest
STORY	What's the STORY?' (Serum testing of representative youngsters)
WHO	World Health Organisation

Abstract

COVID-19 disease (COVID-19) caused by Severe Acute Respiratory Syndrome Coronavirus -2 (SARS-CoV-2) first emerged in December 2019 in Wuhan, China. The World Health Organization (WHO) declared a Public Health Emergency of International Concern on 30th January 2020. Globally, COVID-19 has caused over 700 million cumulative cases and 6.8 million deaths, according to the WHO. Seroprevalence studies evaluating population prevalence of SARS-CoV-2 antibodies have an important role in understanding the spread of and population vulnerability to SARS-CoV-2 infection. Most have been performed in adults, and those in children have predominantly tested samples obtained opportunistically in a clinical context or in school-based populations, both of which bring potential biases.

The primary objective of the 'What's the STORY?' study (Serum Testing of Representative Youngsters) was to establish the feasibility of recruiting a representative sample of the paediatric population in England at a regional and national level. 'STORY' started recruitment in October 2019 ahead of the first reported cases of SARS-CoV-2. To achieve a representative sample the participants enrolled had to closely reflect the characteristics of the population of interest, enabling estimates of the proportion of children and adolescents with SARS-CoV-2 infection-induced antibodies to be extrapolated to the broader population within these age groups. The 'STORY' study was adapted to obtain adequate numbers of samples from groups of special interest. Participant samples were tested for serum antibody concentrations against SARS-CoV-2 spike protein receptor binding domain and nucleocapsid to determine the prevalence of anti-SARS-CoV-2 serum antibodies in children and adolescents aged 0-18 years across England. The serum bank of samples from children and young adults aged 0-24 years collected by the 'STORY' study will be used in future studies related to vaccine-preventable diseases.

The key findings of this thesis include evidence to support the feasibility of a seroepidemiological survey in England in a sample that was both representative of England

and had ‘over-sampling’ in populations of special interest. Recruitment was achieved across the age groups of 0-24 years. This serosurvey was utilised to report SARS-CoV-2 seroepidemiology in children and adolescents aged 0–18 years in England, before widespread immunisation in this population, using samples collected from the end of 2019 through to mid-2021. It also allowed exploration of potential risk factors for COVID-19 seropositivity and the utility of symptom-based indicators of infection.

The findings of ‘STORY’ confirm that SARS-CoV-2 infections in children are generally asymptomatic or mild and rarely progress to severe disease and hospitalisation. Why this is so remains unclear. Analysis of data presented in this thesis explored whether this could be attributed to protection due to pre-existing cross-reactive seasonal coronavirus antibodies. The results showed a high prevalence of pre-existing immunity against seasonal coronaviruses when measuring IgG concentration of antibodies against spike protein but did not confer protection against SARS-CoV-2.

In serum, anti-nucleocapsid antibodies fell below the threshold of positivity more quickly than anti-spike protein receptor binding domain antibodies, suggesting community-based prevalence studies relying on these measures of infection may underestimate the burden of disease. The ‘STORY’ study provides a historically unique insight into the spread of the SARS-CoV-2 infection through the paediatric and adolescent population in England before the availability of vaccines, along with valuable insights into the planning of seroepidemiology studies in this population for future pandemics.

Acknowledgements

The work contained in this thesis is the result of collaborative work between the Oxford Vaccine Group, National Immunisation Schedule Evaluation Consortium (NISEC), UK Paediatric Vaccine Group (UKPVG), UK Health Security Agency (UKHSA), the 'COVID Warriors' research team and the Nuffield department of Medicine.

I am grateful to Matthew Snape for first suggesting the idea of undertaking a Doctorate of Philosophy and then guiding me through the process, helping me navigate the obstacles and opportunities this journey has provided. I would also like to thank Karen Tiley for her patience as I embarked on a steep learning curve of how to use statistical packages. Her feedback and encouragement have kept me on track. I want to extend my sincere thanks to Merryn Voysey and Susie Dunachie, who have shared their knowledge and expertise and whose feedback has improved my work at each stage.

I am thankful for the staff at the Oxford Vaccine Group and the 'STORY' research team without whom this project would not have been possible. I want to thank Iason Vichos and Ella Morey, who coordinated the 'STORY' study across England and worked tirelessly to make targets under challenging circumstances. A special thanks to Tom Waterfield and the 'COVID Warriors' study team, who contributed data to combine with the 'STORY' study. I am also grateful to Stephanie Longet and UKHSA who ran all the samples involved in this work. I would like to express my gratitude for the contribution of all study participants without whom, none of this would have been possible.

I would be remiss in not mentioning Brama Hanumunthadu, who has always made time to discuss and share thoughts and ideas. Finally, I could not have undertaken this journey without my family and friends. The tireless support and emergency childcare provided by Mum and Dad as well as my wonderful husband Tristan, who has kept family life together while I have been distracted. Lastly, thanks to Seth and Isobel, who have kept me grounded and allowed Mummy to work without recrimination in the last few months before submission.

Declarations and contributions

The research presented in this thesis is original work and any contributions not made by the author are clearly stated in the text. After the original proposal to conduct a serological survey on children, my involvement started with writing the study protocol for the feasibility study. This was a collaborative study between UKHSA and the Oxford Vaccine. Once we gained ethical approval, I implemented the study over 13 sites and managed recruitment targets for the overall study. Samples collected during the study were processed at UKHSA and the University of Oxford laboratories. Upon receiving the raw data, I conducted the analysis and subsequently published the findings. The thesis includes data generated from the "What's the STORY?" study in collaboration with NIHR, NISEC, UKCIC and UKHSA and the 'COVID Warriors' study. The author acknowledges the work of the "What's the STORY?" study team and UKHSA for the collection and processing of 'STORY' samples and the 'COVID Warriors' research team for the collection and processing of samples.

This research was funded by the National Institute for Health Research (NIHR) Policy Research Programme and the UK Research and Innovation (UKRI) Medical Research Council (PR-R17-0916-22001, CV220-036, CV220-036/1). The research is supported by the NIHR Oxford Clinical Research Facility and Biomedical Research Centre, the NIHR Clinical Research Network, and the NIHR funded National Immunisation Schedule Evaluation Consortium (PR-R17-0916-22001). The study sponsor was the University of Oxford. The processing of samples was completed by the COVID serology team within Virus Reference Department at UKHSA.

Publications

Published

Ratcliffe H, Tiley KS, Andrews N, Amirthalingam G, Vichos I, Morey E, Douglas NL, Marinou S, Plested E, Aley P, Galiza EP, Faust SN, Hughes S, Murray CS, Roderick M, Shackley F, Oddie SJ, Lees T, Turner DPJ, Raman M, Owens S, Turner P, Cockerill H, Lopez Bernal J, Linley E, Borrow R, Brown K, Ramsay ME, Voysey M, Snape MD. Community seroprevalence of SARS-CoV-2 in children and adolescents in England, 2019-2021. Arch Dis Child. 2022 Jul 20;108(2):123–30. Doi: 10.1136/archdischild-2022-324375. Epub ahead of print. PMID: 35858775; PMCID: PMC9887370.

Content reproduced in thesis with permission. Re-use permitted under CC BY. Published by BMJ. https://creativecommons.org/licenses/by/4.0/

Submitted

Ratcliffe H, Tiley KS, Longet S, Tonry C, Roarty C, Watson C, Amirthalingam G, Vichos I, Morey E, Douglas NL, Marinou S, Plested E, Aley P, Galiza EP, Faust SN, Hughes S, Murray CS, Roderick M, Shackley F, Oddie SJ, Lees T, Turner DPJ, Raman M, Owens S, Turner P, Cockerill H, Lopez Bernal J, Ijaz S, Poh J, Shute J, Linley E, Borrow R, Brown K, Hoschler K, Carroll MW, Klenerman P, Dunachie SJ, Ramsay ME, Voysey M, Waterfield T, Snape MD Serum HCoV-spike specific antibodies do not protect against subsequent SARS-CoV-2 infection in children and adolescents.

Oral presentations

Ratcliffe, H. 2021. First wave community seroprevalence of SARS-CoV-2 in English children and teenagers in the community. European Society for Paediatric Infectious Diseases, 24th May 2021, Virtual conference hosted in Geneva.

Ratcliffe H. 2022 'STORY' SARS-CoV-2 cross-reactive immunity with seasonal coronaviruses in UK children. UK Paediatric Vaccine Group, 10th June 2022, St Georges's University of London

Chapter 1 Introduction

1.1 The emergence of SARS-CoV-2

1.1.1 Epidemiology

COVID-19 disease (COVID-19), caused by Severe Acute Respiratory Syndrome Coronavirus -2 (SARS-CoV-2), first emerged in December 2019 in Wuhan China.[1] Research into the epidemiology of the disease was required to implement effective public health measures. The outbreak spread rapidly, with the World Health Organization (WHO) declaring a Public Health Emergency of International Concern on 30th January 2020.[2] The UK, like many countries, adopted a strategy to limit the spread of SARS-CoV-2 within the population which would reduce the burden on the healthcare system while vaccines were developed and a vaccination programme implemented. Measures taken to limit the spread of SARS-CoV-2 included population-level advice and education, national lockdowns, contact tracing, and the development of vaccines, of which a number have been licensed and offered to the public (Figure 1).[3,4] As of March 2023, the WHO reported over 700 million cumulative cases of COVID-19, with over 6.8 million deaths globally.[5] During the pandemic seroepidemiological studies have contributed to our understanding of the spectrum of disease, symptomatology and risk factors associated with infection of the virus. The identification of high-risk groups was important when implementing a new vaccination campaign.[6]

When COVID-19 emerged, gathering information regarding population susceptibility to infection as rapidly as possible was important. At a time when only symptom-based PCR (polymerase chain reaction) testing was available, seroepidemiology captured information on population exposure and the spectrum of disease. The fastest route to this information was through convenience sampling, e.g. residual samples from patients seeking medical treatment. Historically these surveys have not been representative of the paediatric population. A limitation of this system is that the samples collected may not be representative of the general population, particularly in younger age groups. Paediatric samples available through healthcare

settings are more likely to reflect samples from children with chronic health conditions and, therefore, are not representative of the general paediatric population. As fewer blood samples are obtained from children in healthcare settings, the number of samples available is also lower than from the adult population, making it difficult to achieve sufficient numbers in each age band and region required to stratify by age for a specific antigen. Targeted paediatric serosurveys were required to provide additional information on population exposure to infection, symptoms experienced, and risk factors associated with seropositivity to SARS-CoV-2.

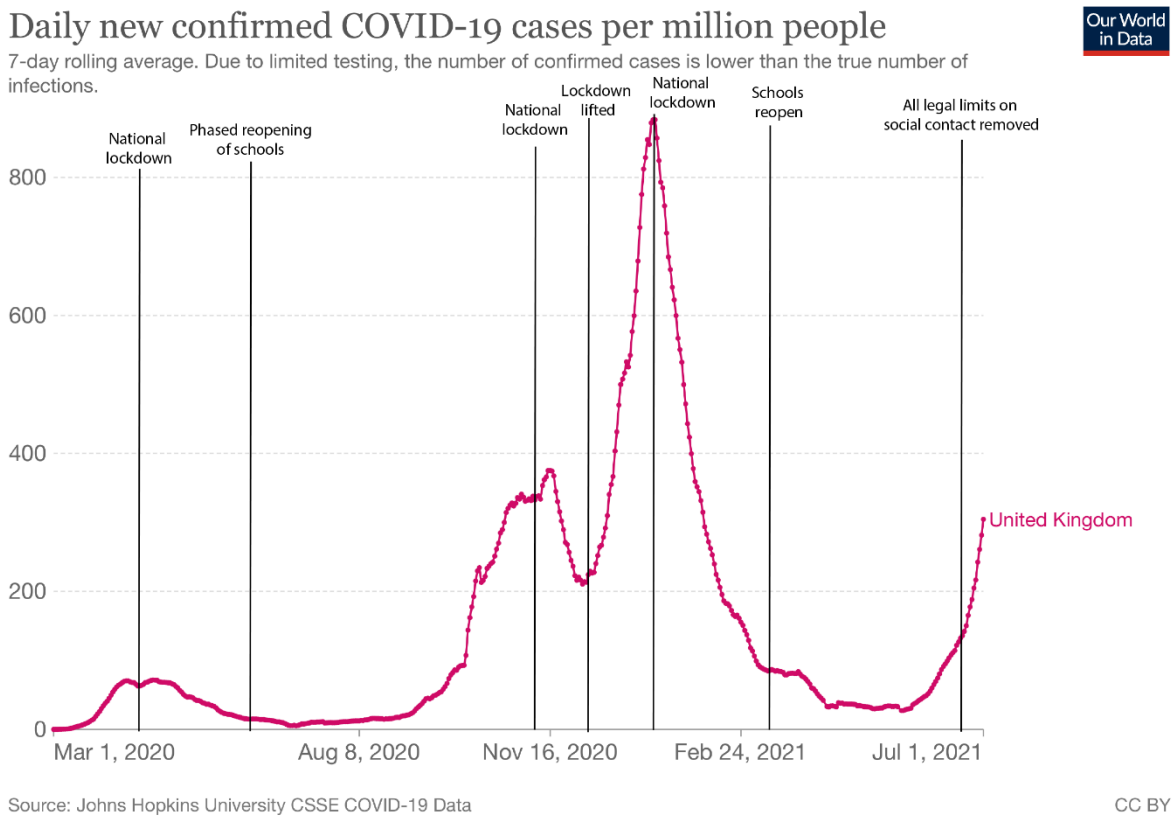


Figure 1 Timeline of COVID-19 cases in the United Kingdom from John Hopkins University CSSE COVID-19 Data adapted to show the timing of measures taken to prevent the spread of disease[3,4]

The understanding of COVID-19 has advanced at an extraordinary rate; this chapter explores the knowledge base that evolved between the beginning of the COVID-19 pandemic and mid 2021, specifically focussing on epidemiology including studies determining the spectrum of disease and transmission within the population. Additionally, the chapter explores the virology

of SARS-CoV-2 up to the emergence of the delta variant looking at the humoral immune response to SARS-CoV-2 and the role of seasonal coronaviruses in protective immunity against SARS-CoV-2. The cross-sectional prospective seroepidemiological study, which is central to this thesis, was called 'What's the STORY?' (Serum Testing of Representative Youngsters hereafter referred to as 'STORY'). The study started just prior to the COVID-19 pandemic, continued through June 2021 and was used to gain an understanding of COVID-19 disease early in the pandemic before the licensing and widespread distribution of vaccines to the paediatric population.

1.1.2 Investigating the emergence of SARS-CoV-2

In an epidemic or pandemic, a variety of epidemiological studies are required to gather data on population exposure, the transmission of the virus and risk factors leading to poor outcomes. When SARS-CoV-2 emerged, governmental public health and research networks had to coordinate to gather data efficiently with mechanisms in place to allow findings to inform public policy both at a national and international level.

1.1.3 Incidence and prevalence

Measurements of incidence and prevalence are required to establish the number of cases within a population. Incidence refers to the number of new cases within a defined time period, whereas prevalence refers to the proportion of a population being affected within a defined time period.[7] Both are important in understanding population exposure to the disease being studied, in this case, SARS-CoV-2. The WHO highlighted the importance of surveillance systems to detect infection, which provide data regarding incidence and also play a role in limiting the spread of disease, which was essential to restoring normal economic and social functioning.[8]

Incidence data in the UK came from multiple methods of surveillance, such as case data from laboratory notifications, clinical coding from hospital databases and results from existing surveillance surveys such as in the UK Health Security Agency (UKHSA) symptom-based PCR testing programme in both the community and hospitals. These data underestimate the

incidence as only symptomatic infections were included in the calculations. Some insight into asymptomatic cases was provided by surveillance programs using lateral flow testing surveillance, which was not symptom-based. These were introduced in April 2021 after the third national lockdown, but they were not mandatory in primary school-aged and preschool-aged children and relied on the individual to carry out the test and log the result correctly.[9]

Seroepidemiology is an epidemiological tool that investigates the prevalence of biomarkers or antibodies to a specific antigen that is unaffected by test-seeking behaviour.[10] By understanding population-level prevalence and incidence of infection, policymakers can make informed decisions about public health measures and on the implementation of vaccination programmes.[10]

During the pandemic, seroepidemiology studies were undertaken to inform prevalence calculations measuring the presence of antibodies to one or more of the structural proteins in SARS-CoV-2 (nucleocapsid, spike and receptor binding domain). Different population sampling methodologies were used, primarily convenience sampling, targeting populations such as healthcare workers, patients accessing healthcare for non-COVID-19 reasons, samples collected as part of blood donation, and sampling in schools. This allowed rapid collection of results to inform estimates on prevalence within these defined populations. The WHO created a global research initiative called the UNITY studies to align national seroprevalence studies,[11] but the focus was the general population, and the paediatric population was typically analysed as one group (<18 years).

When applying these estimates to the general population, caution needs to be taken due to differences in methodology e.g. convenience samples are unlikely to represent the general population as few children access healthcare for blood tests unless there is an underlying medical reason. In school studies, although many children can be sampled within a short period of time, the estimation of prevalence could be confounded due to the clustering of cases within

a school setting or by socio-economic factors related to the schools selected. Ideally, a community study of children recruiting a representative sample from all NHS regions would allow findings to be generalised to the whole of England enabling conclusions to be drawn at a population level.

By achieving representative sampling, these data allow more accurate modelling of the disease leading to insights into its epidemiology and enable healthcare resources to be allocated as appropriate. One way to achieve this in the adult population has been to design and implement a methodology which enables the participant to self-test, avoiding the need for face-to-face contact with the study team which allows for safe serological surveys.[12] This type of study has been trialled in paediatrics and is achievable in school-age children, but pre-school aged children were omitted.[13]

A global standard on how these studies should be designed and reported would allow more valid comparisons to be drawn between studies. Differences in population sampling methodologies can lead to different results which need to be interpreted in context with the population they represent. These differences are evident in early work on COVID-19 seroepidemiology; Badal et al. reported, in a meta-analysis of the paediatric population which included data from both community and hospital settings, that the highest incidence was in children aged 6-14 years.[14] These findings were supported by Götzinger et al. who conducted a large European multinational multicentre cohort study.[15] This contrasted with a German paediatric study by Sorg et al. that reported the highest seroprevalence in children aged 0-4 years. The variations between age groups with the highest seroprevalence reported by each study is likely due to differences in recruitment strategy. Both Badal et al. and Götzinger et al. recruited from community and hospital settings whereas Sorg et al. used convenience sampling from inpatients or outpatients, irrespective of the medical purpose of the clinical contact.[16] In addition, although the time periods of recruitment overlap in each of the studies, Sorg et al.

recruited over a longer period. Not only was the methodology in recruitment different but new strains of SARS-CoV-2 started circulating during the study by Sorg et al. The differences in survey estimates emphasise the importance of understanding the population and the virus being studied.

1.1.4 Spectrum of disease

To gain an understanding of the spectrum of disease, multiple data sources and research studies were required. During the first wave, data from hospital admissions and symptom-based testing became available but these did not inform on mild or asymptomatic infection. At the start of the pandemic, measures such as closure of educational settings including preschool childcare centres were implemented due to fears that children, and especially infants, may be more susceptible to SARS-CoV-2.[8] During the first wave of the pandemic, it was found that COVID-19 tended to have a mild course and an overall favourable prognosis in the paediatric population.[4] A small number of children developed paediatric multisystem inflammatory syndrome which was temporally associated with COVID-19 (MIS-C).[17] Beyond this, little was known regarding the numbers affected by asymptomatic and mild infection and the role of asymptomatic children in transmission.

Seroepidemiology was one approach to understanding the numbers of mild infection not meeting the symptom-based testing criteria, and asymptomatic infections. After the first wave of the pandemic, it became clear that SARS-CoV-2 infection affected all age groups, and there was no age restriction on COVID-19 susceptibility, but the reason for the difference in severity of infection between adults and paediatrics remained unknown.[14] Throughout the pandemic, data regarding the burden of disease in the paediatric population has been limited compared with adult populations.[14] Over three years since the start of the pandemic, data on the direct impact on mortality in children and adolescents from COVID-19 is incomplete. The Max Planck Institute for Demographic Research (MPIDR) COVerAGE database estimated globally a total of 4.1 million COVID-19 deaths in November 2022, of which 0.4% occurred in children

and adolescents under the age of 20 years.[18] The number of deaths across the age bands of 0-9 (8,304/ 1,081,231,571, 0.0008%) and 10-19 years (9,187/ 1,036,658,731, 0.0009%), was roughly equal.[18] Not only is it important to record the information but capturing information in a standardised format allows meaningful comparisons between countries. Sources that informed databases such as MPIDR COVerAGE did not capture relevant information in a consistent manner and frequently lacked a breakdown by age. This could account for a significantly lower estimate for total deaths compared with the WHO estimate of more than 6 million in the same time period.[5] Globally, variations in mortality have been reported with lower death rates in Africa than America and Europe after the first wave.[19,20] Patterns of lower mortality across Africa were perhaps due to poor availability of PCR testing but also more experience with tackling infectious disease outbreaks and different age structures of African populations compared with America and Europe.

Understanding the long-term effects of SARS-CoV-2 infection is an active area of research. Although most children recover from SARS-CoV-2 infection there is a subgroup of children who suffer from long COVID. These children who have confirmed SARS-CoV-2 infection have at least one persisting physical symptom that cannot be explained by an alternative diagnosis, for a minimum duration of 12 weeks after initial testing. These symptoms have an impact on everyday functioning and may continue or develop after COVID infection and may fluctuate or relapse over time.[21,22]

1.1.5 Symptoms

During the first wave of the pandemic, the symptoms reported in the paediatric population varied between studies.[23,24] According to the UK Health Security Agency (UKHSA), symptoms requiring a PCR test included loss of or change in smell or taste sensation, fever, and a new continuous cough.[25] These symptoms were derived from studies in adults, and much less was known about infection in children.

At the end of the first wave, common clinical presentations included fever, cough, and headache, while sore throat, rhinorrhoea, and gastrointestinal symptoms were rare.[14] The symptoms attributed to SARS-CoV-2 have changed during the pandemic as different strains emerged, and the proportion of the vaccinated population increased. COVID-19 was initially considered a pulmonary disease; however, non-pulmonary presentations have been recognised including sepsis, septic shock, acute thrombosis and multisystem inflammatory syndrome.[26]

1.1.6 Transmission

Establishing how SARS-CoV-2 infection is transmitted was crucial for developing public health measures to control the spread of the virus while allowing some normal activities to continue. At the outset influenza protocols were followed in the UK,[27] reflecting pandemic preparedness planning that focused on influenza as the pathogen most likely to cause a pandemic. Children were known to transmit influenza in the community and it was hypothesised that children had the potential to play a similarly important role in the transmission of SARS-CoV-2 infections, despite generally having mild or asymptomatic infections.[28]

Microbiological transmission studies established how SARS-CoV-2 was spread from person to person.[29] Results from multiple epidemiological transmission studies were brought together in systematic reviews which identified that the primary route of SARS-CoV-2 transmission was airborne.[30] Once infected, the nasopharynx contains high concentrations of SARS-CoV-2 even before symptoms start.[26] Children were largely omitted from these studies as severe cases requiring hospitalisation were rare. To understand the transmissibility of the virus, contact-tracing and household studies provided data concerning the timing of transmission and factors increasing or decreasing transmission.[6,31] Measures including lockdowns (local and national), physical distancing, mandatory use of face masks, and hand hygiene were implemented to limit the spread of disease,[32] but interfered with normal transmission patterns making their study

complicated. In mid-2020, the data demonstrated that children were infected, but the majority had either mild symptoms or were asymptomatic.[33]

It was important to establish the role of asymptomatic transmission before the widespread opening of educational institutions, and some data was provided by studies within households and small defined communities. In early 2020, a ski resort in Austria was reported to have been a super-spreading event.[34] The community was quarantined, and the local population underwent serological testing, which showed lower seroprevalence amongst children (27%) than adults (42%).[34] Adults were thought to have a higher seroprevalence due to a higher exposure through occupations at the resort. Of seropositive participants, 80% had not been diagnosed with a COVID-19 infection, and this percentage was higher in children and asymptomatic adults.[34] During the same time period, children with SARS-CoV-2 were reported to contract the infection from adult contacts and were less likely to transmit the infection.[35]

Case reports involving transmission from symptomatic children to household contacts were rare.[36] A case series of household contacts which investigated 31 clusters showed that only three involved a child as a possible source case.[37] School-based transmission studies in Australia involving 18 index cases reported two secondary infections, of which one was from an infected teacher.[38] The SKIDS study was a prospective, cross-sectional study of infection and cluster outbreaks in schools in England conducted as schools started to reopen in the UK in 2020 which reported that staff had higher incidence of infection than students and that the emphasis should be on controlling community transmission to protect educational settings.[39]

A systematic review evaluating the effect of school closures on the control of SARS-CoV-2 observed no meaningful impact on population-level viral transmission dynamics.[40] Further supportive observational data emerged from Sweden where the prevalence of SARS-CoV-2 in children remained low despite educational settings remaining open throughout the pandemic.[41]

Understanding the paediatric populations' role in transmission was important to allow schools

to stay open, which was beneficial for both children's well-being and enabling parents to return to work.[14,39]

1.1.7 Risk factors associated with SARS-CoV-2 infection and severe COVID-19 disease

Establishing shared characteristics from patients with severe COVID-19 disease admitted to hospital identified potential risk factors for further study.[42] As cases of SARS-CoV-2 infection in the UK increased, surveillance data became available from symptom-based testing programmes which linked with primary care data in large cross-sectional studies allowing risk factors associated with SARS-CoV-2 symptomatic infections to be identified. Research networks with the ability to link in with surveillance programmes and access to anonymised patient data are a powerful resource in understanding the epidemiology of SARS-CoV-2 infection.

The Oxford Royal College of General Practitioners (RCGP) Research and Surveillance Centre programme identified risk factors associated with severe disease as increasing age, male sex, increasing deprivation, urban location, and black ethnicity.[43] A limitation of this study is that results relied on symptom-based testing, and the full spectrum of infection was not captured. Seroepidemiological surveys and routine swabbing studies in adults captured the full spectrum of disease but these studies did not include children.[44,45] There was an established association between minority ethnic groups and severe COVID-19 outcomes in adults, including hospitalisations and mortality.[44] A combination of health and sociodemographic factors were thought to play a role in the differences seen between populations. Paediatric populations at this time had not been extensively studied but had the potential to shed light on the interplay between health and sociodemographic factors as children are less likely to have an underlying chronic diseases attributable to risks associated with socio-economic deprivation including smoking, obesity and type 2 diabetes.[46] This area of research was challenging as, according to Saatci et al. there was unequal testing across minority ethnic groups, with more tests being

carried out among white children.[46] This is despite Asian children being significantly more likely to have COVID-19 hospital and ICU admissions compared with white children.[46]

Such underrepresentation of participants from minority ethnic groups in research is a long-standing issue, and disentangling factors which disproportionately affect minority ethnic groups such as deprivation, occupation and chronic health conditions such as coronary vascular disease and diabetes is challenging.[47] Health inequalities have been observed in participation of adults and children within existing screening programmes and vaccination uptake.[48,49] There can be community-wide distrust of healthcare establishments due to institutional discrimination, which contributes to inequalities in healthcare.[50] Studies focusing on these inequalities have shown that all public health interventions must adapt to engage communities from minority ethnic groups to address inequalities in healthcare.

1.2 The SARS-CoV-2 virus and the body's immune response to infection

1.2.1 Coronaviruses and the identification of SARS-CoV-2

Coronaviruses are a large family of positive single-stranded RNA viruses which in humans primarily cause respiratory tract infections.[51] Coronaviruses which affect humans are broadly divided into alphacoronaviruses (including seasonal coronaviruses 229E and NL63) and betacoronaviruses which contain both seasonal coronaviruses (HKU9 and OC43) and coronaviruses causing more severe infections such as Middle East respiratory syndrome (MERS), Severe acute respiratory syndrome -1 (SARS-CoV-1) and SARS-CoV-2.[51] SARS-CoV-2 emerged in December 2019, reported as an outbreak of pneumonia of unknown origin in Wuhan, China.[52] Using bronchoalveolar lavage (BAL) fluid from cases and inoculating human airway epithelial cells and Vero E6 and Huh7 cell lines, SARS-CoV-2, the virus causing COVID-19, was identified.[53]

1.2.2 Virology of SARS-CoV-2

SARS-CoV-2 contains four major structural proteins, spike (S), envelope (E), membrane (M) and nucleocapsid (N) and 16 non-structural proteins (Nsps).[54] There are two parts of the spike protein, the receptor binding domain (RBD) and the N-terminal domain (NTD). The genomic sequences share 79.6% sequence identity to SARS-CoV-1.[55]

The spike protein mediates entry into the host cells by the RBD binding to the angiotensin-converting enzyme 2 (ACE2) receptor.[54,56] The S2 subunit is formed by a fusion peptide (FP), a heptad repeat 1 (HR1), a central helix (CH), a connector domain (CD), a heptad repeat 2 (HR2), a transmembrane domain (TM) and a cytoplasmic tail (CT).[54] The S2 subunit is responsible for the fusion of the hosts and viral membranes, allowing entry of the virus into the host's cells.[57] The mechanism for fusion is by host proteases cleaving the spike glycoprotein at the border between the S1 and S2 subunits (S2 cleavage site).[54,56,58] There is a higher binding affinity between the RBD of SARS-CoV-2 and ACE2 compared with severe acute respiratory syndrome coronavirus 1 (SARS-CoV-1) and ACE2, which resulted in SARS-CoV-2 being more infectious within the population.[54,59]

An additional major difference between SARS-CoV-1 S and SARS-CoV-2 S is the cleavage site for the host cell protease furin located between the S1 and S2 subunits.[54] This cleavage site is not present in other betacoronaviruses found in humans. It has been suggested that the high virulence of SARS-CoV-2 results from furin proteases as it plays a key role in SARS-CoV-2 replication and pathogenesis.[60,61]

1.2.3 Variants of SARS-CoV-2

Since the emergence of SARS-CoV-2 in December 2019, the virus has continually evolved. Global bodies such as the WHO, in collaboration with research networks and governmental bodies, have been monitoring and assessing new variants. With data from genomic sequencing being produced at an extraordinary rate, there have been several different naming systems emerging within the literature. The most established systems for naming and tracking the

evolution of SARS-CoV-2 are the WHO label, GISAID clade (Global Initiative on Sharing Avian Influenza Data), Nextstrain clade[62] and Pango lineage.[63]

Mutations for SARS-CoV-2 occur at rate of approximately two mutations per month, but a higher rate of mutation occurs in immunocompromised patients who demonstrate chronic viral shedding including mutations in the S protein.[2,64]

The WHO defines a variant of interest (VOI) as a SARS-CoV-2 variant with genetic changes that are predicted or known to affect virus characteristics such as transmissibility, disease severity, immune escape, diagnostic or therapeutic escape.[65] In addition, the strain identified has to cause significant community transmission or multiple COVID-19 clusters, in multiple countries with increasing relative prevalence alongside an increasing number of cases over time.[65]

The WHO defines a Variants of concern (VOC) as a SARS-CoV-2 variant that meets the definition of a VOI and, through a comparative assessment, demonstrates one or more of the following changes at a degree of global public health significance:

- Increase in transmissibility or detrimental change in COVID-19 epidemiology.[65]
- Increase in virulence or change in clinical disease presentation.[65]
- Decrease in effectiveness of public health and social measures or available diagnostics, vaccines, therapeutics.[65]

The data in this thesis was collected between October 2019 and June 2021 therefore, the variants of concern discussed were those circulating in the UK between those dates (Table 1 and Figure 2).

Table 1 SARS-CoV-2 strains circulating at the time of data collection for the 'STORY' and 'COVID Warrior' studies. Table recreated from the WHO Tracking of SARS-CoV-2 variants[65]

WHO label	Pango lineage•	GISAID clade	Nextstrain clade	Earliest documented samples	Date of designation
Alpha	B.1.1.7	GRY	20I (V1)	United Kingdom, Sep-2020	VOC: 18-Dec-2020 Previous VOC: 09-Mar-2022
Beta	B.1.351	GH/501Y.V2	20H (V2)	South Africa, May-2020	VOC: 18-Dec-2020 Previous VOC: 09-Mar-2022
Gamma	P.1	GR/501Y.V3	20J (V3)	Brazil, Nov-2020	VOC: 11-Jan-2021 Previous VOC: 09-Mar-2022
Delta	B.1.617.2	G/478K.V1	21A, 21I, 21J	India, Oct-2020	VOI: 4-Apr-2021 VOC: 11-May-2021 Previous VOC: 7-Jun-2022

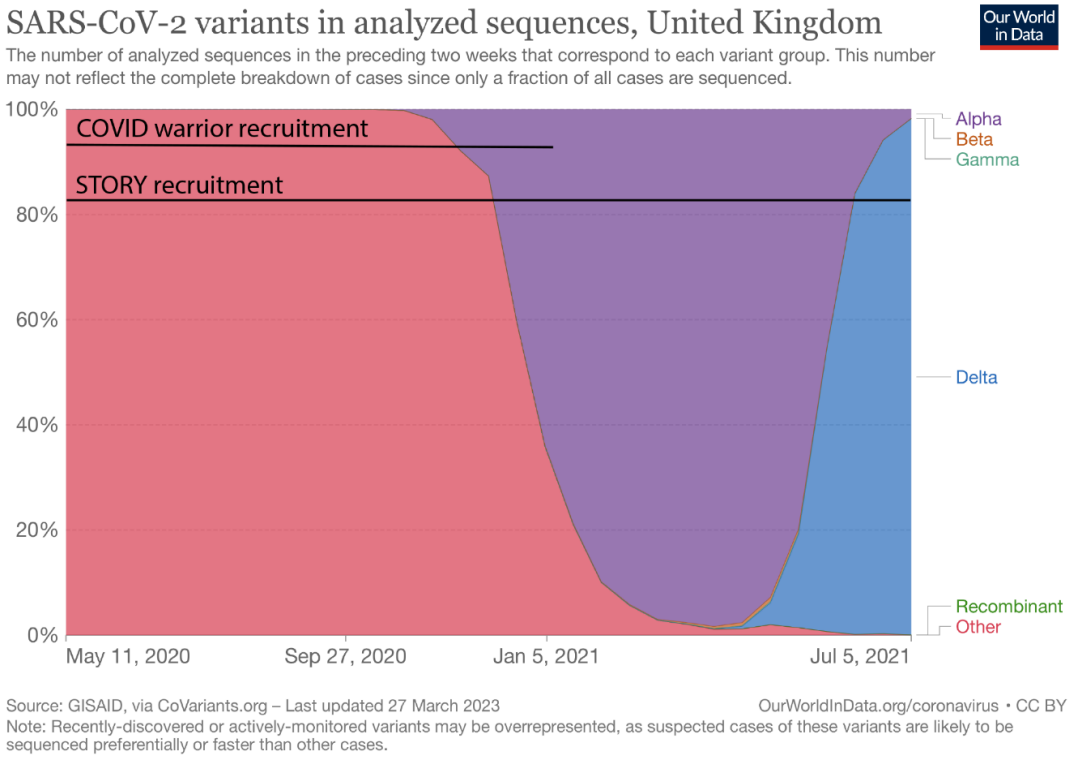


Figure 2 Timeline of SARS-CoV-2 variants circulating within the UK adapted to show recruitment periods for 'STORY' and 'COVID Warrior' studies. Data from John Hopkins University CSSE COVID-19 Data[3]

1.3 The immune response to SARS-CoV-2 infection

1.3.1 Overview of innate and adaptive immune responses to SARS-CoV-2 infection

The first line of defence against pathogens is the innate immune system, generating a fast response in both primary infections and reinfections of the same pathogen.[66] The response is detectable within hours after infection with a pathogen and there are three main purposes of the innate response;

1. to contain viral replication within infected cells,
2. to produce a local tissue environment which is hostile to the virus, which includes recruitment of effector cells,
3. to prime the adaptive immune response.[67]

The adaptive immune system plays an important role in the control and clearance of most viral infections in humans. There are three major cell types: B cells, which produce antibodies, CD4+ T cells, which are helper cells with effector functions, and CD8+ T cells, which kill infected cells.[68] The adaptive immune system is slow compared with the innate immune response, taking 6-10 days after priming to generate a big enough response to control a viral infection. This is because it has to select and expand virus-specific cells from large pools of naïve B cells and T cells specific for different molecular structures and sequences.[66] T helper cells, cytotoxic T cells and effector B cells work together to clear infected cells and circulating virus once sufficient populations are generated.

The innate immune response plays an important role in activating the adaptive immune system. If the innate response is delayed or avoided in COVID-19, the virus replicates without constraint leading to the adaptive response being delayed. Most infections are asymptomatic or mild as the infection is cleared by robust T cell[69] and antibody responses.[70-72]

It has been suggested that the combination of both a slow innate and adaptive immune response is responsible for the observed increased mortality and morbidity in older populations.[66]

Most individuals infected by SARS-CoV-2 have detectable antibodies over the threshold of positivity within 5 to 15 days after the onset of symptoms, with around 90% seroconverting by day 10.[73,74] Infected individuals mount anti-spike IgG, IgA, and IgM responses at the same time.[75,76] Neutralising antibodies develop over the same timeframe as seroconversion.[66] Multiple studies have reported neutralising antibodies against SARS-CoV-2 being developed from naïve B cells instead of from pre-existing cross-reactive B cells responses.[77-80]

Understanding the kinetics of antibody responses over time is important in understanding the level of population immunity. The persistence of circulating SARS-CoV-2 IgG titres being well maintained for 3–4 months was well documented by mid to late 2020.[81,82] Ward et al observed ~25% of cases becoming seronegative over the course of 6 months.[44] Circulating and saliva Spike IgA titres decline faster than IgG titres within the first three months.[75,83] Importantly, these early data were derived from adult studies, and the paediatric antibody responses had not been studied in detail. With antibody concentrations being linked to the severity of the disease,[84] there were concerns that a sustained antibody response would not be seen within the paediatric population due to the typical mild presentation in younger age groups.

1.3.2 Cross-reactive immunity with seasonal coronaviruses

Human coronaviruses (HcoV) include OC43, HKU1, 229E and NL63, usually cause common cold symptoms and co-circulate endemically with other respiratory infections.[85] Coinfections are frequently observed, resulting in a broad spectrum of clinical presentations.[85] Multiple studies have reported that cross-reactive circulating anti-Spike cross-neutralizing antibodies are rare.[76,86,87] However, cross-reactive T-cell memory has been reported,[70,88] which could provide pre-existing immunity.[66] In 2020 epidemiological evidence from large cohorts showed signs of potential pre-existing immunity in SARS-CoV-2 unexposed healthy donors.[89]

A laboratory-confirmed HCoV infection within the previous three years, has been shown to reduce the risk of intensive care admission when adjusting for confounding factors such as age.[67] This suggestion of cross-protective immunity from HCoV to SARS-CoV-2, combined with potentially higher seasonal coronavirus exposure in the paediatric population has been proposed as a mechanism for the reduced severity of COVID-19 in children compared with adults.[67]

Aldridge et al. analysed results from the Flu Watch study, a prospective community cohort study which reported a higher incidence of HCoV infections in children <5 years of age compared with other paediatric and adult age groups.[90,91] This trend could be due to patterns of testing rather than a greater risk of infection and could not establish the true burden of disease as many of these infections are asymptomatic or mild.[90] Nickbakhsh et al. reported a higher prevalence of OC43 among preschool children compared with older age groups. It has been proposed that the reduced severity of COVID disease in children could be explained by a sustained level of OC43 immune-mediated protection, with the increased vulnerability of adults to COVID-19 due to waning immunity against OC43.[85]

1.3.3 Differences in immune responses with age

At birth, the adaptive immune system is immature, and the innate responses are dampened, making newborns more susceptible to infection. This is because during pregnancy, the foetus is exposed to non-shared maternal antigens and requires immune tolerance towards them.[92] At birth, the environment changes, and the body is suddenly exposed to a plethora of pathogens and antigens, which require a fast defence mechanism.[93] Early defences include antibodies transferred transplacentally in the third trimester, but this protection is short-lived and lasts up to 6 months.[94] The innate immune system can respond to a broad range of pathogens, but adaptive immune responses are suboptimal due to the limited exposure to pathogens during pregnancy.[94] The innate and adaptive immune systems mature during

infancy and early childhood as they are exposed to various pathogens and immunisations. The repeated encounters with pathogens and immunisations produce faster and more effective immune responses, leading to long-term protection through immunological memory.[92] As the adaptive immune system matures, there is a shift from relying heavily on the innate immune system to a balanced coordination between the innate and adaptive responses. During childhood there is a shift in the balance between Th1 and Th2 responses. In the neonatal period there is a bias towards Th2 responses and suboptimal Th1 responses and B-cell differentiation, which results in a vulnerability to acute respiratory and gastroenteritis infections.[94,95] The bias towards Th2 responses shifts over the first few years of life. As children develop the Th1 and Th2 profile becomes balanced and by puberty the adaptive immune is fully developed producing an effective immune response against both intracellular and extracellular threats.[92]

In adults, severe COVID-19 causes damage in the lungs and other organs which leads to vascular damage and progression to ARDS and multi-organ failure.[96] Autopsy results in COVID-19 patients have shown inflammatory mononuclear cell infiltration and large numbers of lymphocytes. There was significant damage to the alveoli resulting in pneumocyte desquamation and hyaline membrane formation.[97-101] These studies support COVID-19 pathogenesis being the result of direct virus replication and the host response.[96]

The clinical outcomes seen in children and adolescents are very different to elderly adults. Early in the pandemic it was noted that the clinical course in the paediatric population was of an asymptomatic or mild infection.[33] In rare instances, a multisystem inflammatory syndrome (MIS-C) was observed in previously well children.[102] This has not been observed in the adult population. The variation in disease severity has made the differences in immune response between children and adults an area of intense research activity.

In asymptomatic or mild SARS-CoV-2 infection, cytokine profiles do not differ dramatically when comparing asymptomatic children with SARS-CoV-2 to symptomatic infected children with mild disease.[103] Cohen et al. showed total IFN γ , $CD4^+$ and $CD8^+$ T cell responses against the viral structural proteins were lower in SARS-CoV-2 infected children compared with adults.[104] In addition T cell responses from infected children at memory timepoints (>14 days post infection) had lower maximal activation compared with negative controls (uninfected children), suggesting that T cell activation was refractory based on recent infection and that SARS-CoV-2 infected children have dampened T cell responses.[105]

In response to mild SARS-CoV-2 infection children demonstrate a robust serological response,[102] with similar anti-spike IgG concentrations observed in adults and children with mild COVID-19.[102] In children less than 10 years of age, anti-SARS-CoV-2 antibody neutralising activity was shown to be higher than in adolescents and young adults up to 4 months after acute infection.[102]

Paediatric studies have demonstrated robust innate immune responses to SARS-CoV-2 infection.[106] Compared with adults, Pearce et al. demonstrated children had an increased expression of genes downstream of type I and II IFN in patients with similar viral loads.[106]

By comparing children and adults, it has been established that the immune response to mild SARS-CoV-2 infection is similar.[102] Differences appear in severe COVID-19 where adult responses are dysregulated, and children have greater IFN signalling, widespread T cell activation, and activation of a broad humoral response reactive to common viruses as well as SARS-CoV-2.[102]

1.4 Diagnostics

1.4.1 PCR testing

Real-time reverse transcription-polymerase chain reaction (RT-PCR) is the gold standard for SARS-CoV-2 diagnosis used in the UK in both community settings through track and trace and in healthcare settings.[107] RT-PCR detects the presence of SARS-CoV-2 nucleic acids in

nasopharyngeal and oropharyngeal fluids.[108] Nucleic acid tests have higher sensitivity and specificity than serum antibody testing.[108] Results from RT-PCR testing take up to 72 hours from the collection of the specimen.[109] The viral RNA from the sample undergoes reverse transcription into complementary DNA (cDNA) strands. These DNA strands are then amplified using *Taq* DNA polymerase, allowing detection of the nucleic acids.[110] Limitations of this method include false negatives due to viral evolution, sample storage, low-quality specimens, high cost, and wait times.[108]

1.4.2 Rapid antigen testing

Point of care testing (POC) has played a pivotal role in the timely prevention and control of the pandemic within the UK. This rapid assay detects the presence of viral antigens expressed by SARS-CoV-2.[108] Then antigen binds to antibodies affixed to a paper strip in the plastic casing of the test kit. A detectable signal can be generated visually within 30 minutes of sampling.[111] POC testing has several advantages over labour and time-intensive traditional testing methods given they are faster, with high sensitivity and specificity and cost less than PCR testing.[112] An important limitation is that the viral antigen which the test detects is only expressed when the virus is replicating therefore POC tests are most useful in acute or early infection.[111] The effectiveness of testing depends on the concentration of the virus in the specimen and the quality of the specimen collected.[111]

1.4.3 Serological testing

Both IgM and IgG responses can be measured. IgM levels are not very useful in the context of seroepidemiology as the response is only detectable for a short period of time after the acute infection. IgG assays target either the spike protein, nucleocapsid or receptor binding domain. IgG is detectable from 6-10 days after infection and is maintained above the threshold of positivity for a long period, potentially years later.[113] Spike or nucleocapsid specific antibodies

are the most frequently assayed. Spike assays detect both infection and those who have received a vaccine, whereas nucleocapsid detects recent infection only.

1.4.4 Saliva/ oral fluid

Studies have described saliva as a useful tool in the surveillance of SARS-CoV-2 infection.[75,83,114] Collection can be carried out by the individual without assistance. However, careful instructions need to be given to ensure the correct timings with respect to the last meal and drink ingested.[114] Producing a sufficient volume has proved challenging, especially in younger age groups. Saliva is made up of different biological components, including proteins, electrolytes, nucleic acids from multiple sources both local and systemic.

1.5 Development of a paediatric prospective serological study in England

The UK Health Security Agency (UKHSA) has an ongoing seroprevalence programme to assess population-level immunity against vaccine preventable infectious diseases using residual serum samples from participating laboratories in England. Samples are anonymised, and individual vaccination status is unknown, therefore vaccination data are derived from vaccination programmes and known immunisation coverage at a population level.

An alternative approach to assessing population level immunity is to use active collection of new samples recruited from the community. One example of such a programme is the national seroepidemiological surveys run every ten years in the Netherlands, consisting of a prospective collection of serum samples from a representative cross-section of society.[115] This approach overcomes the limitations of using residual serum samples and would potentially allow a more accurate assessment of the prevalence of antibodies within the population against the infectious disease of interest.

