Infection at mucosal surfaces:

SARS-CoV-2

Katherine Emary

Green Templeton College

University of Oxford

This dissertation is submitted for the degree of Doctor of Philosophy

Trinity 2022

“At first people refuse to believe that a strange new thing can be done. Then they begin to hope it can be done. Then they see it can be done. Then it is done”

— Frances Hodgson Burnett, The Secret Garden

Abstract

The pandemic caused by the SARS-CoV-2 virus has had a profound international impact, causing widespread morbidity and mortality as well as substantial indirect secondary consequences of the restrictions implemented to constrain the virus. During 2020 the University of Oxford developed a SARS-CoV-2 vaccine using a simian adenoviral vector (ChAdOx1 nCoV-19). Pre-clinical studies demonstrated reduced viral load in lower respiratory tract samples in vaccinated animals, but no difference in the upper respiratory samples. In the first part of this thesis, I investigate the impact of ChAdOx1 nCoV-19 vaccination on SARS-CoV-2 infection in the upper respiratory tract and stool of trial participants.

Vaccine trial participants aged 18 and over from 19 UK study sites were randomised to receive the ChAdOx1 nCoV-19 vaccine or a comparator vaccine (MenACWY). From May 2020 to June 2021 participants were asked to perform weekly self-swabs of their nose and throat in addition to attending their study site for swabbing if they experienced any primary SARS-CoV-2 symptoms (fever of 37.8°C or above, cough, shortness of breath, ageusia, anosmia). Samples underwent NAAT testing and were retrieved for genome sequencing. A selected group of trial participants were asked to provide stool samples during acute SARS-CoV-2 infection.

Adherence to weekly sampling was high and over 10% of participants tested positive during the study period. Infections in vaccinated participants were less likely to be symptomatic, were less severe, had higher cycle threshold values and a shorter duration of NAAT positive swabs compared to controls. A novel variant B.1.1.7 with notable spike protein mutations emerged during the study period for which vaccine efficacy for primary symptomatic Covid-19 infection was lower at 58.3% (95% CI 29.1-75.5) than for a preceding, common variant,

B.1.177 (80.7%, 95% CI 67.1-88.6). Symptomatic infection was more likely to be associated with detectable RNA in stool however vaccination status was not related with detectable RNA in stool. There was no demonstrable difference in amino acid or nucleotide substitutions between sequences obtained from control or vaccinated participants for infections caused by B.1.177 or B.1.1.7. Intrahost viral diversity was detected in 90% (135/150) sequences with 93% of iSNVs occurring in only a single participant. This work demonstrates the impact of ChAdOx1 nCoV-19 vaccination on viral dynamics and efficacy against B.1.1.7 and emphasises the value of intensive surveillance of vaccine trial participants.

In Part 2 of this thesis (concerning work conducted prior to the SARS-CoV-2 pandemic, presented in Appendix 4), I investigate the population dynamics of *Salmonella* Typhi using a controlled human infection model. Participants were exposed to equal proportions of two genetically similar but distinct strains and 50% of individuals were diagnosed with typhoid. Both strains were detected at diagnosis from all participants who became bacteraemic showing no evidence of a single cell bottleneck. This study illustrates how co-infection within a controlled human infection model may be used to develop further understanding into the fundamentals of population pathogenesis.

Publications

The following are publications that have resulted from work carried out within this thesis and related projects.

Related to work associated with or specifically included in this thesis:

Emary KRW, Golubchik, T et al. 'Efficacy of ChAdOx1 nCoV-19 (AZD1222) vaccine against SARS-CoV-2 variant of concern 202012/01 (B.1.1.7): an exploratory analysis of a randomised controlled trial'

Lancet. 2021 Apr 10; 397(10282): 1351–1362.

Voysey M, Clemens SAC, Madhi SA (…), **Emary KRW**, et al. 'Safety and efficacy of the ChAdOx1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: an interim analysis of four randomised controlled trials in Brazil, South Africa, and the UK'

Lancet. 2021 Jan 9; 397(10269): 99–111. doi: 10.1016/S0140-6736(20)32661-1

Correction in: Lancet. 2021 9-15 January; 397(10269): 98.