In the most recent survey carried out in 2006/7 Klis et al. recruited Dutch inhabitants (aged 0–79 years), identified from the national population register from 40 municipalities throughout the country.[115] Oversampling took place in areas with low vaccine coverage and migrant populations. Over 17,000 individuals were invited to participate. Individuals received a letter

of invitation together with a brochure containing information on the study, a questionnaire, an informed consent form, and a prescheduled appointment form for blood donation at a local clinic. Participants were offered a gift voucher. The highest response rates (both 38%) were in women aged 10–49 years and 50–79 years. Male and female children aged 0-9 years had response rates of 27% and 26% respectively.[115] Overall, the response rate was 32% (6,386 serum samples), which was lower than the 50% response rate achieved in the first (1995/96) Dutch seroepidemiological study.[116] This was possibly due to greater travelling distances to clinic in the most recent study.[115]

In response to the concerns about the potential limitations of the UK residual sample approach and advantages of an active sample collection program, the Oxford Vaccine Group in collaboration with UKHSA conducted a pilot study called "What's the STORY?", NRES Reference-19/LO/1040 ('STORY'). The aim was to assess the feasibility of establishing a national seroepidemiology survey in England in individuals aged 0–24 years, focusing initially on diphtheria and Group C invasive meningococcal disease through a prospective multicentre cohort study. 'STORY' started in October 2019 however, because of the emergence of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection causing Coronavirus disease 2019 (COVID-19),[3] the study was adapted to address questions regarding SARS-CoV-2 infection within the paediatric population. As a result, the evaluation of population immunity against diphtheria and Group C invasive meningococcal disease has been delayed and will not be discussed as part of this thesis.

1.6 Aims and objectives of the thesis

This thesis aims to evaluate the feasibility and added public health benefit of an English population-based seroepidemiological programme in children and young adults aged 0 to 24 years. In addition, it aims to determine the prevalence of SARS-CoV-2 infections among children and adolescents between October 2019 and June 2021.

Specifically, this thesis will:

1. establish if the proposed methodology succeeded in recruiting a representative sample at a national and local level in terms of ethnicity, community identity and socioeconomic background. Therefore, providing a store of sera from a representative section of children and young adults between the ages of 0-24 years with appropriate consent, demographic, and immunisation details available for future testing of immunity against other infectious diseases of relevance to UK immunisation schedule and public health. (Chapter 3)
2. determine the variation of SARS-CoV-2 infections by age, ethnicity and geography. (Chapter 4)
3. establish the kinetics of antibodies responses specific to SARS-CoV-2 following infection in a paediatric population. (Chapter 5)
4. determine the relationship between serum and salivary antibodies against SARS-CoV-2. (Chapter 5)
5. determine whether prior exposure to seasonal coronaviruses provides pre-existing immunity to SARS-CoV-2 infection. (Chapter 6)

Chapter 2 Methodology used in the 'What's the STORY?' study

2.1 Introduction

2.1.1 Aims

The aim of this chapter is to describe the methodology used within the 'STORY' study, including:

1. Study design
2. Recruitment
3. Data collection
4. Laboratory methods

2.2 Study design

2.2.1 Study sites

England is divided into seven NHS regions (Figure 3)

- East of England
- London
- Midlands
- North-East & Yorkshire
- North-West
- South-East
- South-West

The aim was to collect blood samples from children and young adults between the ages of 0-24 years from seven NHS regions allowing results to be generalised to England as a whole.

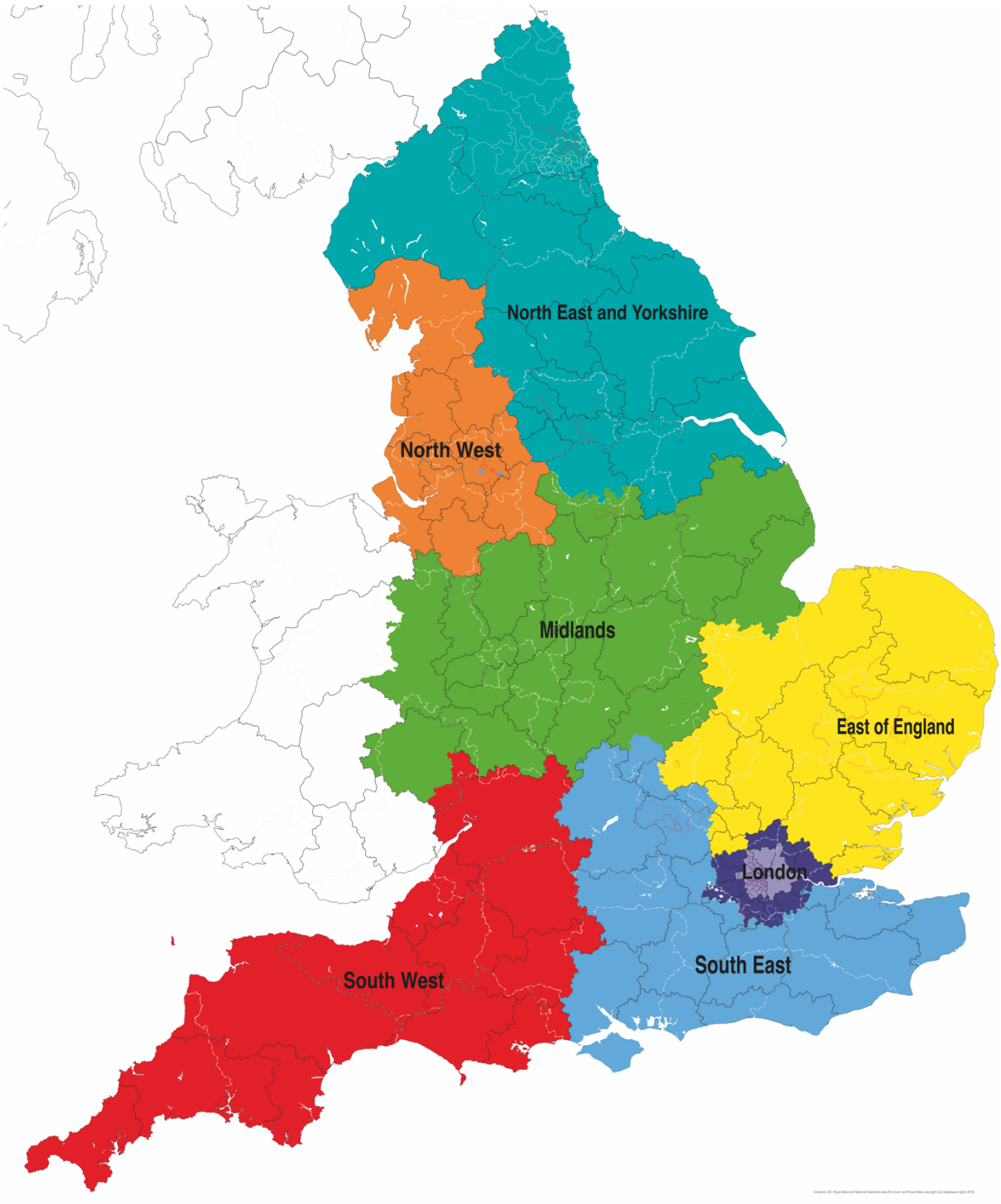


Figure 3 A map to illustrate the seven NHS regions in England

The study was a repeat cross sectional seroepidemiological study.

The study sites included (Figure 4)

- University of Oxford
- Sheffield Children's Hospital NHS trust

- Bradford Teaching Hospitals NHS Foundation Trust
- Leeds teaching Hospitals NHS trust
- University Hospitals Bristol NHS Foundation Trust
- University of Southampton NHS Foundation trust
- Royal Manchester Children's Hospital – Manchester University NHS trust
- St George's University Hospitals NHS Trust
- University of Nottingham Health Service
- University Hospitals Plymouth NHS Trust
- The Newcastle Upon Tyne Hospitals NHS Foundation Trust
- Imperial College Healthcare NHS Trust
- West Suffolk NHS Foundation Trust

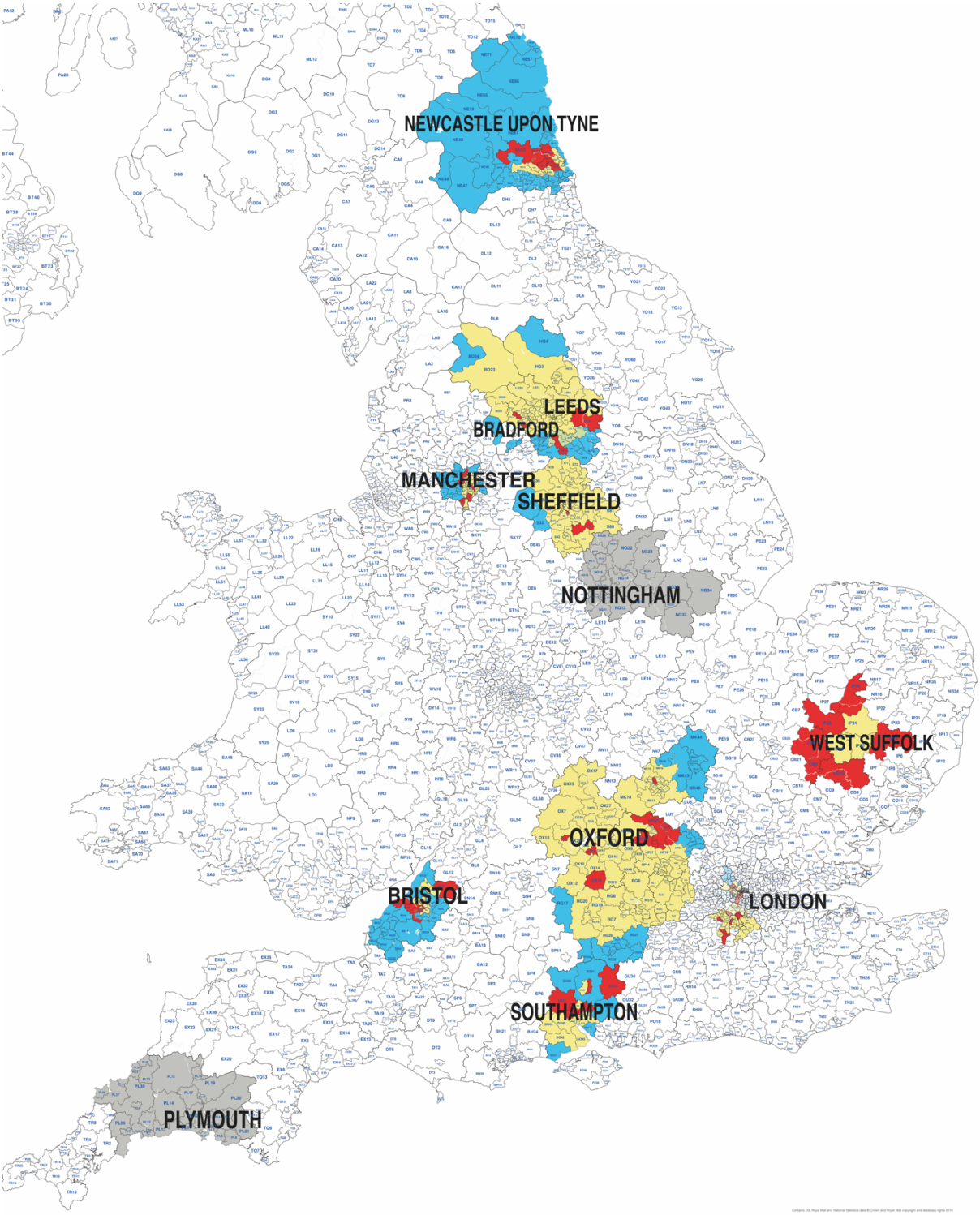


Figure 4 shows sites recruiting for the 'STORY' study. Postcode districts around each site were divided into – postcode districts selected by UKHSA for targeted recruitment (red), postcode districts within the site's catchment area but not selected (yellow) or out of range of the study site to recruit participants (blue). Plymouth and Nottingham's regions (grey) were not divided into postcode districts for recruitment.

The ‘Yorkshire and Humber region’ included three sites (Sheffield, Leeds, and Bradford) and the ‘London region’ included two sites (Imperial and St Georges).

2.2.2 Recruitment

Potential participants were contacted by mail-out invitation letters with study information, through GDPR (General Data Protection Regulation**)** compliant bulk mailing systems (Docmail), using extracts from NHS England Databases, the Child Health Information Service (CHIS) or through the Clinical Research Network. We expected 5-10% of those contacted by invitation letter to respond.

Other methods of promoting the study included website-based advertising, social media, radio, printed publications, contacting families registered with a study site research database and poster advertisements. Potential participants were also identified by local study teams through their visits for other research studies. Study information was also disseminated through GP practices and health visitors. Participants interested in taking part contacted their local site to arrange visits.

In cases where English was not the parent/guardian or participant’s first language, the study information was provided in written form in the other languages (Table 2). For languages other than those listed below, language line was available by telephone allowing interpretation into any language required for informed consent to be obtained.

Table 2 The languages in which study information was translated

French	Somali
Portuguese	Urdu
Spanish	Punjabi
Arabic	Bengali
Persian/ Farsi	Nepalese

Overall recruitment was monitored by the Oxford Vaccine Group to ensure recruitment from each region was spread over all ages with adjustments to overall targets made as necessary.

The aim was to ensure the sample was broadly representative of the region according to Index of Multiple Deprivation (IMD) scores which is a measure of socioeconomic status. Other details collected in the questionnaire such as ethnicity, community identity and the third version of the Family Affluence Scale (FASiii) were used to assess how successful this strategy was for each region.

The original recruitment strategy (hereon referred to as 'group one' which excluded Nottingham and Plymouth) aimed to select a representative sample from each NHS region. Each group one site defined their region (the catchment area the local NHS trust served shown in Figure 4 as red, yellow and blue postcode districts combined). From the regions, each site stated which postcode districts they could feasibly recruit from (Figure 4 red and yellow postcode districts). Local knowledge was used to classify postcode districts into urban or rural. UKHSA generated a list of all postcodes in recruiting regions and determined the quintiles of IMD within that region using the English indices of deprivation 2019 database, available through the Ministry of Housing, Communities and Local Government.[117] For each site UKHSA randomly selected up to seven postcode districts from the feasible postcode districts. Across the selected postcode districts each quintile had between 15% and 25% of postcodes (Table 3, Figure 5 red postcode districts). The postcode district is the first two letters and number e.g.

OX	**3**	**7**	**LE**
Area	District	Sector	Unit

Table 3 UKHSA selected postcodes by site

Oxford	HP17	HP22	HP23	MK13	OX1	OX11	OX28
St Georges	CR4	KT18	KT6	SM3	SW11	SW20	
Southampton	SO16	SO23	SO24	SO50	SO51	SO52	
Bristol	BS2	BS20	BS3	BS37	BS41	BS8	
Manchester	M1	M12	M20	M23	M25	M6	
Yorkshire and Humber	BD15	BD6	LS25	S14	S43	WF2	
Suffolk	CB8	CB9	CO10	IP14	IP24	IP28	
Imperial	SW1	SW3	NW6	NW8	W11	WC2	
Newcastle	NE3	NE12	NE13	NE20	NE23	NE28	

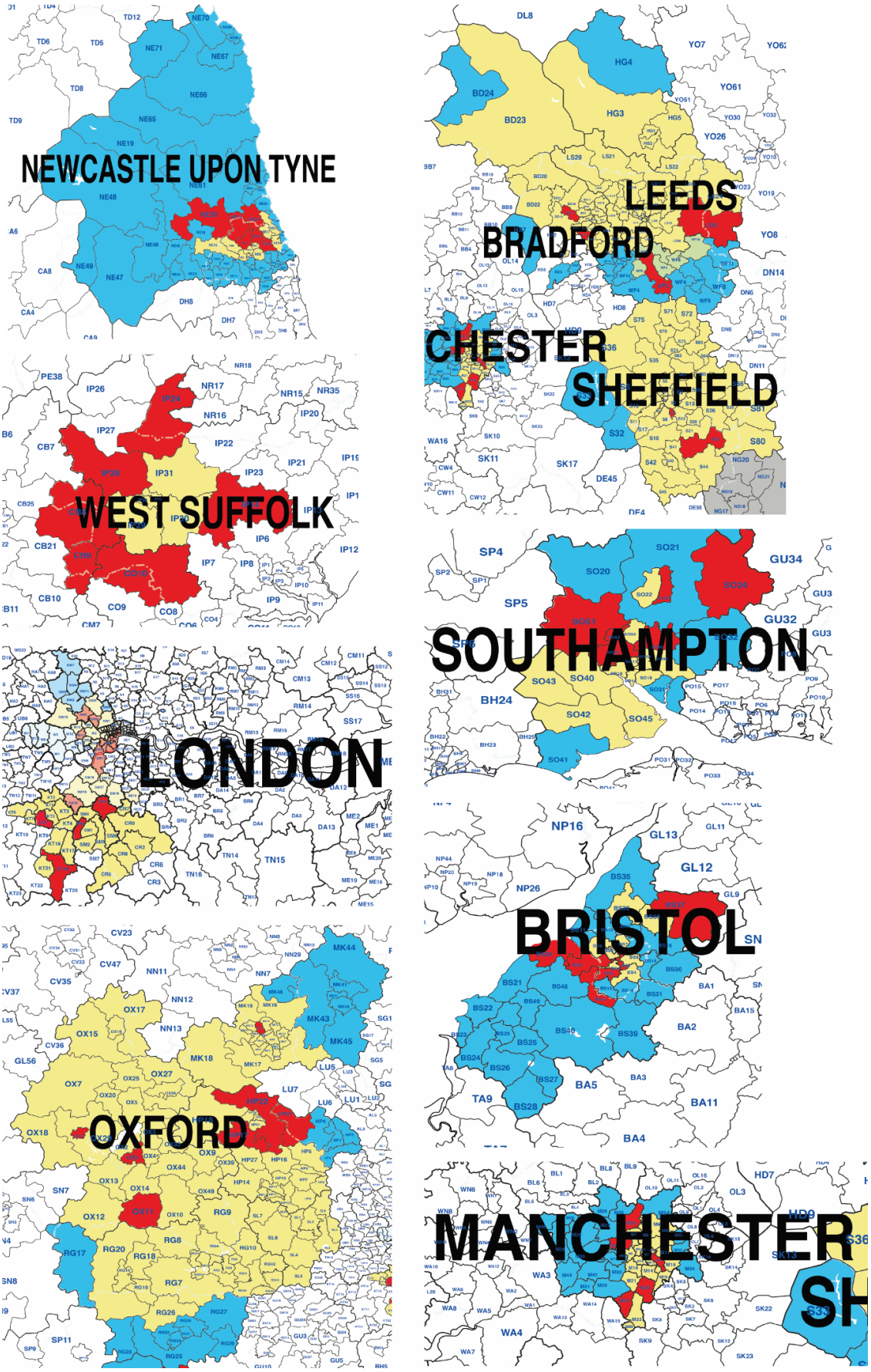


Figure 5 shows each site using postcode districts in targeting recruitment. Each site was divided into – postcode districts selected by UKHSA for targeted recruitment (red), postcode districts within the site's catchment area but not selected (yellow) or out of range of the study site to recruit participants (blue).

NHS digital and CHIS databases using GDPR-UK compliant bulk mailing systems sent study information leaflets to a proportion of households with children living within UKHSA selected postcode districts. In the information that participants (and/or their families) received it was made clear that they may not hear from the site immediately, as the aim was to achieve a representative sample for that region. Therefore, every participant who registered interest in taking part in the study was stored on a local study database containing basic contact details and age. As recruitment progressed this was searched to see if any of the participants were from postcodes with quintiles that were under represented in the sample.

Parents/legal guardians that indicated they did not want to take part in the study and/or receive further communication about the study were not included in any subsequent contact lists. Imperial, Newcastle and Suffolk were added as sites using the group one recruitment strategies as recruitment slowed. Many sites had competing vaccine studies and by adding Newcastle and Suffolk all NHS regions in England were represented.

Due to the rapid expansion of the study in response to the COVID-19 pandemic, Plymouth and Nottingham were added as sites and used different recruitment strategies. Recruitment in Plymouth and Nottingham was not restricted by postcode districts, instead sites used established local methods of recruitment. Group one sites were also allowed to expand recruitment into postcode districts not originally selected for recruitment. Participants recruited from non-selected postcode districts were placed in group two.

Table 4 Summary of recruitment strategies by group.

Group	Type
1	Participants recruited from selected postcode districts, with the aim of recruiting across the socioeconomic spectrum.
2	Participants recruited outside the selected postcode districts. All sites contributed to group two which was added to increase the numbers of recruited participants.
3	Enhanced recruitment of participants identifying as belonging to a minority ethnic group as a result of an interim analysis which showed participants identifying as belonging to a minority ethnic group were at higher risk of seropositivity than white participants.

2.2.3 Sample Size

Sample size was determined by Professor Andrews from UKHSA based on the original study objectives for creating a seroepidemiological survey.

2.2.4 Group one

The sample size of group one was determined based on the primary objectives as well as the secondary objective to have a store of sera for future testing of immunity.

Historically serosurveys such as those carried out as part of the European Seroepidemiology Network (ESEN) have had target sample sizes per age band of interest of 100-200. Age bands have usually been one-year bands until adult ages. For this study this was taken as one-year age bands to age 21 then a single band for young adults between the ages 22-24 years. For the secondary objective, 100 participants per age band allowed for estimates of seroprevalence to be determined with levels of precision as shown in Table 5.

Table 5 Precision of seroprevalence estimates

Prevalence	95% CI Observed around the estimate with 100 per age band (2300 total)
10%	4.9-17.6
20%	12.7-29.2
30%	21.2-40.0
40%	30.3-50.3
50%	39.8-60.2
60%	49.7-69.7
70%	60.0-78.8
80%	70.8-87.3
90%	82.4-95.1

2.2.5 Representativeness of participants sampled

Within each study site representativeness was assessed by comparison to local population demographics. Each study region planned to have approximately 383 individuals recruited into group one. Overall, all 2,300 samples were to be available for analysis in group one. With these numbers the precision of estimates of characteristics in the population and differences that would be significant to population data are given in Table 6. For example, if a specific ethnic

group comprised 20% of the survey sample, then in a region this would have a 95% CI of 16.1-24.4% and would allow a proportion in the population of <14.3% or >26.6% to be detected as significantly different.

Table 6 Assessment of representativeness, precision and detectable differences

	Region (N=383)		Overall (N=2300)	
Prevalence of characteristic in sample	95% confidence interval	Prevalence below, above this in population that would significantly differ (80% power, 5% significance, assuming population much larger than sample (>=10 fold))	95% confidence interval	Prevalence below, above this in population that would significantly differ (80% power, 5% significance, assuming population much larger than sample (>=10 fold))
5%	3.0-7.7	2.3,9.1	4.1-6.0	3.8,6.5
10%	7.2-13.5	6.0,15.3	8.8-11.3	8.2,12.0
15%	11.6-19.0	10.0,21.0	13.6-16.5	12.9,17.3
20%	16.1-24.4	14.3,26.6	18.4-21.7	17.6,22.5
25%	20.7-29.7	18.8,32.0	23.2-26.8	22.4,27.7
30%	25.5-34.9	23.3,37.3	28.1-31.9	27.2,32.9

2.2.6 Sample size for SARS-CoV-2 incidence based on change in prevalence

Combining group one and two, the plan was to have over 3500 blood samples from 2800 to 3500 participants. At least 3200 samples were to be from participants 19 years of age or younger. Younger age groups were not captured in other seroepidemiological surveys, and by including participants aged 18-19 years comparisons could be made with adult studies.

Table 7 below shows the precision (95% CI) of estimates of change in prevalence (incidence) with 100 samples at each of two time points. For example, with a prevalence of 10% at baseline and 30% at the next time point the incidence is 20% with 95% CI 9.3-30.7.

Table 7 Precision of estimates of change in prevalence with 100 samples at two time points.

	Baseline prevalence		
Prevalence	**0%**	**10%**	**20%**
At next time point			
10%	10 (4.1, 15.9)		
20%	20 (12.2, 27.8)	10 (0.2, 19.8)	
30%	30 (21.0, 39.0)	20 (9.3, 30.7)	10 (0.0, 21.9)
40%	40 (30.4, 49.6)	30 (18.7, 41.3)	20 (7.6, 32.4)
50%	50 (40.2, 59.8)	40 (28.6, 51.4)	30 (17.4, 42.6)
60%	60 (50.4, 69.6)	50 (38.7, 61.3)	40 (27.6, 52.4)
70%	70 (61.0, 79.0)	60 (49.3, 70.7)	50 (38.1, 61.9)

2.2.7 Analysis populations

The analysis population for recruitment rates was all individuals invited; for representativeness this was all individuals providing a sample and, for seroprevalence, it was all individuals providing a sample for which a laboratory result was obtained.

2.2.8 Group 3

In response to the COVID-19 pandemic, 'STORY' samples were used to determine the seroepidemiology of SARS-CoV-2 in children and young adults living in England through 2020/2021. Early in the pandemic risk factor analyses within the adult population showed higher rates of COVID-19 disease in non-white minority ethnic groups.[44] Therefore, attempts were made within 'STORY' to enhance recruitment from non-white minority ethnic groups within paediatric populations.

Due to the higher rates of SARS-CoV-2 infection reported in non-white minority ethnic adults,[44] group 3 was created to enhance numbers of participants identifying as non-white to establish whether a similar trend was seen in the paediatric population. Following discussions with UKHSA a target of 300 participants, with at least 100 from the Black/ African/Caribbean/Black British community was set. This was based on preliminary data suggesting rates of seropositivity in adult males, identifying as belonging to minority ethnic groups, were double that of the white community, and rates in those identifying as Black/ African/ Caribbean/Black British were three-fold higher.[44] Based on these proportions, and an estimate of 4% seropositivity in the white community, the above numbers would provide 80% power to detect similar differences in the paediatric populations.

2.2.9 Longitudinal cohort

A repeat cross-sectional longitudinal cohort sub study was also incorporated within this study to provide information regarding antibody decline in both serum and oral fluid (OF) in children and adolescents aged between 0–19 years. Participants within this cohort provided up to four blood samples (and three oral fluid samples) allowing analysis of the change in SARS-CoV-2

specific antibody concentrations over time, and the relationship between serum and oral fluid antibodies. The minimal interval between samples was two months. The longitudinal sampling cohort comprised both seropositive and seronegative participants. To boost the number of participants with longitudinal samples, our data were combined with data from the 'COVID Warriors' seroepidemiological survey[24] (REC Reference-20/HRA/1731 and Reference 19147TW-SW). 'COVID Warriors' recruited participants from five UK centres between April 2020 and January 2021. Participants were children aged 2-15 years whose parents were National Health Service (NHS) employees. Potential participants were recruited via internal intranet advertisements and email circulars.[24] Participants provided a blood sample at baseline, two and six months.

Early in the pandemic, it became clear that children were less severely affected by SARS-CoV-2 infection than adults. Hypotheses to date have included differences in humoral immune responses to SARS-CoV-2 infections in adults and children or differences in baseline humoral immunity to seasonal coronaviruses.[118] Alternate hypotheses suggest differences in cellular immunity such as T cell responses or cytokine responses in adults and children.[119,120] Both the 'STORY' and 'COVID Warriors' longitudinal cohorts provided an opportunity to assess whether baseline immunity to seasonal coronaviruses was protective against SARS-CoV-2 infection within the paediatric population. Datasets from both studies were compared within a set of 74 pre-pandemic serum samples collected from children aged 0-13 years in 2009/2010 (EudraCT Number: 2009-014719-11) and 2017/2018 (EudraCT Number: 2017-004732-11).

2.2.10 Screening and Eligibility Assessment

Given the low-risk nature of the procedures involved in the study, there was no formal requirement for screening prior to the first study visit. However, sites did contact potential participants to discuss any questions they may have had about the study, check the inclusion and exclusion criteria, and discuss arrangements for the visit to allow minimal contact with

potential participants and their families during the pandemic. Appointments were carried out either in a pre-arranged clinic or in the participant's own home.

2.2.11 Informed Consent

The parent/legal guardian of the participant or the participant themselves, if able to consent, personally signed and dated the latest approved version of the informed consent form. Assent forms were collected for participants between the ages of 11 to 15 years.

A written version and verbal explanation of the Study Information leaflet and Informed Consent was presented to the participant or parent/legal guardian of the participant detailing:

- the exact nature of the study
- what it involved for the participant
- the implications and constraints of the protocol
- the known side effects and any risks involved in taking part
- sample handling – participants were informed that anonymised samples taken during the course of study may be shared with study collaborators.

Individual results were not shared with participants – the study aimed not to actively enrol individuals with known or suspected COVID-19, but instead provide a series of snapshots of the general population to describe seroprevalence to SARS-COV-2 at a population level.

It was clearly stated that the participant was free to withdraw from the research study at any time for any reason without prejudice to future care, without affecting their legal rights and with no obligation to give a reason for withdrawal. Data captured up until that point was kept unless the participant stated they wished this data to be withdrawn.

If for any reason a participant gave consent but either blood was not obtained, or the sample was less than the minimum sample required (<0.5ml) they were considered a failed enrolment

and were not counted towards the recruitment numbers, i.e. they were replaced by another participant.

In addition to the informed consent for the provision of serum and the first visit of this study, participants (or their parents/guardians) were also given the opportunity to provide (optional) consent for:

1. Donation of the blood clot left after centrifugation of whole blood to the biobank at the Oxford Vaccine Group. This would allow for extraction of DNA to interrogate the influence of donor's genotype on vaccine/infection induced immunity, and other aspects related to the interdependence between genetics and immunity against infectious diseases.
2. Participation in the longitudinal cohort, in which additional blood samples were taken along with an oral fluid sample and repeat administrations of the COVID-19 questionnaire.

2.2.12 Collection of sample

Blood (and, if applicable, oral fluid) samples were collected in specially designated clinics or home visits. This approach was adapted to suit different age groups depending on local needs. For example, specially designated clinics run during school holiday periods or weekends were appropriate for school aged children. Collection of samples took into account infection control measures in place in response to the COVID-19 outbreak.

Samples were taken in line with local standard operating procedures. A local anaesthetic cream or spray was offered to child participants prior to venepuncture. If the initial attempt at venepuncture was not successful verbal consent was sought for a further attempt at that visit. No more than two attempts were made in one visit. Maximum blood volumes were based on 0.8ml/kg in line with guidance given by the European Commission[121] of public health (Table 8). The volume of blood for each age group was based on the weight of a child on the 0.4th

centile on the female UK-WHO growth chart.[122] A participant was only considered enrolled once a blood sample >0.5ml was taken.

Table 8 Maximum (target) blood volumes by age

Age	Maximum (target) volume of blood (mls)
<2 months	2ml
2-6 months	3ml
6-12 months	5ml
1-2 years	6ml
3-7 years	10ml
8-11 years	15ml
12-14 years	20ml
15 -24 years	30mls

2.2.13 Data capture and Questionnaire

The case report form contained basic demographics such as age and postcode in addition to answers to inclusion/exclusion criteria, blood sampling and vaccination histories. Basic demographic characteristics were collected via a questionnaire and included: sex, GP details, ethnic group, association with communities of special interest (e.g. faith communities), vaccination history and personal and household history of recent respiratory or coronavirus infections. To allow comparison of the sample to the region, standardised questions from the census and FASii were included. Responses to the individual-level UK Census[123] and FASiii questions[124] for children and adolescents aged 0-15 years (scored out of a total of 13)[125] were used to assess the appropriateness of the LSOA-level IMD and the IDACI (the proportion of children aged 0-15 years living in income deprived families) scores for our study. These were captured using REDcap version 12.0.19.

2.2.14 Vaccination history

Vaccination histories were verified during serum sample collection using the Red Book or other vaccination records, or by checking with the general practitioner or the CHIS database. Contemporaneous vaccination history was also obtained from CHIS for children aged 0-5 years, and historical vaccination history could be obtained via GP records for participants aged 2–24 years. A limitation of this method of collection is the quality and completeness of vaccination history in GP records which varies by practice, particularly for children in older age groups who may have been vaccinated at a different practice or at school.

2.2.15 Reimbursement

Participants were offered a £20 voucher per visit as reimbursement for travelling to the study clinic. If they were seen at home, there was no reimbursement. Based on the average EU income levels, the European Society for Paediatric Gastroenterology Hepatology and Nutrition (ESPGHAN) suggest that an incentive of up to the value of 30 Euros was considered acceptable for children and adolescents, and may be offered as cash, vouchers or gifts or toys.[126] Other studies in adolescent age groups in the UK have had approval to use £10 book vouchers, and incentives such as these, or age appropriate gifts or toys in preference to cash for recruiting adolescents and younger children.

2.3 Laboratory methods

2.3.1 Serological assays

This project used multiple different serological tests for the detection of SARS-CoV-2 antibodies. Tests included Enzyme-linked immunosorbent assay (ELISA), electro-chemiluminescent immunoassay (ECLIA) and chemiluminescent immunoassay (CLIA).[127-129] Samples were sent from sites to UKHSA laboratories in Colindale and the University of Oxford to Professor Miles Carroll's group. Samples were processed and results fed back for analysis. ELISA, ECLIA, CLIA assays all used a sandwich technique to detect antibodies against SARS-CoV-2. These assays are carried out in well plates which allow multiple samples to be measured in one run. These are sandwich assays, so named because they trap antibody with the

participants serum between the antigen in the well and the enzyme linked antibodies.[130] Each type of sandwich assay follows these key steps.[130]

1. The well is coated by the antigen of interest e.g. Spike protein, nucleocapsid or receptor binding domain.
2. The wells are then filled with dilutions of the participants serum. If antibodies are present against the target antigen then they will bind to the antigen within the well. The well is then washed out to remove all unbound antibodies.
3. A solution of animal antibody against human antibodies is added; this antibody is covalently conjugated to an enzyme. Wells are then washed to remove any unbound enzyme conjugated antibody.
4. A read out from the analyser depends on the type of assay. An ELISA looks for colour change when a chromogenic enzyme substrate is added. This can be quantified by the analyser. A ECLIA applies voltage to the electrode then induces a chemiluminescent emission which was measured by a photomultiplier. A CLIA assay uses a chemical reaction to generate chemiluminescence.

Assays used in this study are detailed in Table 9.

Table 9 Serological assays

Assay	Manufacturer	Type of assay	Target	Sample size	Threshold of positivity		Sensitivity	Specificity
Roche Elecsys® Anti-SARS-CoV-2 IgG spike assay	Roche Diagnostics GmbH	ECLIA[131,132]	spike protein receptor binding domain[131,132]	sample volume 12 µL. Sample vol + dead space 100µL[131,132]	>=0.8U/ml[131,132]		95.5% (95%CI 93.2-97.1) [131,132]	100% (95% CI 99.1-100) [131,132]
Roche Elecsys® Anti-SARS-CoV-2 IgG nucleocapsid assay	Roche Diagnostics GmbH	ECLIA[129]	nucleocapsid (N) protein[129]	sample volume 20 µL Sample vol + dead space 100µL[129]	>=1.0 COI (Cut off index)		83.9% (95%CI 74.8-90.7) [129]	100% (95% CI 99.1-100) [129]
Receptor binding domain UKHSA assay	UKHSA	ELISA	Receptor binding domain		>=5 T/N[133]*		93.0% (95% CI 88.1-96.3) [133]	98.0% (95% CI 97.0-98.8) [133]
DiaSorin LIAISON® SARS CoV-2 S1/S2 IgG assay	DiaSorin LIAISON®	CLIA[127]	S1 and S2[127]		<12AU/ml negative, <=12 <15 equivocal, >= 15 positive[127]		64.0% (95%CI 53.8-73.4) [127]	97.7% (95% CI 95.8-99.0) [127]
					Adult	Paediatric		
MSD Vplex immunoassay	MesoScale Discovery	ELISA[134-136]	SARS-CoV-2 Spike		450 AU/ml[135]	1160 AU/ml[135]	90.8 % (86.0–94.1)[135]	97.4 % (95%CI 94.1 to 98.9)[135]
MSD Vplex immunoassay	MesoScale Discovery	ELISA[134-136]	SARS-CoV-2 RBD		815 AU/ml[135]	1169 AU/ml[135]	78.1% (71.8–83.3)[135]	92.3% (95%CI 87.6 to 95.3)[135]
MSD Vplex immunoassay	MesoScale Discovery	ELISA[134-136]	SARS-CoV-2 Nucleocapsid		3364 AU/ml[135]	3874 AU/ml[135]	73.0% (66.3–78.7)[135]	92.8 % (95%CI 88.2 to 95.7)[135]
IgG Capture ELISA NP	UKHSA[114]	ELISA[114]	Nucleopcapsid protein[114]		>= 1.0 COI[137]		80% (95%CI 71-80) [137]	99% (95%CI 98-100) [137]
IgG Capture ELISA S	UKHSA[114]	ELISA[114]	Spike protein[114]		>= 1.0 COI[137]		78% (95%CI 68-86) [137]	99% (95%CI 98-100) [137]

*T/N ratio – samples analysed in duplicate. Optical density evaluated at 450 nm (OD_{450}) data by dividing average OD_{450} values for individual samples by average OD_{450} of a known calibrator with negative antibody levels (T/N ratio). [133]

2.3.2 Crevicular fluid

Oral fluid (OF) was collected using the *OracolTM swab.*[114,137] Participants were asked to swab their gum line for 2 minutes.[138] Swabs were shipped to UKHSA at Colindale. OF was extracted from the foam swab by centrifugation to remove cellular debris. An ELISA targeting Nucleocapsid and Spike proteins was undertaken.

2.3.3 DNA storage

When processing serum samples a clot containing DNA is separated from the serum. Extraction and storage of the clots will allow for extraction of DNA at a future date to interrogate the

influence of donor's genotype on vaccine/infection induced immunity. The DNA samples obtained within the 'STORY' study were added to the Oxford Vaccine Centre's existing biobank of stored biological samples. to facilitate further research on immunisation, immunity and infectious diseases.

2.4 Approvals

Approvals were sought through NHS ethics committees. 'STORY' Ethical approval was obtained from the London-Surrey Research Ethics Committee (REC Reference-19/LO/1040) 'COVID Warriors' Ethical approval was obtained from the London-Chelsea Research Ethics Committee (REC Reference-20/HRA/1731) and the Belfast Health & Social Care Trust Research Governance (Reference 19147TW-SW).

Chapter 3 A comparison of the 'STORY' study sample demographics with those of children and teenagers living in England

3.1 Aims

This chapter aims to evaluate whether the 'STORY' study met the following objectives.

1. Demonstrate the feasibility of running an English population based seroepidemiology programme in children and young adults aged 0-24 years
2. Evaluate whether the methodology was successful in recruiting a representative sample at a national and local level in terms of ethnicity, community identity and socioeconomic background
3. Demonstrate how successful the serum store was in terms of collecting samples from children and young adults between the ages of 0-24 years with appropriate consent and information such as information demographics, and immunisation details available for future testing of immunity against other infectious diseases of relevance to UK immunisation schedule and public health.

3.2 Capturing sample demographics

The primary objective of 'STORY' was to recruit a representative sample of the paediatric population in England at a regional and national level. In order to achieve this the participants enrolled had to closely reflect the characteristics of the population of interest.[139] In the 'STORY' sample, the characteristics of interest included sex, ethnicity, community membership, migrant population, and socioeconomic background. A representative sample would allow for estimates of immunity within the paediatric population to be extrapolated to population-level immunity within these age groups.[139] Due to the COVID-19 pandemic, recruitment methods changed as the study was rapidly expanded to provide information regarding SARS-CoV-2 infections in children. Here, a balance was struck between having a representative sample of children and young adults in England and obtaining adequate numbers of samples from groups of special interest. In the context of COVID-19, the group of interest

was children from minority ethnic groups. This chapter explores whether a representative population sample was enrolled in the study using the methodology outlined in chapter two.

3.2.1 Methods

The participant characteristics were summarised using descriptive statistics and compared to national and regional data. This included comparing recruitment methods, response rates, failed enrolment rates and participant characteristics. Comparisons with national and regional data were made to determine the representativeness of the participants recruited and whether there were differences in participants' characteristics attained from group one and group two recruitment methods.

The original recruitment method (group one) was designed to recruit a representative sample of socioeconomic status from each site. Group one used a strategy of targeting postcode districts containing the breadth of the socioeconomic spectrum defined by IMD quintiles generated using the English indices of deprivation 2019 database (available through the Ministry of Housing, Communities and Local Government).[117] Group two recruitment utilised existing volunteer databases and social media-based advertisements to recruit volunteers (without targeting across IMD quintiles). From December 2020 onwards, both strategies were adapted to over-sample participants who identified as belonging to a minority ethnic group. Group one used CHIS extracts of participants living within UKHSA-selected postcodes to approach potential participants who identified as belonging to an ethnic minority group. Group two used local contacts within the community, e.g., through newsletters and email circulars to associations for minority ethnic groups.

Participants' full postcodes were gathered to determine the IMD and IDACI (Income Deprivation Affecting Children Index) distribution by region. These ranks are based on small areas defined as Lower-Layer Super Output Areas (LSOA) in England, typically representing 1,500 residents or 650 households. Each postcode within the UK is ranked from one (the most deprived area) to 32,844 (the least deprived area).[117] IMD is determined by seven domains

which are weighted as follows: Income Deprivation (22.5%), Employment Deprivation (22.5%), Education, Skills and Training Deprivation (13.5%), Health Deprivation and Disability (13.5%), Crime (9.3%), Barriers to Housing and Services (9.3%) and Living Environment Deprivation (9.3%).[140,141] In addition to IMD, a supplementary index, the Income Deprivation Affecting Children Index (IDACI), was used. IDACI measures the proportion of children living in income-deprived households and has been validated for children aged 0-15 years.[141] Information from NHS Digital was used to define all postcodes within an NHS region.[142] The English indices of deprivation 2019 database, available through the Ministry of Housing, Communities and Local Government[117], provided IMD and IDACI quintiles for each postcode within the NHS region. NHS regional data were then compared with IMD and IDACI data for the site's region (postcodes districts within the site's NHS trusts catchment area) so that those districts selected within the catchment area had a quintile distribution that was representative of their region. As IMD and IDACI assess deprivation at a postcode rather than household level,[141]participant socioeconomic data was collected using the study questionnaire. The study questionnaire included questions derived from the census (2011) and Family Affluence Scale iii (FASiii).[125] The results comparing IMD and IDACI of participants recruited by recruitment strategy were compared at the national and regional levels. The sample's overall representativeness was assessed using questionnaire measures and indices of deprivation, which were then compared with national and regional level data published by the Office for National Statistics.

3.3 Results

3.3.1 Recruitment methods

Most participants recruited were contacted through GDPR-compliant bulk mailing databases to UKHSA-selected postcode districts (group one recruitment). As recruitment methods were diversified, a participant's recruitment method was recorded in their case report forms (CRF). Table 10 shows the number of participants recruited by each recruitment method at a time when

all methods were being utilised. Mailouts were the most successful method of recruitment, recruiting 68% of participants. Social media and GP texts resulted in very few responses across all sites, but is not possible to know how many potential participants were contacted through these methods. Pharmacies were used in London and Oxford and yielded no responses. Plymouth and Nottingham, who were not included in the mailout recruitment, used alternative recruitment methods through local advertising and healthy volunteer databases.

Table 10 The number of participants recruited by each recruitment method as recorded on the participants case report form (CRF)

Site	Mail out	Via GP	Social media	Other *
Bradford	45 (90.0%)	0 (0.0%)	1 (2.0%)	4 (8.0%)
Bristol	232 (85.6%)	1 (0.4%)	1 (0.4%)	37 (13.7%)
Leeds	101 (90.2%)	0 (0.0%)	1 (0.9%)	10 (8.9%)
Manchester	289 (99.7%)	0 (0.0%)	1 (0.3%)	0 (0.0%)
Oxford	191 (86.4%)	6 (2.7%)	8 (3.6%)	16 (7.2%)
Sheffield	70 (95.9%)	0 (0.0%)	3 (4.1%)	0 (0.0%)
Southampton	238 (97.9%)	0 (0.0%)	1 (0.4%)	4 (1.6%)
St George's	284 (93.4%)	7 (2.3%)	2 (0.7%)	11 (3.6%)
Nottingham	N/A	3 (0.8%)	14 (3.7%)	357 (95.5%)
Plymouth	N/A	2 (0.8%)	78 (29.8%)	182 (69.5%)
Newcastle	127 (80.4%)	0 (0.0%)	18 (11.4%)	13 (8.2%)
Imperial	71 (67.6%)	12 (11.4%)	6 (5.7%)	16 (15.2%)
Suffolk	110 (91.7%)	0 (0.0%)	9 (7.5%)	1 (0.8%)
Overall	1758 (68.1%)	31 (1.2%)	143 (5.5%)	651 (25.2%)

**Other refers to methods such as email circulars, contacting healthy volunteer lists and utilising pre-existing local recruitment strategies. Nottingham and Plymouth recruited primarily by these methods*.

Table 11 shows the response rate to the mail out recruitment strategy by site. The response rate was 4.8% overall and varied between sites from 2.0-12.7% (mean 5.2%, median 4.7%). The mail out cost per participant recruited was £17.64 with a total cost of £31,002.48 recruiting 1,758 participants. The total cost included the database search and the cost of printing and distributing the letters. Social media adverts cost £312.78 in total, and 413 participants were recruited at £2.18 per participant.

Table 11 Summarises the number of mail-out invitations sent and responses received per site.

Site	Invitations	Responses	Response rate (%)
Bradford	8,644	352	4.1
Bristol	21,220	1,901	9.0
Leeds	6,778	595	8.8
Manchester	22,196	945	4.3
Oxford	28,761	1,354	4.7
Sheffield	7,290	413	5.7
Southampton	20,584	415	2.0
St George's	39,490	885	2.2
Nottingham	NA		
Plymouth	NA		
Newcastle	6,306	689	10.9
Imperial	12,116	376	3.1
Suffolk	4,790	607	12.7
Overall	178,175	8,532	4.8

3.3.2 Participant characteristics

The study recruited 2,963 participants (Table 12). Of those enrolled 2,899 (98%) completed their questionnaires to give additional information describing participant demography, and 2,901 (98%) provided vaccine histories. The vaccine history was complete including both immunisations given in healthcare and within school programmes in 2,369 (81%) participants. In this section the data is described at a national level (England) and a regional level (NHS regions defined by NHS commissioning groups Figure 3).[142]

Overall, 1,533 (52% 95%CI 49.9-53.5) participants were female which was above the national estimate for England (47%)[143] for children and young adults between the ages of 0-24 years; this was also the case for NHS regions (47%).[143] East of England, South-East and South-West were representative in terms of sex with confidence intervals overlapping the regional estimate (Table 12).

Overall 2397 (81%, 95%CI 79.4-82.3) participants identified as white or from a white minority ethnic group (Table 12). In the 2011 census 87% of the population identified as White,[144] but estimates specific to the population aged between 0-24 years are lower (75%).[143] The

proportion of participants from each ethnic background varied by region. Table 12 compares ethnicity as reported by participants across regions with population estimates for children and young adults aged 0-24 years.[143] Comparing to regional estimate for the population between 0-24 years, London and the Midlands sites over recruited white participants and sites in the North-West under recruited this cohort.

Participants identifying as Asian were under recruited across the study (5%, 95 CI 4.77-6.42) compared with national estimates (13%). Asian participants were under recruited across all regions apart from the North-West and South-West. Similarly, the black population was under recruited in the study (2%, 95% CI 1.43-2.41) compared with national estimates (5%). London, South-East and South-West under recruited participants identifying as black.

Mixed heritage participants were over recruited within the study (8%, 95% CI 7.27-9.25) compared with national estimates of 6%.[143] Regions which over recruited included the North-East & Yorkshire, North-West and South-East. Participants who identified from other minority ethnic groups were in line with both national and regional estimates apart from Midlands who under recruited for the region.

Table 12 Shows participant characteristics by site and reference regional and national estimates.

	East of England n (%, 95%CI)	London n (%, 95%CI)	Midlands n (%, 95%CI)	North-East & Yorkshire n (%, 95%CI)	North-West n (%, 95%CI)	South-East n (%, 95%CI)	South-West n (%, 95%CI)	Overall (England) n (%, 95%CI)
Female (STORY)	97 (47.3, 40.5-54.2)	181 (52.5, 47.2-57.7)	130 (55.6, 49.1-61.8)	261 (53.7, 49.2-58.1)	226 (55.1, 50.3-59.9)	382 (49.2, 45.7-52.7)	255 (50.5, 46.1-54.9)	1533 (51.7, 49.9-53.5)
Female (Census)[143]	46.5%	46.60%	46.8%	46.9%	47.0%	46.6%	46.6%	46.7%
Ethnicity								
***White (STORY)**	169 (82.4, 76.6-87.1)	236 (68.4, 63.3-73.1)	198 (84.6, 79.4-88.7)	420 (84.6 ,79.4-88.7)	287 (70.0, 65.4-74.3)	635 (81.6, 78.7-84.2)	452 (89.5, 86.5-91.9)	2397 (80.9, 79.4-82.3)
***White (ONS estimates)[143]**	82.8%	46.3%	74.3%	80.6%	80.1%	82.3%	90.1%	74.7%
****Asian (STORY)**	5 (2.4, 1.0-5.8)	36 (10.4, 7.6-14.1)	13 (5.6, 3.2-9.4)	17 (3.5, 2.2-5.6)	51 (12.4, 9.6-16.0)	30 (3.9, 2.7-5.5)	12 (2.4, 1.4-4.1)	164 (5.5, 4.8-6.4)
****Asian (ONS estimates)[143]**	7.7%	23.24%	15.2%	12.6%	12.2%	8.9%	3.7%	12.9%
*****Black (STORY)**	2 (1.0, 0.2-3.9)	20 (5.8, 3.8-8.8)	4 (1.7, 0.6-4.5)	7 (1.4, 0.7-3.0)	16 (3.9, 2.4-6.2)	4 (0.5, 0.2-1.4)	2 (0.4, 0.1-1.6)	55 (1.9, 1.4-2.4)
*****Black (ONS estimates)[143]**	3.7%	17.2%	3.9%	2.0%	2.5%	2.6%	1.9%	5.4%
Mixed (STORY)	10 (4.9, 2.6-8.9)	42 (12.2, 9.1-16.1)	16 (6.8, 4.2-10.9)	31 (6.4, 4.5-8.9)	43 (10.5 7.9-13.9)	70 (9.0, 7.2-11. 2)	31 (6.1, 4.4-8.6)	243 (8.2, 7.3-9.3)
Mixed (ONS estimates)[143]	5.1%	9.8%	5.6%	3.7%	4.2%	5.2%	3.9%	5.6%
Other (STORY)	3 (1.5, 0.5-4.5)	11 (3.2, 1.8-5.7)	0	4 (0.8, 0.3-2.2)	9 (2.2, 1.1-4.2)	5 (0.6, 0.3-1.5)	2 (0.4, 0.1-1.6)	34 (1.1, 0.8-1.6)
Other (ONS estimates)[143]	0.7%	3.5%	1.0%	1.1%	1.0%	0.9%	0.5%	1.3%
No data (STORY)	16 (7.8, 4.8-12.4)	0	3 (1.3, 0.4-3.9)	7 (1.4, 0.7-3.0)	4 (1.0, 0.4-2.6)	34 (4.4, 3.1-6.1)	6 (1.2, 0.5-2.6)	70 (2.4, 1.9-3.0)
Religion †								
Christian (STORY)	75 (36.6, 30.2-43.4)	132 (38.3, 33.3-43.5)	90 (38.3, 33.3-43.5)	195 (40.1, 35.8-44.6)	136 (33.2, 28.8-37.9)	274 (35.2, 31.9-38.7)	151 (29.9, 26.1-34.1)	1053 (35.5, 33.8-37.3)
Christian (Census)[144,145]								59.6%

No religion (STORY)	98 (47.8, 41.0-54.7)	135 (39.1, 34.1-44.4)	107 (45.7, 39.4-52.2)	232 (47.7, 43.3-52.2)	148 (36.1, 31.6-40.9)	402 (51.7, 48.2-55.2)	302 (59.8, 55.5-64.0)	1424 (48.1, 46.3-49.9)
No religion (Census)[144,145]								31.9%
Muslim (STORY)	6 (2.9, 1.3-6.4)	19 (5.5, 3.5-8.5)	9 (3.9, 2.0-7.3)	22 (4.5, 3.0-6.8)	56 (13.7, 10.7-17.4)	9 (1.2, 0.6-2.2)	4 (0.8, 0.3-2.1)	125 (4.2, 3.6-5.0)
Muslim (Census)[144,145]								4.7%
Hindu (STORY)	0	13 (3.8, 2.2-6.4)	2 (0.9, 0.2-3.4)	3 (0.6, 0.2-1.9)	5 (1.2, 0.5-2.9)	10 (1.3, 0.7-2.4)	2 (0.4, 0.1-1.6)	35 (1.2, 0.9-1.6)
Hindu (Census)[144,145]								1.6%
Jewish (STORY)	0	3 (0.9, 0.3-2.7)	0	0	22 (5.4, 3.6-8.0)	0	0	25 (0.8, 0.6-1.3)
Jewish (Census)[144,145]								0.5%
Other (STORY)	2 (1.0, 0.2-3.9)	2 (0.6, 0.1-2.3)	0	1 (0.2, 0.0-1.5)	2 (0.5, 0.1-1.9)	2 (0.3, 0.1-1.0)	2 (0.4, 0.1-1.6)	11 (0.4, 0.2-0. 7)
Other (Census)[144,145]								0.4%
Buddhist (STORY)	1 (0.5, 0.1-3.4)	2 (0.6, 0.1-2.3)	1 (0.4, 0.1-3.0)	0	0	3 (0.4, 0.1-1.2)	2 (0.4, 0.1-1.6)	9 (0.3, 0.2-0.6)
Buddhist (Census)[144,145]								0.5%
Sikh (STORY)	0	0	1 (0.4, 0.1-3.0)	0	0	1 (0.1, 0.0-0.9)	1 (0.2, 0.0-1.4)	3 (0.1, 0.0-0.3)
Sikh (Census)[144,145]								0.8%
No data (STORY)								278 (8.0, 7.0-9.0)

*White including white minorities
** Asian (Bangladeshi, Chinese, Indian, Pakistani, Other Asian background)
***Black (African, Caribbean, Other Black background)
†Religion (this question was optional). Census religion data was not available at a regional level.

The study recruited across all religions, but the proportion of Christians recruited when compared with the 2011 census was lower (Table 12). More participants identified as having no religion compared with census data. It is striking that participants from Muslim, Hindu and Jewish religions were not under represented when compared with census data.

Failed enrolments were participants who agreed to take part in the study and met the inclusion and exclusion criteria but were unable to provide an adequate blood sample (0.5ml of blood or 200μL sera). Failed enrolment rates varied from 1% at Imperial to 11% at Manchester. Although all ages were affected by failed enrolments for 12/13 sites the mean and median age for failed enrolments was under 10 years (Figure 6 and Table 13). To recruit the target number of participants in each age band there was an over recruitment of 10-20% in children under 5 years, 5-10% in children aged 6–11 years and 1-6% in participants over the age of 11 years.

Table 13 Numbers of enrolment and failed enrolment per site

	Invited for study visit	**Enrolled**	**Failed enrolment**			
Site	Number	Number	Number (%)	Age range	Mean age	Median age
Bradford	151	137	14 (9.3)	1-24 years	5	5
Bristol	425	380	45 (10.6)	0-21 years	6	3
Leeds	162	156	6 (3.7)	1-8 years	4	5
Manchester	460	410	50 (10.9)	2-22 years	6	7
Oxford	502	478	24 (4.8)	0-21 years	6	4
Sheffield	130	123	7 (5.3)	0-17 years	9	9
Southampton	352	340	12 (3.4)	0-19 years	6	6
St George's	303	295	8 (2.6)	0-18 years	7	9
Nottingham	153	149	4 (2.6)	0-17 years	11	14
Plymouth	130	124	6 (4.6)	0-15 years	8	7
Newcastle	162	153	9 (5.6)	0-13 years	3	3
Imperial	105	104	1 (1.0)	4 years old	4	4
Suffolk	120	114	6 (5.0)	0-19 years	9	9

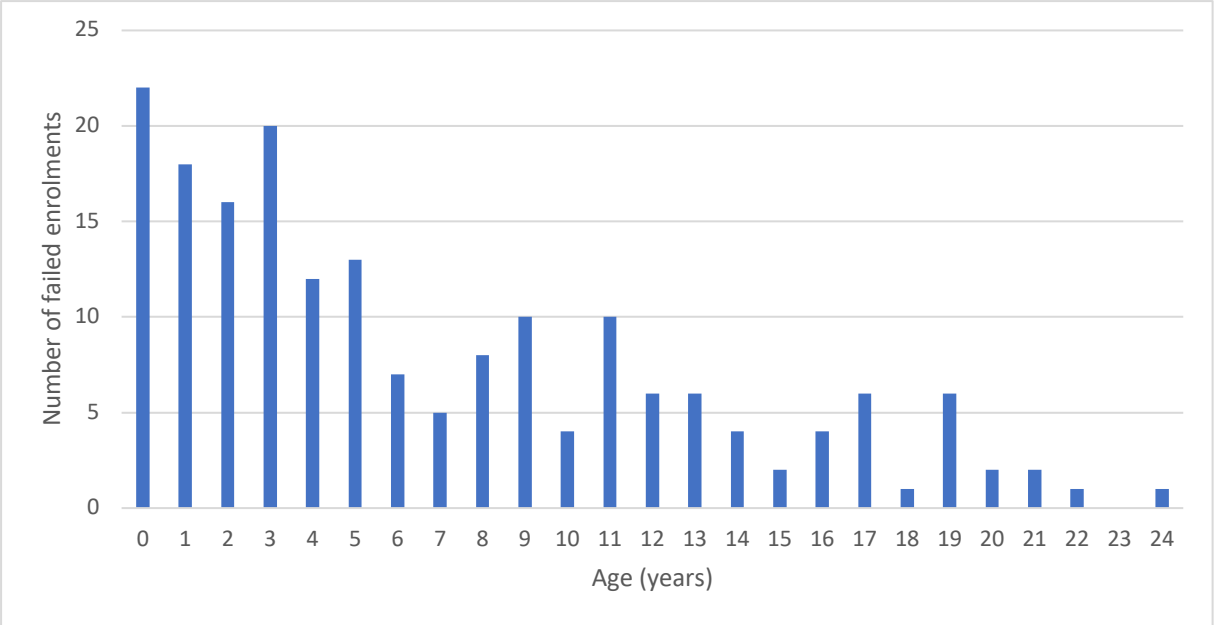


Figure 6 Number of failed enrolments from all sites by age (years)

Combining all sites and methods of recruitment Figure 7, Figure 8 and Table 14 shows there were more than 100 participants recruited in each age group up to age 20 years. The original aim was to achieve this from group one recruitment only. This was achieved in each age group except for children aged 4 years where only 95 participants were recruited (Figure 7 & Figure 8). Due to insufficient numbers in young adults aged 20 years, the 21-24 year age band was broadened to include participants aged 20-24 years for analysis.

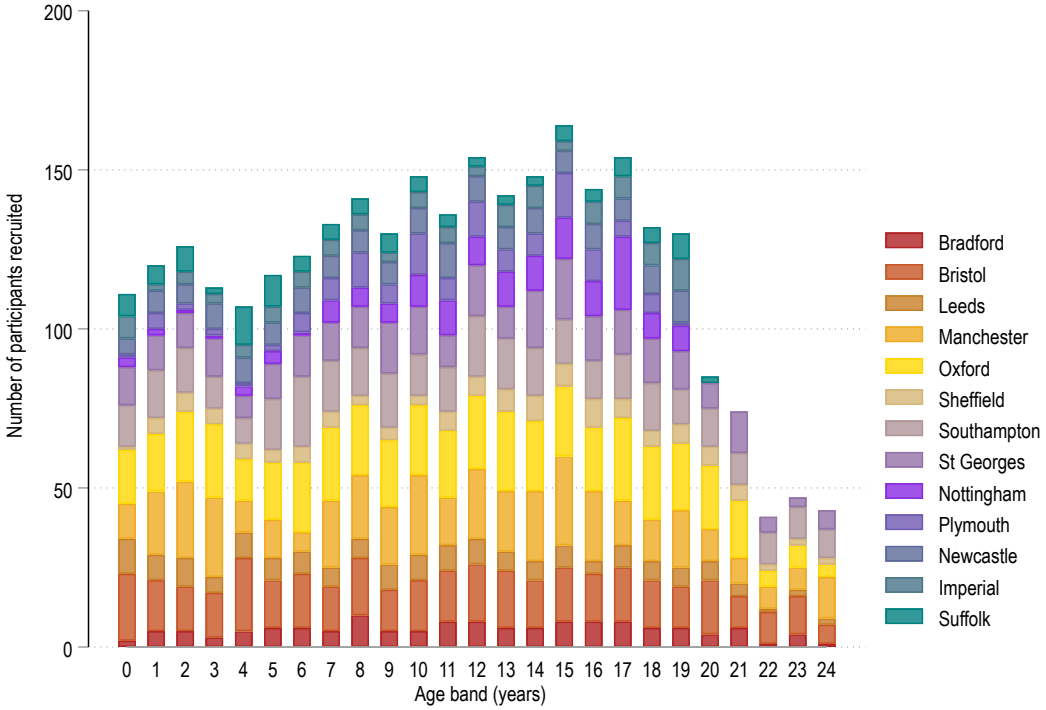


Figure 7 Number of participants per age band recruited by site

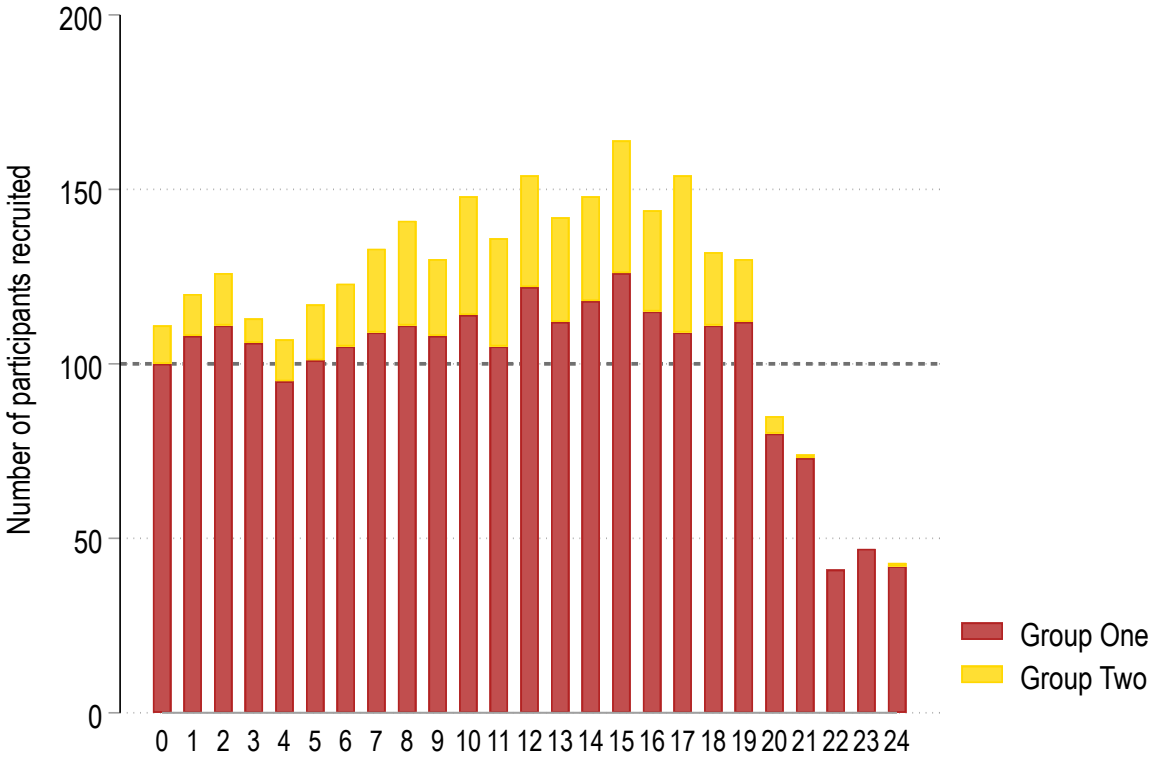


Figure 8 Number of participants recruited per age band by recruitment strategy

Table 14 Recruitment by NHS region and age band

Age band	East of England	London	Midlands	North-East & Yorkshire	North-West	South-East	South-West	Total
0	8	19	4	18	11	29	22	111
1	9	11	5	22	20	32	21	120
2	11	12	5	22	24	36	16	126
3	5	15	5	17	25	30	16	113
4	15	9	7	22	10	19	25	107
5	13	12	7	22	12	34	17	117
6	10	14	5	22	6	43	23	123
7	10	15	10	20	21	36	21	133
8	7	14	8	24	20	39	29	141
9	9	18	9	22	18	35	19	130
10	9	15	13	21	25	36	29	148
11	7	14	15	29	15	33	23	136
12	9	16	12	27	22	39	29	154
13	10	15	15	22	19	36	25	142
14	5	23	19	20	22	37	22	148
15	10	18	16	26	28	35	31	164
16	12	17	19	21	22	28	25	144
17	8	18	28	23	14	41	22	154
18	6	19	11	23	13	39	21	132
19	14	19	9	27	18	29	14	130
20	6	8	6	10	10	28	17	85
21	6	12	3	12	8	23	10	74
22-24	6	12	3	14	27	41	28	131
Total	205	345	234	486	410	778	505	2,963

3.3.3 Socioeconomic representativeness of the population at a national and regional level

3.3.3.1 National level

To determine the representativeness of the 'STORY' study in terms of socioeconomic status, 'STORY' participants postcode IMD and IDACI were compared with the general population in England. Figure 9 illustrates the distribution of postcodes across each quintile for IMD and IDACI in England (dark blue) and across all recruiting sites (orange). The selected postcode districts for the group one sampling strategy are shown in grey. Figure 9 illustrates how study participants recruited to group one (yellow) and group two (light blue) as defined in section 2.2.2, compare with England as a region.

Quintiles were derived from the 2019 deprivation indices.[117] Not all postcodes listed in the NHS digital database published in 2022 were present in the 2019 deprivation indices[117] (data unavailable for 34768/2232890 1.6% of postcodes) and therefore these postcodes have been omitted for the region.

For IMD and IDACI the quintile distribution of participants was within 10% of the recruiting regions for all but quintile five at a national level. Ideally the IMD and IDACI distribution would be 20% for each quintile for region, sites, and participants. Each quintile of IMD and IDACI was represented but both recruitment strategies (group one and two) recruited the highest proportion of participants from IMD quintile five and from IDACI quintiles four and five (Figure 9). There was little difference between group one and group two recruitment in terms of deprivation, demonstrated by overlapping confidence intervals across the majority of quintiles.

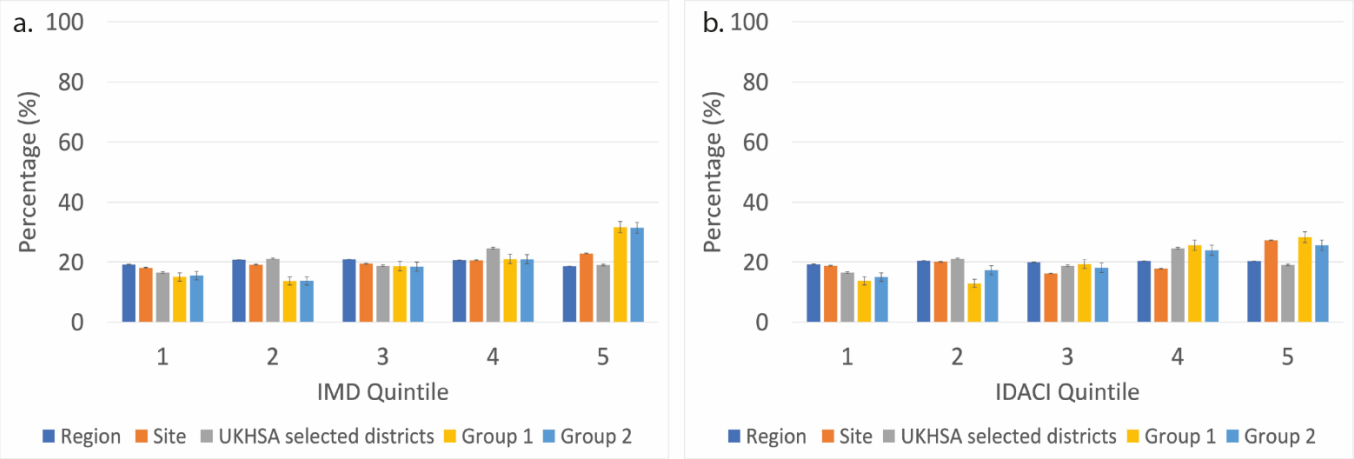


Figure 9 Distribution of English postcodes across(a) IMD and (b) IDACI quintiles in all NHS regions combined (I.e. whole of England), 'STORY' study sites, selected postcode districts for 'STORY' recruitment, and participants recruited to 'STORY' according to recruitment methods as described in section 2.2.2 (Group one and Group two). Error bars show 95% confidence intervals (CI)

3.3.3.2 Regional level

Figures 10-16 show the variation in the different regions and the populations recruited within each region. Region refers to the NHS region as defined by NHS commissioning regions[142] and site refers to all recruiting sites located within the NHS region. The postcodes were distributed differently across the quintiles in each region. The North-West, Midlands and North-East & Yorkshire had a higher proportion of postcodes living in the lower (more deprived) quintiles compared with the South of England where the proportion of households living in the higher quintiles was higher. In group one each region recruited from all quintiles, but London, South-East and South-West over recruited in higher quintiles and under recruited in lower quintiles. North-East & Yorkshire over recruited in higher quintiles but recruited a representative sample in lower quintiles. The Midlands, and North-West over recruited in lower quintiles and under recruited in higher quintiles. On a national level, recruitment over quintiles was representative. For group one and group two recruitment methods confidence intervals frequently overlapped for quintiles within regions.

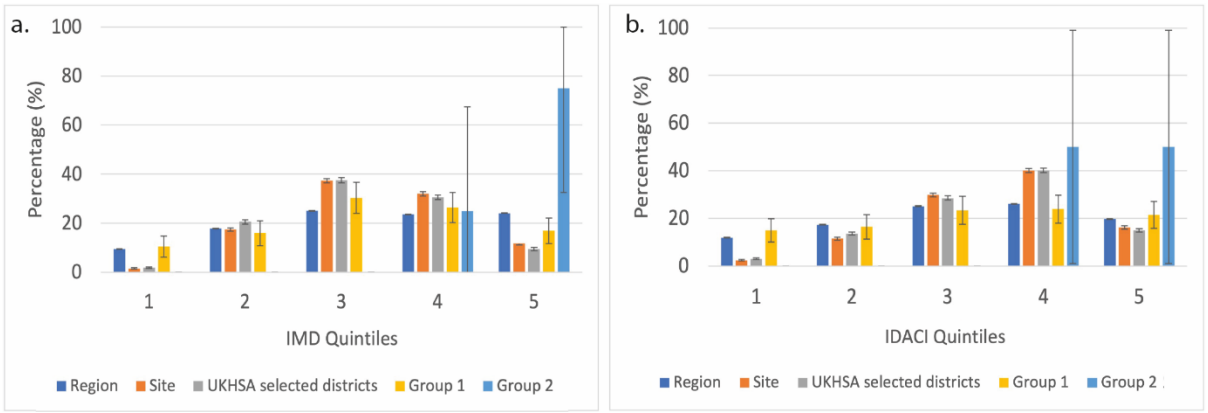


Figure 10 Distribution of population over (a) IMD and (b) IDACI quintiles for the East of England according to postcode for NHS region, site, UKHSA selected postcode districts and participants recruited by group one and group two recruitment methods. Error bars show 95% confidence intervals (CI)

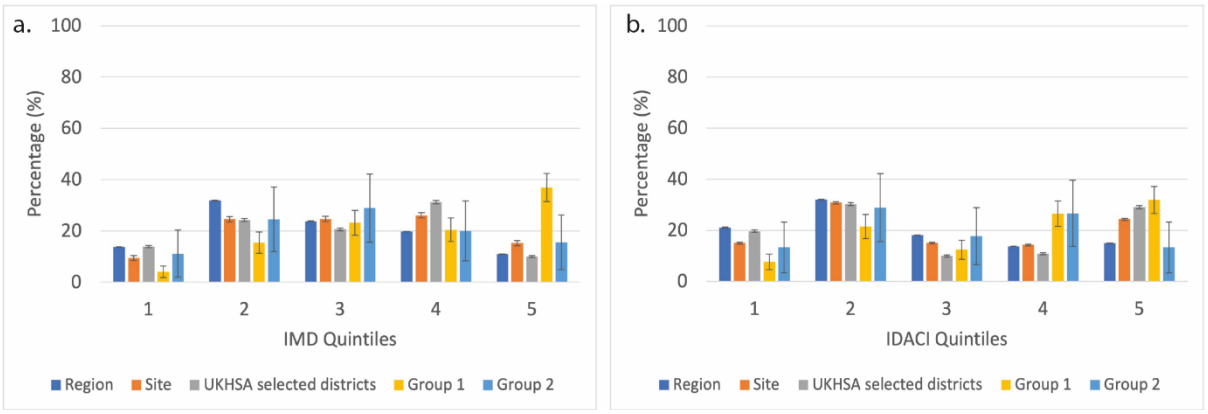


Figure 11 Distribution of population over (a) IMD and (b) IDACI quintiles for London according to postcode for NHS region, site, UKHSA selected postcode districts and participants recruited by group one and group two recruitment methods. Error bars show 95% confidence intervals (CI)

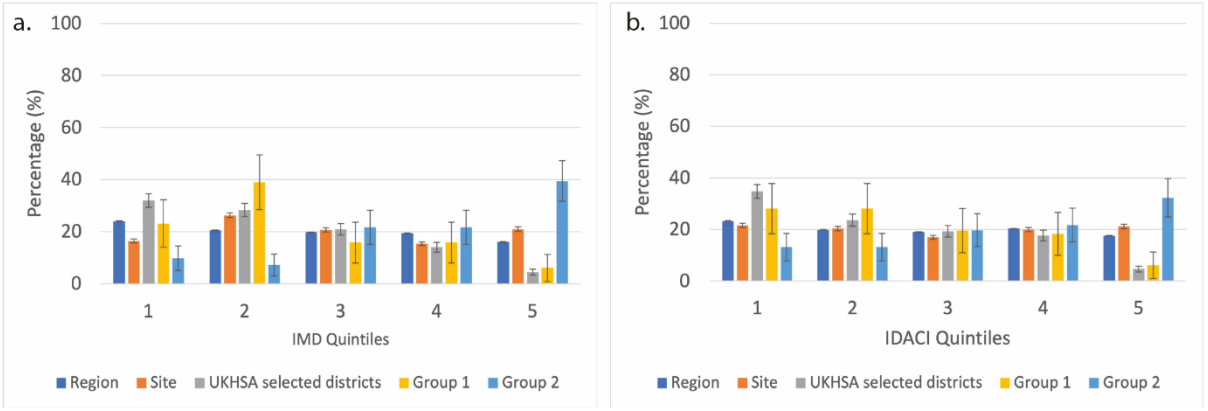


Figure 12 Distribution of population over (a) IMD and (b) IDACI quintiles for the Midlands according to postcode for NHS region, site, UKHSA selected postcode districts and participants recruited by group one and group two recruitment methods. Error bars show 95% confidence intervals (CI)

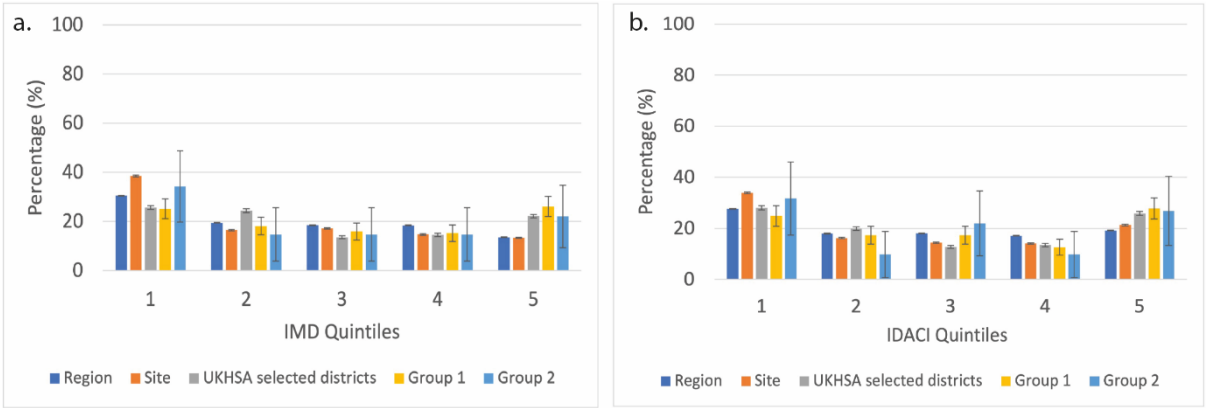


Figure 13 Distribution of population over (a) IMD and (b) IDACI quintiles for the North-East and Yorkshire & Humber region according to postcode for NHS region, site, UKHSA selected postcode districts and participants recruited by group one and group two recruitment methods. Error bars show 95% confidence intervals (CI)

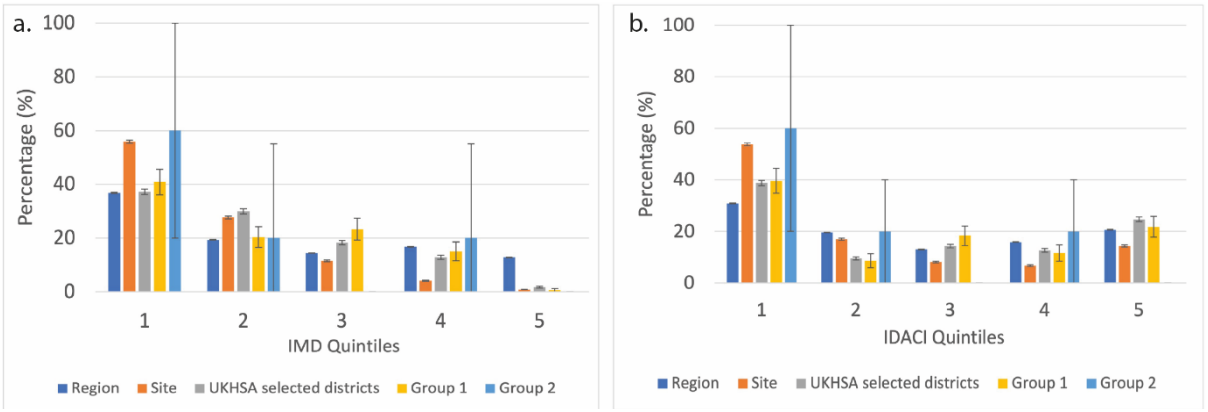


Figure 14 Distribution of population over (a) IMD and (b) IDACI quintiles for the North-West according to postcode for NHS region, site, UKHSA selected postcode districts and participants recruited by group one and group two recruitment methods. Error bars show 95% confidence intervals (CI)

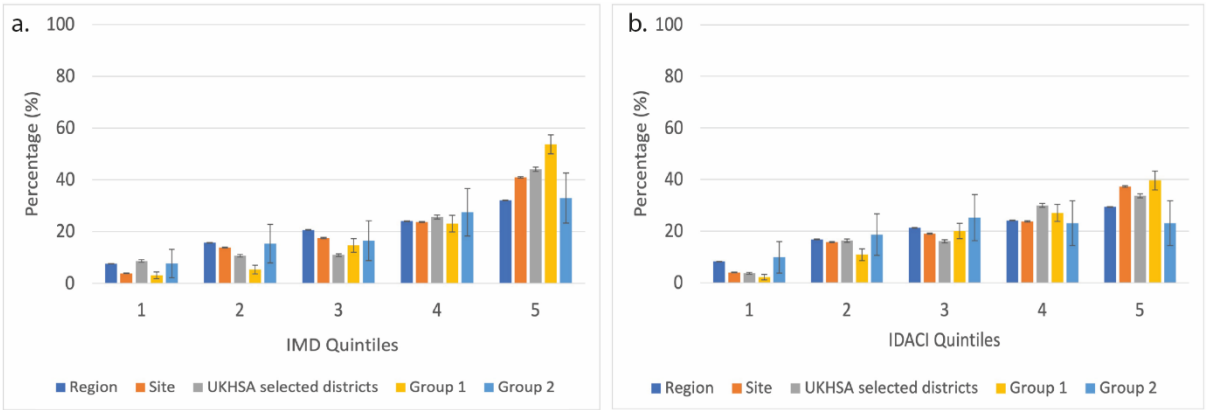


Figure 15 Distribution of population over (a) IMD and (b) IDACI quintiles for the South-East according to postcode for NHS region, site, UKHSA selected postcode districts and participants recruited by group one and group two recruitment methods. Error bars show 95% confidence intervals (CI)

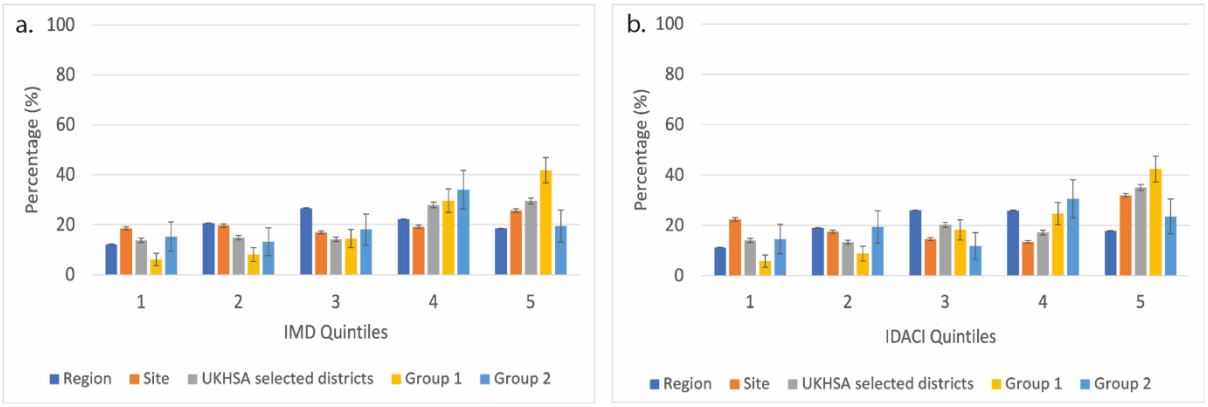


Figure 16 Distribution of population over IMD quintiles for the South-West according to postcode for NHS region, site, UKHSA selected postcode districts and participants recruited by group one and group two recruitment methods. Error bars show 95% confidence intervals (CI)

Table 15 shows recruited participants by IDACI and IMD quintiles and by 'region', as a measure of socioeconomic deprivation in participants aged 0-24 years. IMD and IDACI quintiles were derived from postcodes of participants. The proportion of participants in each IMD and IDACI quintile was similar (within 2.5%) across quintiles for the study overall, and within 13.5% of their local IMD quintile across most regions apart from the North-West and London. The North-West had <1% participants recruited in the least deprived IMD quintile vs 21% participants in the least deprived IDACI quintile. Responses to selected UK Census questions and FASiii scores were generally in agreement with IMD and IDACI quintiles. The smallest average number of rooms/bedrooms, lowest number of families owning their own home (outright or with mortgage) and lowest FASiii scores were in the most deprived quintiles. The highest average number of rooms/bedrooms, highest number of families owning their own home (outright or with mortgage) and highest FASiii scores were in the least deprived quintiles (Table 15).

Table 15 Summary of measures of socioeconomic deprivation in children and young adults aged 0-24 years. IMD and IDACI derived from postcode of participant. FASiii and Census 2011 Home ownership and house size data were obtained by questionnaire. Quintile 1 was most deprived and quintile 5 was most affluent.

		Quintile				
		1	**2**	**3**	**4**	**5**
Overall	**Number of families recruited in each IMD quintile**	437	400	553	645	923
	Number of families that own their own home in each IMD quintile	180 (41.2%)	222 (55.5%)	353 (63.8%)	422 (65.4%)	648 (70.2%)
	Mean number of rooms in home in each IMD quintile (census)	5.7	6	6.6	7.1	7.4
	Mean Number of bedrooms in home in each IMD quintile (census)	3	3.2	3.5	3.7	3.8
	Mean FASiii score in each IMD quintile	6.7	7.9	8.7	9.1	9.7
	Number of families recruited in each IDACI quintile	454	422	542	636	904
	Number of families that own their own home in each IDACI quintile	180 (39.6%)	233 (55.2%)	348 (64.2%)	432 (67.9%)	632 (69.9%)
	Mean number of rooms in home in each IDACI quintile (census)	5.7	6.1	6.7	6.9	7.5
	Mean Number of bedrooms in home in each IDACI quintile (census)	3	3.3	3.5	3.6	3.9
	Mean FASiii score in each IDACI quintile	6.8	8	8.8	9.1	9.7
East of England	**Number of families recruited in each IMD quintile**	21	32	61	54	37
	Number of families that own their own home in each IMD quintile	5 (23.8%)	16 (50.0%)	33 (54.1%)	37 (68.5%)	26 (70.3%)
	Mean number of rooms in home in each IMD quintile (census)	6.6	6.7	6.9	7.3	8
	Mean Number of bedrooms in home in each IMD quintile (census)	3.1	3.5	3.3	3.8	3.8
	Mean FASiii score in each IMD quintile	6.7	8	8.2	9.3	9.9
	Number of families recruited in each IDACI quintile	30	33	47	50	45
	Number of families that own their own home in each IDACI quintile	8 (26.7%)	15 (45.5%)	28 (59.6%)	35 (70.0%)	31 (68.9%)
	Mean number of rooms in home in each IDACI quintile (census)	6.7	6.1	7.4	7.5	7.7
	Mean Number of bedrooms in home in each IDACI quintile (census)	3.1	3.4	3.7	3.6	3.8
	Mean FASiii score in each IDACI quintile	6.7	7.7	9.3	8.9	9.6
London	**Number of families recruited in each IMD quintile**	17	57	82	70	117
	Number of families that own their own home in each IMD quintile	3 (17.6%)	19 (33.3%)	50 (61.0%)	48 (68.6%)	77 (65.8%)
	Mean number of rooms in home in each IMD quintile (census)	4.8	5.2	5.6	6.6	6.9
	Mean Number of bedrooms in home in each IMD quintile (census)	2.5	3	3.1	3.5	3.8
	Mean FASiii score in each IMD quintile	6.3	7.6	8.2	9.3	9.8
	Number of families recruited in each IDACI quintile	29	77	45	91	101
	Number of families that own their own home in each IDACI quintile	5 (17.2%)	36 (46.8%)	30 (66.7%)	57 (62.6%)	69 (68.3%)
	Mean number of rooms in home in each IDACI quintile (census)	4.5	5.5	6.2	6.3	7
	Mean Number of bedrooms in home in each IDACI quintile (census)	2.6	3	3.3	3.5	3.8
	Mean FASiii score in each IDACI quintile	7	7.7	8.4	9	10
Midlands	**Number of families recruited in each IMD quintile**	34	43	46	46	65
	Number of families that own their own home in each IMD quintile	19 (55.9%)	22 (51.2%)	33 (71.7%)	36 (78.3%)	59 (90.8%)
	Mean number of rooms in home in each IMD quintile (census)	6.5	6.2	6.5	6.9	7.8
	Mean Number of bedrooms in home in each IMD quintile (census)	3.4	3.3	3.5	3.8	4
	Mean FASiii score in each IMD quintile	8.1	8.2	9	9.3	9.8

	Number of families recruited in each IDACI quintile	43	43	46	48	54
	Number of families that own their own home in each IDACI quintile	22 (51.2%)	26 (60.5%)	32 (69.6%)	41 (85.4%)	48 (88.9%)
	Mean number of rooms in home in each IDACI quintile (census)	6.4	6.4	6.3	7.3	7.8
	Mean Number of bedrooms in home in each IDACI quintile (census)	3.3	3.5	3.5	3.8	4
	Mean FASiii score in each IDACI quintile	8.1	8.6	8.8	9.5	9.8
North-East and Yorkshire	**Number of families recruited in each IMD quintile**	125	86	76	73	124
	Number of families that own their own home in each IMD quintile	57 (45.6%)	59 (68.6%)	54 (71.1%)	56 (76.7%)	90 (72.6%)
	Mean number of rooms in home in each IMD quintile (census)	5.6	6.2	6.9	7	7.5
	Mean Number of bedrooms in home in each IMD quintile (census)	3	3.2	3.6	3.7	3.8
	Mean FASiii score in each IMD quintile	6.7	7.6	8.8	9.4	9.3
	Number of families recruited in each IDACI quintile	123	81	86	60	134
	Number of families that own their own home in each IDACI quintile	56 (45.5%)	53 (65.4%)	60 (69.8%)	41 (68.3%)	106 (79.1%)
	Mean number of rooms in home in each IDACI quintile (census)	5.6	6.2	6.8	7	7.4
	Mean Number of bedrooms in home in each IDACI quintile (census)	3	3.2	3.6	3.6	3.8
	Mean FASiii score in each IDACI quintile	6.7	7.7	8.8	9.1	9.3
North-West	**Number of families recruited in each IMD quintile**	168	83	94	62	2
	Number of families that own their own home in each IMD quintile	71 (42.3%)	55 (66.3%)	66 (70.2%)	45 (72.6%)	1 (50.0%)
	Mean number of rooms in home in each IMD quintile (census)	5.7	6.2	7.2	7.9	8
	Mean Number of bedrooms in home in each IMD quintile (census)	3	3.3	3.8	4.2	4.5
	Mean FASiii score in each IMD quintile	6.6	8.4	9.1	8.7	9.5
	Number of families recruited in each IDACI quintile	163	36	74	47	89
	Number of families that own their own home in each IDACI quintile	68 (41.7%)	23 (63.9%)	51 (68.9%)	32 (68.1%)	64 (71.9%)
	Mean number of rooms in home in each IDACI quintile (census)	5.7	6.6	7.1	6.3	7.6
	Mean Number of bedrooms in home in each IDACI quintile (census)	3	3.3	3.6	3.7	4
	Mean FASiii score in each IDACI quintile	6.7	8.3	8.5	8.5	9.1
South-East	**Number of families recruited in each IMD quintile**	28	51	116	184	399
	Number of families that own their own home in each IMD quintile	6 (21.4%)	24 (47.1%)	63 (54.3%)	101 (54.9%)	258 (64.7%)
	Mean number of rooms in home in each IMD quintile (census)	5.7	5.9	6.4	7.5	7.4
	Mean Number of bedrooms in home in each IMD quintile (census)	2.8	3.1	3.4	3.8	3.8
	Mean FASiii score in each IMD quintile	6.4	7.9	8.8	9.7	9.7
	Number of families recruited in each IDACI quintile	24	92	161	207	294
	Number of families that own their own home in each IDACI quintile	4 (16.7%)	47 (51.1%)	87 (54.0%)	130 (62.8%)	184 (62.6%)
	Mean number of rooms in home in each IDACI quintile (census)	5.3	6.1	6.8	7.2	7.8
	Mean Number of bedrooms in home in each IDACI quintile (census)	2.5	3.2	3.6	3.7	3.9
	Mean FASiii score in each IDACI quintile	6	8.3	9.1	9.5	9.9
South-West	**Number of families recruited in each IMD quintile**	44	48	78	156	179
	Number of families that own their own home in each IMD quintile	19 (43.2%)	27 (56.3%)	54 (69.2%)	99 (63.5%)	137 (76.5%)
	Mean number of rooms in home in each IMD quintile (census)	5.3	5.6	6.9	6.5	7.3
	Mean Number of bedrooms in home in each IMD quintile (census)	3.2	3.2	3.6	3.5	3.9

Mean FASiii score in each IMD quintile	6.4	7.6	8.9	8.4	9.8
Number of families recruited in each IDACI quintile	42	60	83	133	187
Number of families that own their own home in each IDACI quintile	17 (40.5%)	33 (55.0%)	60 (72.3%)	96 (72.2%)	130 (69.5%)
Mean number of rooms in home in each IDACI quintile (census)	5.3	6.6	6.2	6.7	7.2
Mean Number of bedrooms in home in each IDACI quintile (census)	3	3.5	3.3	3.6	3.8
Mean FASiii score in each IDACI quintile	6.5	7.9	8.5	8.6	9.6

**FASiii – family affluence scale is a measure of socioeconomic status used in children and adolescents. Scores range from 1 (most deprived) to 13 (least deprived)*

***Census questions – questions from 2011 census measuring socioeconomic status regarding house size and home ownership*

3.3.4 Country of birth

Out of 2899 enrolled participants who listed their country of birth, 2732 (94.2%) participants were born in the UK, 60 (2.1%) had migrated from an EU country and 107 (3.7%) were born in a non-EU country. In contrast ONS reports that 15.7% of the English population (all ages) were born outside the UK in 2021.[146,147] Of all persons who reported their place of birth as outside the UK the ONS reports that 5.5% were born in the EU, 10.1% were born in non-EU countries (Table 16).[146] Of note is that the ONS estimates are based on all age groups within a population, rather than the paediatric/young adult sub-section recruited to this study, necessitating caution when comparing these data.

According to the ONS in 2021 around half of the non-UK-born population lived in either London or the South-East.[146] In contrast, the non-UK born 'STORY' participants were recruited mainly in the South-East and North-West. Fewer participants recruited in London were from the non-UK born population compared with the ONS data (Table 16).