Ramasamy MN, Minassian AM, Ewer KJ, (…) Emary KRW, et al. 'Safety and immunogenicity of ChAdOx1 nCoV-19 vaccine administered in a prime-boost regimen in young and old adults (COV002): a single-blind, randomised, controlled, phase 2/3 trial'

Lancet. 2020 Dec 19; 396(10267): 1979–1993. doi: 10.1016/S0140-6736(20)32466-1

Correction in: Lancet. 2020 Dec 19; 396(10267): 1978, Lancet. 2021 Apr 10; 397(10282): 1350.

Voysey M, Clemens SAC, Madhi SA (…), Emary KRW, et al. 'Single-dose administration and the influence of the timing of the booster dose on immunogenicity and efficacy of ChAdOx1 nCoV-19 (AZD1222) vaccine: a pooled analysis of four randomised trials'

Lancet. 2021 Mar 6; 397(10277): 881–891. doi: 10.1016/S0140-6736(21)00432-3

Correction in: Lancet. 2021 Mar 6; 397(10277): 880.

Related to associated projects:

Clemens SAC, Folegatti PM, Emary KRW, et al. 'Efficacy of ChAdOx1 nCoV-19 (AZD1222) vaccine against SARS-CoV-2 lineages circulating in Brazil'

Nat Commun. 2021; 12: 5861.

Hodgson SH, (…) Emary KRW, Pollard AJ. 'What defines an efficacious COVID-19 vaccine? A review of the challenges assessing the clinical efficacy of vaccines against SARS-CoV-2'

Lancet Infect Dis. 2021 Feb; 21(2): e26–e35.

Vanderslott S, Emary KRW, et al. 'Vaccine nationalism and internationalism: perspectives of COVID-19 vaccine trial participants in the United Kingdom'

BMJ Glob Health. 2021; 6(10): e006305.

Vanderslott S, Palmer A, (…) Emary KRW, et al. 'Co-producing Human and Animal Experimental Subjects: Exploring the Views of UK COVID-19 Vaccine Trial Participants on Animal Testing'

Science, Technology, & Human Values. November 2021.

COVID-19 impact statement

My original DPhil project was to test a novel oral vaccine for *Salmonella* Paratyphi A with a controlled human infection model of *Salmonella* Paratyphi A. During vaccine manufacture, I ran a small challenge study looking at population dynamics within a *Salmonella* Typhi challenge model to gain experience of the infection model. Whilst this study was running, I was also planning and preparing the submission package for the *Salmonella* Paratyphi A project, the manufacture of which was delayed during 2019 and only was finally produced in February 2020. At this time all non-COVID work was paused, and the *Salmonella* Paratyphi A study was unable to be reviewed by the University, external ethics committees or proceed to recruitment. The study was finally granted REC approval in October 2021 and opened to participants in March 2022 but does not form part of my thesis.

At the onset of the pandemic, I changed to work full time on the COVID-19 vaccine trials investigating ChAdOx1 nCoV-19 vaccination and SARS-CoV-2 infection. As well as working as a clinician on these trials from inception until April 2022, I developed and ran three sub-studies of the main vaccine trials. A change in focus from a bacterial to viral pathogen, staff leaving during the pandemic and other significant commitments for supervisors also necessitated changes of supervisors in 2021.

This thesis is therefore divided into two parts:

Part 1 of this thesis covers work completed during two COVID-19 vaccine trials encompassing longitudinal testing of participants, genomic analysis of isolates and stool analysis.

Part 2 of this thesis is presented in appendix 4 and focuses on a *Salmonella* Typhi study of pathogen dynamics.

Whilst there are some common themes between the two parts, these are separate pieces of academic work and should be considered as such.

During the pandemic I also project managed and performed data collection and analysis of a large mixed methods qualitative study on COV001 participants exploring motivations and experiences of being a first in human vaccine trial participant during the pandemic. This will also not be included in my thesis but has led to two publications so far with further publications planned. I also co-managed a cohort of twenty medical students during 2020

who were seconded from medical school to the trial. I supervised their training and work on the vaccine trials, provided support and mentorship. They made an invaluable contribution to trial work during this time.

Finally, but equally critically, COVID-19 has impacted my home life in terms of access to childcare during working hours and multiple periods of isolation, testing and latterly illness for COVID-19 within our household.