Table 16 A comparison of the 'STORY' sample and ONS data of migrant populations across the NHS regions

SAMPLE	East of England	London	Midlands	North-East	North-West	South-East	South-West	England
EU	3 (1.8%)	16 (9.6%)	4 (2.4%)	4 (2.4%)	13 (7.8%)	13 (7.8%)	7 (4.2%)	60 (2.1%)
Non-EU	4 (2.4%)	17 (10.2%)	9 (5.4%)	6 (3.6%)	21 (12.6%)	35 (21.0%)	15 (9.0%)	107 (3.7 %)
Combined	7 (4.2 %)	33 (19.8 %)	13 (7.8 %)	10 (6.0 %)	34 (20.4 %)	48 (28.7 %)	22 (13.2 %)	167 (5.8%)
				ONS dataset[147]				
EU	373000 (4.3%)	995000 (11.4%)	958000 (6.1%)	259000 (3.0%)	260000 (3.0%)	456000 (5.2%)	211000 (2.4%)	3084000 (5.5%)
Non-EU	438000 (5.0%)	2352000 (26.8%)	912000 (10.4%)	417000 (4.8%)	469000 (5.4%)	830000 (9.5%)	259000 (3.0%)	5677000 (10.1%)
Combined	811000 (9.3%)	3347000 (38.2%)	1870000 (16.5%)	676000 (7.7%)	729000 (8.3%)	1286000 (14.7%)	470000 (5.4%)	8761000 (15.7%)

3.3.5 Household size and childcare in the under five years population

The mean household size in the 'STORY' sample was 3.8 (mode 4, median 4 range 1-13). This is larger than the ONS estimates which in 2021 was a mean of 2.4 in the UK.[148]

Out of the 533 families of participants aged less than five years enrolled in the 'STORY' study, 66.8% reported that their child attended childcare. In children under 3 years of age 53% attended a childcare setting. Table 17 shows the hours spent in childcare for participants aged less than five years who attended childcare.

Table 17 Distribution of hours spent in childcare in participants aged less than five years who attended childcare

	Frequency	Percent (%)
1-4 hours	3	0.8
4-8 hours	18	5.1
8-12 hours	24	6.7
12-20 hours	88	24.7
21-30 hours	127	35.7
30 + hours	96	27.0

3.4 Discussion

Nationally the children recruited were representative for sex and socioeconomic status according to IMD. When this study was conceived, it was designed to evaluate population immunity to vaccine preventable diseases and not an evolving pandemic. With the emergence of SARS-CoV-2 the study was rapidly adapted during a time of intense research

activity across the country. Recruiting amid high competition for resources from vaccine studies across all sites, and in the context of intermittent national lockdowns, provided additional challenges which would not be seen during a routine national serosurvey. Serosurveys are public health tools to help understand levels of immunity/IgG across different age groups within the population and can influence health policy to deploy health resources to where they are most required.[149] Conducting the study as the SARS-CoV-2 pandemic emerged altered priorities in 'STORY', including a new requirement to continue recruiting over time in order to capture the developing pandemic in children and over sample from groups of special interest. To answer questions about COVID-19 in children and young people the study expanded rapidly to include seven new sites and allowed different recruitment strategies to be utilised (group two).

Strengths of the study include the recruitment strategy employed in 'STORY'. It was successful in generating responses and recruiting large numbers from all target age groups. From experience of other mail-out studies run through the Oxford Vaccine Group, we expected 5-10% of those contacted would respond. Sites response rates were between 2.0-12.7% (mean 5.2%, median 4.7%). Therefore, the response rate was within the expected range. The 'STORY' study achieved the target of enrolling over 100 participants from the community in each one-year age band from 0-19 years of age. The response rate was high across all ages which allowed over recruitment of younger participants where failed enrolment rates were higher. Failed recruitment refers to participants who consented to the study but venepuncture was unsuccessful and did not achieve the minimum volume of 0.5mls of blood. To achieve enrolment targets in children between the ages of 0-5 years, over recruitment of 20% was required. This was reduced to approximately 10% in participants aged 6–10 years and 5% in participants aged 11–19 years.

Each age group was recruited from all seven NHS regions which enables generalisations to be made to England as a whole. When assessing representativeness of the paediatric population in terms of socioeconomic status at a national level according to IMD and IDACI the 'STORY' study was broadly successful, although with over recruitment in the least deprived quintile. At a regional level, recruitment was less representative according to IMD or IDACI, therefore results should be treated with care when interpreting data at a regional level e.g., vaccine responses in response to an outbreak within a specific region. Recruiting across the socioeconomic spectrum is key in assessing the impact of social determinants of health and to reduce health inequalities.[150,151] Early in the pandemic people from minority ethnic groups were of interest as data appeared to show they were at higher risk.[44] By altering recruitment strategy 'STORY' was able to recruit above the national average in terms of participation from minority ethnic groups.[143] The number of participants recruited varied between each of the minority ethnic groups with many identifying as mixed ethnicities. Although overall the study recruited above the national average, on a regional level the sample was less representative of minority ethnic groups. For example, London has a diverse community where both Black and Asian participants were under recruited.[143] Many public engagement strategies such as meeting in community centres of places of worship were not available to study teams due to restrictions on social contact. In the adult studies such as the PRINCIPLE trial, pharmacies were successful in increasing population recruitment from groups of interest[152] but this was not the case within the 'STORY' study. Different sites had different links with the community and some sites recruited more successfully from ethnic minorities as there were better ties to the local community. With social distancing measures in place it was difficult to engage these communities on a personal level; although there were many community action groups in place through sports and schools, these were difficult to engage without an in-person

meeting. In these circumstances the most effective method of recruitment was to send a mail-out to participants who identified as belonging to a minority ethnic group. To achieve this, enquiries were made as to the nature of the demographic material held on each database (individual GP surgeries within selected postcode districts, NHS digital and CHIS) and how it was categorised. This differed considerably from database to database. Gathering this information allowed a more targeted mail out to those under represented groups and enhanced recruitment within them. However, mailouts were an expensive process and with the end of recruitment fast approaching not all sites were able to put approvals in place in time to use targeted mail outs.

Disparities in healthcare across minority ethnic groups are well documented,[47,153] and the relative under recruitment from minority ethnic groups in regions with diverse communities, limits further study of the inter-dependencies between ethnicity and other linked health-relevant demographic factors, such as socioeconomic status, employment type (potentially changing exposure e.g. with COVID-19) and urban or rural location. Lack of diversity in clinical research is a well-recognised challenge.[154] For future seroepidemiological studies time and resources should be spent when study designs are being conceived to tackle behaviours which make it less likely that a representative sample will be recruited.[155] The COVID-19 pandemic and movements such as Black lives matter have raised the profile of health inequalities across minority ethnic groups. Farooqi et al. have developed a tool kit aimed at increasing participation of minority ethnic populations in research, describing how the researchers must be culturally competent to engage with minority ethnic groups.[156] To achieve this there must be public engagement with the communities of interest when the study is in the planning stages and communications must be culturally appropriate.

Unlike for residual sampling methods, 'STORY' participants' immunisation histories were available, with data coming from multiple sources including the participant (in the form of the parent-held health record, i.e. the 'red book' or equivalent), CHIS and GPs. Individuals' vaccine histories show how many children were up to date with their vaccinations for their age and these data will inform the analysis when assessing population immunity against vaccine preventable diseases such as diphtheria and meningococcal capsular group C. This will address the limitation of the residual sample method where assumptions on known population immunisation coverage in the region are used for analysis.[19,157] Although consent for vaccine histories was obtained, the lack of a complete vaccination record for some participants led to partial data collection particularly in the older age groups. Records from red books and GPs were incomplete for vaccinations given through school vaccination programmes. To improve this there would need to be a central immunisation database such as the Nation Immunisation management service (NIMS) which currently holds data on COVID-19 influenza vaccination programmes but no other routine immunisation programmes.[158]

The 'STORY' study limitations include changes in methodology which could mean that the population being recruited at the end of the study was different to the earlier time points e.g. by a skewing of recruitment of ethnic minorities towards the end of the study. This needs to be considered when examining reasons for changes in seroprevalence over time. The rate of recruitment over time varied overall and by region due to competition from vaccine trials across sites and limited responses after a delay in mail-outs. The responses to the group one recruitment strategy were high, however at a national level according to socioeconomic status the representativeness of the sample by recruitment strategy according to IMD was as successful in group two as group one. Therefore, the careful selection of postcode districts to mail out does not appear to have resulted in a more

representative sample than non-selective approaches such as advertising and promotion on social media. As with other studies that have the intention of recruiting a representative sample, it is more common to over recruit participants from more affluent backgrounds and therefore, assessment of socioeconomic status at screening could be a useful tool ensuring equal recruitment across the socioeconomic spectrum rather than a targeted mail out method.[12]

Characteristics of participants within 'STORY' where the representativeness compared with the national England population could not be determined included religion, country of birth and household size. These characteristics did not have age-specific breakdowns within children and young adults between 0-24 years in the census data and therefore are being compared with the general population of England. In the general population, different religions have been shown to have different age profiles in England.[145] Christians who are under represented within the 'STORY' sample tend to have an older age profile with 22% being 65 years and over whereas for children with Muslim faith, who were well-represented, 48% are under 25 years old[145] and would therefore be included within the 'STORY' sample. The differences in population profiles with age can account for the differences in household size. Households with young children are likely to be above the national average for the general population as recruiting children biases the sample towards families rather than the general population where adults are more likely to live in households of one or two people. The 'STORY' study had a lower percentage of participants born outside of the UK when compared with ONS data for the general population in the UK.[146] The highest proportion of England's non-UK born individuals are between the ages of 26-64 years[146] and therefore were not in the study target age group.

It is difficult to interpret data with regards to time in childcare in the under 5-year-old cohort in terms of how representative this was of the general population due to lack of

comprehensive data for the UK population. Data can often be found through the number of children claiming 15-hour entitlement, which is available to every child of pre-school age regardless of parental working status. In 2021 in children aged 3- and 4 years there was the lowest uptake for the entitlement since measurements began in 2011 which is likely due to changes in circumstances due to the pandemic.[159]

National seroprevalence studies, which evaluate immunity across a broad age range, often analyse data in age groups of at least 10 years which combines infants with 10-year-old children in the same group. This can lead to a skewed view of the age group especially since younger age groups are often under represented within the sample.[11,160-170] Younger children of pre-school age may differ in their risk of exposure and immunological responses compared with school age children. Under representation of younger children could result in either under or overestimation of antibodies against specific antigens in these wide age groups, illustrating the need for paediatric-specific seroprevalence studies both within the context of the pandemic and when evaluating population immunity against vaccine preventable disease.

Seroprevalence studies specifically looking at seroprevalence of SARS-CoV-2 within the paediatric population used a variety of different methodologies such as recruitment through schools and through residual samples from healthcare encounters.[39,171-183] Each recruitment strategy has its limitations, e.g. recruiting from schools can result in misleading seroprevalence estimates due to clustering effects, while participants presenting to healthcare within the paediatric population are less likely to be representative of the general paediatric population as in general most children have few healthcare encounters.[171,183]

The concurrent conduct of three paediatric seroprevalence studies during the pandemic allows comparison of the impact different methods deployed and the impact this had on the demographic information obtained and representativeness of the relevant population.

As well as ‘COVID Warriors’ (described in section 2.2.9**Error! Reference source not found.**) a School Infection Survey (SIS) was conducted by UKHSA. This aimed to assess transmission within educational settings by recruiting from selected schools. Each study collected basic demographics such as age and gender but ethnicity was not recorded in ‘COVID Warriors’ and socioeconomic status was less widely reported in the context of the region being sampled in SIS.[24,39,184] The later study aimed to assess transmission within educational settings by recruiting from selected schools. The socioeconomic status of participants was assessed using IMD and whether students were receiving free school meals.[185] The SIS recruited from all NHS regions within England and presented data at a national level by age (primary school and secondary school). At a national level each quintile was represented with higher representation from the most deprived backgrounds compared with STORY. In serostudies designed to evaluate population immunity against vaccine preventable diseases it is important that all ages within an age group are represented if attempting to stratify by age against specific antigens.[186] Residual sample methods have few paediatric samples for each age group and those that are from paediatric populations tend to be from children with underlying health conditions.[157,181] The ‘STORY’ sampling method, although logistically more challenging, addresses many of the limitations of the convenience sampling methods (residual samples, schools and children of healthcare workers). The participants are recruited from the community and the majority are healthy with no underlying health conditions.

Overall, the study design recruited participants between the ages of 0-24 years, creating a biobank that captured many of the characteristics of the population it aims to represent. On a national level the study recruited across the soicioeconomic spectrum and recruited above the national average in terms of participation from minority ethnic groups. There was no significant difference between recruitment methodologies in terms of the participants

recruited at a national level suggesting that an integrated approach of all methods would be appropriate with a screening tool at the point of recruitment to ensure a representative sample is being collected.[12] However, whether this method of prospective sampling will provide added public health benefit is yet to be seen as results for vaccine responses such as diphtheria and meningococcal capsular group C need to be tested and compared with residual sampling methods data.

Chapter 4 Community seroprevalence of SARS-CoV-2 in children and adolescents in England, 2019-2021

4.1 Aims

The aim of this chapter is to report SARS-CoV-2 seroepidemiology in children and adolescents aged 0–18 years in England, prior to widespread immunisation in this population, from samples collected from the end of 2019 through to mid-2021. In addition, we explore potential risk factors for COVID-19 seropositivity and the utility of symptom-based indicators of infection. This chapter is based on Ratcliffe H, Tiley KS, Andrews N, *et al* Community seroprevalence of SARS-CoV-2 in children and adolescents in England, 2019–2021 *Archives of Disease in Childhood* 2023;**108:**123-130.

4.2 Introduction

Seroprevalence studies evaluating population prevalence of SARS-CoV-2 antibodies have an important role in understanding the spread of and population vulnerability to SARS-CoV-2 infection. The majority have been performed in adults,[44,187] and those in children have predominantly tested samples obtained opportunistically in a clinical context or in school-based populations,[24,39,188] both of which bring potential biases.

4.3 Methods

4.3.1 Study design

This was a cross-sectional seroprevalence study recruiting participants from 13 sites distributed across all seven NHS regions in England, conducted between October 2019 and June 2021. This chapter contains data for participants aged 0-18 years recruited according to the methodology outlined in section 2.2. Table 17 shows the NHS region, site and the date the first participant was recruited.

Table 18 Recruiting sites NHS region(s) and postcode districts where applicable

Site	NHS region(s)	Date of first recruit	Selected postcode districts						
Bradford	Yorkshire and Humber	31/3/2020	BD15	BD6					
Bristol	South-West	9/4/2020	BS2	BS20	BS3	BS37	BS41	BS8	
Imperial	London	8/2/2021	SW1	SW3	NW6	NW8	W11	WC2	
Leeds	Yorkshire and Humber	5/6/2020	LS25	WF2					
Manchester	North-West	12/6/2020	M1	M12	M20	M23	M25	M6	
Newcastle	North-East	7/1/2021	NE3	NE12	NE13	NE20	NE23	NE28	
Nottingham*	Midlands	10/7/2020							
Oxford	South-East East of England	15/10/2019	HP17	HP22	HP23	MK13	OX1	OX11	OX28
Plymouth*	South-West	3/6/2020							
Sheffield	Yorkshire and Humber Midlands	20/4/2020	S14	S43					
Southampton	South-East	12/4/2020	SO16	SO23	SO24	SO50	SO51	SO52	
St Georges London	London South-East	26/3/2020	CR4	KT18	KT6	SM3	SW11	SW20	
West Suffolk	East of England	21/1/2021	CB8	CB9	CO10	IP14	IP24	IP28	

**recruitment not restricted by postcode district*

Due to emerging differences in COVID-19 infection in minority ethnic groups in adult studies,[44,189] an enhanced recruitment strategy for ethnic minority groups started in January 2021.

4.3.2 Data collection

Participants or their parent/guardian recorded (electronically or on a paper form) responses to questions regarding the participant's demographics, as well as selected questions from the UK Census 2011 relating to accommodation and employment, and from the FASiii[124] relating to socioeconomic status as described in section 2.2.13.

In response to the pandemic, from February 2020, questionnaires were adapted to ask if participants and/or their household had experienced any potential COVID-19 symptoms (fever, dry cough, shortness of breath, muscle aches, feeling tired, loss of appetite). This was further adapted in July 2020 to enable description of which symptoms participants and household contacts had experienced, self-reported results of relevant PCR or antibody testing. Those participants who were already enrolled were approached retrospectively to collect missing data from the updated questionnaire. Receipt of a COVID-19 vaccine and

vaccination date were recorded either at the visit from personal written documentation or afterwards from GP or CHIS records.

4.3.3 Measurement of serum antibodies

Blood samples were analysed for SARS-CoV-2 specific antibody responses using the Roche Elecsys Anti-SARS-CoV-2 serological assays for the detection of anti-SARS-CoV-2 IgG spike protein receptor binding domain (RocheS) and nucleocapsid (RocheN) antibodies in serum/plasma samples using electrochemiluminescent immunoassays (ECLIA).[129,132] Both assays report high sensitivity and specificity (Table 19). [129,132]

Table 19 Sensitivity, specificity and thresholds of positivity for anti-SARS-CoV-2 IgG spike protein receptor binding domain antibodies (RocheS) and anti-SARS-CoV-2 IgG nucleocapsid antibodies (RocheN) [129,132]

	RocheS	***RocheN***
Threshold of positivity	0·8U/ml	1·0 COI
Sensitivity	95·5% (95%CI 93·2-97·1)	83·9% (95% CI 74·8-90·7)
Specificity	100% (95% CI 99·1-100)	100% (95% CI 99·1-100)

COI – cut off index
U/ml-*units per millilitre*

4.4 Statistical analysis

4.4.1 Seroprevalence

The unweighted observed prevalence for RocheS and RocheN separately was calculated as n+/N for children aged 0-18 years, where n+ was the number of individuals who tested positive, and N was the total number of individuals tested with an available result. Unweighted prevalence was calculated for each of six time periods (Figure 17), [4,190] overall, by age group (0-4 years, 5-9 years, 10-14 years, 15-18 years) and by region. Prevalence estimates for each time period were standardised by age group, ethnicity and NHS region using the Stata stdize command. Population estimates of demographic variables were determined from NewETHPOP- Evaluation, Revision and Extension of Ethnic Population Projections.[143]

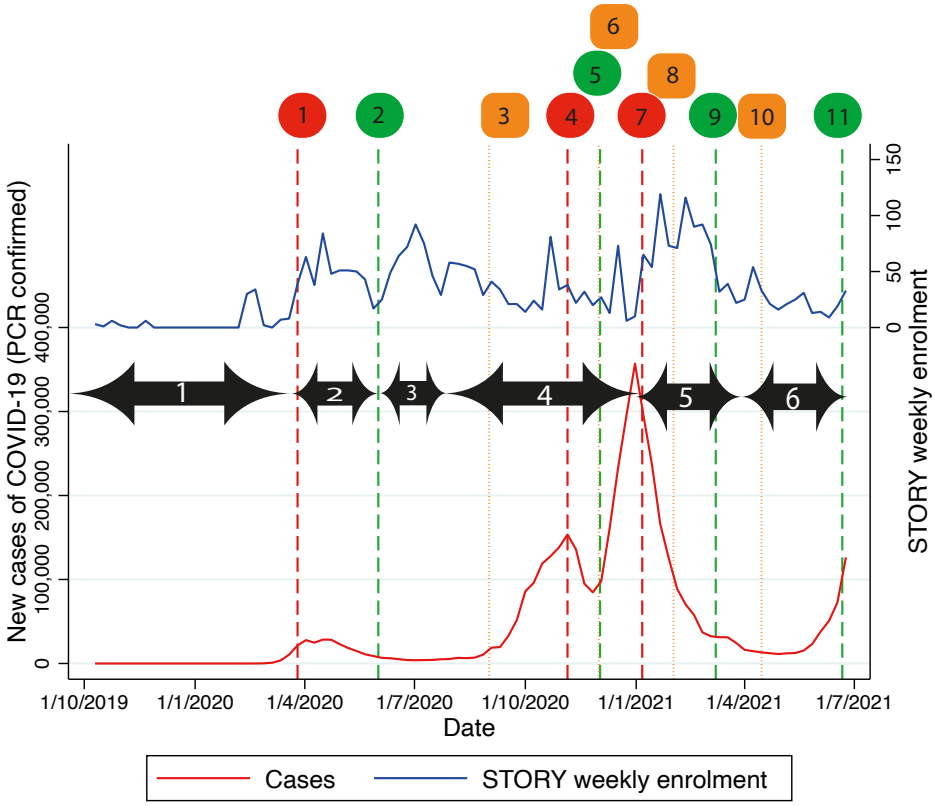


Figure 17 Timeline for COVID-19 cases in England and recruited sample by week for 'STORY' during the same time-period. The numbered arrows denote the time periods used for analysis. Reference lines refer to actions taken by the government in England

1. *First national lockdown came into force (26th March 2020). Schools closed with only children of key workers attending school[4]*

2. *Phased reopening of schools (1st June 2020). Pupils aged 5, 6 and 11 years returned to school. Pupils aged 16- and 18-year-were allowed to attend in limited times[4]*

3. *Variant of concern – B.1.1.7 (Alpha) first detected in the UK and sequenced in September 2020[190]*
4. *Second national lockdown came into force (5th November 2020). Schools closed with only children of key workers attending school[4]*
5. *Second national lockdown came to an end (2nd December 2020) [4]*
6. *Variant of concern B.1.351 (Beta) variant first detected in South Africa and was first sequenced in December 2020. [190]*
7. *Third national lockdown came into force (6th January 2021). Schools closed with only children of key workers attending school[4]*
8. *Variant of concern – P.1 (Gamma) first detected in Japan in travellers from Brazil in January 2021 and was first detected in the UK in February 2021[190]*
9. *Primary and secondary schools reopen in England (8th March 2021) [4]*
10. *Variant of concern B.1.617.2 (Delta) variant first detected in India were first detected in the United Kingdom in mid-April 2021[190]*
11. *All legal limits on social contact removed (21st June 2021)*

4.4.2 Risk factors

Age group, sex, NHS region, time period, ethnicity (grouped as white, and minority ethnic groups), index of multiple deprivation (IMD) or income deprivation affecting children index (IDACI) deprivation quintile (for comparison) and urban/rural classifications were

analysed in univariate and multivariable logistic regression models. A separate model tested the presence of a healthcare worker in the family as a risk factor.

4.4.3 Symptoms

Symptoms associated with seropositivity were explored for children aged 0-9 years and children and adolescents aged 10-18 years separately to optimise statistical power while allowing discrimination of differences between participants attending early years and primary school educational settings compared with secondary school and higher education. A backwards stepwise regression approach was applied whereby variables with the highest p-value were sequentially excluded and model Akaike Information Criterion (AIC) values were compared until a model with the lowest AIC value had been reached. Sex, a non-significant variable, was included to show it did not influence the model. Highly correlated symptoms were grouped e.g. gastrointestinal symptoms included diarrhoea, vomiting and abdominal pain.

Participants reported to be vaccinated before their visit were excluded from all analyses. Analyses were carried out in Stata version 17.[191]

4.5 Results

The study recruited 2,963 participants 0-24 years between October 2019 and June 2021, 2,542 of whom were aged 0-18 years. Of these, 2,540 were COVID-19 vaccine naïve prior to their visit. RocheS and RocheN results were available for 2,477/2,540 (98%) and 2,475/2,540 (97%) participants, respectively. Of those with ethnicity specified, 17% were non-white ethnic groups (Table 20). When comparing household sizes between white and non-white minority ethnic groups, the average was the same, but the range was broader for non-white minority ethnic groups when compared with white participants (Table 21).

Table 20 Summary of participants aged 0-18 years tested for COVID-19 with RocheS-Anti-SARS-CoV-2 IgG spike protein receptor binding domain antibodies and RocheN-Anti-SARS-CoV-2 IgG nucleocapsid antibodies assays for the seroprevalence analysis, symptoms analysis (a full symptom history available) and risk factor analysis (demographics known)

	Seroprevalence		Symptoms		Risk factor	
	RocheS	RocheN	RocheS	RocheN	RocheS	RocheN
Age group						
0-4 years	532 (21·5%)	529 (21·4%)	289 (21·6%)	287 (21·5%)	516 (21·4%)	513 (21·3%)
5-9 years	636 (25·7%)	636 (25·7%)	344 (25·7%)	344 (25·7%)	620 (25·7%)	620 (25·7%)
10-14 years	721 (29·1%)	721 (29·1%)	377 (28·2%)	377 (28·2%)	705 (29·2%)	705 (29·3%)
15-18 years	588 (23·7%)	589 (23·8%)	329 (24·6%)	329 (24·6%)	571 (23·7%)	572 (23·7%)
Sex						
Female	1230 (49·7%)	1230 (49·7%)	689 (51·5%)	688 (51·5%)	1198 (49·7%)	1198 (49·7%)
Male	1247 (50·3%)	1245 (50·3%)	650 (48·7%)	649 (48·5%)	1214 (50·3%)	1212 (50·3%)
NHS region						
East of England	171 (6·9%)	171 (6·9%)	96 (7·2%)	96 (7·2%)	156 (6·5%)	156 (6·5%)
London	289 (11·7%)	289 (11·7%)	181 (13·6%)	181 (13·5%)	287 (11·9%)	287 (11·9%)
Midlands	210 (8·5%)	211 (8·5%)	138 (10·3%)	138 (10·3%)	207 (8·6%)	208 (8·6%)
North-East and Yorkshire	419 (16·9%)	419 (16·9%)	232 (17·3%)	232 (17·4%)	410 (17·0%)	410 (17·0%)
North-West	331 (13·4%)	331 (13·4%)	141 (10·5%)	141 (10·5%)	326 (13·5%)	326 (13·5%)
South-East	633 (25·6%)	631 (25·5%)	260 (19·4%)	258 (19·3%)	607 (25·2%)	605 (25·1%)
South-West	424 (17·1%)	423 (17·1%)	291 (21·7%)	291 (21·8%)	419 (17·4%)	418 (17·3%)
Time period						
Pre-pandemic (01 Oct 2019-31 Mar 2020)	125 (5·0%)	125 (5·1%)	32 (2·4%)	32 (2·4%)	125 (5·2%)	125 (5·2%)
First wave (01 Apr 2020-31 May 2020)	335 (13·5%)	335 (13·5%)	74 (5·5%)	74 (5·5%)	334 (13·8%)	334 (13·9%)
Post first wave (01 Jun 2020-31 Aug 2020)	516 (20·8%)	515 (20·8%)	111 (8·3%)	111 (8·3%)	507 (21·0%)	506 (21·0%)
Schools reopening and second wave (01 Sep 2020-31 Dec 2020)	464 (18·7%)	464 (18·7%)	298 (22·3%)	297 (22·2%)	456 (18·9%)	456 (18·9%)
Post second wave (01 Jan 2021-31 Mar 2021)	771 (31·1%)	770 (31·1%)	617 (46·1%)	616 (46·1%)	736 (30·5%)	735 (30·5%)
Emergence of delta variant (01 Apr 2021-30 Jun 2021)	266 (10·7%)	266 (10·7%)	207 (15·5%)	207 (15·5%)	254 (10·5%)	254 (10·5%)
Ethnicity						
White (White and White minorities)	1991 (80·4%)	1989 (80·4%)	1073 (80·1%)	1071 (80·1%)	1987 (82·4%)	1985 (82·4%)
Black (African, Caribbean, Other Black background)	51 (2·1%)	51 (2·1%)	29 (2·2%)	29 (2·2%)	51 (2·1%)	51 (2·1%)

Asian (Bangladeshi, Pakistani, Indian, Chinese, Arab, other Asian background and other ethnic group)	144 (5·8%)	144 (5·8%)	99 (7·4%)	99 (7·4%)	144 (6·0%)	144 (6·0%)
Multiple ethnic backgrounds	231 (9·3%)	231 (9·3%)	134 (10·0%)	134 (10·0%)	230 (9·5%)	230 (9·5%)
Not specified	60 (2·4%)	60 (2·4%)	4 (0·3%)	4 (0·3%)		
IMD quintile						
Most deprived 1	357 (14·4%)	357 (14·4%)	179 (13·4%)	179 (13·4%)	352 (14·6%)	352 (14·6%)
2	337 (13·6%)	336 (13·6%)	178 (13·3%)	177 (13·2%)	328 (13·6%)	327 (13·6%)
3	461 (18·6%)	461 (18·6%)	247 (18·4%)	247 (18·5%)	446 (18·5%)	446 (18·5%)
4	535 (21·7%)	534 (21·6%)	289 (21·6%)	289 (21·6%)	518 (21·5%)	517 (21·5%)
Least deprived 5	782 (31·6%)	782 (31·6%)	441 (33·0%)	440 (32·9%)	768 (31·8%)	768 (31·9%)
Not specified	5 (0·2%)	5 (0·2%)	5 (0·4%)	5 (0·4%)		
IDACI quintile						
Most deprived 1	374 (15·1%)	374 (15·1%)	188 (14·0%)	188 (14·1%)	368 (15·3%)	368 (15·3%)
2	348 (14·0%)	347 (14·0%)	194 (14·5%)	193 (14·4%)	342 (14·2%)	341 (14·1%)
3	451 (18·2%)	451 (18·2%)	225 (16·8%)	225 (16·8%)	433 (18·0%)	433 (18·0%)
4	549 (22·2%)	541 (21·9%)	305 (22·8%)	305 (22·8%)	538 (22·3%)	537 (22·3%)
Least deprived 5	750 (30·3%)	750 (30·3%)	422 (31·5%)	421 (31·5%)	731 (30·3%)	731 (30·3%)
Not specified	5 (0·2%)	5 (0·2%)	5 (0·4%)	5 (0·4%)		
Urban/rural						
Urban	1984 (80·1%)	1982 (80·1%)	1076 (80·4%)	1075 (80·4%)	1933 (80·1%)	1931 (80·1%)
Rural	493 (19·9%)	493 (19·9%)	263 (19·6%)	262 (19·6%)	479 (19·9%)	479 (19·9%)
Member of family is a healthcare worker						
Yes	628 (25·4%)	627 (25·3%)	421 (31·5%)	421 (31·5%)	626 (26·0%)	625 (25·9%)
No	1269 (51·2%)	1268 (51·2%)	804 (60·0%)	802 (60·0%)	1262 (52·3%)	1261 (52·3%)
Not specified	580 (23·4%)	580 (23·4%)	114 (8·5%)	114 (8·5%)	524 (21·7%)	524 (21·7%)
Total	2477	2475	1339	1337	2412	2410

Table 21 shows a summary of household sizes for participants by ethnicity.

	Minority ethnic group	White
Mean	4.0	3.8
Median	4	4
Mode	5	4
Range	1-32	1-13

The proportion of children aged 0-15 years in each IMD and IDACI quintile was similar (within 1·5%) for the study overall, and within 10% of their local IMD quintile across the majority of regions with the exception of the North-West (<1% in least deprived IMD quintile vs 21% in least deprived IDACI quintile) and South-East (Table 22). Responses to selected UK Census questions for children aged 0-15 years were generally in agreement with IMD and IDACI quintiles but did not clearly differentiate between the scoring systems.

In total 628/2,477 (25%) were children of healthcare workers.

Table 22 Comparing children aged 0-15 years on IDACI and IMD by region as a measure of socioeconomic deprivation in children IMD and IDACI derived from postcode of participant, FASiii and Census 2011 home ownership and house size questions data were collected on the questionnaire. Quintile 1 was most deprived and quintile 5 was most affluent.

		Quintile				
		1 (most deprived)	**2**	**3**	**4**	**5 (least deprived)**
Overall	**Number of families recruited in each IMD quintile**	303	280	409	453	663
	Number of families that own their own home in each IMD quintile	50·8	69·4	77·9	81·7	85·8
	Mean number of rooms in home in each IMD quintile (census)	5·6	6·1	6·6	7·0	7·3
	Mean Number of bedrooms in home in each IMD quintile (census)	2·9	3·2	3·5	3·6	3·8
	Mean FASiii score in each IMD quintile	6·5	7·7	8·7	8·9	9·5
	Number of families recruited in each IDACI quintile	315	304	481	476	632
	Number of families that own their own home in each IDACI quintile	49·8	66·4	79·8	81·6	87·7
	Mean number of rooms in home in each IDACI quintile (census)	5·6	6·1	6·6	6·8	7·5
	Mean Number of bedrooms in home in each IDACI quintile (census)	2·9	3·2	3·5	3·6	3·8
	Mean FASiii score in each IDACI quintile	6·6	7·7	8·6	8·9	9·5
East of England	**Number of families recruited in each IMD quintile**	16	22	44	40	25
	Number of families that own their own home in each IMD quintile	30·8	59·1	65·8	86·1	100·0
	Mean number of rooms in home in each IMD quintile (census)	5·8	6·6	6·8	7·4	7·8
	Mean Number of bedrooms in home in each IMD quintile (census)	2·9	3·5	3·3	3·9	3·7
	Mean FASiii score in each IMD quintile	6·4	7·6	7·8	9·3	9·7
	Number of families recruited in each IDACI quintile	24	22	30	40	31

	Number of families that own their own home in each IDACI quintile	33·3	59·1	72·0	88·6	93·1
	Mean number of rooms in home in each IDACI quintile (census)	6·2	5·2	7·2	7·1	8·0
	Mean Number of bedrooms in home in each IDACI quintile (census)	3·0	3·4	3·8	3·5	3·8
	Mean FASiii score in each IDACI quintile	6·5	7·2	9·2	8·5	9·7
London	**Number of families recruited in each IMD quintile**	9	41	61	48	79
	Number of families that own their own home in each IMD quintile	22·2	36·6	73·8	83·3	84·6
	Mean number of rooms in home in each IMD quintile (census)	4·8	5·1	5·9	6·7	6·8
	Mean Number of bedrooms in home in each IMD quintile (census)	2·3	2·9	3·2	3·5	3·8
	Mean FASiii score in each IMD quintile	5·8	7·5	8·5	9·3	9·7
	Number of families recruited in each IDACI quintile	18	55	35	61	69
	Number of families that own their own home in each IDACI quintile	27·8	50·9	74·3	80·3	88·2
	Mean number of rooms in home in each IDACI quintile (census)	4·6	5·3	6·1	6·4	7·0
	Mean Number of bedrooms in home in each IDACI quintile (census)	2·5	2·9	3·3	3·6	3·9
	Mean FASiii score in each IDACI quintile	7·2	7·5	8·3	9·2	10·1
Midlands	**Number of families recruited in each IMD quintile**	17	30	34	30	44
	Number of families that own their own home in each IMD quintile	58·8	67·9	84·8	86·7	95·3
	Mean number of rooms in home in each IMD quintile (census)	6·8	6·3	6·4	7·1	7·8
	Mean Number of bedrooms in home in each IMD quintile (census)	3·6	3·4	3·5	3·7	4·0
	Mean FASiii score in each IMD quintile	7·5	8·3	8·8	9·1	9·8
	Number of families recruited in each IDACI quintile	23	32	31	35	34
	Number of families that own their own home in each IDACI quintile	52·2	75·9	83·9	91·2	97·1
	Mean number of rooms in home in each IDACI quintile (census)	6·6	6·4	6·3	7·3	7·9
	Mean Number of bedrooms in home in each IDACI quintile (census)	3·4	3·5	3·4	3·9	4·0
	Mean FASiii score in each IDACI quintile	7·7	8·5	8·7	9·5	9·7
North-East and Yorkshire	**Number of families recruited in each IMD quintile**	88	59	61	57	89
	Number of families that own their own home in each IMD quintile	58·8	82·5	81·7	91·1	88·6
	Mean number of rooms in home in each IMD quintile (census)	5·5	6·3	6·8	7·1	7·4
	Mean Number of bedrooms in home in each IMD quintile (census)	2·9	3·2	3·6	3·6	3·7
	Mean FASiii score in each IMD quintile	6·8	7·2	8·7	9·3	9·0
	Number of families recruited in each IDACI quintile	83	57	68	41	105
	Number of families that own their own home in each IDACI quintile	60·0	73·2	83·3	87·5	92·3
	Mean number of rooms in home in each IDACI quintile (census)	5·6	6·3	6·8	6·7	7·4
	Mean Number of bedrooms in home in each IDACI quintile (census)	3·0	3·2	3·5	3·5	3·7
	Mean FASiii score in each IDACI quintile	6·7	7·3	8·6	8·8	9·1
North-West	**Number of families recruited in each IMD quintile**	119	54	71	51	2
	Number of families that own their own home in each IMD quintile	51·7	86·5	81·7	85·7	50·0
	Mean number of rooms in home in each IMD quintile (census)	5·7	6·5	7·3	7·8	8·0
	Mean Number of bedrooms in home in each IMD quintile (census)	2·9	3·4	3·9	4·2	4·5
	Mean FASiii score in each IMD quintile	6·5	8·4	9·0	8·7	9·5

	Number of families recruited in each IDACI quintile	116	27	50	41	63
	Number of families that own their own home in each IDACI quintile	52·2	74·1	87·5	70·0	91·9
	Mean number of rooms in home in each IDACI quintile (census)	5·7	6·0	7·3	6·4	8·1
	Mean Number of bedrooms in home in each IDACI quintile (census)	3·0	3·2	3·7	3·7	4·3
	Mean FASiii score in each IDACI quintile	6·5	8·0	8·3	8·5	9·3
South-East	**Number of families recruited in each IMD quintile**	23	40	81	112	293
	Number of families that own their own home in each IMD quintile	27·3	63·2	73·7	76·9	80·4
	Mean number of rooms in home in each IMD quintile (census)	5·6	5·9	6·1	7·1	7·3
	Mean Number of bedrooms in home in each IMD quintile (census)	2·7	3·1	3·3	3·5	3·9
	Mean FASiii score in each IMD quintile	6·3	7·6	8·4	9·4	9·4
	Number of families recruited in each IDACI quintile	19	68	104	152	206
	Number of families that own their own home in each IDACI quintile	23·5	65·7	72·6	77·7	82·5
	Mean number of rooms in home in each IDACI quintile (census)	5·3	6·1	6·5	7·0	7·5
	Mean Number of bedrooms in home in each IDACI quintile (census)	2·5	3·2	3·5	3·6	3·8
	Mean FASiii score in each IDACI quintile	5·5	8·0	8·7	9·2	9·6
South-West	**Number of families recruited in each IMD quintile**	31	34	57	115	131
	Number of families that own their own home in each IMD quintile	54·8	75·8	83·6	76·1	91·3
	Mean number of rooms in home in each IMD quintile (census)	5·1	5·8	6·9	6·4	7·3
	Mean Number of bedrooms in home in each IMD quintile (census)	2·8	3·4	3·5	3·5	3·8
	Mean FASiii score in each IMD quintile	6·1	7·5	8·5	8·2	9·5
	Number of families recruited in each IDACI quintile	32	43	63	106	124
	Number of families that own their own home in each IDACI quintile	51·6	71·4	85·2	84·8	85·8
	Mean number of rooms in home in each IDACI quintile (census)	5·2	6·5	6·4	6·7	7·2
	Mean Number of bedrooms in home in each IDACI quintile (census)	2·9	3·3	3·3	3·7	3·7
	Mean FASiii score in each IDACI quintile	6·3	7·6	8·5	8·6	9·2

4.5.1 Comparison of assays

Of the 2,472 participants with results for both assays, 218 (9%) were both positive, and 2,215 (90%) both negative with 40 (1%) having discordant results (Figure 18). The majority (35/40, 88%) of discordant results were RocheS positive and RocheN negative.

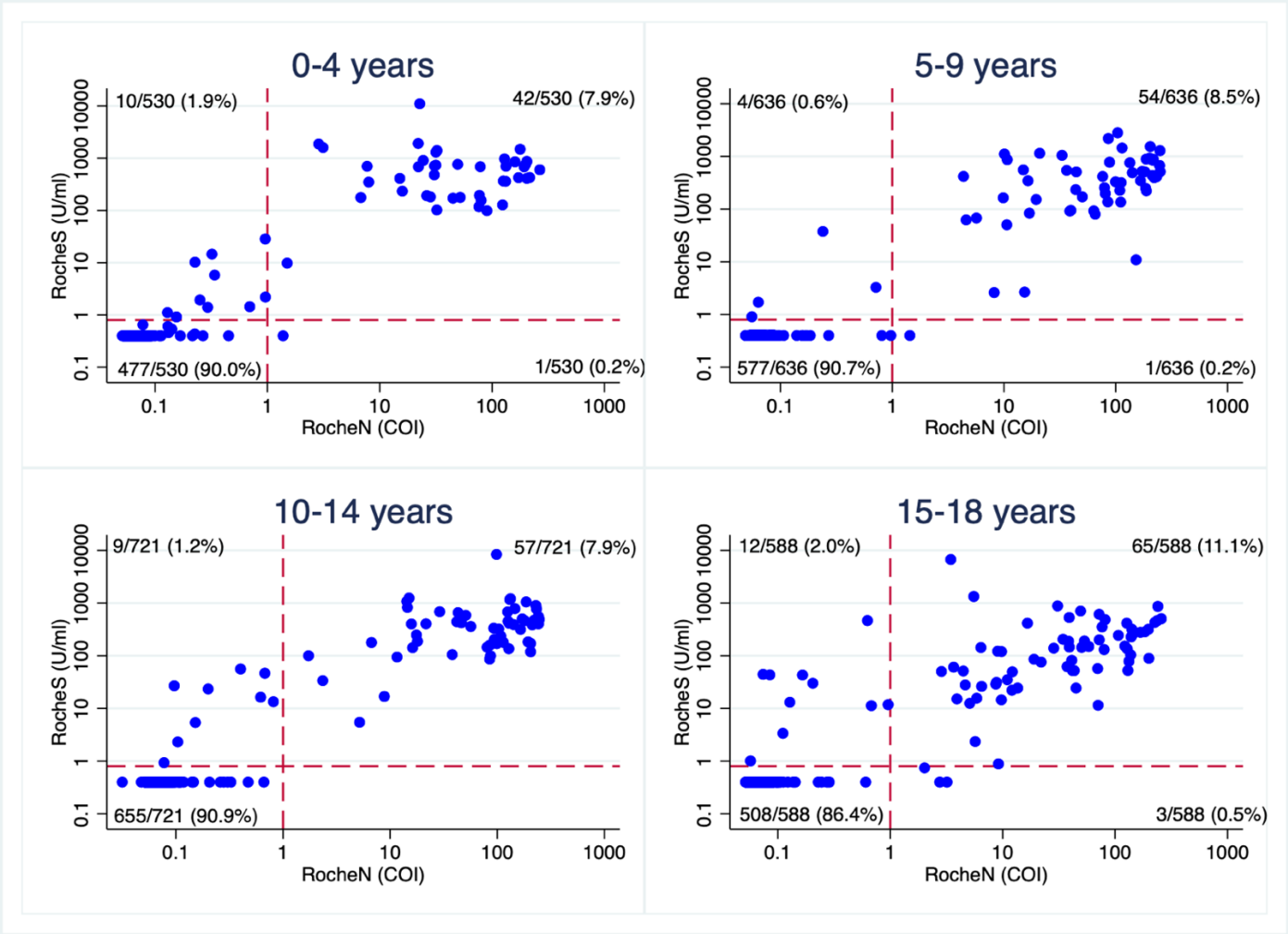


Figure 18 A comparison of results by the RocheN-Anti-SARS-CoV-2 IgG nucleocapsid antibodies assay and RocheS-Anti-SARS-CoV-2 IgG spike protein receptor binding domain antibodies assay by age

** Assay positivity thresholds indicated as red dashed lines (0.8 U/ml for Spike and 1 COI for Nucleocapsid)*
***COI – cut off index, U/ml (units per millilitre)*

4.5.2 Seroprevalence

Overall seroprevalence, adjusted for age, ethnicity and NHS region, increased over time (Figure 19, Table 23) from zero in October 2019-March 2020 to 19.9% (95% CI 16·1%-23·7%) and 16.8% (95% CI 13·1%-20·6%) on RocheS and Roche N respectively in April 2021-June 2021.

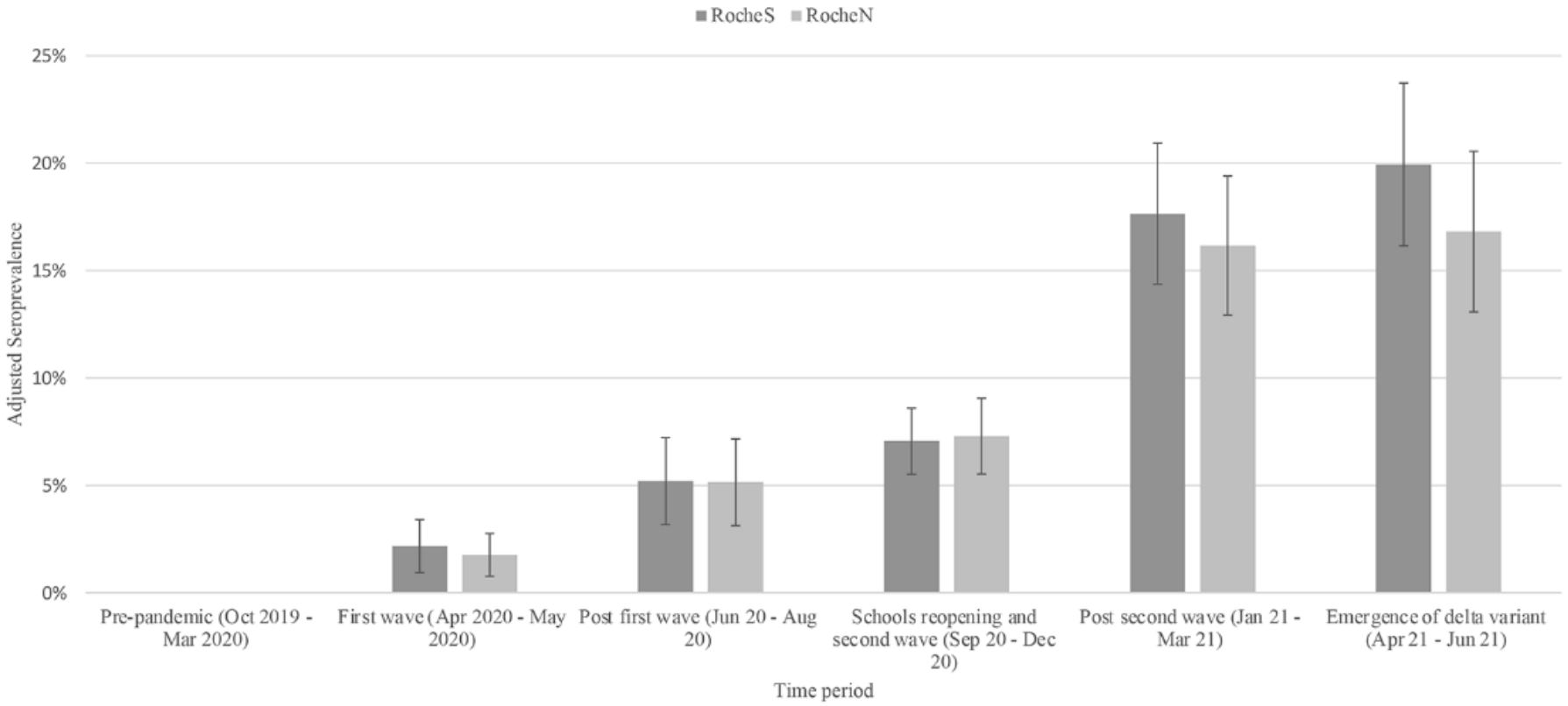


Figure 19 Overall SARS-CoV-2 seroprevalence (RocheS-Anti-SARS-CoV-2 IgG spike protein receptor binding domain antibodies and RocheN-Anti-SARS-CoV-2 IgG nucleocapsid antibodies) by time period October 2019 to June 2021 in England, adjusted for age, NHS region and ethnicity. Error bars indicate 95% confidence intervals

Table 23 SARS-CoV-2 seroprevalence (RocheS-Anti-SARS-CoV-2 IgG spike protein receptor binding domain antibodies and RocheN-Anti-SARS-CoV-2 IgG nucleocapsid antibodies) by time period October 2019 to June 2021 in England, adjusted for age, region and ethnicity

	RocheS			RocheN		
Period	**N**	**Crude rate**	**Adjusted rate**	**N**	**Crude rate**	**Adjusted rate**
Pre-pandemic (Oct 19-Mar 20)	125	0·0%	0·0%	125	0·0%	0·0%
First wave (Apr 20-May 20)	335	3·9%	2·2% (0·9%-3·4%)	335	3·6%	1·8% (0·8%-2·8%)
Post first wave (Jun 20-Aug 20)	516	3·9%	5·2% (3·2%-7·2%)	515	3·7%	5·1% (3·1%-7·2%)
Schools reopening and second wave (Sep 20-Dec 20)	464	6·9%	7·1% (5·5%-8·6%)	464	6·0%	7·3% (5·5%-9·1%)
Post second wave (Jan 21-Mar 21)	771	17·5%	17·6% (14·4%-20·9%)	770	15·2%	16·2% (12·9%-19·4%)
Emergence of delta variant (Apr 21-Jun 21)	266	19·9%	19·9% (16·1%-23·7%)	266	17·7%	16·8% (13·1%-20·6%)

For all age groups seroprevalence remained relatively stable (10% or below) until September-December 2020 ('schools reopening and second wave' period), when adolescents aged 15-18 years increased to 23% (95% CI 17%-29%) 21% (95% CI 15%-26%) on RocheS and RocheN respectively, followed by an increase in all age groups from January 2021 onwards (Figure 20, Table 22). By April-June 2021 ('emergence of delta variant' period) the adjusted seroprevalence was 20%, varying from 14% in children aged 0-4 years to 33% in adolescents aged 15-18 years on Roche S (Figure 20, Table 24).

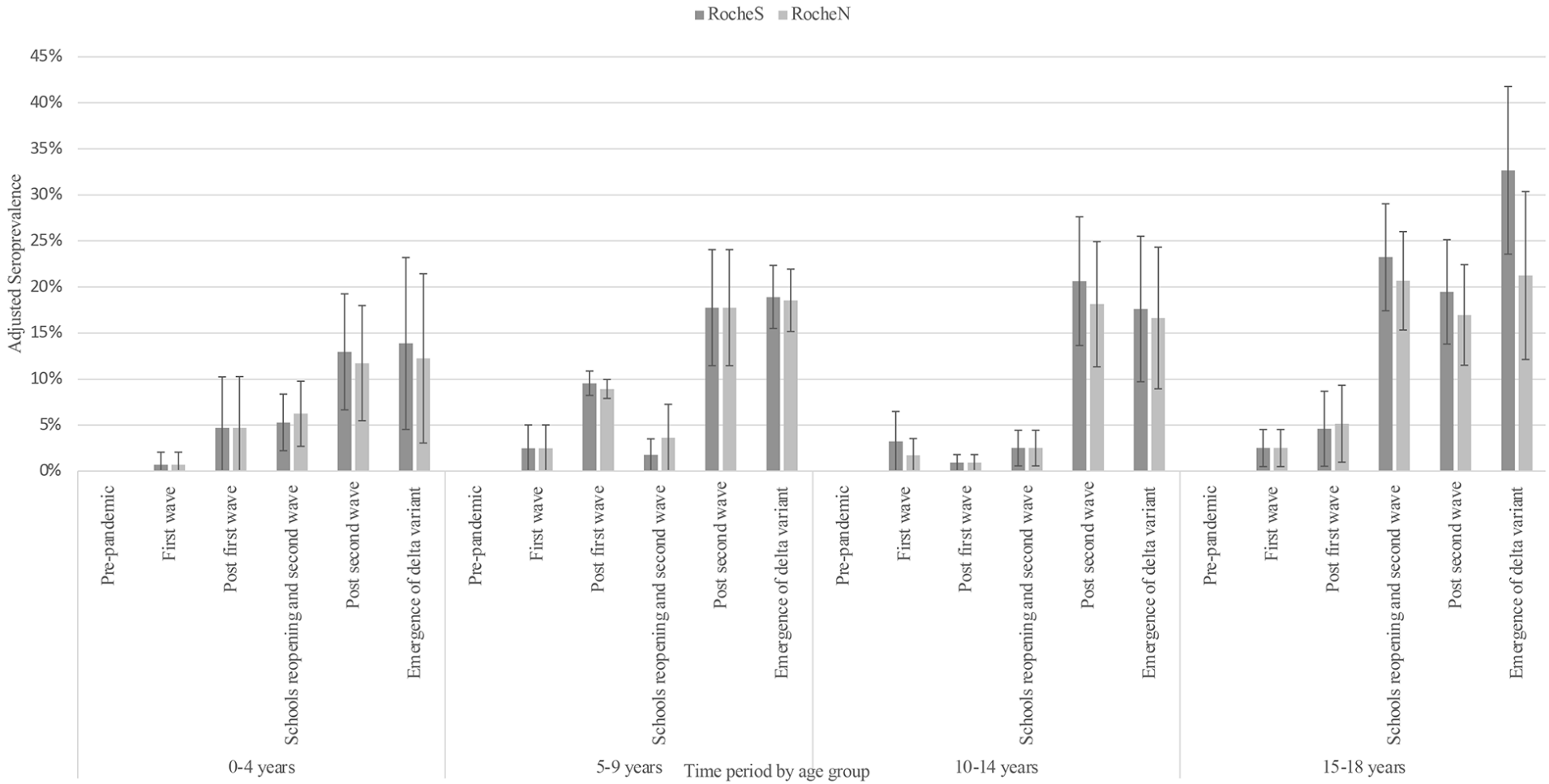
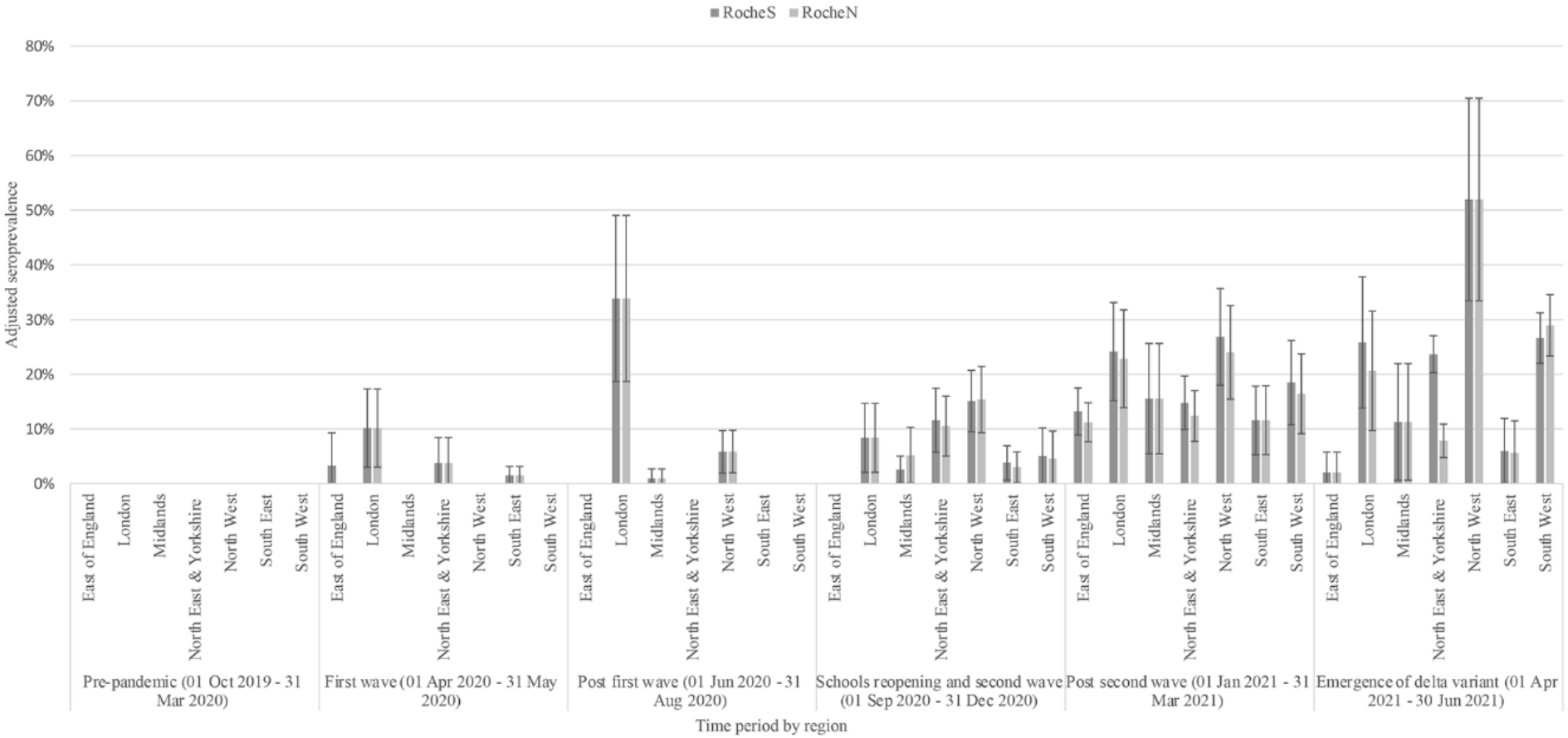


Figure 20 SARS-CoV-2 seroprevalence (RocheS-Anti-SARS-CoV-2 IgG spike protein receptor binding domain antibodies and RocheN-Anti-SARS-CoV-2 IgG nucleocapsid antibodies) by age group and time period October 2019 to June 2021 in England, adjusted for NHS region and ethnicity. Error bars indicate 95% confidence intervals

Table 24 SARS-CoV-2 seroprevalence (RocheS-Anti-SARS-CoV-2 IgG spike protein receptor binding domain antibodies and RocheN-Anti-SARS-CoV-2 IgG nucleocapsid antibodies) by age group and time period October 2019 to June 2021 in England, adjusted for NHS region and ethnicity

		RocheS			RocheN		
		N	Crude	Adjusted	N	Crude	Adjusted
0-4 years	**Pre-pandemic (Oct 19-Mar 20)**	25	0·0%	0·0%	25	0·0%	0·0%
	First wave (Apr 20-May 20)	62	1·6%	0·7% (0·0%-2·1%)	62	1·6%	0·7% (0·0%-2·1%)
	Post first wave (Jun 20-Aug 20)	64	4·7%	4·7% (0·0%-10·2%)	63	4·8%	4·7% (0·0%-10·3%)
	Schools reopening and second wave (Sep 20-Dec 20)	77	5·2%	5·3% (2·2%-8·4%)	76	6·6%	6·2% (2·7%-9·8%)
	Post second wave (Jan 21 -Mar 21)	200	13·5%	13·0% (6·7%-19·3%)	199	10·1%	11·7% (5·5%-18·0%)
	Emergence of delta variant (Apr 21-Jun 21)	104	16·3%	13·9% (4·5%-23·2%)	104	13·5%	12·2% (3·1%-21·4%)
5-9 years	**Pre-pandemic (Oct 19-Mar 20)**	31	0·0%	0·0%	31	0·0%	0·0%
	First wave (Apr 20-May 20)	83	3·6%	2·5% (0·0%-5·0%)	83	3·6%	2·5% (0·0%-5·0%)
	Post first wave (Jun 20-Aug 20)	144	4·9%	9·5% (8·2%-10·9%)	144	3·5%	8·9% (7·9%-10·0%)
	Schools reopening and second wave (Sep 20-Dec 20)	119	3·4%	1·8% (0·0%-3·5%)	119	3·4%	3·6% (0·0%-7·3%)
	Post second wave (Jan 21 -Mar 21)	187	15·0%	17·8% (11·5%-24·1%)	187	15·0%	17·8% (11·5%-24·1%)
	Emergence of delta variant (Apr 21-Jun 21)	72	22·2%	18·9% (15·5%-22·3%)	72	20·8%	18·5% (15·2%-21·9%)
10-14 years	**Pre-pandemic (Oct 19-Mar 20)**	41	0·0%	0·0%	41	0·0%	0·0%
	First wave (Apr 20-May 20)	108	3·7%	3·2% (0·0%-6·5%)	108	2·8%	1·7% (0·0%-3·5%)
	Post first wave (Jun 20-Aug 20)	174	2·3%	0·9% (0·0%-1·8%)	174	2·3%	0·9% (0·0%-1·8%)
	Schools reopening and second wave (Sep 20-Dec 20)	152	5·3%	2·5% (0·6%-4·4%)	152	5·3%	2·5% (0·6%-4·4%)
	Post second wave (Jan 21 -Mar 21)	195	21·5%	20·6% (13·6%-27·6%)	195	17·9%	18·1% (11·3%-24·9%)
	Emergence of delta variant (Apr 21-Jun 21)	51	15·7%	17·6% (9·7%-25·5%)	51	13·7%	16·6% (8·9%-24·3%)
15-18 years	**Pre-pandemic (Oct 19-Mar 20)**	28	0·0%	0·0%	28	0·0%	0·0%
	First wave (Apr 20-May 20)	82	6·1%	2·5% (0·5%-4·5%)	82	6·1%	2·5% (0·5%-4·5%)
	Post first wave (Jun 20-Aug 20)	134	4·5%	4·6% (0·5%-8·7%)	134	5·2%	5·1% (1·0%-9·3%)
	Schools reopening and second wave (Sep 20-Dec 20)	116	13·8%	23·2% (17·4%-29·0%)	117	9·4%	20·7% (15·3%-26·0%)
	Post second wave (Jan 21 -Mar 21)	189	20·1%	19·5% (13·8%-25·1%)	189	18·0%	17·0% (11·5%-22·4%)
	Emergence of delta variant (Apr 21-Jun 21)	39	30·8%	32·7% (23·6%-41·8%)	39	28·2%	21·2% (12·1%-30·4%)

In June-August 2020, the highest age and ethnicity adjusted seroprevalence was recorded in London (34%). In April-June 2021 the highest seroprevalence was in the North-West (52%) (Figure 21 and Table 25).



Enhanced recruitment for minority ethnic groups started in Jan 2021.
Data showed differing results between RocheS and RocheN in the North-East and Yorkshire (April-June 2021). After comparing the raw data between groups without finding significant differences the most likely explanation is a chance variation within these data

Figure 21 SARS-CoV-2 seroprevalence (RocheS-Anti-SARS-CoV-2 IgG spike protein receptor binding domain antibodies and RocheN-Anti-SARS-CoV-2 IgG nucleocapsid antibodies) by region and time period October 2019 to June 2021 in England, adjusted for age and ethnicity. Error bars indicate 95% confidence intervals.

Table 25 SARS-CoV-2 seroprevalence (RocheS-Anti-SARS-CoV-2 IgG spike protein receptor binding domain antibodies and RocheN-Anti-SARS-CoV-2 IgG nucleocapsid antibodies) by region and time period October 2019 to June 2021 in England, adjusted for age and ethnicity·

		RocheS			RocheN		
		N	Crude	Adjusted	N	Crude	Adjusted
Pre-pandemic (Oct 19-Mar 20)	**East of England**	21	0·0%	0·0%	21	0·0%	0·0%
	London	10	0·0%	0·0%	10	0·0%	0·0%
	Midlands	0	0·0%	0·0%	0	0·0%	0·0%
	North-East & Yorkshire	2	0·0%	0·0%	2	0·0%	0·0%
	North-West	0	0·0%	0·0%	0	0·0%	0·0%
	South-East	92	0·0%	0·0%	92	0·0%	0·0%
	South-West	0	0·0%	0·0%	0	0·0%	0·0%
First wave (Apr 20-May 20)	**East of England**	16	6·3%	3·3% (0·0%-9·3%)	16	6·3%	0·0%
	London	55	12·7%	10·2% (3·0%-17·3%)	55	12·7%	10·2% (3·0%-17·3%)
	Midlands	27	0·0%	0·0%	27	0·0%	0·0%
	North-East & Yorkshire	64	3·1%	3·8% (0·0%-8·4%)	64	3·1%	3·8% (0·0%-8·4%)
	North-West	0	0·0%	0·0%	0	0·0%	0·0%
	South-East	134	2·2%	1·5% (0·0%-3·1%)	133	2·3%	1·5% (0·0%-3·1%)
	South-West	39	0·0%	0·0%	39	0·0%	0·0%
Post first wave (Jun 20- Aug 20)	**East of England**	9	0·0%	0·0%	9	0·0%	0·0%
	London	14	21·4%	33·9% (18·7%-49·1%)	14	21·4%	33·9% (18·7%-49·1%)
	Midlands	46	2·2%	0·9% (0·0%-2·7%)	46	2·2%	0·9% (0·0%-2·7%)
	North-East & Yorkshire	73	0·0%	0·0%	73	0·0%	0·0%
	North-West	127	6·3%	5·8% (1·9%-9·7%)	125	6·4%	5·9% (2·0%-9·8%)
	South-East	97	0·0%	0·0%	94	0·0%	0·0%
	South-West	39	0·0%	0·0%	39	0·0%	0·0%
Schools reopening and second wave (Sep 20- Dec 20)	**East of England**	11	0·0%	0·0%	11	0·0%	0·0%
	London	12	16·7%	8·4% (2·0%-14·7%)	12	16·7%	8·4% (2·0%-14·7%)
	Midlands	77	6·5%	2·6% (0·1%-5·0%)	75	5·3%	5·2% (0·1%-10·3%)
	North-East & Yorkshire	62	11·3%	11·6% (5·7%-17·4%)	61	9·8%	10·5% (5·0%-16·0%)
	North-West	61	11·5%	15·1% (9·5%-20·7%)	61	11·5%	15·4% (9·3%-21·4%)
	South-East	127	3·9%	3·8% (0·7%-6·9%)	122	4·1%	3·0% (0·2%-5·8%)
	South-West	114	5·3%	5·1% (0·0%-10·2%)	114	5·3%	4·6% (0·0%-9·6%)
Post second wave (Jan 21-Mar 21)	**East of England**	88	9·1%	13·2% (8·9%-17·5%)	76	10·5%	11·2% (7·7%-14·8%)
	London	115	26·1%	24·1% (15·1%-33·1%)	115	26·1%	22·8% (13·9%-31·8%)
	Midlands	49	14·3%	15·5% (5·4%-25·6%)	48	14·6%	15·5% (5·4%-25·6%)
	North-East & Yorkshire	193	16·1%	14·8% (9·9%-19·7%)	187	16·6%	12·4% (7·7%-17·0%)
	North-West	116	28·4%	26·8% (18·0%-35·7%)	115	27·8%	24·0% (15·4%-32·6%)
	South-East	120	10·8%	11·5% (5·3%-17·8%)	109	11·0%	11·6% (5·3%-17·9%)
	South-West	90	14·4%	18·5% (10·8%-26·2%)	89	14·6%	16·4% (9·1%-23·7%)
	East of England	26	7·7%	2·0% (0·0%-5·8%)	23	4·3%	2·0% (0·0%-5·8%)

Emergence of delta variant (Apr 21-Jun 21)	**London**	83	24·1%	25·8% (13·8%-37·8%)	83	24·1%	20·6% (9·7%-31·5%)
	Midlands	11	18·2%	11·3% (0·7%-21·9%)	11	18·2%	11·3% (0·7%-21·9%)
	North-East & Yorkshire	25	24·0%	23·7% (20·3%-27·0%)	25	24·0%	7·8% (4·8%-10·9%)
	North-West	27	40·7%	51·9% (33·4%-70·5%)	26	42·3%	51·9% (33·4%-70·5%)
	South-East	63	11·1%	6·0% (0·1%-11·9%)	57	10·5%	5·6% (0·0%-11·5%)
	South-West	31	16·1%	26·6% (22·0%-31·2%)	31	16·1%	28·9% (23·3%-34·6%)

4.5.3 Presence of symptoms in seropositive participants

Overall, 48% (95% CI 42-55%) and 45% (95% CI 38-52%) of seropositive participants reported no symptoms as measured by RocheS and RocheN (Figure 22 and Figure 23). Fever was reported in 29% and 25% of seropositive participants on RocheS and RocheN respectively. In a univariate analysis fever was predictive of SARS-CoV-2 positivity in aged 10-18 year only (Figure 24 & Figure 25). No solicited symptoms were individually predictive of seropositivity in children aged 0-9 years.

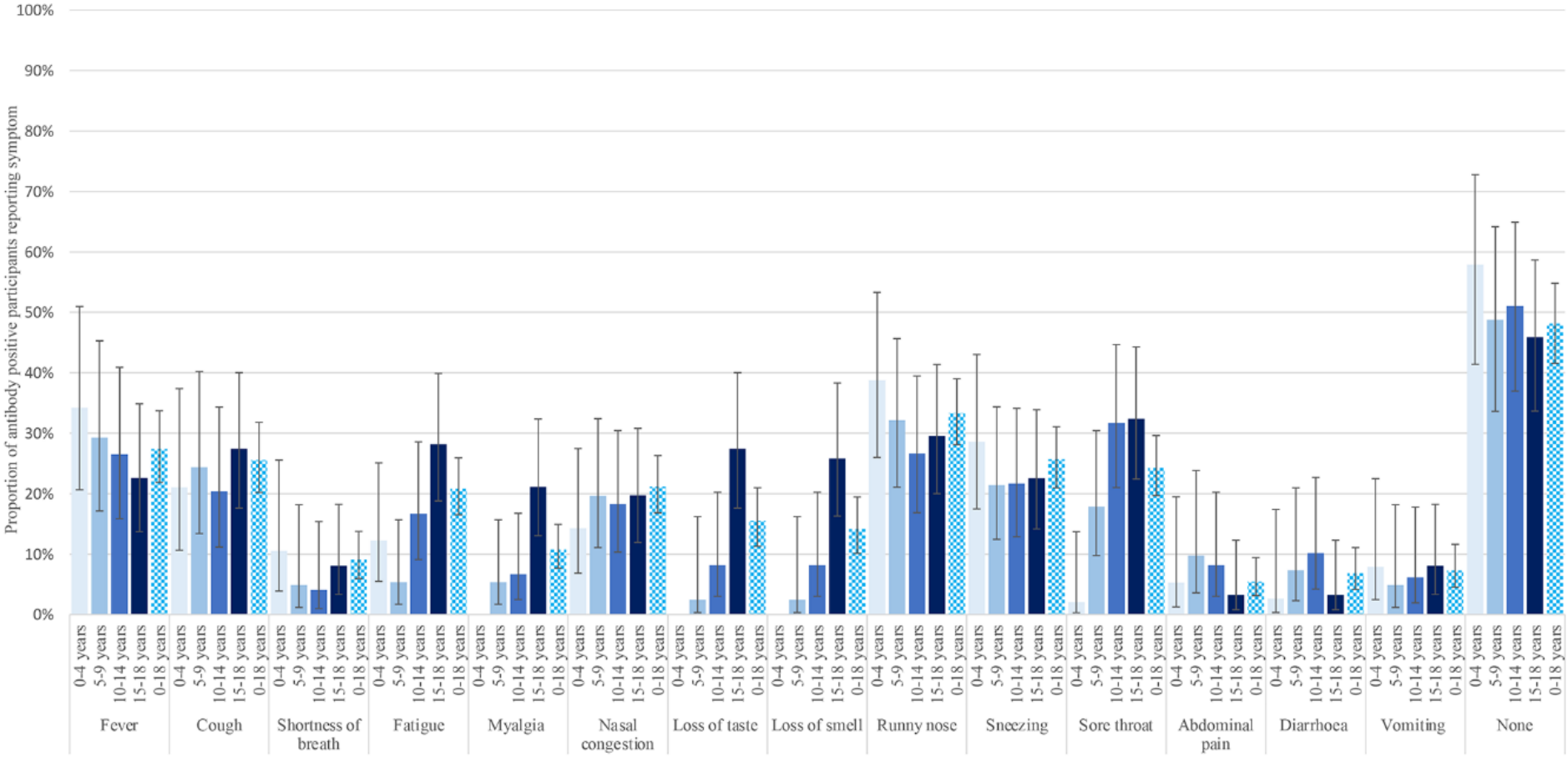


Figure 22 Summary of symptoms reported by participants seropositive on RocheS-Anti-SARS-CoV-2 IgG spike protein receptor binding domain antibodies assays by age group

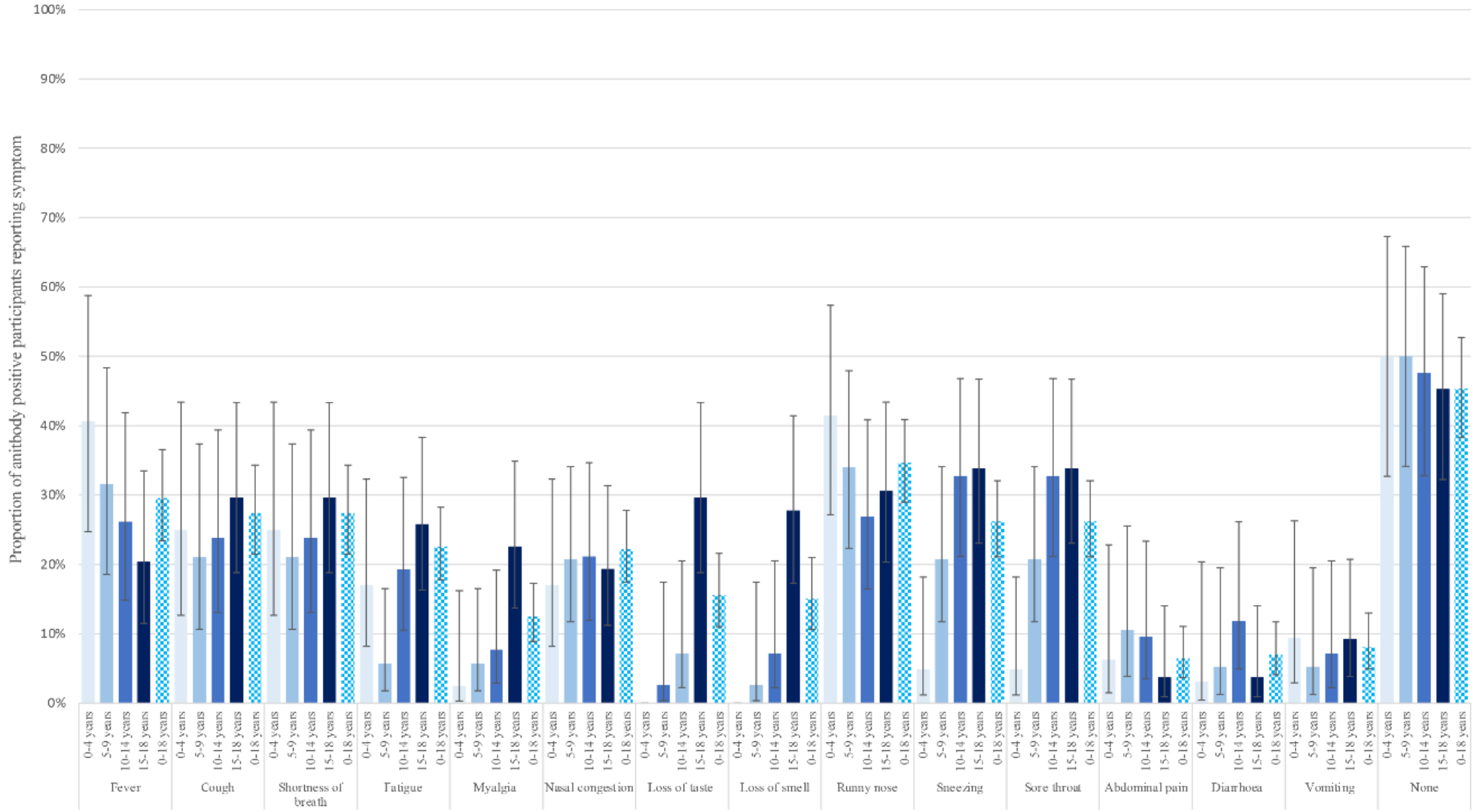


Figure 23 Summary of symptoms reported by participants seropositive on RocheN-Anti-SARS-CoV-2 IgG nucleocapsid antibodies assays by age group

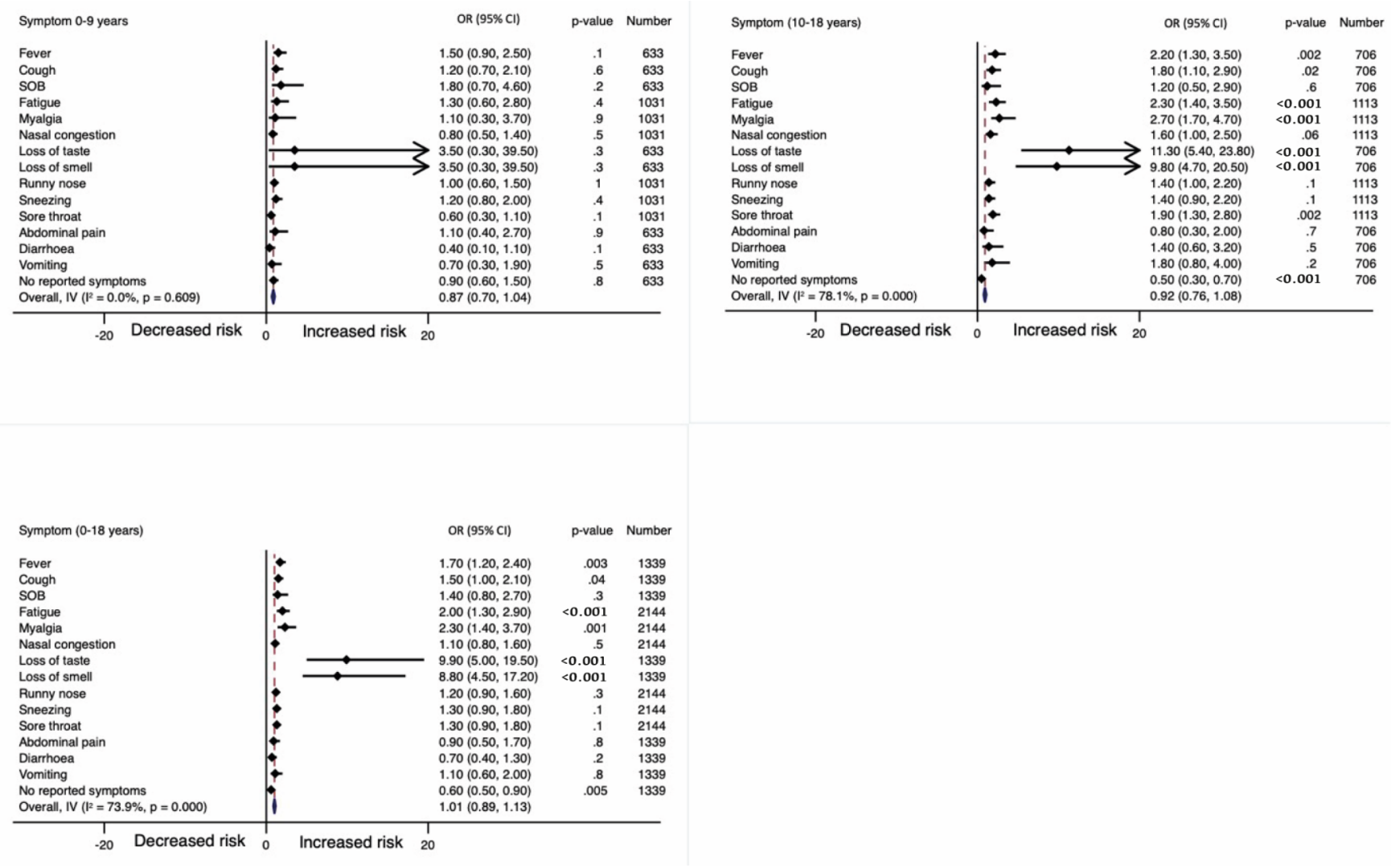
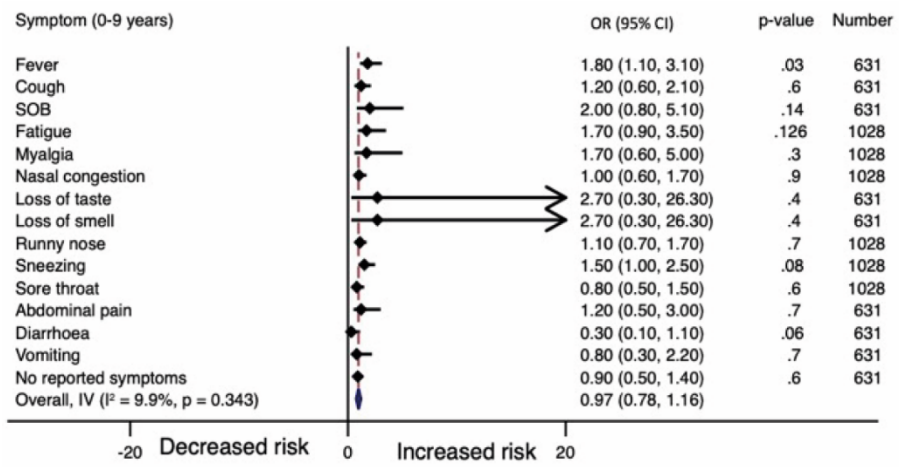


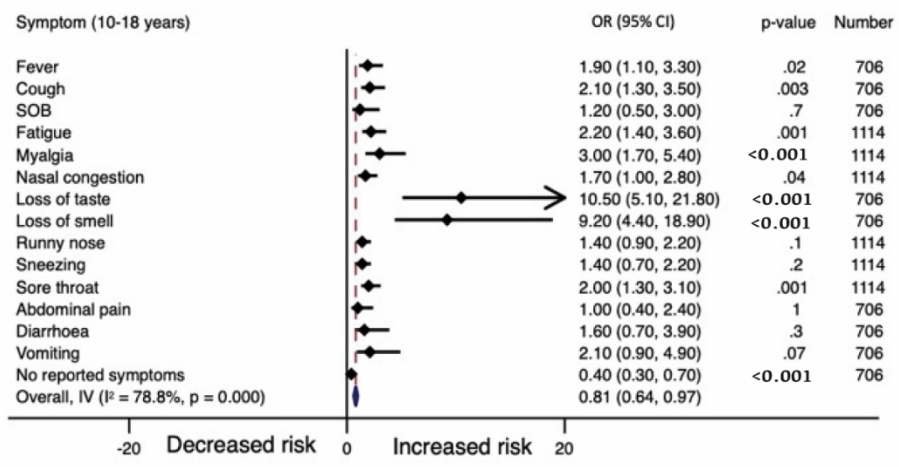
Figure 24 Univariate analysis of symptoms by age band (RocheS-Anti-SARS-CoV-2 IgG spike protein receptor binding domain antibodies)

Table 26 Logistic regression models to determine symptoms associated with positivity the RocheS-Anti-SARS-CoV-2 IgG spike protein receptor binding domain antibodies assay results. Symptoms with the highest p-value were sequentially excluded and model Akaike Information Criterion (AIC) values were compared until a model with the lowest AIC value had been reached. Separate models created for 0-9 years and 10-18 years.

RocheS		***OR (95% CI)***	***p-value***
Age group: 0-9 years (N=663 observations)			
	Male	1·0 (0·6-1·7)	0·89
	Fever	1·7 (1·0-2·9)	0·05
	Gastrointestinal symptoms	0·6 (0·3-1·1)	0·09
Age group: 10-18 years (N=873 observations)			
	Male	0·9 (0·6-1·3)	0·57
	Fever	1·6 (1·0-2·7)	0·05
	Loss of taste and smell	9·2 (5·1-16·3)	<0·001

OR: odds ratio





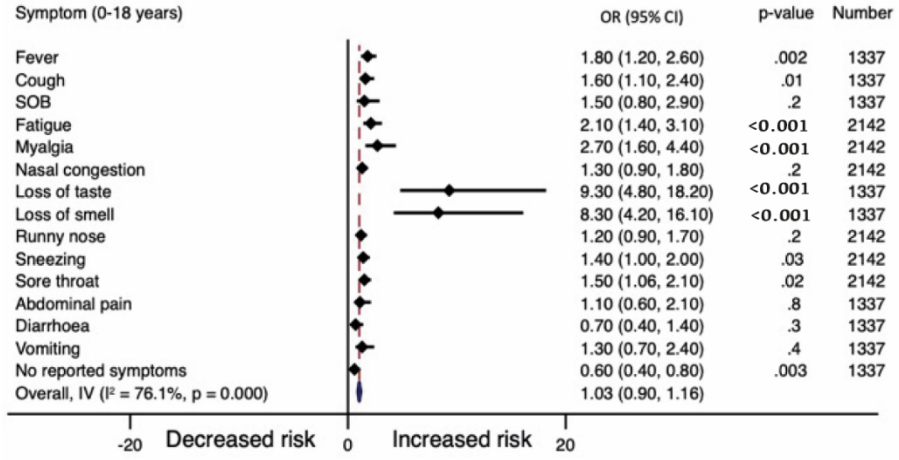


Figure 25 Univariate analysis of symptoms by age band (RocheN-Anti-SARS-CoV-2 IgG nucleocapsid antibodies)

Table 27 Logistic regression models to determine symptoms associated with positivity the RocheN-Anti-SARS-CoV-2 IgG nucleocapsid antibodies assay result. Symptoms with the highest p-value were sequentially excluded and model Akaike Information Criterion (AIC) values were compared until a model with the lowest AIC value had been reached· Separate models created for 0-9 years and 10-18 years

	RocheN OR (95% CI)	***p-value***
Age group: 0-9 years (N=631 observations)		
Male	1·1 (0·6-1·8)	0·80
Fever	2·1 (1·2-3·6)	0·009
Gastrointestinal symptoms	0·5 (0·3-1·1)	0.09
Age group: 10-18 years (N=873 observations)		
Male	0·9 (0·6-1·4)	0·65
Cough	1·6 (1·0-2·6)	0·07
Loss of taste or smell	8·0 (4·5-14·4)	<0·001

OR: odds ratio

4.5.4 Risk factor analysis

The odds of a RocheS seropositive result in different age groups compared with the 10-14 year age group was not significant (p=0·08), but in 2 out of 3 multivariable analyses age group was a significant factor (p=0.01) with higher odds of a positive result in the 15-18 years age group (OR 1·5 95% CI 1.1-2.2) (Table 28

Table 28). NHS region was a significant factor in all analyses ($p<0.001$). The North-West and London NHS regions as well as North-East and Yorkshire (in 2 out of 3 multivariable analyses) showed a higher seropositivity compared to South-East (baseline). The odds of a seropositive result was significantly higher for children living in urban areas compared with rural areas in 2 out of 3 multivariable analyses. Children in minority ethnic groups had significantly increased odds of a seropositive result compared with white children in univariate analysis ($p<0.001$), which remained significant, though less so in 2 out of 3 multivariable analyses ($p=0·04$). In the univariate analysis a significant trend towards higher odds of seropositivity in the areas with higher deprivation was seen however this was not statistically significant in the multivariable analysis when using either IDACI and IMD. Having a healthcare worker in the family increased the odds of seropositivity OR 1·6 (95% CI 1·2-2·2 $p=0·001$, multivariable analysis) but the increased odds in minority ethnic groups was no longer significant in this analysis.

Table 28 Univariate and multivariable logistic regression models to establish risk of SARS-CoV-2 seropositivity on Roche Elecsys Anti-SARS-CoV-2 serological assays for the detection of anti-SARS-CoV-2 IgG spike protein receptor binding domain antibodies (RocheS-Anti-SARS-CoV-2 IgG spike protein receptor binding domain antibodies) in children aged 0-18 years. Multivariable analysis adjusts for age, sex, NHS region, time period, ethnicity, socioeconomic deprivation (using either Index of multiple deprivation (IMD) or Income Deprivation Affecting Children Index (IDACI)), urban/rural geography and presence of healthcare workers within the household.

	Univariate		Multivariable					
			IMD* deprivation quintiles		IDACI †** deprivation quintiles		IMD and incl. HCW†,	
Number of participants in model			2287		2287		1886	
	Odds ratio (95% CI)	**LR test (p value)**	**Odds ratio (95% CI)**	**LR test‡ (p value)**	**Odds ratio (95% CI)**	**LR test (p value)**	**Odds ratio (95% CI)**	**LR test (p value)**
Age group								
0-4 years	1·1 (0·7-1·6)	0·08	0·8 (0·5-1·2)	0·01	0·8 (0·5-1·2)	0·01	0·8 (0·5-1·2)	0·05
5-9 years	1·0 (0·7-1·4)		1·0 (0·6-1·4)		1·0 (0·6-1·4)		0·9 (0·6-1·4)	
10-14 years	1 (ref)		1 (ref)		1 (ref)		1 (ref)	
15-18 years	1·5 (1·0-2·1)		1·5 (1·1-2·2)		1·5 (1·1-2·2)		1·4 (1·0-2·2)	
Sex								
Female	1 (ref)	0·7	1 (ref)	0.9	1 (ref)	1.0	1 (ref)	1.0
Male	0·9 (0·7-1·2)		1·0 (0·8-1·3)		1·0 (0·8-1·3)		1·0 (0·7-1·3)	
NHS region								
East of England	1·5 (0·7-3·0)	<0·001	1·1 (0·5-2·4)	<0·001	1·1 (0·5-2·3)	<0·001	1·0 (0·4-2·3)	<0·001
London	6·0 (3·7-9·5)		3·1 (1·9-5·2)		3·1 (1·8-5·2)		3·0 (1·7-5·3)	
Midlands	1·7 (0·9-3·2)		1·2 (0·6-2·5)		1·2 (0·6-2·4)		1·1 (0·5-2·4)	
North-East and Yorkshire	2·7 (1·6-4·3)		1·9 (1·1-3·2)		1·8 (1·1-3·1)		1·8 (1·0-3·2)	
North-West	4·7 (2·9-7·5)		2·9 (1·7-5·2)		2·8 (1·6-4·9)		2·8 (1·6-5·0)	
South-East	1 (ref)		1 (ref)		1 (ref)		1 (ref)	
South-West	1·8 (1·0-3·0)		1·6 (0·9-2·8)		1·6 (0·9-2·8)		1·6 (0·9-2·9)	
Time period								
Pre-pandemic (01 Oct 2019-31 Mar 2020)		<0·001		<0·001		<0·001		<0·001
First wave (01 Apr 2020-31 May 2020)	0·2 (0·1-0·3)		0·2 (0·1-0·4)		0·2 (0·1-0·4)			
Post first wave (01 Jun 2020-31 Aug 2020)	0·2 (0·1-0·3)		0·2 (0·1-0·3)		0·2 (0·1-0·3)		0·2 (0·1-0·3)	
Schools reopening and second wave (01 Sep 2020-31 Dec 2020)	0·4 (0·2-0·5)		0·4 (0·2-0·6)		0·4 (0·3-0·6)		0·4 (0·3-0·6)	
Post second wave (01 Jan 2021-31 Mar 2021)	1 (ref)		1 (ref)		1 (ref)		1 (ref)	
Emergence of delta variant (01 Apr 2021-30 Jun 2021)	1·2 (0·8-1·7)		1·0(0·7-1·5)		1·0 (0·7-1·6)		1·0 (0·7-1·6)	
Ethnicity								
White	1 (ref)	<0·001	1 (ref)	0·04	1 (ref)	0·04	1 (ref)	0·07
Minority ethnic group††	2·5 (1·8-3·3)		1·4 (1·0-2·0)		1·4 (1·0-2·0)		1·4 (1·0-2·0)	

IMD deprivation quintile								
Most deprived 1	2·5 (1·7-3·7)	<0·001	1·4 (0·8-2·2)	0·1			1·6 (0·9-2·6)	0·04
2	1·6 (1·0-2·5)		0·8 (0·5-1·2)				0·7 (0·4-1·2)	
3	1·1 (1·0-2·3)		1·0 (0·6-1·5)				1·1 (0·7-1·7)	
4	1·5 (1·0-2·2)		1·2 (0·8-1·9)				1·3 (0·8-2·0)	
Least deprived 5	1 (ref)		1 (ref)				1 (ref)	
IDACI deprivation quintile								
Most deprived 1	1·9 (1·3-2·8)	<0·001			1·2 (0·8-1·9)	0·1		
2	1·1 (0·7-1·6)				0·6 (0·4-1·0)			
3	1·2 (0·8-1·7)				1·0 (0·6-1·5)			
4	1·0 (0·7-1·5)				0·9 (0·6-1·4)			
Least deprived 5	1 (ref)				1 (ref)			
Urban/rural								
Rural	0·3 (0·2-0·5)	<0·001	0·6 (0·3-1·0)	0·03	0·6 (0·3-1·0)	0·03	0·6 (0·4-1·1)	0·07
Urban	1 (ref)		1 (ref)		1 (ref)		1 (ref)	
HCW †,								
No	1 (ref)	0·002					1 (ref)	
Yes	1·5 (1·2-2·0)						1·6 (1·2-2·2)	0·001

**IMD Index of Multiple Deprivation*
***IDACI Income Deprivation Affecting Children Index (IDACI) measures the proportion of all children aged 0 to 15 living in income deprived families*
†,fewer results available for this analysis due to incomplete data in HCW (healthcare worker) field (see Table 2)
††,Minority ethnic group includes all minority groups apart from white minorities.

‡Likelihood ratio for each risk factor calculated with overall p value displayed

The RocheN analysis (Table 29) showed similar results on univariate and multivariable analyses as RocheS.

Table 29 Univariate and multivariable logistic regression models to establish risk of SARS-CoV-2 seropositivity on Roche Elecsys Anti-SARS-CoV-2 serological assays for the detection of anti-SARS-CoV-2 IgG nucleocapsid antibodies (RocheN-Anti-SARS-CoV-2 IgG nucleocapsid antibodies) in children aged 0-18 years. Multivariable analysis adjusts for age, sex, NHS region, time period, ethnicity, socioeconomic deprivation (using either Index of multiple deprivation (IMD) or Income Deprivation Affecting Children Index (IDACI)), urban/rural geography and presence of healthcare workers within the household.

	Univariate		**Multivariable**					
			IMD* deprivation quintiles		**IDACI** deprivation quintiles**		**IMD and incl· HCW†**	
Number of participants in model			2285		2285		1886	
	Odds ratio (95%CI)	**LR test (p value)**	**Odds ratio (95% CI)**	**LR test‡ (p value)**	**Odds ratio (95% CI)**	**LR test (p value)**	**Odds ratio (95% CI)**	**LR test (p value)**
Age group								
0-4 years	1·0 (0·7-1·6)	0·1	0·8 (0·5-1·3)	0·02	0·8 (0·1-1·3)	0·02	0·8 (0·5-1·3)	0·07
5-9 years	1·1 (0·7-1·6)		1·1 (0·7-1·7)		1·1 (0·7-1·7)		1·1 (0·7-1·6)	
10-14 years	1 (ref)		1 (ref)		1 (ref)		1 (ref)	
15-18 years	1·5 (1·1-2·2)		1·6 (1·1-2·4)		1·6 (1·1-2·4)		1·5 (1·0-2·2)	
Sex								
Female	1 (ref)	0.7	1 (ref)	0.9	1 (ref)	0.9	1 (ref)	0.9
Male	0·9 (0·7-1·2)		1·0 (0·7-1·3)		1·0 (0·7-1·3)		1·0 (0·7-1·3)	
NHS region								

East of England	1·1 (0·5-2·6)	<0·001	0·8 (0·3-2·0)	<0·001	0·8 (0·3-2·0)	0·001	0·9 (0·3-2·2)	0·001
London	5·5 (3·3-8·9)		2·7 (1·6-4·7)		2·6 (1·5-4·4)		2·7 (1·5-4·8)	
Midlands	1·7 (0·8-3·2)		1·3 (0·6-2·6)		1·2 (0·6-2·5)		1·2 (0·6-2·6)	
North-East and Yorkshire	2·3 (1·3-3·8)		1·5 (0·9-2·7)		1·5 (0·8-2·6)		1·5 (0·8-2·7)	
North-West	4·5 (2·8-7·4)		2·7 (1·5-4·9)		2·6 (1·5-4·6)		2·7 (1·4-4·9)	
South-East	1 (ref)		1 (ref)		1 (ref)		1 (ref)	
South-West	1·7 (0·9-3·0)		1·5 (0·8-2·7)		1·5 (0·8-2·7)		1·5 (0·8-2·7)	
Time period								
Pre-pandemic (01 Oct 2019-31 Mar 2020)		<0·001		<0·001		<0·001		<0·001
First wave (01 Apr 2020-31 May 2020)	0·2 (0·1-0·4)		0·2 (0·1-0·5)		0·2 (0·1-0·5)			
Post first wave (01 Jun 2020-31 Aug 2020)	0·2 (0·1-0·4)	<0·001	0·2 (0·1-0·3)		0·2 (0·1-0·3)		0·2 (0·1-0·3)	
Schools reopening and second wave (01 Sep 2020-31 Dec 2020)	0·4 (0·2-0·6)	<0·001	0·4 (0·3-0·6)		0·4 (0·2-0·6)		0·4 (0·3-0·6)	
Post second wave (01 Jan 2021-31 Mar 2021)	1 (ref)		1 (ref)		1 (ref)		1 (ref)	
Emergence of delta variant (01 Apr 2021-30 Jun 2021)	1·2 (0·8-1·7)	0·3	1·0 (0·7-1·6)		1·0 (0·7-1·6)		1·0 (0·7-1·6)	
Ethnicity								
White	1 (ref)	<0·001	1 (ref)	0·04	1 (ref)	0.04	1 (ref)	0.06
Minority ethnic group ††	2·5 (1·9-3·5)		1·5 (1·0-2·1)		1·5 (1·0-2·1)		1·4 (1·0-2·1)	
IMD deprivation quintile								
Most deprived 1	2·8 (1·8-4·2)	0·002	1·5 (0·9-2·4)	0·02			1·7 (1·0-2·9)	0·009
2	1·4 (0·9-2·3)		0·7 (0·4-1·2)				0·7 (0·4-1·2)	
3	1·5 (1·0-2·3)		0·9 (0·6-1·5)				1·0 (0·6-1·7)	
4	1·7 (1·1-2·5)		1·4 (0·9-2·2)				1·5 (0·9-2·4)	
Least deprived 5	1 (ref)		1 (ref)				1 (ref)	
IDACI deprivation quintile								
Most deprived 1	2·0 (1·3-2·9)	<0·001			1·2 (0·7-1·8)	0·04		
2	1·0 (0·6-1·6)				0·6 (0·4-1·0)			
3	0·8 (0·5-1·3)				0·6 (0·4-1·0)			
4	1·0 (0·7-1·5)				1·0 (0·6-1·5)			
Least deprived 5	1 (ref)				1 (ref)			
Urban/rural								
Rural	0·2 (0·1-0·4)	<0·001	0·4 (0·2-0·8)	0·002	0·4 (0·2-0·7)	<0·001	0·5 (0·2-0·9)	0·009
Urban	1 (ref)		1 (ref)		1 (ref)		1 (ref)	
HCW †								
No	1 (ref)	0·003					1 (ref)	<0·001
Yes	1·5 (1·1-2·0)						1·7 (1·3-2·4)	

**IMD Index of Multiple Deprivation*

***IDACI Income Deprivation Affecting Children Index (IDACI) measures the proportion of all children aged 0 to 15 living in income deprived families*

†,fewer results available for this analysis due to incomplete data in HCW (healthcare worker) field

††,Minority ethnic group includes all minority groups apart from white minorities·

‡Likelihood ratio for each risk factor calculated with overall p value displayed

4.6 Discussion

Here, we report results of a community-based study recruiting a representative cohort of the paediatric population in England from the start of the SARS-CoV-2 pandemic, with extensive characterisation of demography and potential COVID-19 symptoms. This provides a unique assessment of the prevalence and risk factors for naturally acquired antibodies against the SARS-CoV-2 virus in children and adolescents prior to the introduction of widescale immunisation. Notably, approximately one third of adolescents aged 15-18 years had antibodies against SARS-CoV-2 in April-June 2021, prior to immunisation or the further wave of infections in this age group in Autumn 2021.

This study clarifies the evidence base regarding paediatric SARS-CoV-2 seroprevalence, building on other studies in England and the United Kingdom, including SKIDS,[39,188] (recruiting from educational settings) and COVID Warriors[24] (recruiting children of health care workers). Together these seroprevalence studies provide estimates of the levels of seropositivity within the paediatric population, vital for modelling of disease susceptibility and immunisation planning, complementing insights gained from repeat cross-sectional community infection surveys such as REACT-1[187] and the COVID infection survey.[192]

Differences in estimates emerge when comparing studies at similar timepoints. In April-August 2020, overall rates of seropositivity in our data were approximately half those of COVID Warriors,[24] likely due to the elevated risk of seropositivity in household members of healthcare workers. In June 2021 our age-adjusted estimates were higher than the SKIDS study, which reported 11·25% and 12·95% of primary and secondary school pupils being seropositive using oral fluid sampling,[193] compared with 18·9% and 32·7% in similar age groups in our study. This may reflect differences in sensitivity in antibody detection in saliva (75%) as used in the SKIDS study vs serum (95·5%) used in our study.[132,137] Lastly, ONS data in summer 2021

showed a sero-prevalence of 47% in young adults aged 16-24 years[192] compared with 32·7% in adolescents aged 15-18 years in our study, a variance potentially accounted for by higher infection rates observed in young adults aged 18-24 years in the period,[194] along with the possibility of vaccine-induced immunity in this older age group in the ONS data.

More consistency is seen when comparing trends over time. The dramatic increase in sero-positivity evident in September-December 2020 in adolescents aged 15-18 years, followed by all younger age groups (including children aged 0-4 years) in January 2021-March 2021 is consistent with antibody data from SKIDS showing an increase in infection rates in adolescents aged 11-18 years in September-December 2020.[188] A further increase in seropositivity in adolescents aged 15-18 years in April-June 2021 corresponds to the emergence of the Delta variant, and associated increase in adolescent (aged 13 to 17 years) infections demonstrated in REACT-1 antigen data in May-July 2021.[187] COVID-19 seroprevalence varied across regions, with an early (June-August 2020) increase in London, again consistent with 'COVID Warriors' and REACT studies,[24] while the striking increase in seropositivity rates in the North-West from May-June 2021 is reflected in regional data from the COVID-19 infection survey.

Comparing seroprevalence estimates for children and adolescents from 70 countries around the world during a similar period showed considerable variation.[195] There were a number of factors to consider when comparing estimates, including the intensity of the pandemic, the availability of testing, public health measures, vaccination policy and implementation. Overall, studies suggested that seroprevalence rates were lower in children particularly in those younger than 10 years of age.[195] Patterns did emerge from the global data, seroprevalence was higher in older children, children of higher deprivation and those belonging to minority ethnic backgrounds.[195] Adult data from a similar time period showed a global pooled seroprevalence of 9.5%, with the highest seroprevalence recorded in Central and Southern Asia (23%) and the lowest reported in Eastern and South-Eastern Asia (1.62%). The highest seroprevalence estimates were in

persons between the ages of 20-50 years, minority ethnic groups living in countries or regions of deprivation.[196]

Our data showed a greater than two-fold increased odds of seropositivity for participants from a minority ethnic group compared with white participants, which reduced to around 1.5 in the multivariable analysis (including adjustment for socioeconomic status). A limitation of this analysis was it did not include household size as a risk factor however our findingswere consistent with the increased risk of SARS-CoV-2 infection from minority ethnic groups reported in the adult and paediatric populations since the first peak of the pandemic in England. Ward et al. showed a threefold increase in risk of being antibody positive in the adult Black population in England.[44] This trend was also reported by Ladhani et al in primary school children in June 2020.[197] Our data was not powered to look at individual ethnic groups nevertheless, we show that belonging to a minority ethnic group is a significant risk factor in the paediatric population, independent of socio-economic status and adult-type comorbidities. This increased susceptibility requires further study, it may be confounded by the presence of healthcare workers within the household (ethnicity was not significant in the multivariable analysis that included healthcare worker as a factor, though this analysis was based on smaller numbers) or could be social inequalities beyond what was measured by IMD. IMD is not a direct measure of deprivation and overlooks indicators such as household overcrowding, employment status, education and health and disability all of which could lead to the increase in risk.

Unlike COVID Warriors, we did not demonstrate that gastrointestinal symptoms were predictive of seropositivity.[24] This may reflect a symptomatology specific to ancestral lineages A and B which were in circulation when 'COVID Warriors' collected their initial data,[190] or differing methods of data collection. Of note is that in our study symptomatology questionnaires were adapted as the study progressed, with some symptoms collected

retrospectively creating the potential of recall bias. Furthermore, while the age discrepancy in reporting of anosmia and ageusia are striking, this may reflect difficulties in children <10 years communicating these symptoms.

The strengths of this study are that it includes children and adolescents aged 0-18 years and has sampled from all NHS regions in England which allows generalisations to be made to England as a whole. To our knowledge at the time of publication this is the only paediatric study in England that has a community-based recruitment strategy with efforts to recruit a representative sample of the region. The study has several limitations, including having been adapted from a pre-pandemic design. One-quarter of participants were children of healthcare workers which is higher than the general population.[198] Both the RocheS and RocheN assays have primarily been validated in the adult, rather than paediatric, population and there is the possibility that the SARS-CoV-2 nucleocapsid-specific IgG response to infection, and the longevity of that response, may differ in younger age groups.[199]

This study has provided a gold-standard SARS-CoV-2 seroprevalence dataset as part of a unique biobank of serum samples from children and young adults aged 0-24 years with a comprehensive history of immunisation and demography. The willingness of participants and their families to participate in research has created an invaluable resource for understanding COVID-19 and other infectious diseases.

Chapter 5 Persistence of antibody concentrations against SARS-CoV-2 over time

5.1 Introduction

Antibody persistence following SARS-CoV-2 infection is key to understanding potential population immunity.[200] At the time 'STORY' was collecting samples (June 2021), most available data had been derived from severe or hospitalised cases, or from population serosurveys, where most participants are adults, or from vaccination studies. Children and adolescents rarely suffer from severe disease and there are fewer serosurveys focusing on children rather than adults. One study comparing adult and paediatric antibody responses to SARS-CoV-2 infection reported that children predominantly generate IgG responses to spike protein rather than to nucleocapsid.[201] Looking at the persistence of antibody concentrations over time may give insights into the potential for population immunity, which is vital when developing immunisation strategies. This chapter models the persistence of SARS-CoV-2 antibodies in serum and oral fluid in children and adolescents andis based on the submission Ratcliffe H, Tiley KS, Longet S et al. Serum HCoV-spike specific antibodies do not protect against subsequent SARS-CoV-2 infection in children and adolescents.

5.2 Methods

Two or more time-points were available for 1,350 participants across 'COVID Warriors' and 'STORY' studies. The assay used to test the highest number of samples was RocheN with 1,263 participants having had two or more time-points tested on the assay (Table 30). The average time between visits for 'COVID warriors' was 57 days (range 43-193, median 98 and mode 63) and 109 days for 'STORY' (range 17-411, median 97 and mode 63). Within the longitudinal cohort of 'STORY' 83 (16.5%) identified as belonging to a minority ethnic group. 'COVID warriors' did not collect information on ethnicity. Seropositive participants from 'COVID Warriors' (participants with a positive test on either DiaSorin or RocheN

assays) were retested with RocheS to allow the ‘STORY’ and ‘COVID Warrior’ datasets to be combined when modelling antibody decline of spike protein receptor binding domain (RocheS) and nucleocapsid (RocheN) antibodies.

Table 30 Number of repeat samples by age group and assay deployed

Age	COVID Warriors		STORY		
	2 timepoints only	3 timepoints only	2 timepoints only	3 timepoints only	4 timepoints only
0-4 years	82	58	90	37	14
5-9 years	314	251	119	56	28
10-14 years	370	285	161	84	31
15-19 years	82	68	112	61	31
20-24 years			20	7	2
Sex (Female)	416	390	248	122	106
RocheN	790	661	473	219	87
RocheS	66	68	473	218	87
RBD			351	151	24
Diasorin	871	693			

Concentrations of SARS-CoV-2 IgG using the RocheS and RocheN assays were plotted against time with up to four observations per participant. Time from first positive antibody result and highest positive antibody concentration were explored separately. Participants were grouped by age (0-4 years, 5-9 years, 10-14 years, 15-24 years).

A mixed-effects linear regression analysis was carried out with log-transformed serum antibody concentration as the outcome, fixed effects for time from highest positive result as a continuous variable, age group and study, and random intercepts for each individual participant. Different models were fitted and the linear model which had the lowest Akaike Information Criterion (AIC) value was selected. This was demonstrated graphically by plotting adjusted means for antibody concentration over the time course of the study with day 0 defined as the highest positive result. The highest positive results were modelled to give a conservative estimate of the persistence of the immune response.

Oral fluid (OF) samples were collected using Oracol[138] swabs from participants returning for repeat visits within the 'STORY' study. OF contains saliva and gingiva-crevicular fluids which

has transudate from serum. This results in IgG and IgM concentrations at approximately 1/800 and 1/400, respectively, of that which is found in serum.[137,202] These were analysed using a SARS-CoV-2 IgG Enzyme immunoassays (EIA) developed by UKHSA, using SARS-CoV-2 viral nucleoprotein (NP) and spike (S) proteins in IgG isotype capture format.[114,202]

For the OF data, a mixed-effects linear regression analysis was carried out with OF antibody concentration as the outcome, fixed effects for time from highest positive result as a continuous variable, age group and random intercepts for each individual participant using the same methodology used to assess antibody decline in serum. This was demonstrated graphically by plotting adjusted means for antibody concentration at fixed values of days with day 0 defined as the highest positive result.

5.3 Results

5.3.1 Children are predicted to maintain a SARS-CoV-2 stable immune response to spike protein receptor binding domain.

Figure 26a-b shows serum antibody titres over time using the RocheS and RocheN assays respectively. Time-point 0 was defined as the participants' first positive result using manufacturers' definitions (Table 9). The spike protein and receptor binding domain assay and nucleocapsid assay show a similar trend over all age groups, with IgG subsequently falling below the threshold of detection within the study period for 8/175 (4.5%) participants with a positive spike (Figure 26a) compared with 6/181 (3.3%) participants when measured on the nucleocapsid assay (Figure 26b).

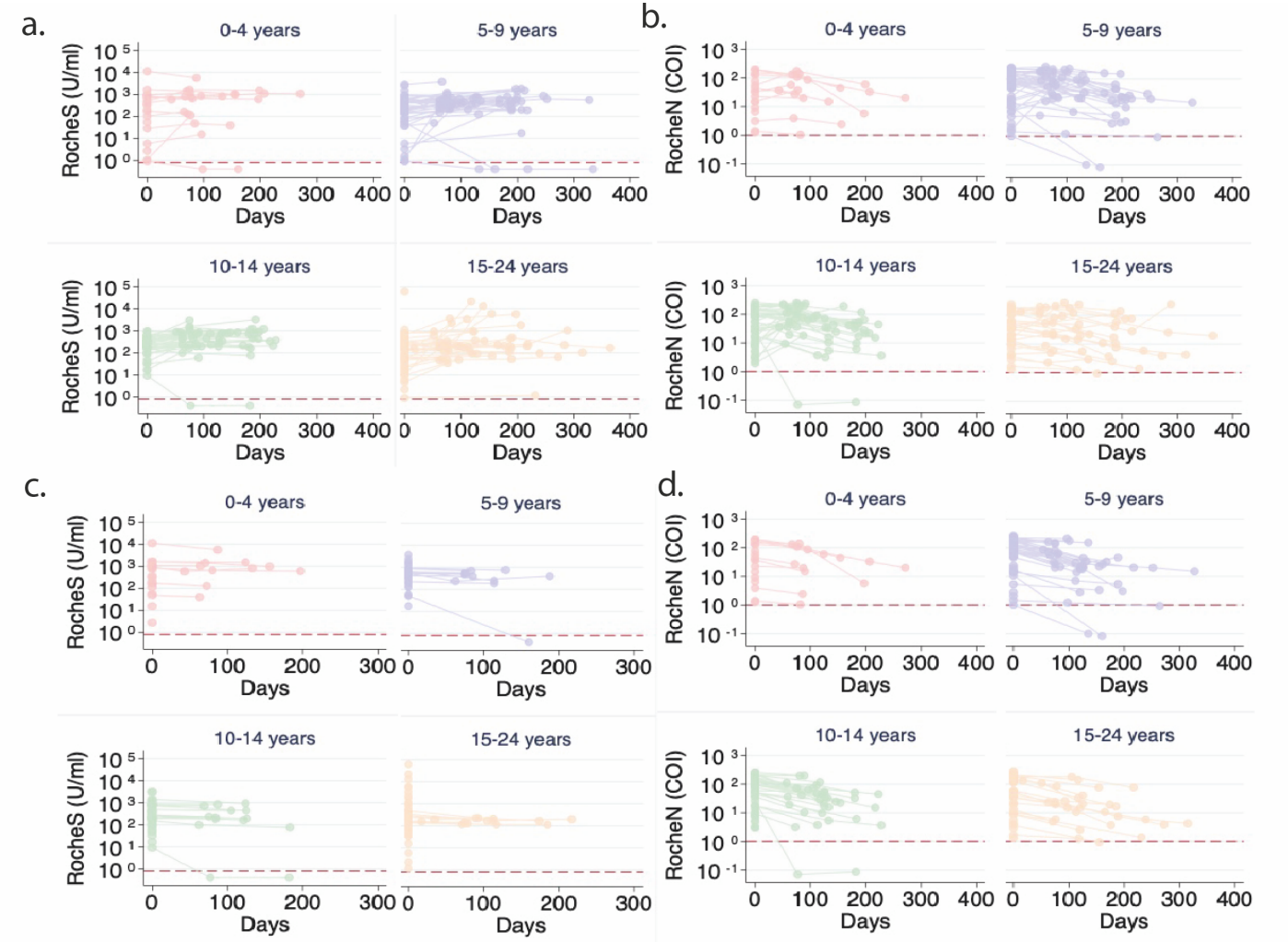


Figure 26 shows results of serological testing for a.Roche Elecsys® Anti-SARS-CoV-2 IgG spike receptor binding domain antibody concentration (U/ml) over time. b. Roche Elecsys® Anti-SARS-CoV-2 IgG nucleocapsid antibody concentration (COI)over time. Time-point 0 is defined as the participant's first positive result (a & b). c. Roche Elecsys® Anti-SARS-CoV-2 IgG spike receptor binding domain antibody concentration (U/ml) over time d. Roche Elecsys® Anti-SARS-CoV-2 IgG nucleocapsid antibody concentration (COI) over time. Time-point 0 is defined as the participant's highest positive result (c & d). Figures a-d are broken down into age groups 0-4 (red), 5-9 (purple),10-14 (green) and 15-24 (orange). The threshold of positivity is shown by the red line on each plot.

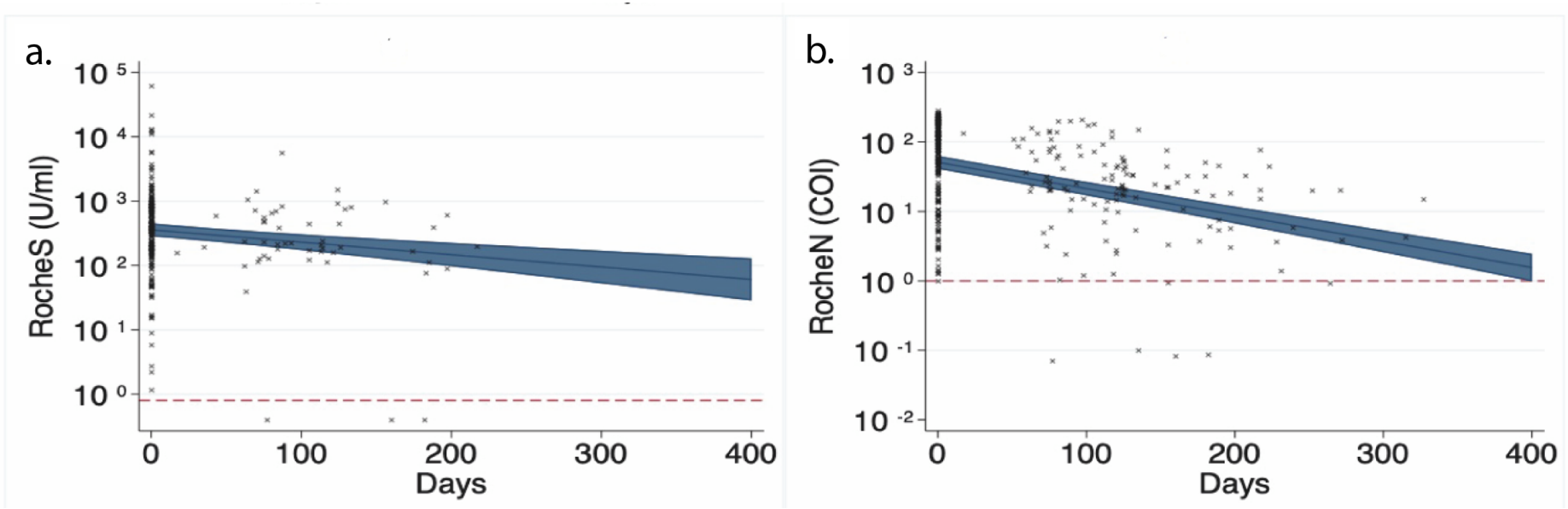


Figure 27a-b Mixed regression model predicting the adjusted means with time point 0 being the highest positive result with individual participant's serum antibody concentrations over time marked with (x) for Roche Elecsys® Anti-SARS-CoV-2 IgG spike antibodies receptor binding domain (U/ml) (a) and Roche Elecsys® Anti-SARS-CoV-2 IgG nucleocapsid antibodies (COI) (b). 95% confidence intervals are shown. The threshold of positivity is indicated by the red line on each plot.

Plotting from the highest positive result for RocheS anti-spike receptor binding domain IgG assay (Figure 26c) shows results from 181 participants with one or more samples taken at times ranging between 0-217 days from the highest antibody value (median 0 days, mean, 27 days). The RocheN anti-nucleocapsid IgG assay (Figure 26Figure 27d) model shows 175 participants with one or more samples taken at times ranging between 0-327 days (median 75 days, mean 80 days) after their highest antibody result.

Figure 27a and b demonstrate antibody decline graphically by plotting adjusted means for antibody concentration over time with day 0 as the highest positive result for RocheS and RocheN respectively. The half-life was 160 days (95% CI 79-241 days) for anti-spike receptor binding domain antibodies and 82 days (95% CI 57-107 days) for anti-nucleocapsid antibodies. Serum anti-nucleocapsid antibodies sero-revert more rapidly than anti-spike receptor binding domain antibodies (Figure 27a-b and Table 31).

Table 31 A comparison of the predicted antibody decline of anti-spike and anti-nucleocapsid IgG concentrations in serum and oral fluid.

		Spike receptor binding domain		**Nucleocapsid**	
		Days (95% CI)	IgG concentration (AU/ml)	Days (95% CI)	IgG concentration (COI)
Serum	Time point 0	0	354.8	0	50.7
	Half life	160 (79-241)	176.2	82 (57-107)	25.1
	Time to sero-revert from positive to negative	1320 (786-1872)	0.8	450 (392-508)	1.0
		Spike		Nucleocapsid	
		Days	IgG concentration (COI)	Days	IgG concentration (COI)
Oral fluid	Time point 0	0	4.1	0	2.9
	Half life	161 (84-240)	2.0	162 (109-213)	1.5
	Time to sero-revert from positive to negative	323 (178-463)	1.0	250 (184-317)	1.0

5.3.2 Oral fluid concentrations of IgG wane more quickly than serum antibodies within the paediatric population

Serum and oral fluid IgG concentrations collected from the same participants at the same time point were concordant for 623/658 (94.6%) anti-spike and 610/662 (92.1%) anti-nucleocapsid results (Figure 28).The number of serum-positive results with a negative oral fluid result was 34/658 (5.2%) and 50/662 (7.6%) for anti-spike and anti-nucleocapsid IgG respectively (Figure 28), while only 1 (0.2%) and 2 (0.3%) oral fluid results were positive with negative serum results for anti-spike and anti-nucleocapsid IgG (respectively) (Figure 28).

Figure 29a- b shows the change in concentrations over time for oral fluid anti-spike IgG and anti-nucleocapsid IgG from the first positive result. When tested on the oral anti-spike IgG assay 4/26 (15%) of participants who provided at least one result after their initial positive result had concentrations falling below the threshold of detection within the study period compared with 8/20 (40%) falling below the threshold for the oral anti-nucleocapsid IgG assay (Figure 29b).

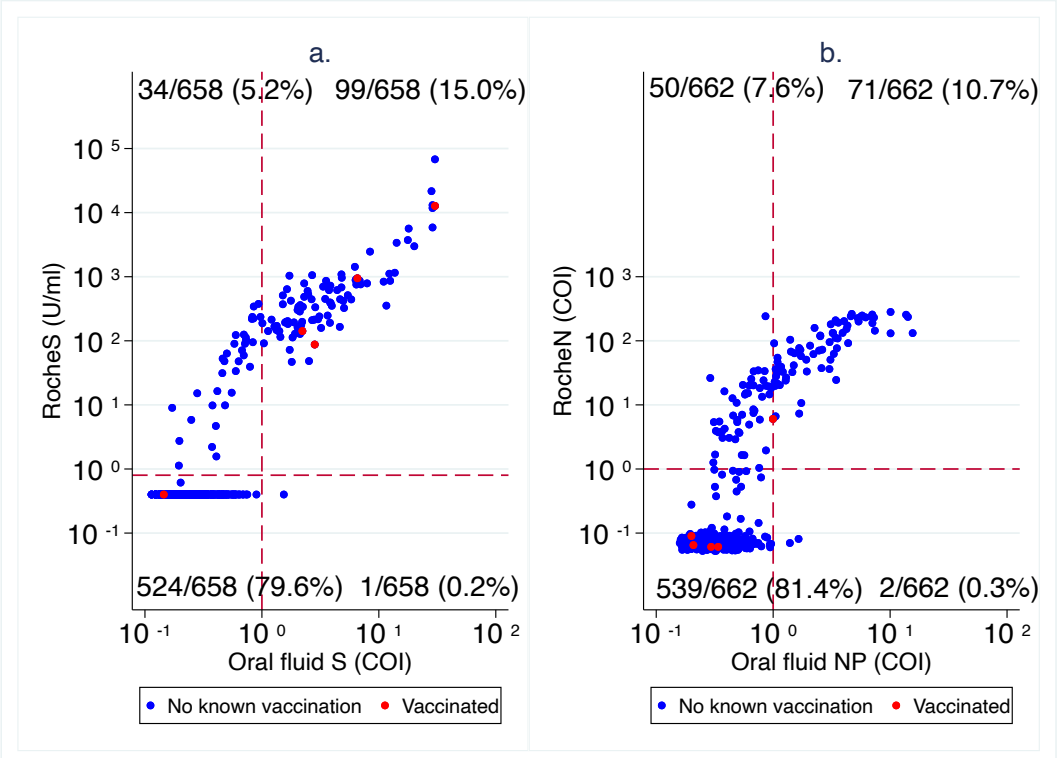


Figure 28 shows matched serum and oral fluid samples for IgG concentration against a. SARS-CoV-2 viral Spike protein in spike protein receptor domain in serum (U/ml) and oral fluid IgG Spike protein (COI) b. SARS-CoV-2 viral nucleocapsid (COI) in serum and nucleopcapsid (COI) in oral fluid (NP). Red markers indicate a sample from a participant who was immunised at the time of sampling. Red dashed lines show the threshold of positivity for each assay.

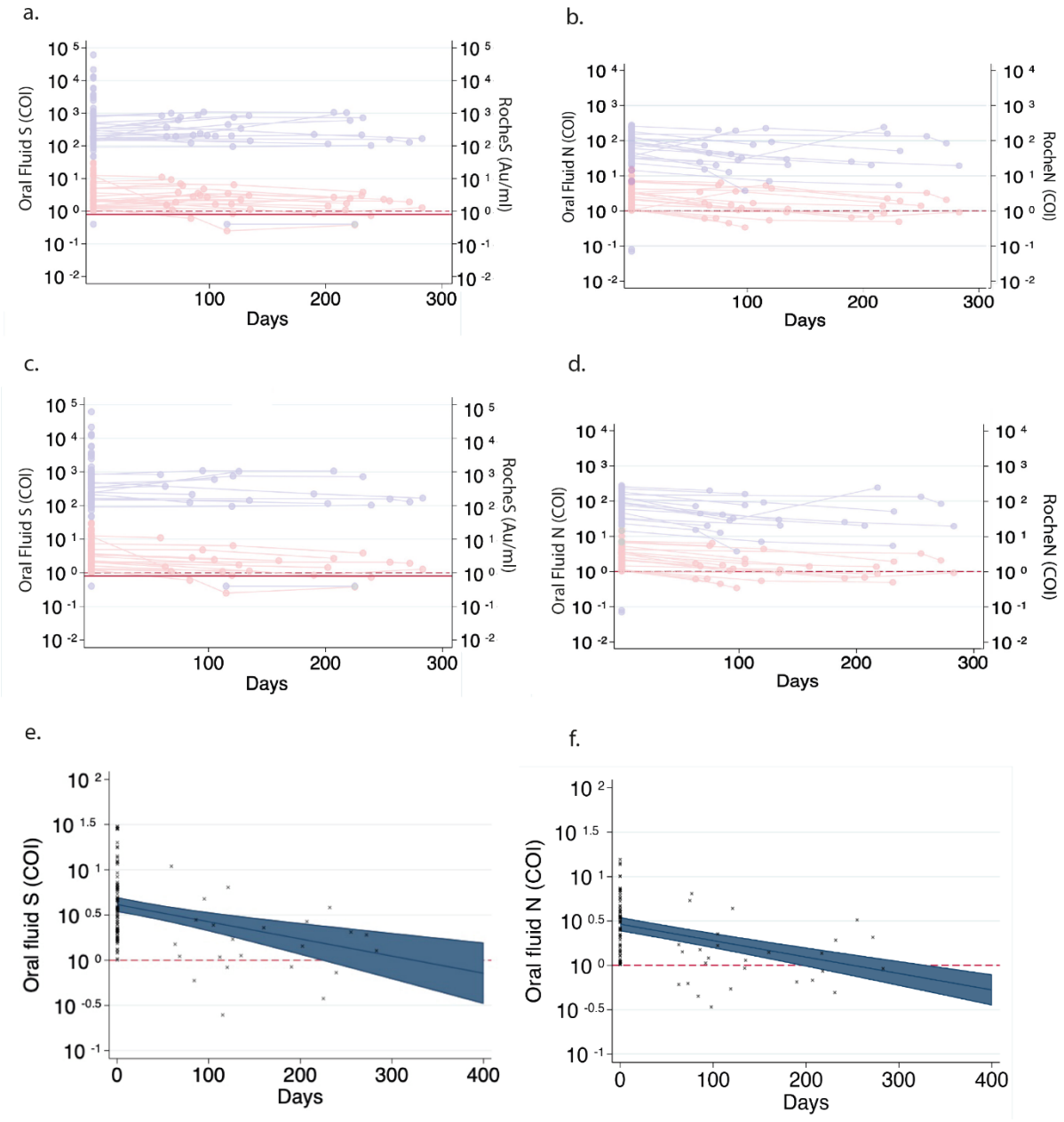


Figure 29 a-d shows both oral fluid (red) and corresponding serum samples where available (blue). a-b demonstrate IgG concentration over time where timepoint 0 is defined as the first positive oral fluid result for oral fluid spike protein (COI) and serum spike protein receptor binding domain (Roche S Au/ml) (a) and oral fluid nucleocapsid (COI) and serum nucleocapsid (Roche N COI) (b). c-d define time-point 0 as the highest positive oral fluid result for oral fluid spike protein (COI) and serum spike protein receptor binding domain (Roche S Au/ml) (c) and oral fluid nucleocapsid (COI) and serum nucleocapsid (Roche N COI) (d). The threshold of positivity is marked by the dashed line (OF) and solid line (serum). These thresholds overlap on nucleocapsid plots. e-f shows a mixed regression model for oral fluid predicting the adjustment means with time point 0 being highest positive result with individual participants antibody concentrations over time marked with (x) for anti-spike protein (COI) (e) and anti-nucleocapsid antibodies (COI) (f). 95% confidence intervals shown. The threshold of positivity indicated by the red line on each plot

Time-point 0 was defined as the highest positive result when calculating adjusted means for antibody titre. The model of oral fluid anti-spike IgG assay contained data from 85 participants with one or more data points taken at times ranging between 0-283 days (median 0 days mean 25 days). The oral fluid anti-nucleocapsid IgG assay model contained data from 64 participants with one or more data points taken at times ranging between 0-283 days (median 0 days, mean 44 days) after their highest antibody result (Figure 29c-d). The models demonstrate that anti-nucleocapsid antibodies in oral fluid are predicted to cross below the threshold of positivity more quickly than anti-spike antibodies but unlike for serum IgG, this difference is not significant (Figure 29e-f and Table 9 **Error! Reference source not found.**).

5.4 Discussion

The 'COVID Warrior' and 'STORY' studies, conducted in a vaccine-naïve population between the beginning of the pandemic and June 2021 show infected children and teenagers were predicted to have minimal waning from the highest anti-spike receptor binding domain serum IgG antibody concentration (half-life of five months) in the absence of vaccine boosting.

While we have demonstrated minimal waning in serum anti-spike receptor binding domain antibodies against SARS-CoV-2 we know this is not protective against breakthrough infections. Vaccinated individuals and those that have previously had COVID-19 are able to be re-infected with the virus however the majority of these individuals have milder symptoms and are protected from more serious illness.[203] It is difficult to separate whether breakthrough infections are due to waning immunity or the emergence of new variants of concern. As of August 2022 the UKHSA had reported over a million reinfections.[203] Understanding antibody responses in terms of magnitude and rate of decline following natural infection is important when considering population immunity and in turn public health policy and vaccine booster campaigns.[204] Dowell et al. showed seropositive children and adults had broadly similar antibody responses against viral proteins. They also demonstrated a trend towards a higher

geometric mean in children when compared with adults, most notably against nucleocapsid and RBD domains.[205]

Our data show that children who are antibody positive for SARS-CoV-2 have evidence of an immune response (IgG antibody concentration in serum) after the highest result for up to 217 days for spike protein receptor binding domain and up to 327 days for nucleocapsid. In contrast to Weisberg et al., our data showed no difference in the number of seropositive children when comparing those seropositive for anti-spike and anti-nucleocapsid.[201] The spike protein receptor binding domain was predicted to cross the threshold of positivity at 1320 days compared with 450 days for nucleocapsid. Nucleocapsid antibodies waning more quickly than spike antibodies has been widely reported in the literature.[205-207] Modelling from the highest positive result rather than the first positive result was intended to give a conservative estimate of the persistence of the immune response independent of reinfections. A limitation of our model is that the number of participants who had multiple data points after their highest positive result was lower for RocheS than RocheN. For both assays many participants had their highest result at their penultimate or last visit therefore there were fewer participants to inform the model as time increased. It is also difficult to rule out whether reinfections occurred between visits. In addition, neither 'COVID Warriors' nor 'STORY' tested for neutralising antibodies and no data was collected on the severity of the participant's illness.

Multiple studies in adults have shown the persistence of antibodies to SARS-CoV-2 infection for up to 18 months regardless of infection severity.[71,205,206] Dowell et.al. assessed the longevity of immune responses for 35 children at six months. All had retained humoral immunity and maintained higher antibody concentrations for anti-spike protein and anti-RBD than their adult group.[205] By 12 months Dowell et al. showed that 16 children had antibody responses to spike protein and RBD which were similar but slightly reduced to those levels seen at 12 months.[205] Studies looking at antibody concentrations over time are further complicated by the emergence

of different SARS-CoV-2 variants. Reinfections are widely reported with different immune responses dependent on strain[208] 'COVID Warriors' and 'STORY' recruited over time periods where ancestorial lineages A and B, alpha, beta, gamma and delta were circulating within the UK.[209]

In keeping with other studies the oral fluid data shows a trend where nucleocapsid antibodies fall below the threshold of positivity more quickly than spike protein[210,211] after their highest positive result however this was not statistically significant as the confidence intervals were overlapping. Although the numbers were small the data also indicated that oral fluid antibodies cross the threshold of positivity more quickly than serum. This is likely a reflection that the oral fluid contained gingiva-crevicular fluids, which has transudate from serum, being tested. Oral fluid testing is useful in large-scale population studies in paediatrics because it is more acceptable to parents and young children than a blood test but its limitations should be considered. To stimulate the crevicular fluid the participant needs to swab the gum line for two minutes.[212] This can be extremely challenging in the preschool age group. Results can also be affected by the amount of saliva produced which in turn can be affected by the time of their last meal.[114] Hoschler et al. found that the sensitivity of the spike and nucleocapsid assay in children was 78% and 80% respectively however in adults this fell to 58% and 67% respectively.[137] The specificity was calculated on whether the participant had been seropositive for SARS-CoV-2 on a serum nucleocapsid assay rather than Polymerase Chain Reaction (PCR) swabs which was the gold standard due to low rates of PCR testing at the time of sampling. This could account for lower sensitivity in adults if this waned more quickly than in children as suggested by Dowell et al.[205] Hoschler et al. reported a high specificity (99%) in both children and adults. Oral fluid samples are useful when investigating levels of circulating SARS-CoV-2 which could be measured at specific time points using the nucleocapsid assay within the paediatric population, however it appears to be less sensitive in adult populations. It

has the added benefit that it can be performed alongside self-swabbing which would also inform on strains circulating at the time.

Chapter 6 Serum HCoV-spike specific antibodies do not protect against subsequent SARS-CoV-2 infection in children and adolescents.

6.1 Introduction

Since the emergence of SARS-CoV-2, it has been noted that infections experienced by children were generally asymptomatic and rarely progressed to severe disease and hospitalisation.[120] Understanding why children are less severely affected is important to further our knowledge of COVID-19 pathogenesis. Hypotheses to date have included differences in humoral immune responses to SARS-CoV-2 infections in adults and children or differences in baseline seasonal coronaviruses' humoral immunity.[118] Alternate theories suggest differences in cellular immunity, such as T-cell or cytokine responses in adults and children.[119,120] Uncertainty remains regarding the role of immunity against seasonal coronaviruses (229E, NL63, OC43 and HKU1) in protecting against SARS-CoV-2 infections. This chapter explores the theory of immunity against seasonal coronaviruses protecting against SARS-CoV-2 infections in children and is based on the submission Ratcliffe H, Tiley KS, Longet S et al. Serum HCoV-spike specific antibodies do not protect against subsequent SARS-CoV-2 infection in children and adolescents.

Alpha and betacoronaviruses are antigenically related, and each have specific humoral, B cell and T cell immune responses.[149] All betacoronaviruses share extensive sequence homology with SARS-CoV-2.[205] Antibody responses to HCoVs have been shown to cross-react with SARS-CoV-2.[205,213,214] Sagar et al. examined electronic medical records to show that recent seasonal coronavirus infections were associated with decreased severity of COVID-19 disease in adults.[89] However, other studies using similar methodologies have produced conflicting results.[89]

Dowell et al. showed that SARS-CoV-2 infection in children boosts humoral responses against alphacoronaviruses (229E and NL63) and betacoronaviruses (OC43, HKU-1),[205] and that

concentrations of antibodies against OC43 and HKU-1 following SARS-CoV-2 infection were significantly higher in children than adults. The implications of these findings are unclear, as it could not be determined if the observed differences pre-dated the SARS-CoV-2 infection. Furthermore, given antibodies to 229E, NL63, HKU-1 and OC43 viruses provide only short-term immunity from reinfection with the same virus[215-217] any cross-protection may be short-lived. Even so, Dowell et al. suggest that the degree of cross-reactivity between seasonal coronaviruses and SARS-CoV-2 explains the age-related differences in the severity of COVID-19 infection.[213]

To understand better immunological responses to SARS-CoV-2 and seasonal coronaviruses, and the inter-relationship between the two, we analysed data from two longitudinal cross-sectional seroprevalence studies 'COVID Warriors'[218] (ClinicalTrials.gov Identifier: NCT04347408) and 'STORY'[219] (ClinicalTrials.gov Identifier: NCT04061382). Both studies were carried out within the UK at a time when widespread immunisation within the paediatric population was not available. A case-control analysis was used to determine whether the presence of detectable antibodies against seasonal coronaviruses influenced the subsequent likelihood of seroconversion for SARS-CoV-2.

6.2 Methods

Serum or plasma samples from participants were collected in both studies and were processed using multiple assays evaluating antibodies against spike protein, nucleocapsid and receptor binding domains (Table 9**Error! Reference source not found.**). Samples were classified a ccording to Roche Elecsys® Anti-SARS-CoV-2 IgG spike protein receptor binding domain assay (RocheS), Roche Elecsys® Anti-SARS-CoV-2 IgG nucleocapsid assay (RocheN), DiaSorin LIAISON® SARS CoV-2 S1/S2 IgG assay (DiaSorin) and a UK Health Security Agency in house receptor binding domain assay (UKHSA RBD) into seropositive samples (defined as a positive result against SARS-CoV-2 as defined by the manufacturers' definition (Table 9**Error! Reference source not found.**)[127,129,220] on one or more of the assays) and s

eronegative samples (defined as a negative result against SARS-CoV-2 across all assays,Table 9**Error! Reference source not found.**).

6.2.1 Unmatched analysis

Samples were pooled into three categories: seropositive, seronegative and pre-pandemic serum samples. Pre-pandemic samples were collected from children aged 0-13 years in 2009/2010 (EudraCT Number: 2009-014719-11) and 2017/2018 (EudraCT Number: 2017-004732-11). All samples were retested using the Meso Scale Discovery (MSD) V-PLEX platform to quantify IgG concentrations to SARS-CoV-2 **(**receptor binding domain, nucleocapsid and spike protein), alpha coronaviruses (229E and NL63 spike protein) and beta coronaviruses (SARS-CoV-1, MERS, HKU-1, and OC43 spike protein). Groups were compared using the Mann Whitney U test as data were not normally distributed, and results with value of $P<0.01$ were considered statistically significant as multiple comparisons were made.

6.2.2 Matched case control analysis

Pre-existing immune responses to seasonal coronaviruses have been postulated as a reason why children are less severely affected by SARS-CoV-2 infection. Consequently, we carried out a matched case-control analysis of 38 cases and 87 controls aged 2-19 years where two time points were available (baseline and seroconversion visit). A seroconversion 'case' was defined by a positive result (above the thresholds of positivity shown in Table 9 **Error! Reference s ource not found.**) for SARS-CoV-2 on any study assay (RocheS, Diasorin, RocheN or RBD) at the second or third visit, following previously negative visit result(s). A control was defined as a child who did not seroconvert for SARS-CoV-2 and was matched on:

- date of the matched cases first positive result (date +/- 15 days)
- date of a negative preceding visit (date +/- 15 days)
- study site (COVID Warriors) or NHS region (STORY)

- Age (COVID Warriors +/- two years) ‘STORY’ (within an age band +/- two years Figure 30)

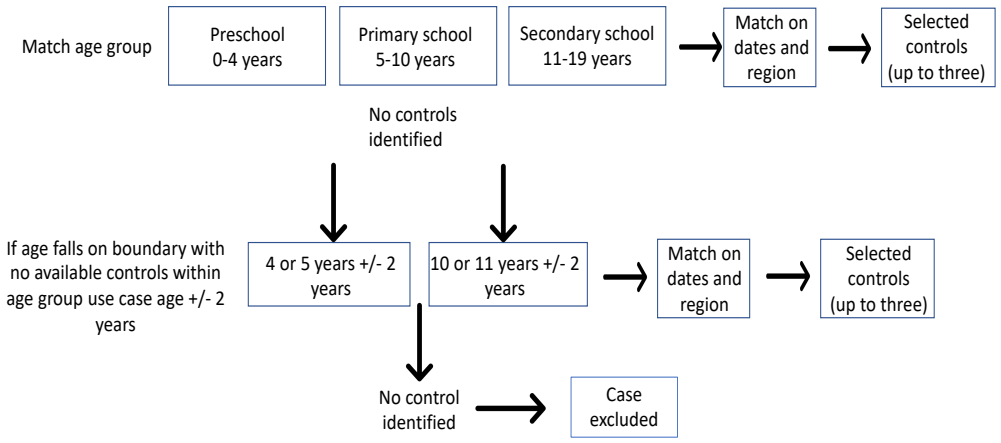


Figure 30 Age matching within the 'STORY' dataset

Between one and three controls were identified for each case.

Each case and control had a baseline (V1) visit followed by a seroconversion (V2) visit, and samples were tested on the MSD platform. Conditional logistic regression was used to determine if baseline (V1) level of seasonal coronavirus antibody could predict SARS-CoV-2 seroconversion (case/control status), adjusting for study.

6.3 Results

6.3.1 Assay concordance

Figure 31 shows concordance for negative results of 84 to 90% across all assays when detecting the same SARS-CoV-2 antigen, with the MSD platform identifying between 5 and 9% more positive results than the Roche assay (spike and nucleocapsid proteins), the internal PHE assay (RBD) and Diasorin (spike protein).

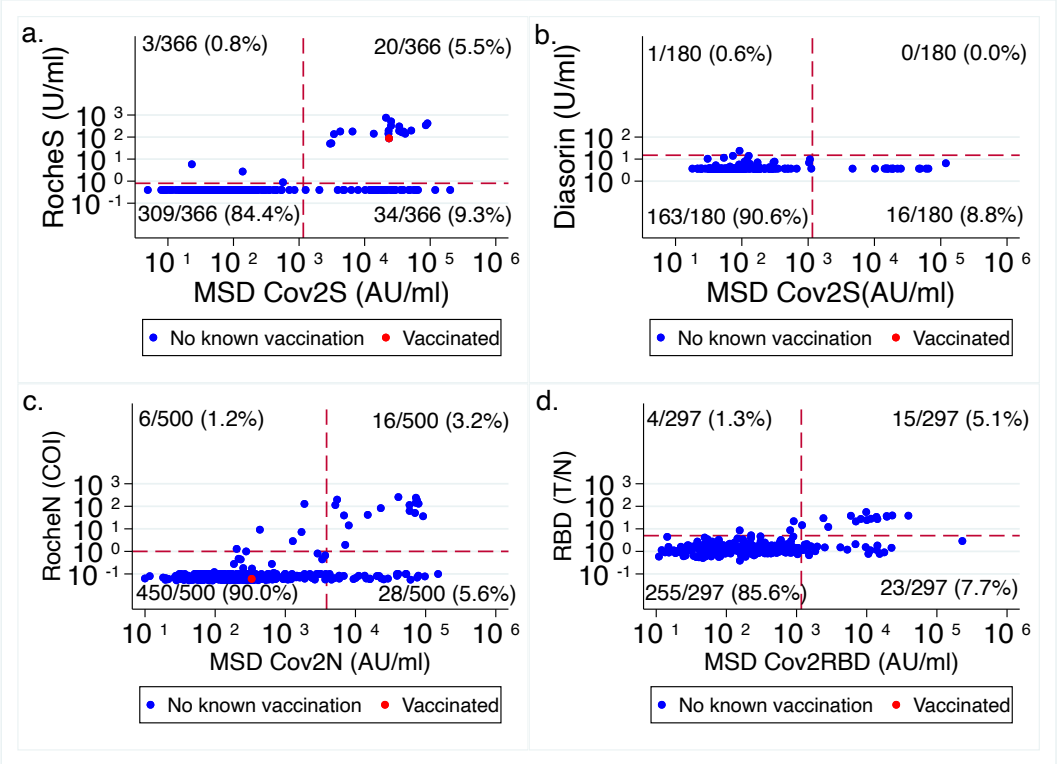


Figure 31 A comparison of assays by a. Comparing RocheS (Spike protein receptor binding domain assay) and MSD Cov2S (spike assay) b. Comparing Diasorin (spike assay) and MSD Cov2S (spike assay) c. Comparing RocheN (nucleocapsid assay) and MSD Cov2N (nucleocapsid assay). d. in-house PHE assay (Receptor binding domain assay) and MSDCov2RBD (Receptor binding domain assay). Red markers are a sample from participants who were immunised at the time of sampling. Red dashed lines show the threshold of positivity for each assay.

6.3.2 SARS-CoV-2 antibodies in children cross-react with Beta HCoV.

A total of 53 children with seropositive samples (range 3-22 years, mean and median 12 years) and 145 children with seronegative samples (range 0-19 years, mean 10 years, median 11 years) were analysed. Figure 32 shows one sample per participant categorised as seropositive, seronegative (positivity defined on original study assays) or pre-pandemic. IgG concentrations against SARS-CoV-2 S, N and RBD as determined by the MSD assay were significantly higher in seropositive participants than in seronegative and pre-pandemic cohorts. SARS-CoV-2 seropositive children also had significantly higher antibody titres against all other seasonal coronavirus spike proteins than did pre-pandemic cohorts and, for the betacoronaviruses, seropositive children had higher antibody titres than seronegative children.

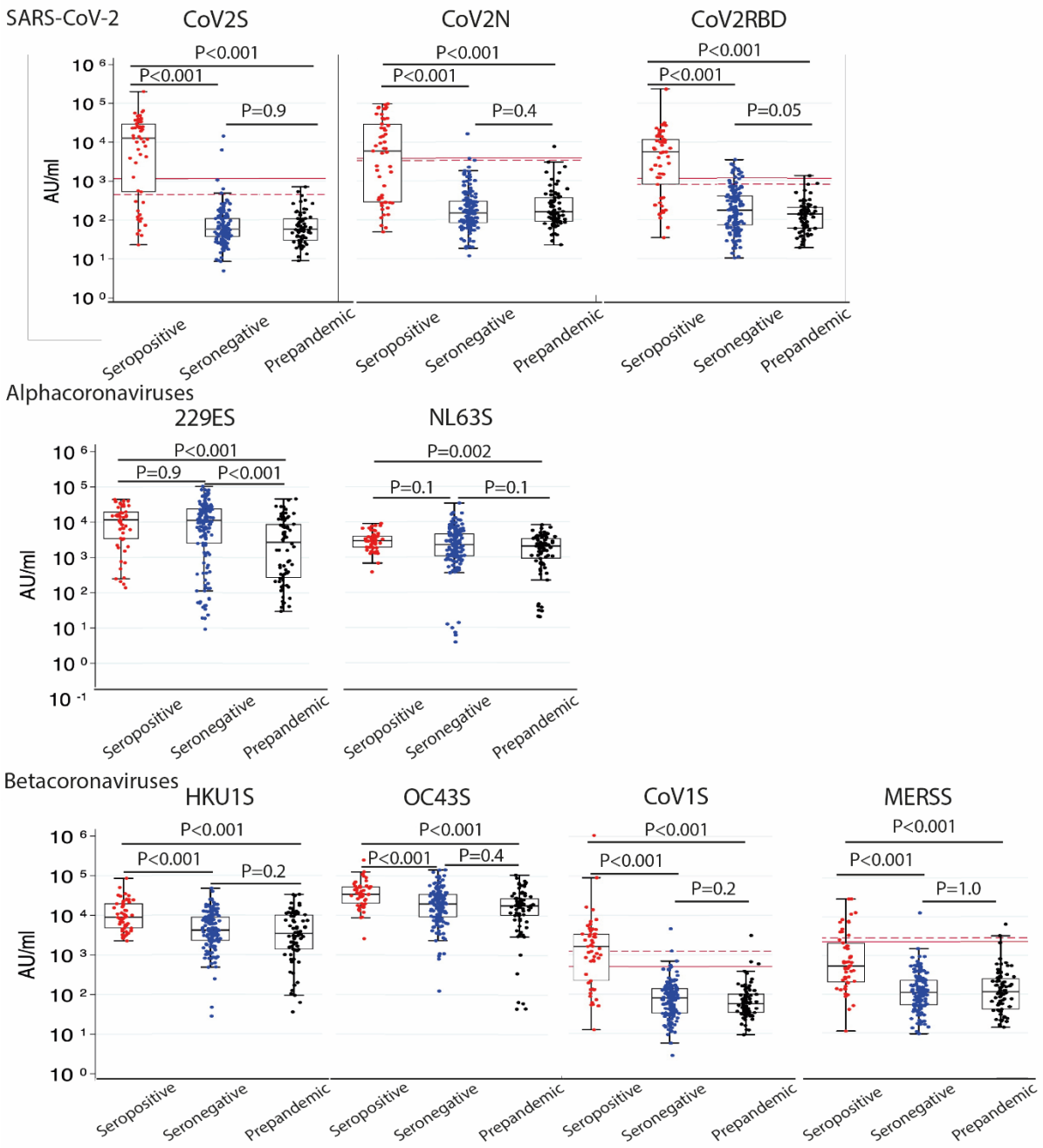


Figure 32 A comparison of seropositive participants for SARS-CoV-2 (red), seronegative participants for SARS-CoV-2 (blue) and pre-pandemic (black) samples (positivity defined by original study assays) tested using the MSD platform. The MSD platform included IgG concentration (Au/ml) for IgG anti-spike, anti-nucleocapsid and anti-receptor binding domain (RBD) are shown for SARS-CoV-2. Anti-spike protein IgG concentration (Au/ml) is shown on for alphacoronaviruses (229E and NL63) and betacoronaviruses (HKU1, OC43, CoV1 and MERS). Mann Whitney U test was used to compare between groups. Threshold of positivity – Solid line (adult cut off) dashed line (paediatric cut off)

Figure 32shows one sample per participant categorised as seropositive, seronegative (positivity defined on original study assays) or pre-pandemic.

IgG concentrations against SARS-CoV-2 S, N and RBD as determined by the MSD assay were significantly higher across both age groups in seropositive participants than in seronegative and pre-pandemic cohorts (Figure 32).

6.3.3 HCoV spike protein antibodies do not confer protection against SARS-CoV-2 infection

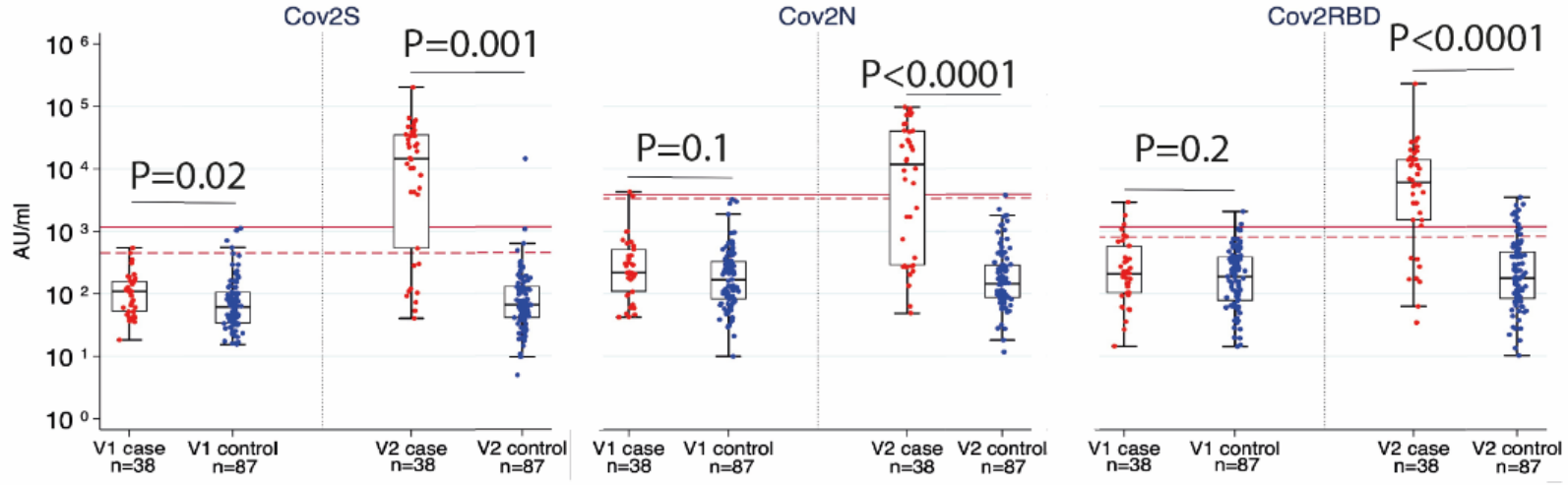
At baseline (V1) cases and controls showed no significant difference in antibody concentrations against any seasonal coronaviruses. 10-fold higher antibody titres were not associated with decreased odds of subsequent infection (seroconversion) with SARS-CoV-2 for any alpha- or beta-coronaviruses (Table 32).

At the seroconversion (V2) visit, concentrations of spike-specific IgG against betacoronaviruses and SARS-CoV-2 were significantly higher in cases than in controls. No significant differences in IgG concentrations against alphacoronaviruses between cases and controls were seen at V2 between cases and controls (Figure 33).

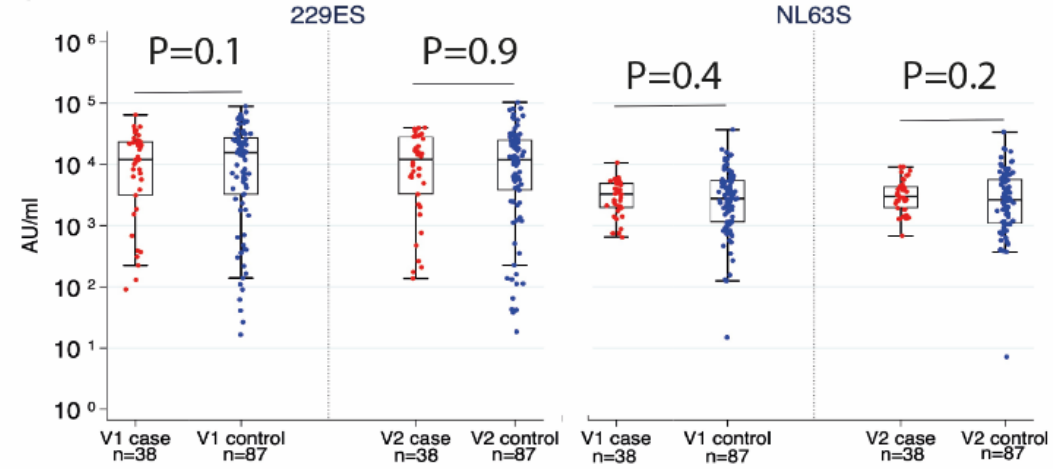
Table 32 Shows the odds being infected per 10-fold higher baseline titre (V1) results with value of P<0.01 were considered statistically significant

		Visit 1 (baseline) Odds ratio (OR) (99%CI)	p-value
Alphacoronaviruses	229E Spike	1.0 (0.6-1.8)	0.9
	NL63 Spike	1.3 (0.4-4.0)	0.5
Betacoronaviruses	HKU1 Spike	1.9 (0.6-6.3)	0.2
	OC43 Spike	1.6 (0.5-4.7)	0.3
	SARS-CoV-1 Spike	1.8 (0.6-5.5)	0.2
	MERS Spike	1.2 (0.4-3.1)	0.7

SARS-CoV-2



Alphacoronaviruses



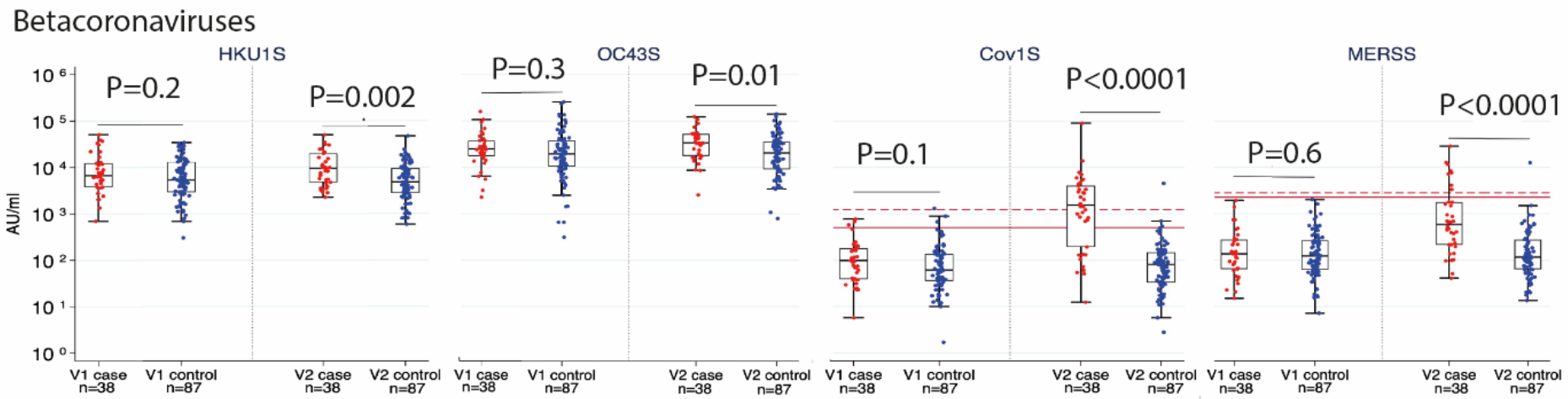


Figure 33 A matched case-control analysis comparing seropositive SARS-CoV-2 cases (red) with SARS-CoV-2 controls (blue) tested on the MSD platform. Seropositivity and seronegativity were defined by original study assays. The MSD platform included IgG concentration (Au/ml) for IgG anti-spike, anti-nucleocapsid and anti-receptor binding domain (RBD) are shown for SARS-CoV-2. Anti-spike protein IgG concentration (Au/ml) is shown on for alphacoronaviruses (229E and NL63) and betacoronaviruses (HKU1, OC43, CoV1 and MERS) Mann Whitney U test was used to compare between baesline (V1) visits and seroconversion (V2) visits.Threshold of positivity solid line (adult cut off) dashed line (paediatric cut off)

6.4 Discussion

These studies conducted in a vaccine-naïve population between the beginning of the pandemic and June 2021 show no evidence of protection against SARS-CoV-2 infection due to antibodies against seasonal coronaviruses. Our data supports and builds on work by Dowell et al. who reported a trend of increased antibodies against both alphacoronaviruses and betacoronaviruses in SARS-CoV-2 seropositive versus seronegative participants, which was statistically significant for OC43 and HKU-1.[205] This increase was reported to be due to antibodies that were cross-reactive with the S2 subunit of the spike protein of SARS-CoV-2, which is more closely related in betacoronaviruses.[205] This concurs with our data demonstrating a significantly higher IgG concentration for betacoronaviruses when comparing seropositive with seronegative samples, and when analysing the V2 samples for cases versus controls, but this was not seen with the alphacoronaviruses.

Although the numbers in the case-control analysis were small, the results build on these cross-sectional comparisons and shows no evidence of higher baseline seasonal coronavirus antibodies conveying protection against SARS-CoV-2 infection. This suggests that the higher antibodies against seasonal coronavirus antibodies seen in Dowell et al. have occurred in response to SARS-CoV-2 infection, implying cross-reactivity but not cross-protection.

Immune imprinting may explain the increase in IgG concentration in betacoronaviruses seen within our case-control analysis. This is where the immune system recalls existing memory cells, rather than stimulating de novo responses when encountering a novel but closely related antigen (back-boosting effect).[214,221] Aydillo et al. reported that in patients with severe COVID-19 there was back-boosting of antibodies against the S2 subunit from betacoronaviruses.[214] Westerhuis et al. also showed back-boosting in OC43 within adult participants when measuring both nucleocapsid and spike antibodies.[221] The limitation of both these studies was that they were in hospitalised patients rather than mild/asymptomatic participants, but the data we

present here suggests this is seen in children with mild infections as well as in adults with severe COVID-19. Lavell et al. studied healthcare workers who were working during the first wave of the pandemic and demonstrated no significant association with HCoV-OC43 spike protein IgG, or with antibodies against other HCoVs. In contrast, when measuring nucleocapsid antibodies, they found that high levels of HCoV-OC43 was an indicator of recent infection and was associated with protection against SARS-CoV-2 infection.[222] Our research focused on antibodies against spike protein, and did not evaluate anti-nucleocapsid antibodies or cell mediated immunity, and these limitations should be born in mind when drawing conclusions about the association of seasonal coronaviruses with protection against SARS-CoV-2 infection. While our findings suggest that immunity against seasonal coronaviruses in children does not provide protection against SARS-CoV-2 infections, it is still possible that it could mitigate the severity of infection. Ortega et al. assessed whether adults with higher anti-HCoV N antibody levels prior to infection were protected against COVID-19 symptoms in participants who seroconverted during their study period. Ortega et al. demonstrated higher levels of anti-nucleocapsid IgA antibodies against OC43 in asymptomatic compared with symptomatic SARS-CoV-2 seropositive participants.[206] Ortega et al. also demonstrated anti-nucleocapsid IgG concentrations against NL63 underwent a higher fold-increase after SARS-CoV-2 infection in asymptomatic than symptomatic seroconverters.[206] It is difficult to distinguish whether the difference in findings between Ortega et al. and our findings when looking at baseline anti-HCoV antibodies are due to differences in methodologies or the age groups of the participants. Other adult studies have reported no cross-protection from seasonal coronaviruses against SARS-CoV-2.[223-225]

The literature shows seropositive children and adults had broadly similar antibody responses against SARS-CoV-2 nucleocapsid. They also demonstrated a trend towards a higher antigen

specific geometric mean IgG concentration in children when compared with adults, most notably against spike protein and RBD domains.[205]

The evidence of lack of cross protection from anti-spike antibodies against seasonal coronaviruses emphasises the need for ongoing research to understand why children are less susceptible to severe infections including more detailed assessments of the differences in cellular immune response between adult and paediatric populations. Children are reported to display fewer symptoms than their adult counterparts despite a similar viral load, but they mount a robust innate immune response which allows better early control of inflammation than adult counterparts.[226] Swadling et al from their work within the adult population suggest that not all individuals develop a PCR or antibody positivity due to T cell responses.[227]

T cells initially primed against seasonal coronaviruses (both alpha and beta coronaviruses) have been shown to exhibit cross-reactivity with SARS-CoV-2 in children.[205] Children and young adults display increased levels of cross-reactive T-cell responses, which may lead to less severe disease when compared with their adult counterparts.[228]

Further clues may be provided by the increase in susceptibility to SARS-CoV-2 following the emergence of omicron variants.[229] Omicron variants have altered the clinical presentation of COVID-19, predominantly affecting the upper respiratory tract rather than the lower respiratory tract, leading to an increased number of children presenting with croup. Children are more susceptible to upper airway infections as they have relatively smaller airways compared with adults.[230] The entry of SARS-CoV2 into host cells has also changed with the new variants, with previous strains entering via ACE2 receptors (less numerous in children than adults)[33], while the method of entry for omicron variants is endocytic.[231] With increased susceptibility to Omicron, a proportional increase of severe disease in children has not been seen[229] suggesting that the method of viral entry was not responsible for the reduced disease severity in children.

With vaccination programmes offering vaccines to all children over five years of age in the UK and with SARS-CoV-2 continuously circulating within the community, the opportunity to study children's immune responses to SARS-CoV-2 in a naïve population becomes more challenging. These data provide insight into a critically important issue.

Chapter 7 Discussion

Assessing the levels of antibodies against specific infectious diseases across a population is a valuable public health tool to assess potential vulnerabilities in the control of infectious diseases. While convenience sampling of residual sera obtained for clinical purposes runs the risk of not being representative of the general paediatric population, and therefore needs to be interpreted with care, community based prospective sampling selects participants on whether they meet specific inclusion criteria. This thesis details a prospective strategy for collecting blood samples in the community and compares targeted vs untargeted recruitment across the socioeconomic spectrum. To date, these samples have provided a unique insight into the prevalence of SARS-CoV-2 infection induced antibodies in the serum of children aged between 0 to 18 years during the COVID-19 pandemic. This has in turn given an indication of 'when and where' SARS-CoV-2 infections rates were increasing in the paediatric populations, and populations at particular risk of infection, as well the proportion of the population with potential reduced risk of infection or significant disease. This has also provided insight into the lack of protection against SARS-CoV-2 infection afforded by pre-existing anti-spike protein antibodies against seasonal coronaviruses, and the strengths and weaknesses of different sample types (serum, saliva) and antibody targets (spike, nucleocapsid) in seroprevalence studies.

7.1 Establishing a nationally representative paediatric serobank

The 'STORY' study met its primary objective of demonstrating the feasibility of running a prospective seroepidemiology study in children and young adults aged 0-24 years in England. Early in the pandemic, convenience sampling from the paediatric population was limited. Healthcare interactions for non-urgent care were kept to a minimum, and very few children were hospitalised due to COVID-19 disease, therefore convenience sampling was not representative of the paediatric population when establishing the prevalence of SARS-CoV-2 antibodies. Samples from 'STORY' provided a unique resource as a prospectively collected

serological survey including participants from all seven NHS regions, allowing generalisations to be made to England as a whole.

The study design recruited participants between the ages of 0-24 years, creating a biobank that captured crucial demographic characteristics of the paediatric population in England and their corresponding immunisation histories. Both group one recruitment (using mail-outs targeted at specific postcodes) and group two recruitment (not postcode targeted) methodologies managed to recruit participants across the socioeconomic spectrum in England. Group two methods of recruitment through word-of-mouth, social media and email circulars had a cheaper per-participant cost than mailouts and were as successful at recruiting across the socioeconomic spectrum, but mailouts resulted in higher response rates and were more effective when targeting under represented groups such as participants from minority ethnic groups. For subsequent paediatric seroepidemiology studies an integrated approach of both methods starting with group two methodology may be appropriate to achieve the most economical recruitment strategy. Electronic consent and a screening tool at the point of response to check that recruitment criteria were being matched with the region prior to enrolment may produce better representation than using a targeting mail out later in the study to target groups who were under represented.[12]

The community based recruiting used in the 'STORY' study was suited to evaluating seroprevalence of an emerging pathogen as this avoids the clustering effect which could potentially be seen with other serosurvey strategies such as school-based recruitment.[39] By collecting bloods by venepuncture rather than finger prick the 'STORY' study was able to collect larger volumes of blood which have allowed a serobank to be created, but in future pandemics the potential for remote testing through finger prick samples is an appealing option as it protects participants and study staff from potential exposure to the pathogen. Finger prick testing was used in the school infection survey,[193] but it was not utilized in preschool children.

The approach could be piloted in in the preschool age group to allow for a more cost effective seroepidemiology survey and as part of a pandemic preparedness strategy.

7.1.1 Insights into COVID-19 epidemiology gained from STORY

From previous sero-studies using convenience sampling early in the pandemic, children were shown to be asymptomatic or have mild symptoms following SARS-CoV-2 infection and ancestoral lineages A and B presented predominantly with gastrointestinal symptoms in the paediatric population.[24] From adult studies it was shown that minority ethnic groups were at increased risk of seropositivity for SARS-COV-2.[44] The 'STORY' study built on these findings using community recruitment to establish the prevalence of SARS-CoV-2 between October 2019 and June 2021 of children aged 0-18 years across seven NHS regions allowing generalisations to be made across England as a whole. The study showed increasing prevalence across all age groups over time. Approximately one-third of aged 15–18 years had evidence of antibodies against SARS-CoV-2 prior to the introduction of widespread vaccination.[219] Consistent with adult studies, minority ethnic groups had higher risk of SARS-CoV-2 seropositivity than white participants (odds ratio 1·4, 95% CI 1·0-2·0), after adjusting for sex, age, region, time period, deprivation, and urban/rural geography.[219]

The apparent differences in SARS-CoV-2 infection rates in white and non-white populations are especially relevant given hospital data showed an excess in mortality in minority ethnic populations.[189] It had been suggested that the differences in mortality data between white and minority ethnic groups were due to the differences in comorbidities in different groups.[232] However the increased infection rates in minority ethnic groups in both the adult and paediatric populations[44,219] suggest that the increased risk is independent of an individual's co-morbidity status and therefore the relationship between higher seropositivity and ethnicity requires further study to understand why certain groups were affected more than others.

Overall, 48% of children who were seropositive and had complete questionnaires listing full symptomatology reported no symptoms.[219] In children <10 years there were no symptoms or symptom clusters that reliably predicted seropositivity.[219] These striking findings demonstrate the limitations of symptom-based testing within the paediatric population to identify individuals with active SARS-CoV-2 infections.

7.1.2 Insights into COVID-19 immunology gained from STORY

A feature of 'STORY' was the collection of repeated samples in both SARS-CoV-2 sero-positive and sero-negative participants, providing a longitudinal assessment of the kinetics of antibody decline after infection, and the immunological predictors of subsequent SARS-CoV-2 infection. The literature suggests that children generate IgG antibodies predominantly against spike protein compared to other SARS-CoV-2 antigens and this bias is more pronounced when comparing adult and paediatric responses to SARS-CoV-2 infection.[201] From the data presented in chapter five, there appears to be no significant difference between serum half-life of anti-spike and anti-nucleocapsid IgG concentrations but the anti-spike response persists for a longer duration before sero-reverting. There is no such difference seen with oral fluid with both anti-spike and anti-nucleocapsid antibodies demonstrating similar antibody decay profiles. Oral fluid samples revert from positive to negative earlier than serum samples making oral fluid sampling a less sensitive tool for seroprevalence estimates.

The assessment of the children sero-converting for SARS-CoV-2 during the study identified no protective effect from pre-existing anti-spike humoral immunity against seasonal coronaviruses, but there was some evidence of cross-reactivity following SARS-CoV-2 infection. This is an important negative finding to inform our understanding of protection against severe SARS-CoV-2 infection within the paediatric population and shows further research is required into alternative explanations as to why the paediatric population is less severely affected than the adult population.

7.2 Challenges in the 'STORY' study

7.2.1 Running a multi-site seroepidemiological study

The 'STORY' study rapidly expanded from 6 to 13 sites in response to the COVID-19 pandemic. Real-time tracking of the number of responses received and enrolled participants across each site was necessary to ensure each age band contained participants from all NHS regions and from demographically representative postcodes. This allowed sharing of targets across sites within NHS regions if recruitment of specific age bands was difficult. By continually monitoring recruitment, mailouts were used to target those demographics which were missing from our sample of recruited participants. This increased the number of participants from under represented age groups by region. For minority ethnic groups, we found that the most effective method of recruitment was targeted mail-outs through CHIS. However, not all sites had a relationship with their local CHIS and most required a PIC agreement which was time-consuming and costly, therefore, not all sites had access to targeted mailouts. Future studies of this nature would greatly benefit from either a more centralised system such as NHS digital containing the same data or a country-wide policy enabling local CHIS to facilitate research with a standardised approach to data protection.

Immunisation histories were gathered from either the participant or their medical records. This process was challenging as the quality of the data recorded was variable. The quality of the data could be improved if systems such as NIMS were expanded to cover all vaccinations. In this way NHS numbers could be linked to records enabling availability of complete and accurate information, including brand of the vaccine when evaluating immunisation schedules. In the event of a public health emergency, anonymised data should be shared within the research community, whilst respecting an individual's right to privacy. At the beginning of the pandemic Dame Caldicott stated that it has "never been more important to get the right information, to the right people, at the right time".[233] Data sharing during the COVID-19 pandemic has had both successes and failures. In terms of collaboration within the research

community and the production of vaccines, data sharing was a success. For example, researchers could access participants COVID-19 vaccine histories through NIMS to estimate vaccine coverage and evaluate the effectiveness of the immunisation programme.[234] These data were readily available for studies recruiting participants after doses had been given in the community such as COM COV-3.[235] In the UK, the IT infrastructure made data collection challenging. For instance the NHS COVID-19 testing app which was developed by Zühlke Engineering did not initially communicate with the NHS app.[236] The difficulties in linking put delays on real-time data to understand how the pandemic was progressing. Despite being hailed as a success, the NHS COVID-19 app adoption rate was 29% and did not meet the generally accepted adoption target of 60%.[237] Reasons why adoption rates of the apps was low should be an area of research when preparing for the next pandemic. It is hoped that the post-COVID reviews of how the use of data within the NHS can best be harnessed to inform public policy do not overlook the importance of obtaining accurate data in paediatric and adolescent populations, and the specific challenges and requirements (e.g. obtaining immunisation histories) that this entails.

7.2.2 Processing of serological samples during a pandemic

The bulk of 'STORY' serum samples were analysed by UKHSA. The advantage was that a standardised approach was taken for sample processing and assays being used allowing the data to be compared with other serological surveys and vaccination studies. The disadvantage of this approach was that the large number of samples to be processed within a limited number of laboratories resulted in some delays in provision of results. Therefore, in preparing for the next pandemic, a strategy to include other research facilities in serum testing allowing more samples to be tested in non UKHSA laboratories should be in place allowing more timely provision of results and analyses. Alternatively, UKHSA would require investment to allow a streamlined process to communicate results for each study to their research teams around the country.

7.2.3 Representation from populations of interest

The original study protocol had no strategy for over recruiting groups of interest who are under represented in research such as minority ethnic groups, migrant populations or groups who do not vaccinate due to personal beliefs, and instead focussed on recruiting set numbers from each age band and from across the socioeconomic spectrum. When the COVID-19 pandemic began the recruitment targets were increased and when it became clear that minority ethnic groups were at higher risk of SARS-CoV-2 infection[238] the study was adapted to enhance recruitment of participants identifying as belonging to a minority ethnic group. Community-based sampling was maintained but schools were deliberately not used for recruitment as we did not want results to be biased by clusters of infection. Importantly, the goals of 'STORY' were not to evaluate viral transmission, a linked but separate question which lends itself well to school (or family) based designs.

Lack of diversity in clinical research is a well-recognised challenge[154] and one that needed to be met to successfully address the objectives of STORY. Engaging populations who typically do not engage with research when the study did not offer them a direct benefit such as a vaccine was challenging. Due to social distancing measures, initial plans of engaging communities by direct contact such as community gatherings through places of worship and community clubs were not possible. There were few public engagement groups who were able to help us gain insight into the lower uptake of research from minority ethnic groups. At the beginning of the pandemic, we were unable to establish whether among minority ethnic groups this was hesitancy to engage rather than refusal, or perhaps members of these communities had other priorities such as juggling childcare whilst maintaining key worker roles or more pressing financial concerns. The reasons for under representation of minority ethnic groups in health and social care research in the UK are complex.[156] They include researchers focussing on more general organisation of a trial rather than monitoring the demographics of a sample. Enhanced recruitment of populations of interest can increase the cost of a trial with additional costs in

translations required for informed consent.[155] Researchers may have unconscious biases against certain populations e.g. lower socioeconomic status or minority ethnic groups leading to decisions to not enrol potential participants. Certain populations may not trust health professionals and therefore are less likely to come forward to participate in research.[155] Time and resources should be spent when study designs are being conceived to tackle behaviours which make it less likely that a representative sample will not be recruited.[155] To this end, Farooqi et al. have developed a tool kit aimed at increasing participation of minority ethnic populations in research, describing how the researchers must be culturally competent to engage with minority ethnic groups.[156] To achieve this there must be public engagement with the communities of interest when the study is in the planning stages and communications must be culturally appropriate. When the 'STORY' study was in the planning stage, there were very few minority ethnic community PPI (Patient and Public Involvement) groups within our research network, those that did exist did not have capacity to review study materials within the required time frame. The experience with 'STORY' was that those sites who already had links with minority ethnic groups in their local study area were more effective at recruiting from populations of interest. Overall, 17% of participants were recruited from minority ethnic groups, but these participants were disproportionately recruited towards the end of the study which created the potential to distort infection rates in an evolving pandemic that had to be corrected for in the analysis.

During the pandemic, minority ethnic groups were more likely to be unwilling to have a vaccine against COVID-19. A key finding from research into vaccine hesitancy was that trust in public sector officials has a key role to play in vaccination uptake.[239] It is the role of public health officials to communicate key messages and tackle misinformation and strengthen public health literacy to improve uptake. Progress was made during the pandemic in vaccine studies such as Com-Cov which recruited 25% of participants from minority ethnic groups.[240].

However, this study allowed early access to vaccines which may have influenced participants decisions to take part. Further research into how to address inequalities in healthcare and engage populations who are not represented in research is paramount.

7.2.4 The role of translations in aiding recruitment in minority ethnic groups

An early priority in establishing the 'STORY' study was to ensure that written translations of key study documents were provided to all sites and available online, along with language line (a verbal translation service provided over the phone). A key learning point from this study was that despite translations being costly and time-consuming to set up they were of little additional benefit and were not utilised by either researchers or participants as they were not required in taking consent from participants. In part this could reflect discrepancies between spoken and written language e.g. second and third-generation British citizens whose grandparents or parents migrated from non-English speaking countries may converse in the language of their originating country, but not read and write in these languages, and therefore written material was not helpful to them. School aged children will most likely be attending an English-speaking school and therefore the majority of our potential participants will be able to read and write in English however the same cannot be said for their parent or guardian. Farooqi et al. call for culturally competent researchers who preferably speak the language of the population of interest and criticise written documents which are word for word translations as the meaning of the text can be lost through literal translations from one language to another.[156] This reflected feedback from team members who were bilingual and reviewed translations that had been translated through an accredited translation service. The research ethics service in the UK has no guidance on how to translate documents however both the FDA (Food and Drug Administration) and Australian ethics committee state that any translated study documents have to be translated from the English version and back translated for each language with no flexibility where literal translations might change the meaning of the text.[241,242] Perhaps our translations were not used as they were not culturally competent translations. If repeating this

study when social contact restrictions are not in place, then time and effort should be spent with both PPI groups who belong to the population of interest and the ethics committee to produce more culturally competent research materials giving identical messages. Having researchers who are culturally competent across all minority groups is not feasible but engagement with local communities to change views on research and their involvement in local PPI groups will allow research groups to become more culturally competent within their local community.

7.2.5 Building ties to the community in groups who are not well represented in the population – both adults and paediatrics

As life returns to normal with the advent of vaccines and social restrictions being lifted it is important to build on the momentum started during the pandemic to engage with communities who are often under represented in UK research. There are inequalities in health within the UK which were demonstrated prior to the pandemic by life expectancy in England stalling.[243] Strategies already exist to tackle inequalities such as infant mortality but these have made limited progress.[244] During the early stages of the pandemic mortality data showed black and other minority ethnic groups had a higher mortality rate than their white counterparts.[232] According to the WHO "Health inequities are differences in health status or in the distribution of health resources between different population groups, arising from the social conditions in which people are born, grow, live, work and age. Health inequities are unfair and could be reduced by the right mix of government policies".[245] Lower socioeconomic status has also been linked to poorer health which results in significant social and economic costs both to the individual and to wider society.[244]

Communication with the public is key in implementing public health strategies. No matter how good a public health intervention might be, without public support it will fail. Public health strategies also need to be realistic; individuals will have their own priorities which are in direct conflict with a health strategy and will result in individuals failing to adhere to guidance, e.g.

if families have to choose between following a health policy which pushes them into poverty as they are unable to work then individuals will not adhere to the policy. The UK does not have the infrastructure to police a lockdown and adherence to this relied on the goodwill of the population. A lesson from the lockdown strategy in response to SARS-CoV-2 is that there has been a profound impact on the populations' mental health. Future strategies will be required to factor in the adverse effects of lockdowns and a balance needs to be struck between health protection from the pandemic versus excess deaths from other causes.

Now that social restrictions have been lifted, the time to act in addressing health inequalities is here. Our experiences from the 'STORY' study indicated that having established links to minority communities makes a difference to recruitment from these communities. In 'STORY' there were two sites recruiting in London. One had close ties to the community the other did not. The feedback from the study who had close ties to the community improved our research materials and recruited a high proportion of participants from minority ethnic groups. The other site recruited predominantly white participants even through the local population was diverse.

In a moral society, we act to protect the most vulnerable individuals, which was the purpose of lockdown. However, the role of children in the transmission of SARS-CoV-2 is a key question that was not answered in this thesis and has not been adequately addressed in other studies conducted through the pandemic. In the UK schools were closed on the 20th of March 2020 just seven weeks after the first case of COVID-19 was reported in the UK.[246] Lockdown and the closing of schools have had a profound impact on children and young adults. The effects of social isolation, school closures and family stress as a direct result of the virus or as a consequence of the economic repercussions of the pandemic has negatively affected children and adolescents.[247] When discussing pandemic preparedness, a strategy to establish the role of transmission in young people quickly needs to be in place with appropriate collaboration between public health and the department of education. This plan must be ready to be

implemented at very short notice. In the COVID-19 pandemic when our understanding of SARS-CoV-2 was limited, initial steps were based on previous flu protocols where the key role of children in transmission is well established.[27] There was political pressure to close schools both from the general public and from teachers' unions. Schools were closed with very little evidence to support that children were affected by the disease and that they were the drivers for transmission. By closing schools, transmission was prevented but the importance of school transmission could not be ascertained, which made returning to school a contentious issue. While schools were closed more children were forced into poverty and those children who were most vulnerable were put at risk.

One way of achieving transmission studies is to embed research involving children within the educational system. This would involve establishing long-term ties between research groups and schools. During the pandemic, NISEC (National Immunisation Schedule Evaluation Consortium https://nisec.ac.uk), which involves a number of different research groups, worked together to run many of the post-licensure COVID-19 vaccine trials in the UK. Pre-existing research agreements allowed the group to move at speed as there was already a relationship of cooperation between research groups working on routine studies in England. Due to previous collaborations staff from all groups were familiar with the challenges of delivering large studies and therefore the delivery of studies was done quickly and efficiently. If similar relationships were embedded within schools then transmission could be investigated to give up to date knowledge at the beginning of the pandemic so an informed decision on school closures could be made. To do this, schools should become an active part in research which would increase children and young people's public health literacy. A network of schools would be required in each of the NHS regions with a mix of rural/urban schools, state and private schools. Young people should be aware of relevant research and given the opportunity to take part. Research which could be routinely carried out in schools would be surveillance studies such as

pneumococcal carriage studies, qualitative research examining different health beliefs and motivations regarding why young people take part in research, seroprevalence studies interested in vaccine response as well as studies tackling questions regarding vaccine hesitancy within the population. Within a transmission protocol an established threshold should be defined to end the study if it becomes clear that children and adolescents are driving transmission.

It is also important to compare data with other countries which may adopt a different public health strategy. For instance, Sweden did not close schools during the pandemic and therefore had observational data for comparison. In terms of transmission studies, ideally the design would be a repeat cross-sectional study with samples to identify infection being provided at regular (e.g. weekly) intervals. Samples could be throat swabs or saliva samples for PCR testing. The additional benefit of saliva sampling, as piloted in children during the SARS-CoV-2 pandemic is that it would allow both PCR and antibody responses to be investigated, albeit with the lower sensitivity for seroprevalence as outlined in section 5.3.2.[248] Diaries using both closed and open questions should ascertain symptoms to help guide testing strategies and determine the number of asymptomatic infections. Testing would ideally test both students and the adults in their lives such as parents and teachers so that the direction of transmission can be determined. The logistics of distributing test kits, sending tests back and receiving results must be convenient to encourage compliance. PCR was the gold standard, and early in the pandemic it was a test which was available and more quickly validated than methods of antibody testing. This will give an idea of incidence within the local population and allow the characteristic of clusters to be examined. If a positive swab was taken, then there should be an optional consent for blood tests to look at antibody responses over time from the first swab. A study protocol of this nature is challenging as it will need to adapt to emerging knowledge about the virus without compromising data collected earlier in the study

prior to adaptations. Alongside this there is the additional challenge that the likely removal of any child identified as being infected from the school environment would disrupt the 'natural' patterns of transmission, and potentially undermine the study objective.

7.3 Further work

7.3.1 Applications for samples

The 'STORY' study has produced a unique serobank which is a valuable resource containing samples with relevant immunisations histories for children and young adults aged 0-24 years. The 'STORY' study was designed and implemented in collaboration with UKHSA at a time when COVID-19 was circulating but widespread immunisation of the paediatric population had not been implemented. With community-based recruitment rather than convenience sampling, the serobank is more representative of the general paediatric population and is available to researchers to evaluate population immunity against other vaccine-preventable diseases. Testing is already underway for anti-meningococcal C and anti-diphtheria IgG. This could be expanded to look at other diseases of interest as the public health need arises. Overall, the 'STORY' study has demonstrated the feasibility and public health benefit of a prospective serosurvey in the paediatric population. A repeat sero-survey in the post pandemic era would be feasible and beneficial to monitor changes to population immunity to infectious diseases.

1. Sandberg T. B Cell Responses to Human Flavivirus Vaccination and SARS-Cov-2 Infection [Ph.D.]. Sweden: Karolinska Institutet (Sweden); 2022.
2. Organization WH. Coronavirus disease (COVID-19) pandemic. 2020. https://www.who.int/europe/emergencies/situations/covid-19 (accessed 29th December 2022).
3. Edouard Mathieu HR, Lucas Rodés-Guirao, Cameron Appel, Charlie Giattino, Joe Hasell, Bobbie Macdonald, Saloni Dattani, Diana Beltekian, Esteban Ortiz-Ospina and Max Roser. "Coronavirus Pandemic (COVID-19)". Published online at OurWorldInData.org; 2020.
4. analysis IfG. Timeline of UK government coronavirus lockdowns. 2021. https://www.instituteforgovernment.org.uk/charts/uk-government-coronavirus-lockdowns (accessed 19 Jan 2022).
5. Organization WH. WHO Coronavirus Dashboard. 2022. https://covid19.who.int (accessed 29th December 2022).
6. Lipsitch M, Swerdlow DL, Finelli L. Defining the Epidemiology of Covid-19 — Studies Needed. *New England Journal of Medicine* 2020; **382**(13): 1194-6.

7. McAnally HB. An Epidemiologic Perspective. In: McAnally HB, ed. Opioid Dependence : A Clinical and Epidemiologic Approach. Cham: Springer International Publishing; 2018: 3-19.
8. Organization WH. Surveillance strategies for COVID-19 human infection: interim guidance, 10 May 2020: World Health Organization, 2020.
9. Care DoHaS, MP TRHBJ, MP TRHMH. Everyone in England will be able to access free, regular, rapid coronavirus (COVID-19) testing from 9 April, the government has announced. gov.uk; 2021.
10. Cutts FT, Hanson M. Seroepidemiology: an underused tool for designing and monitoring vaccination programmes in low- and middle-income countries. *Trop Med Int Health* 2016; **21**(9): 1086-98.
11. Ahmad AM, Shahzad K, Masood M, Umar M, Abbasi F, Hafeez A. COVID-19 seroprevalence in Pakistan: a cross-sectional study. *BMJ open* 2022; **12**(4): e055381.
12. Aatresh AV, Cummings K, Gerstein H, et al. Development of at-home sample collection logistics for large-scale seroprevalence studies. *PloS one* 2021; **16**(11): e0258516.
13. Walker S, Diamond I, Rourke E, Farrar J, Bell J, Newton J. Incidence of COVID-19 (SARS-CoV-2) infection and prevalence of immunity to COVID-19 (SARS-CoV-2) in the UK general population as assessed through repeated cross-sectional household surveys with additional serial sampling and longitudinal follow-up - an Office for National Statistics Survey. The University of Oxford and Oxford University Hospitals NHS Foundation Trust 2019; 2022.
14. Badal S, Thapa Bajgain K, Badal S, Thapa R, Bajgain BB, Santana MJ. Prevalence, clinical characteristics, and outcomes of pediatric COVID-19: A systematic review and meta-analysis. *Journal of Clinical Virology* 2021; **135**: 104715.
15. Götzinger F, Santiago-García B, Noguera-Julián A, et al. COVID-19 in children and adolescents in Europe: a multinational, multicentre cohort study. *The Lancet Child & Adolescent Health* 2020; **4**(9): 653-61.
16. Sorg A-L, Bergfeld L, Jank M, et al. Cross-sectional seroprevalence surveys of SARS-CoV-2 antibodies in children in Germany, June 2020 to May 2021. *Nature communications* 2022; **13**(1): 3128.
17. Cathal R, Thomas W. Review and future directions for PIMS-TS (MIS-C). *Archives of Disease in Childhood* 2023; **108**(4): e2.
18. Riffe T, Acosta E, Team C-D. Data Resource Profile: COVerAGE-DB: a global demographic database of COVID-19 cases and deaths. *International Journal of Epidemiology* 2021; **50**(2): 390-+.
19. Adetifa IMO, Uyoga S, Gitonga JN, et al. Temporal trends of SARS-CoV-2 seroprevalence during the first wave of the COVID-19 epidemic in Kenya. *Nature communications* 2021; **12**(1): 3966.
20. Stringhini S, Wisniak A, Piumatti G, et al. Seroprevalence of anti-SARS-CoV-2 IgG antibodies in Geneva, Switzerland (SEROCoV-POP): a population-based study. *The Lancet* 2020; **396**(10247): 313-9.
21. Morello R, Mariani F, Mastrantoni L, et al. Risk factors for post-COVID-19 condition (Long Covid) in children: a prospective cohort study. *eClinicalMedicine* 2023; **59**.
22. Terence S, Benjamin A, Manjula DN, et al. Long COVID (post-COVID-19 condition) in children: a modified Delphi process. *Archives of Disease in Childhood* 2022; **107**(7): 674.
23. Viner RM, Mytton OT, Bonell C, et al. Susceptibility to SARS-CoV-2 Infection Among Children and Adolescents Compared With Adults: A Systematic Review and Meta-analysis. *JAMA Pediatr* 2020.

24. Waterfield T, Watson C, Moore R, et al. Seroprevalence of SARS-CoV-2 antibodies in children: a prospective multicentre cohort study. *Archives of Disease in Childhood* 2020: archdischild-2020-320558.
25. NHS. Get tested for coronavirus (COVID-19). 2021. https://www.nhs.uk/conditions/coronavirus-covid-19/testing/get-tested-for-coronavirus/ (accessed 19th January 2022.
26. Conlon CP, Firth JD. COVID-19 Disease. Oxford Textbook of Medicine: Oxford University Press; 2020. p. 0.
27. Care DoHaS. UK Pandemic preparedness. In: Care DoHaS, editor.: gov.uk; 2020.
28. Williams PCM, Howard-Jones AR, Hsu P, et al. SARS-CoV-2 in children: spectrum of disease, transmission and immunopathological underpinnings. *Pathology* 2020; **52**(7): 801-8.
29. Liu Y, Hu Y, Deng X, et al. Seroepidemiology of mumps in the general population of Jiangsu province, China after introduction of a one-dose measles-mumps-rubella vaccine. *Sci Rep* 2015; **5**: 14660.
30. Noorimotlagh Z, Jaafarzadeh N, Martínez SS, Mirzaee SA. A systematic review of possible airborne transmission of the COVID-19 virus (SARS-CoV-2) in the indoor air environment. *Environ Res* 2021; **193**: 110612.
31. Singanayagam A, Hakki S, Dunning J, et al. Community transmission and viral load kinetics of the SARS-CoV-2 delta (B.1.617.2) variant in vaccinated and unvaccinated individuals in the UK: a prospective, longitudinal, cohort study. *The Lancet Infectious Diseases* 2022; **22**(2): 183-95.
32. Hollingsworth TD, Klinkenberg D, Heesterbeek H, Anderson RM. Mitigation strategies for pandemic influenza A: balancing conflicting policy objectives. *PLoS computational biology* 2011; **7**(2): e1001076.
33. Ladhani SN, Amin-Chowdhury Z, Davies HG, et al. COVID-19 in children: analysis of the first pandemic peak in England. *Archives of Disease in Childhood* 2020: archdischild-2020-320042.
34. Knabl L, Mitra T, Kimpel J, et al. High SARS-CoV-2 seroprevalence in children and adults in the Austrian ski resort of Ischgl. *Communications medicine* 2021; **1**(1): 4.
35. Danis K, Epaulard O, Bénet T, et al. Cluster of coronavirus disease 2019 (COVID-19) in the French Alps, February 2020. *Clinical Infectious Diseases* 2020; **71**(15): 825-32.
36. Zhen-Dong Y, Gao-Jun Z, Run-Ming J, et al. Clinical and transmission dynamics characteristics of 406 children with coronavirus disease 2019 in China: a review. *Journal of Infection* 2020; **81**(2): e11-e5.
37. Zhu Y, Bloxham CJ, Hulme KD, et al. Children are unlikely to have been the primary source of household SARS-CoV-2 infections. 2020.
38. Zhu Y, Yuan Z, Zhao Q, Chen G, Xu B. Seroprevalence of hepatitis A virus antibody in a population aged 0-30 years in Shanghai, China: implications for hepatitis A immunization. *Epidemiol Infect* 2013; **141**(3): 556-62.
39. Ismail SA, Saliba V, Lopez Bernal J, Ramsay ME, Ladhani SN. SARS-CoV-2 infection and transmission in educational settings: a prospective, cross-sectional analysis of infection clusters and outbreaks in England. *The Lancet Infectious Diseases*; (21): 344–53.
40. Viner RM, Russell SJ, Croker H, et al. School closure and management practices during coronavirus outbreaks including COVID-19: a rapid systematic review. *The Lancet Child & Adolescent Health* 2020; **4**(5): 397-404.
41. Hildenwall H, Luthander J, Rhedin S, et al. Paediatric COVID-19 admissions in a region with open schools during the two first months of the pandemic. *Acta Paediatrica (Oslo, Norway: 1992)* 2020; **109**(10): 2152-4.

42. Yi Y, Lagniton PNP, Ye S, Li E, Xu RH. COVID-19: what has been learned and to be learned about the novel coronavirus disease. *Int J Biol Sci* 2020; **16**(10): 1753-66.
43. de Lusignan S, Dorward J, Correa A, et al. Risk factors for SARS-CoV-2 among patients in the Oxford Royal College of General Practitioners Research and Surveillance Centre primary care network: a cross-sectional study. *Lancet Infect Dis* 2020; **20**(9): 1034-42.
44. Ward H, Atchison CJ, Whitaker M, et al. Antibody prevalence for SARS-CoV-2 in England following first peak of the pandemic: REACT2 study in 100,000 adults. *medRxiv* 2020: 2020.08.12.20173690.
45. Niedzwiedz CL, O'Donnell CA, Jani BD, et al. Ethnic and socioeconomic differences in SARS-CoV-2 infection: prospective cohort study using UK Biobank. *BMC Medicine* 2020; **18**(1): 160.
46. Saatci D, Ranger TA, Garriga C, et al. Association Between Race and COVID-19 Outcomes Among 2.6 Million Children in England. *JAMA Pediatrics* 2021; **175**(9): 928-38.
47. Mackey K, Ayers CK, Kondo KK, et al. Racial and Ethnic Disparities in COVID-19-Related Infections, Hospitalizations, and Deaths A Systematic Review. *ANNALS OF INTERNAL MEDICINE* 2021; **174**(3): 362-+.
48. Fisher KA, Bloomstone SJ, Walder J, Crawford S, Fouayzi H, Mazor KM. Attitudes toward a potential SARS-CoV-2 vaccine: a survey of US adults. *Annals of internal medicine* 2020; **173**(12): 964-73.
49. England PH. National Immunisation Programme: health equity audit. 2019.
50. Dodds C, Fakoya I. Covid-19: ensuring equality of access to testing for ethnic minorities. British Medical Journal Publishing Group; 2020.
51. Ciotti M, Angeletti S, Minieri M, et al. COVID-19 Outbreak: An Overview. *Chemotherapy* 2019; **64**(5-6): 215-23.
52. Ciotti M, Ciccozzi M, Terrinoni A, Jiang W-C, Wang C-B, Bernardini S. The COVID-19 pandemic. *Critical Reviews in Clinical Laboratory Sciences* 2020; **57**(6): 365-88.
53. Zhu N, Zhang D, Wang W, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med* 2020; **382**(8): 727-33.
54. Wang M-Y, Zhao R, Gao L-J, Gao X-F, Wang D-P, Cao J-M. SARS-CoV-2: Structure, Biology, and Structure-Based Therapeutics Development. *Frontiers in Cellular and Infection Microbiology* 2020; **10**.
55. Zhou P, Yang XL, Wang XG, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 2020; **579**(7798): 270-3.
56. Walls AC, Park Y-J, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* 2020; **181**(2): 281-92.e6.
57. Li F. Structure, Function, and Evolution of Coronavirus Spike Proteins. *Annu Rev Virol* 2016; **3**(1): 237-61.
58. Wrapp D, Wang N, Corbett KS, et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* 2020; **367**(6483): 1260-3.
59. Shang J, Ye G, Shi K, et al. Structural basis of receptor recognition by SARS-CoV-2. *Nature* 2020; **581**(7807): 221-4.
60. Sternberg A, Naujokat C. Structural features of coronavirus SARS-CoV-2 spike protein: Targets for vaccination. *Life sciences* 2020; **257**: 118056.
61. Johnson BA, Xie X, Bailey AL, et al. Loss of furin cleavage site attenuates SARS-CoV-2 pathogenesis. *Nature* 2021; **591**(7849): 293-9.
62. Hadfield J, Megill C, Bell SM, et al. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* 2018; **34**(23): 4121-3.

63. Rambaut A, Holmes EC, O'Toole Á, et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nature Microbiology* 2020; **5**(11): 1403-7.
64. Choi JY, Smith DM. SARS-CoV-2 Variants of Concern. *Yonsei Med J* 2021; **62**(11): 961-8.
65. Organization WH. Tracking SARS-CoV-2 variants. 2022. https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/ (accessed 28th December 2022).
66. Sette A, Crotty S. Adaptive immunity to SARS-CoV-2 and COVID-19. *Cell* 2021; **184**(4): 861-80.
67. Sette A, Crotty S. Pre-existing immunity to SARS-CoV-2: the knowns and unknowns. *Nature Reviews Immunology* 2020; **20**(8): 457-8.
68. Vitali L, Merlini A, Galvagno F, Proment A, Sangiolo D. Biological and Exploitable Crossroads for the Immune Response in Cancer and COVID-19. *Biomedicines* 2022; **10**(10).
69. Sekine T, Perez-Potti A, Rivera-Ballesteros O, et al. Robust T cell immunity in convalescent individuals with asymptomatic or mild COVID-19. *bioRxiv* 2020: 2020.06.29.174888.
70. Grifoni A, Weiskopf D, Ramirez SI, et al. Targets of T cell responses to SARS-CoV-2 coronavirus in humans with COVID-19 disease and unexposed individuals. *Cell* 2020; **181**(7): 1489-501. e15.
71. Dan JM, Mateus J, Kato Y, et al. Immunological memory to SARS-CoV-2 assessed for up to 8 months after infection. *Science* 2021; **371**(6529).
72. Peng Y, Mentzer AJ, Liu G, et al. Broad and strong memory CD4(+) and CD8(+) T cells induced by SARS-CoV-2 in UK convalescent individuals following COVID-19. *Nat Immunol* 2020; **21**(11): 1336-45.
73. Long Q-X, Tang X-J, Shi Q-L, et al. Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections. *NATURE MEDICINE* 2020; **26**(8): 1200-+.
74. Moderbacher CR, Ramirez SI, Dan JM, et al. Antigen-specific adaptive immunity to SARS-CoV-2 in acute COVID-19 and associations with age and disease severity. *Cell* 2020; **183**(4): 996-1012. e19.
75. Isho B, Abe KT, Zuo M, et al. Persistence of serum and saliva antibody responses to SARS-CoV-2 spike antigens in COVID-19 patients. *Science immunology* 2020; **5**(52): eabe5511.
76. Suthar MS, Zimmerman MG, Kauffman RC, et al. Rapid generation of neutralizing antibody responses in COVID-19 patients. *Cell Reports Medicine* 2020; **1**(3): 100040.
77. Ng KW, Faulkner N, Cornish GH, et al. Preexisting and de novo humoral immunity to SARS-CoV-2 in humans. *Science* 2020: eabe1107.
78. Nguyen-Contant P, Embong AK, Kanagaiah P, et al. S protein-reactive IgG and memory B cell production after human SARS-CoV-2 infection includes broad reactivity to the S2 subunit. *MBio* 2020; **11**(5): e01991-20.
79. Shrock E, Fujimura E, Kula T, et al. Viral epitope profiling of COVID-19 patients reveals cross-reactivity and correlates of severity. *Science* 2020; **370**(6520): eabd4250.
80. Song SK, Lee DH, Nam JH, et al. IgG Seroprevalence of COVID-19 among Individuals without a History of the Coronavirus Disease Infection in Daegu, Korea. *Journal of Korean medical science* 2020; **35**(29): e269.
81. Gudbjartsson DF, Norddahl GL, Melsted P, et al. Humoral Immune Response to SARS-CoV-2 in Iceland. *N Engl J Med* 2020; **383**(18): 1724-34.
82. Wajnberg A, Amanat F, Firpo A, et al. Robust neutralizing antibodies to SARS-CoV-2 infection persist for months. *Science* 2020; **370**(6521): 1227-30.

83. Thomas A, Oliver E, Baum H, et al. Evaluation of isotype specific salivary antibody assays for detecting previous SARS-CoV-2 infection in children and adults. medRxiv; 2022.
84. Garcia-Beltran WF, Lam EC, Astudillo MG, et al. COVID-19-neutralizing antibodies predict disease severity and survival. *Cell* 2021; **184**(2): 476-88.e11.
85. Nickbakhsh S, Ho A, Marques DFP, McMenamin J, Gunson RN, Murcia PR. Epidemiology of Seasonal Coronaviruses: Establishing the Context for the Emergence of Coronavirus Disease 2019. *The Journal of Infectious Diseases* 2020; **222**(1): 17-25.
86. Amanat F, Stadlbauer D, Strohmeier S, et al. A serological assay to detect SARS-CoV-2 seroconversion in humans. *Nature medicine* 2020; **26**(7): 1033-6.
87. Tan H-X, Lee WS, Wragg KM, et al. Adaptive immunity to human coronaviruses is widespread but low in magnitude. *Clinical & translational immunology* 2021; **10**(3): e1264.
88. Braun J, Loyal L, Frentsch M, et al. SARS-CoV-2-reactive T cells in healthy donors and patients with COVID-19. *Nature* 2020; **587**(7833): 270-4.
89. Sagar M, Reifler K, Rossi M, et al. Recent endemic coronavirus infection is associated with less-severe COVID-19. *J Clin Invest* 2021; **131**(1).
90. Hayward AC, Fragaszy EB, Bermingham A, et al. Comparative community burden and severity of seasonal and pandemic influenza: results of the Flu Watch cohort study. *The Lancet Respiratory Medicine* 2014; **2**(6): 445-54.
91. Aldridge RW, Lewer D, Beale S, et al. Seasonality and immunity to laboratory-confirmed seasonal coronaviruses (HCoV-NL63, HCoV-OC43, and HCoV-229E): results from the Flu Watch cohort study. *Wellcome Open Res* 2020; **5**: 52.
92. Simon AK, Hollander GA, McMichael A. Evolution of the immune system in humans from infancy to old age. *Proc Biol Sci* 2015; **282**(1821): 20143085.
93. Vogel K, Pierau M, Arra A, et al. Developmental induction of human T-cell responses against Candida albicans and Aspergillus fumigatus. *Sci Rep* 2018; **8**(1): 16904.
94. Basha S, Surendran N, Pichichero M. Immune responses in neonates. *Expert Rev Clin Immunol* 2014; **10**(9): 1171-84.
95. Pierau M, Lingel H, Vogel K, Arra A, Brunner-Weinzierl MC. CTLA-4-competent conventional T cells back up regulatory T cells to restrain memory T-helper type 2 cell responses. *Allergy* 2020; **75**(3): 684-7.
96. Lowery SA, Sariol A, Perlman S. Innate immune and inflammatory responses to SARS-CoV-2: Implications for COVID-19. *Cell Host & Microbe* 2021; **29**(7): 1052-62.
97. Barton LM, Duval EJ, Stroberg E, Ghosh S, Mukhopadhyay S. COVID-19 Autopsies, Oklahoma, USA. *American Journal of Clinical Pathology* 2020; **153**(6): 725-33.
98. Liu J, Li S, Liu J, et al. Longitudinal characteristics of lymphocyte responses and cytokine profiles in the peripheral blood of SARS-CoV-2 infected patients. *EBioMedicine* 2020; **55**: 102763.
99. Tian S, Xiong Y, Liu H, et al. Pathological study of the 2019 novel coronavirus disease (COVID-19) through postmortem core biopsies. *Modern Pathology* 2020; **33**(6): 1007-14.
100. Xu Z, Shi L, Wang Y, et al. Pathological findings of COVID-19 associated with acute respiratory distress syndrome. *The Lancet Respiratory Medicine* 2020; **8**(4): 420-2.
101. Yao X-H, He Z-C, Li T-Y, et al. Pathological evidence for residual SARS-CoV-2 in pulmonary tissues of a ready-for-discharge patient. *Cell Research* 2020; **30**(6): 541-3.
102. Chou J, Thomas PG, Randolph AG. Immunology of SARS-CoV-2 infection in children. *Nature Immunology* 2022; **23**(2): 177-85.
103. Jia R, Wang X, Liu P, et al. Mild Cytokine Elevation, Moderate CD4(+) T Cell Response and Abundant Antibody Production in Children with COVID-19. *Virol Sin* 2020; **35**(6): 734-43.

104. Cohen CA, Li APY, Hachim A, et al. SARS-CoV-2 specific T cell responses are lower in children and increase with age and time after infection. *Nature Communications* 2021; **12**(1): 4678.
105. Cohen R, Ashman M, Taha M-K, et al. Pediatric Infectious Disease Group (GPIP) position paper on the immune debt of the COVID-19 pandemic in childhood, how can we fill the immunity gap? *Infectious Diseases Now* 2021; **51**(5): 418-23.
106. Pierce CA, Sy S, Galen B, et al. Natural mucosal barriers and COVID-19 in children. *JCI insight* 2021; **6**(9).
107. Santos-Hovener C, Neuhauser HK, Rosario AS, et al. Serology- and PCR-based cumulative incidence of SARS-CoV-2 infection in adults in a successfully contained early hotspot (CoMoLo study), Germany, May to June 2020. *Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin* 2020; **25**(47).
108. Kevadiya BD, Machhi J, Herskovitz J, et al. Diagnostics for SARS-CoV-2 infections. *Nature Materials* 2021; **20**(5): 593-605.
109. NHS. How you get your NHS coronavirus (COVID-19) test result. 2022. https://www.nhs.uk/conditions/coronavirus-covid-19/testing/test-results/how-you-get-your-test-result/ (accessed 1 Jan 2023 2023).
110. Udugama B, Kadhiresan P, Kozlowski HN, et al. Diagnosing COVID-19: the disease and tools for detection. *ACS nano* 2020; **14**(4): 3822-35.
111. Oraganization WH. Advice on the use of point-of-care immunodiagnostic tests for COVID-19. 2020. https://www.who.int/docs/default-source/coronaviruse/sb-2020-1-poc-immunodiagnostics-2020-04-08-e.pdf?sfvrsn=4c26ac39_2 (accessed 3rd January 2023).
112. Wang C, Liu M, Wang Z, Li S, Deng Y, He N. Point-of-care diagnostics for infectious diseases: From methods to devices. *Nano Today* 2021; **37**: 101092.
113. Hou H, Wang T, Zhang B, et al. Detection of IgM and IgG antibodies in patients with coronavirus disease 2019. *Clinical & translational immunology* 2020; **9**(5): e1136.
114. Hoschler K, Ijaz S, Andrews N, et al. SARS-CoV-2 IgG detection in human oral fluids. *medRxiv* 2021: 2021.07.07.21260121.
115. van der Klis FRM, Mollema L, Berbers GAM, de Melker HE, Coutinho RA. Second national serum bank for population-based seroprevalence studies in the Netherlands. *Netherlands The Journal of Medicine* 2009; **67**(7): 301-8.
116. De Melker HE, Conyn-Van Spaendonck MAE. Immunosurveillance and the evaluation of national immunization programmes: a population-based approach. 1998.
117. Ministry of Housing CaLG. English indices of deprivation 2019.
118. Kang CK, Shin HM, Park WB, Kim HR. Why are children less affected than adults by severe acute respiratory syndrome coronavirus 2 infection? *Cell Mol Immunol* 2022; **19**(5): 555-7.
119. Balasubramanian S, Rao NM, Goenka A, Roderick M, Ramanan AV. Coronavirus Disease 2019 (COVID-19) in Children - What We Know So Far and What We Do Not. *Indian Pediatr* 2020; **57**(5): 435-42.
120. Warner S, Richter A, Stamataki Z, Kelly D. Understanding COVID-19: are children the key? *BMJ Paediatr Open* 2021; **5**(1): e001063.
121. Commission E. Ethical considerations for clinical trials on medicinal products conducted with the paediatric population. 2008. https://health.ec.europa.eu/system/files/2016-11/ethical_considerations_en_0.pdf (accessed 20th June 2020).
122. RCPCH. Growth Charts. 2009. https://www.rcpch.ac.uk/resources/growth-charts (accessed 20th June 2020).
123. Statistics OoN. Household questionnaire England. 2011. https://www.ons.gov.uk/file?uri=/census/censustransformationprogramme/consultations/the2

021censusinitialviewoncontentforenglandandwales/2011censusquestionnaireenglandh1.pdf (accessed 17 Jan 2022).
124. Hartley JEK, Levin K, Currie C. A new version of the HBSC Family Affluence Scale - FAS III : Scottish qualitative findings from the international FAS Development Study Child Indicators Research 2016. p. pp. 233-45.
125. Hobza V, Hamrik Z, Bucksch J, De Clercq B. The Family Affluence Scale as an Indicator for Socioeconomic Status: Validation on Regional Income Differences in the Czech Republic. *International journal of environmental research and public health* 2017; **14**(12): 1540.
126. Fidler Mis N, Kennedy K, Fewtrell M, et al. Role of Incentives in Long-term Nutritional and Growth Studies in Children. *J Pediatr Gastroenterol Nutr* 2018; **67**(6): 767-72.
127. Duggan J, Andrews N, Bown A, Brooks T, Migchelsen S. Evaluation of LIAISON SARS-CoV-2 S1/S2 IgG for detection of Anti-SARS-CoV-2 antibodies. In: PHE, editor. Assets Publishing Service .gov.uk; 2020.
128. PHE. Evaluation of Roche Elecsys Anti- SARS-CoV-2 serology assay for the detection of anti-SARS-CoV-2 antibodies, 2020.
129. Duggan J, Andrews N, Brooks T, Bown A, Migchelsen S. Evaluation of Roche Elecsys Anti- SARS-CoV-2 serology assay for the detection of anti-SARS-CoV-2 antibodies. In: England PH, editor. assets publishing service; 2020.
130. Horlock C. Enzyme-linked immunosorbent assay (ELISA). https://www.immunology.org/public-information/bitesized-immunology/experimental-techniques/enzyme-linked-immunosorbent-assay (accessed 5th September 2022).
131. Ltd RDI. Elecsys® Anti-SARS-CoV-2 S Immunoassay for the quantitative determination of antibodies to the SARS-CoV-2 spike protein. Elecsys® Anti-SARS-CoV-2 S Package Insert 2020-09, V10; Material Numbers 09289267190 and 09289275190: Roche Diagnostics International Ltd; 2020.
132. Duggan J, Otter A, Andrews N, Brooks T. Evaluation of Roche Elecsys Anti- SARS-CoV-2 S serology assay for the detection of anti-SARS-CoV-2 S antibodies. In: England PH, editor. 11 March 2021 ed. assets publishing service; 2021.
133. Andrews N. Assessment of sens and spec of RBD assay PHE; 2020.
134. Diagnostics MS. Immunoassays. 2022. https://www.mesoscale.com/en/technical_resources/our_technology/our_immunoassays (accessed 06 September 2022).
135. Johnson M, Wagstaffe HR, Gilmour KC, et al. Evaluation of a novel multiplexed assay for determining IgG levels and functional activity to SARS-CoV-2. *J Clin Virol* 2020; **130**: 104572.
136. Discovery MS. MSD multi-spot assay system - COVID-19 Serology Kits. In: Diagnostics MS, editor. wwwmesoscalecom; 2022.
137. Hoschler K, Ijaz S, Andrews N, et al. SARS Antibody Testing in Children: Development of Oral Fluid Assays for IgG Measurements. *Microbiol Spectr* 2022: e0078621.
138. (WHO) WHO. Manual for the laboratory diagnosis of measles and rubella virus infection, second edition. World Health Organization; 2007.
139. Grafström A, Schelin L. How to Select Representative Samples. *Scandinavian Journal of Statistics* 2014; **41**(2): 277-90.
140. Gill B. The English Indices of Deprivation 2015. In: Government DfCaL, editor. Assets publishing service gov.uk; 2015.
141. Ministry of housing calg. The English Indices of Deprivation 2019 (IoD2019) - Statistical release. In: Ministry of housing calg, editor. Ministry of housing, communities and local government; 2019.

142. Statistics TOfN. ODS postcode files. 8 September 2022 ed. NHS digital: NHS digital; 2022.
143. Wohland P BM, Rees P, Norman P, Lomax N, Clark S. NewETHPOP- Evaluation, Revision and Extension of Ethnic Population Projections.
144. O'Brien R, Potter-Collins A. 2011 Census analysis: Ethnicity and religion of the non-UK born population in England and Wales: 2011. 2011.
145. Statistics OfN. How religion has changed in England and Wales. Office for National Statistics; 2015.
146. Observatory TM. Migrants in the UK: An Overview. 2022. https://migrationobservatory.ox.ac.uk/resources/briefings/migrants-in-the-uk-an-overview/ (accessed 12th November 2022).
147. Stickney C. Population of the UK by country of birth and nationality. Office for National Statistics; 2021.
148. Guy P. Families and households. Office for National Statistics; 2022.
149. Huang AT, Garcia-Carreras B, Hitchings MDT, et al. A systematic review of antibody mediated immunity to coronaviruses: kinetics, correlates of protection, and association with severity. *Nat Commun* 2020; **11**(1): 4704.
150. Abel GA, Barclay ME, Payne RA. Adjusted indices of multiple deprivation to enable comparisons within and between constituent countries of the UK including an illustration using mortality rates. *BMJ Open* 2016; **6**(11): e012750.
151. Marmot M. Social determinants of health inequalities. *Lancet* 2005; **365**(9464): 1099-104.
152. Dorward J, Yu L-M, Hayward G, et al. Colchicine for COVID-19 in adults in the community (PRINCIPLE): a randomised, controlled, adaptive platform trial. *medRxiv* 2021: 2021.09.20.21263828.
153. Feehan AK, Velasco C, Fort D, et al. Racial and Workplace Disparities in Seroprevalence of SARS-CoV-2, Baton Rouge, Louisiana, USA. *Emerging infectious diseases* 2021; **27**(1).
154. Nutbeam D, Lloyd JE. Understanding and Responding to Health Literacy as a Social Determinant of Health. *Annual Review of Public Health* 2021; **42**(1): 159-73.
155. Hussain-Gambles M, Atkin K, Leese B. Why ethnic minority groups are under-represented in clinical trials: a review of the literature. *Health & social care in the community* 2004; **12**(5): 382-8.
156. Farooqi A, Jutlla K, Raghavan R, et al. Developing a toolkit for increasing the participation of black, Asian and minority ethnic communities in health and social care research. *BMC Medical Research Methodology* 2022; **22**(1): 17.
157. Wagner KS, White JM, Andrews NJ, et al. Immunity to tetanus and diphtheria in the UK in 2009. *Vaccine* 2012; **30**(49): 7111-7.
158. SCW. NHS National Immunisation Management Service. 2022. https://www.scwcsu.nhs.uk/services/nhs-immunisation-management-service/.
159. Statistics EE. Education provision: children under 5 years of age. gov.uk; 2022.
160. Pollán M, Pérez-Gómez B, Pastor-Barriuso R, et al. Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study. *The Lancet* 2020; **396**(10250): 535-44.
161. Silva AAMd, Lima-Neto LG, Azevedo CdMPESd, et al. Population-based seroprevalence of SARS-CoV-2 and the herd immunity threshold in Maranhao. *Revista de saude publica* 2020; **54**: 131.
162. Alvarez-Antonio C, Meza-Sanchez G, Calampa C, et al. Seroprevalence of anti-SARS-CoV-2 antibodies in Iquitos, Peru in July and August, 2020: a population-based study. *The Lancet Global health* 2021; **9**(7): e925-e31.

163. Bajema KL, Wiegand RE, Cuffe K, et al. Estimated SARS-CoV-2 Seroprevalence in the US as of September 2020. *JAMA internal medicine* 2021; **181**(4): 450-60.
164. Bendavid E, Mulaney B, Sood N, et al. COVID-19 antibody seroprevalence in Santa Clara County, California. *INTERNATIONAL JOURNAL OF EPIDEMIOLOGY* 2021; **50**(2): 410-9.
165. Bhuiyan TR, Akhtar M, Akter A, et al. Seroprevalence of SARS-CoV-2 antibodies in Bangladesh related to novel coronavirus infection. *IJID Regions (Online)* 2022; **2**: 198-203.
166. Biggs HM, Harris JB, Breakwell L, et al. Estimated Community Seroprevalence of SARS-CoV-2 Antibodies - Two Georgia Counties, April 28-May 3, 2020. *MMWR Morbidity and mortality weekly report* 2020; **69**(29): 965-70.
167. Bushnik T, Earl S, Clark J, Cabot J. COVID-19 infection in the Canadian household population. *Health reports* 2022; **33**(4): 24-33.
168. Diaz-Velez C, Failoc-Rojas VE, Valladares-Garrido MJ, et al. SARS-CoV-2 seroprevalence study in Lambayeque, Peru. June-July 2020. *PeerJ* 2021; **9**: e11210.
169. Dickson E, Palmateer NE, Murray J, et al. Enhanced surveillance of COVID-19 in Scotland: population-based seroprevalence surveillance for SARS-CoV-2 during the first wave of the epidemic. *Public health* 2021; **190**: 132-4.
170. Doi A, Iwata K, Kuroda H, et al. A cross-sectional follow up study to estimate seroprevalence of coronavirus disease 2019 in Kobe, Japan. *Medicine* 2021; **100**(48): e28066.
171. Bogogiannidou Z, Speletas M, Vontas A, et al. Repeated Leftover Serosurvey of SARS-CoV-2 IgG Antibodies in Greece, May to August 2020. *Vaccines* 2021; **9**(5).
172. Blankenberger J, Haile SR, Puhan MA, et al. Prediction of Past SARS-CoV-2 Infections: A Prospective Cohort Study Among Swiss Schoolchildren. *Frontiers in pediatrics* 2021; **9**: 710785.
173. Boey L, Roelants M, Merckx J, et al. Age-dependent seroprevalence of SARS-CoV-2 antibodies in school-aged children from areas with low and high community transmission. *European journal of pediatrics* 2022; **181**(2): 571-8.
174. Bordas A, Soriano-Arandes A, Subirana M, et al. Study protocol for monitoring SARS-CoV-2 infection and its determinants in Catalonia (Spain): an observational and participatory research approach in a Sentinel Network of Schools. *BMJ open* 2022; **12**(1): e055649.
175. Bullis SSM, Grebber B, Cook S, et al. SARS CoV-2 seroprevalence in a US school district during COVID-19. *BMJ paediatrics open* 2021; **5**(1): e001259.
176. Fontanet A, Tondeur L, Grant R, et al. SARS-CoV-2 infection in schools in a northern French city: a retrospective serological cohort study in an area of high transmission, France, January to April 2020. *Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin* 2021; **26**(15).
177. Kern A, Kuhlmann PH, Matl S, et al. Surveillance of Acute SARS-CoV-2 Infections in Elementary Schools and Daycare Facilities in Bavaria, Germany (09/2020-03/2021). *Frontiers in pediatrics* 2022; **10**: 888498.
178. Kirsten C, Kahre E, Blankenburg J, et al. Seroprevalence of SARS-CoV-2 in German secondary schools from October 2020 to July 2021: a longitudinal study. *Infection* 2022: 1-8.
179. Lopez L, Nguyen T, Weber G, et al. Seroprevalence of anti-SARS-CoV-2 IgG Antibodies in the Staff of a Public School System in the Midwestern United States. *medRxiv : the preprint server for health sciences* 2020.
180. Ulyte A, Radtke T, Abela IA, et al. Seroprevalence and immunity of SARS-CoV-2 infection in children and adolescents in schools in Switzerland: design for a longitudinal, school-based prospective cohort study. *Int J Public Health* 2020; **65**(9): 1549-57.

181. Dingens AS, Crawford KHD, Adler A, et al. Serological identification of SARS-CoV-2 infections among children visiting a hospital during the initial Seattle outbreak. *medRxiv : the preprint server for health sciences* 2020.
182. Bloomfield M, Pospisilova I, Cabelova T, et al. Searching for COVID-19 Antibodies in Czech Children-A Needle in the Haystack. *Frontiers in pediatrics* 2020; **8**: 597736.
183. Brinkmann F, Diebner HH, Matenar C, et al. Longitudinal Rise in Seroprevalence of SARS-CoV-2 Infections in Children in Western Germany-A Blind Spot in Epidemiology? *Infectious disease reports* 2021; **13**(4): 957-64.
184. Ladhani S. PHE: Prospective active national surveillance of preschools and primary schools for SARS-CoV-2 infection and transmission, 21 August 2020. 2020.
185. Halliday KE, Nguipdop-Djomo P, Oswald WE, et al. The COVID-19 Schools Infection Survey in England: Protocol and Participation Profile for a Prospective Observational Cohort Study. *JMIR Res Protoc* 2022; **11**(11): e34075.
186. Sepúlveda N, Drakeley C. Sample size determination for estimating antibody seroconversion rate under stable malaria transmission intensity. *Malar J* 2015; **14**: 141.
187. Elliott P, Haw D, Wang H, et al. Exponential growth, high prevalence of SARS-CoV-2, and vaccine effectiveness associated with the Delta variant. *Science* 2021; **374**(6574): eabl9551.
188. Mensah AA, Sinnathamby M, Zaidi A, et al. SARS-CoV-2 infections in children following the full re-opening of schools and the impact of national lockdown: Prospective, national observational cohort surveillance, July-December 2020, England. *J Infect* 2021; **82**(4): 67-74.
189. de Lusignan S, Joy M, Oke J, et al. Disparities in the excess risk of mortality in the first wave of COVID-19: Cross sectional study of the English sentinel network. *J Infect* 2020; **81**(5): 785-92.
190. England PH. National COVID-19 surveillance reports. 2022. https://www.gov.uk/government/publications/national-covid-19-surveillance-reports (accessed 19 Jan 2022).
191. StataCorp. Stata Statistical Software: Release 17. . College Station: TX: StataCorp LLC; 2021.
192. Coronavirus (COVID-19) Infection Survey, antibody and vaccination data, UK: 8 December 2021. In: (DHSC) OatDfHaSC, editor. ons.gov.uk; 2021.
193. A J. COVID-19 Schools Infection Survey, England: Round 6, pupil antibody data, June 2021. 2021. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/covid19schoolsinfectionsurveyengland/round6pupilantibodydatajune2021 (accessed 27 Jan 2022).
194. Statistics OfN. Coronavirus (COVID-19). 2022. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases (accessed 26 Jan 2022).
195. Naeimi R, Sepidarkish M, Mollalo A, et al. SARS-CoV-2 seroprevalence in children worldwide: a systematic review and meta-analysis. *eClinicalMedicine* 2023; **56**.
196. Rostami A, Sepidarkish M, Fazlzadeh A, et al. Update on SARS-CoV-2 seroprevalence: regional and worldwide. *Clin Microbiol Infect* 2021; **27**(12): 1762-71.
197. Ladhani SN, Baawuah F, Beckmann J, et al. SARS-CoV-2 infection and transmission in primary schools in England in June-December, 2020 (sKIDs): an active, prospective surveillance study. *Lancet Child Adolesc Health* 2021; **5**(6): 417-27.
198. Workforce Team ND. NHS Workforce Statistics - November 2021 (Including selected provisional statistics for December 2021). In: Digital N, editor.; 2022.

199. Roarty C, Tonry C, McFetridge L, et al. Kinetics and seroprevalence of SARS-CoV-2 antibodies in children. *The Lancet Infectious diseases* 2020: S1473-3099(20)30884-7.
200. Ladage D, Rösgen D, Schreiner C, et al. Persisting Antibody Response to SARS-CoV-2 in a Local Austrian Population. *Frontiers in Medicine* 2021; **8**.
201. Weisberg SP, Connors TJ, Zhu Y, et al. Distinct antibody responses to SARS-CoV-2 in children and adults across the COVID-19 clinical spectrum. *Nat Immunol* 2021; **22**(1): 25-31.
202. Ijaz S, Dicks S, Jegatheesan K, et al. Mapping of SARS-CoV-2 IgM and IgG in gingival crevicular fluid: Antibody dynamics and linkage to severity of COVID-19 in hospital inpatients. *J Infect* 2022; **85**(2): 152-60.
203. UKHSA. GOV.UK Coronavirus (COVID-19) in the UK. https://coronavirus.data.gov.uk/details/cases?areaType=nation&areaName=England.
204. Zhang S, Xu K, Li C, et al. Long-Term Kinetics of SARS-CoV-2 Antibodies and Impact of Inactivated Vaccine on SARS-CoV-2 Antibodies Based on a COVID-19 Patients Cohort. *Frontiers in Immunology* 2022; **13**.
205. Dowell AC, Butler MS, Jinks E, et al. Children develop robust and sustained cross-reactive spike-specific immune responses to SARS-CoV-2 infection. *Nat Immunol* 2022; **23**(1): 40-9.
206. Ortega N, Ribes M, Vidal M, et al. Seven-month kinetics of SARS-CoV-2 antibodies and role of pre-existing antibodies to human coronaviruses. *Nature Communications* 2021; **12**(1): 4740.
207. Dan JM, Mateus J, Kato Y, et al. Immunological memory to SARS-CoV-2 assessed for up to 8 months after infection. *Science* 2021; **371**(6529): eabf4063.
208. Faulkner N, Ng KW, Wu MY, et al. Reduced antibody cross-reactivity following infection with B.1.1.7 than with parental SARS-CoV-2 strains. *eLife* 2021; **10**: e69317.
209. Vöhringer HS, Sanderson T, Sinnott M, et al. Genomic reconstruction of the SARS-CoV-2 epidemic in England. *Nature* 2021; **600**(7889): 506-11.
210. Cohen JI, Dropulic L, Wang K, et al. Comparison of Levels of Nasal, Salivary, and Plasma Antibody to Severe Acute Respiratory Syndrome Coronavirus 2 During Natural Infection and After Vaccination. *Clinical Infectious Diseases* 2023; **76**(8): 1391-9.
211. Dhakal S, Yu T, Yin A, et al. Reconsideration of Antinucleocapsid IgG Antibody as a Marker of SARS-CoV-2 Infection Postvaccination for Mild COVID-19 Patients. *Open Forum Infect Dis* 2023; **10**(1): ofac677.
212. UKHSA. How to take an oral fluid swab. In: UKHSA, editor. gov.uk; 2014.
213. Lavinder JJ, Ippolito GC. Boosted immunity to the common cold might protect children from COVID-19. *Nature Immunology* 2022; **23**(1): 8-10.
214. Aydillo T, Rombauts A, Stadlbauer D, et al. Immunological imprinting of the antibody response in COVID-19 patients. *Nature Communications* 2021; **12**(1): 3781.
215. Edridge AWD, Kaczorowska J, Hoste ACR, et al. Seasonal coronavirus protective immunity is short-lasting. *Nature Medicine* 2020; **26**(11): 1691-3.
216. Galanti M, Shaman J. Direct Observation of Repeated Infections With Endemic Coronaviruses. *The Journal of Infectious Diseases* 2020; **223**(3): 409-15.
217. Kiyuka PK, Agoti CN, Munywoki PK, et al. Human Coronavirus NL63 Molecular Epidemiology and Evolutionary Patterns in Rural Coastal Kenya. *J Infect Dis* 2018; **217**(11): 1728-39.
218. Corr M, Christie S, Watson C, et al. Seroprevalence of SARS-CoV-2 antibodies in children of United Kingdom healthcare workers: a prospective multicentre cohort study protocol. *BMJ Open* 2020; **10**(11): e041661.
219. Ratcliffe H, Tiley KS, Andrews N, et al. Community seroprevalence of SARS-CoV-2 in children and adolescents in England, 2019-2021. *Arch Dis Child* 2022.

220. Duggan J, Andrews N, Bown A, Brooks T, Migchelsen S. Evaluation of DiaSorin LIAISON SARS- CoV-2 S1/S2 IgG serology assay for the detection of anti-SARS-CoV-2 antibodies. In: England PH, editor. Assets publishing service; 2020.
221. Kohler H. Novel vaccine concept based on back-boost effect in viral infection. *Vaccine* 2015; **33**(29): 3274-5.
222. Lavell AHA, Sikkens JJ, Edridge AWD, et al. Recent infection with HCoV-OC43 may be associated with protection against SARS-CoV-2 infection. *iScience* 2022; **25**(10): 105105.
223. Westerhuis BM, de Bruin E, Chandler FD, et al. Homologous and heterologous antibodies to coronavirus 229E, NL63, OC43, HKU1, SARS, MERS and SARS-CoV-2 antigens in an age stratified cross-sectional serosurvey in a large tertiary hospital in The Netherlands. *medRxiv* 2020: 2020.08.21.20177857.
224. Gombar S, Bergquist T, Pejaver V, et al. SARS-CoV-2 infection and COVID-19 severity in individuals with prior seasonal coronavirus infection. *Diagnostic Microbiology and Infectious Disease* 2021; **100**(2): 115338.
225. Anderson EM, Goodwin EC, Verma A, et al. Seasonal human coronavirus antibodies are boosted upon SARS-CoV-2 infection but not associated with protection. *medRxiv* 2020: 2020.11.06.20227215.
226. Vono M, Huttner A, Lemeille S, et al. Robust innate responses to SARS-CoV-2 in children resolve faster than in adults without compromising adaptive immunity. *Cell Rep* 2021; **37**(1): 109773.
227. Swadling L, Diniz MO, Schmidt NM, et al. Pre-existing polymerase-specific T cells expand in abortive seronegative SARS-CoV-2. *Nature* 2022; **601**(7891): 110-7.
228. Moss P. The T cell immune response against SARS-CoV-2. *Nature Immunology* 2022; **23**(2): 186-93.
229. Chun JY, Jeong H, Kim Y. Identifying susceptibility of children and adolescents to the Omicron variant (B.1.1.529). *BMC Medicine* 2022; **20**(1): 451.
230. Peacock TP, Brown JC, Zhou J, et al. The altered entry pathway and antigenic distance of the SARS-CoV-2 Omicron variant map to separate domains of spike protein. *bioRxiv* 2022: 2021.12.31.474653.
231. Zhao H, Lu L, Peng Z, et al. SARS-CoV-2 Omicron variant shows less efficient replication and fusion activity when compared with Delta variant in TMPRSS2-expressed cells. *Emerg Microbes Infect* 2022; **11**(1): 277-83.
232. Yehia BR, Winegar A, Fogel R, et al. Association of race with mortality among patients hospitalized with coronavirus disease 2019 (COVID-19) at 92 US hospitals. *JAMA network open* 2020; **3**(8): e2018039-e.
233. Caldicott F. Data sharing during this public health emergency2020. https://www.gov.uk/government/speeches/data-sharing-during-this-public-health-emergency (accessed.
234. Tessier E, Edelstein M, Tsang C, et al. Monitoring the COVID-19 immunisation programme through a national immunisation Management system - England's experience. *Int J Med Inform* 2023; **170**: 104974.
235. Kelly E, Greenland M, Whalley PD, et al. Reactogenicity, immunogenicity and breakthrough infections following heterologous or fractional second dose COVID-19 vaccination in adolescents (COM-COV-3): A randomised controlled trial. 41st Annual Meeting of the European Society for Paediatric Infectious Diseases (ESPID); 2023.
236. Rae PM. COVID-19 Series: Future Pandemics, WHO Pandemic Treaty - Episode 108. In: Stingelin-Giles DN, editor. Royal Society of Medicine 2023.
237. Pandit JA, Radin JM, Quer G, Topol EJ. Smartphone apps in the COVID-19 pandemic. *Nature Biotechnology* 2022; **40**(7): 1013-22.

238. Ward H, Atchison CJ, Whitaker M, et al. Antibody prevalence for SARS-CoV-2 in England following first peak of the pandemic: REACT2 study in 100,000 adults. medRxiv; 2020. p. 2020.08.12.20173690.
239. Chaudhuri K, Chakrabarti A, Chandan JS, Bandyopadhyay S. COVID-19 vaccine hesitancy in the UK: a longitudinal household cross-sectional study. *BMC Public Health* 2022; **22**(1): 104.
240. Liu X, Shaw RH, Stuart ASV, et al. Safety and immunogenicity of heterologous versus homologous prime-boost schedules with an adenoviral vectored and mRNA COVID-19 vaccine (Com-COV): a single-blind, randomised, non-inferiority trial. *The Lancet* 2021; **398**(10303): 856-69.
241. Administration USFaD. Title 21 - Food and Drugs Chapter - Food and drug administration department of health and human services Subchapter A - General . In: Services UDoHaH, editor. CFR- Code of Federal Regulations; 2023.
242. Diseases NIoAaI. ClinRegs- Aggregating clinical research regulations from around the globe. National Institute of Allergy and Infectious Diseases; 2023.
243. Raleigh V. What is happening to life expectancy in England?2022. https://www.kingsfund.org.uk/publications/whats-happening-life-expectancy-england (accessed 1st February 2023).
244. Exworthy M, Blane D, Marmot M. Tackling health inequalities in the United Kingdom: the progress and pitfalls of policy. *Health Serv Res* 2003; **38**(6 Pt 2): 1905-21.
245. Organization WH. Health inequities and their causes. 2018. https://www.who.int/news-room/facts-in-pictures/detail/health-inequities-and-their-causes (accessed 1st February 2023).
246. Aspinall E. COVID-19 Timeline. 2022. https://bfpg.co.uk/2020/04/covid-19-timeline/.
247. Bai M-S, Miao C-Y, Zhang Y, Xue Y, Jia F-Y, Du L. COVID-19 and mental health disorders in children and adolescents (Review). *Psychiatry Research* 2022; **317**: 114881.
248. Godfrey K, Bagust L, Baird J, et al. Evaluation of the expanded Southampton pilot study (Phase 2) for use of saliva-based lamp testing in asymptomatic populations: Final report, 16th November 2020: University of Southampton, 2020.