Acknowledgements

My time as a DPhil student, as for many, has been a challenge; however, the nature of it was unimaginable when I started nearly four years ago. I started off with the anticipation I would be involved in two bacterial studies with a total number of participants being no more than 100 with lots of laboratory work. The reality became studying a viral disease not identified until 4 terms into my studies, going through multiple lockdowns restricting non-COVID-19 studies and lab work, being involved in multiple COVID-19 studies with over 10,000 participants, organising over 350,000 swabs (not to mention the many swabs my family took for every cough and fever!) and being part of a team that developed a successful vaccine. It feels like many years have been packed into the past two. It has required perseverance, a flexible mindset, plenty of hope and the help of those acknowledged below.

This thesis would not exist without the enormous commitment of the participants in these studies. I am humbled by their generosity and willingness to be involved despite the acknowledged risks and particularly in the case of COVID-19, many unknowns when the projects were started. Your trust and dedication to participating in these studies is amazing, for which me (and the world) owe you a debt of thanks.

I have received an enormous amount of support to complete this work for which I am truly grateful. Special thanks to my supervisors:

Professor Sir Andrew Pollard encouraged me and gave me the freedom and backing to pursue opportunities. I would like to thank Professor Sir Pollard for continuing to supervise me despite his many significant commitments during the pandemic, for his vision, clarity of thought and particularly for the many insights and opportunities he has given me.

Professor Maheshi Ramasamy and Dr Tanya Golubchik, both of whom agreed to supervise me as the focus of my DPhil had to change. Professor Ramasamy has provided regular supervision which has been immensely valuable, endured endless conversations about swabs and stool, has demanded the best of me and helped me develop as a researcher and writer. Dr Golubchik has been very supportive and patient as I learned to code and realised I had to be more realistic as to what I could achieve in the timeframe. To you both my heartfelt gratitude. Professor Moxon provided me with fascinating conversations about the origins and history of the bottleneck hypothesis, its implications and valuable observations about the study design and data. Professor MacLennan for providing a broader context and much

needed perspective throughout. Dr Jenny Hill supervised the typhoid work before leaving OVG and could not have been more supportive as I found my feet and was extremely patient teaching me the laboratory techniques.

My college supervisor Siân Rees has been an absolute rock through some difficult times. Thank you for being a sounding board, dining companion and kindred soul. I feel very lucky to have met you during this process, thank you for asking the right questions that helped me reframe my thoughts.

The OVG team are inspirational, and it has been a privilege to work with you all. The world knows of your achievements and having been behind the scenes I know your hard work and dedication are phenomenal. None of these projects would have been possible without each and every cog in the machine playing their part. Thank you to everyone for all that you do to try and improve the lives of children and adults through research.

OVG swab team, Jessica Larwood and Lucky Khan - my profound thanks for all that you did to make things run smoothly, I couldn't have done it without you. The multiple COG members, BMS staff, clinicians, and individuals around the UK who I spoke to and asked if they would help retrieve individual swabs from busy hospital laboratories and get them to sequencing - I know this created extra work at a tiring time, thank you for your kindness and willingness to help. Merryn Voysey, Dan Phillips, and Melanie Greenland for being supportive and patient in helping me manage and analyse the enormous amount of data. I have learnt a huge amount courtesy of you.

I also had the good fortune to work with the Lighthouse Teams - the brilliant Dr Jodie Hay, Dr Tony Cox OBE and Farah Tahmasebi, Alex Hope from Deloitte and Martin Shanahan and Amina Elmi from DHSC. Thank you for providing so much support to these projects and working tirelessly in all your many roles but always remaining cheerful and helpful in the face of further requests. To my colleagues across the pond Jessica Bonnell from AZ and Lina Survilaite from IHMA who supported the stool project - it was a pleasure working with you both, thank you for all your help. Emily Thornton and Dana Costigan thank you for being so welcoming over at the WIMM and for your enthusiasm and dedication to helping with the stool project, it was a pleasure to work with you.

Thank you to the OVG adult fellows that have helped with the projects I have been involved in from being on call, to doing endpoint assessments and dealing with swab queries. Thank you also for providing much needed respite in terms of cups of tea, good quality office

chocolate, head-clearing walks around the Churchill site and laughter: Arabella Stuart, Naina McCann, Kitty Smith, Blanché Oguti, Brama Hanumunthadu, Nasir Kanji, Rob Shaw, Iain Turnbull, Jonathan Kwok plus all the COVID Jenner (particularly Susanne Hodgson, Hazel Morrison, Mimi Hou and Suzy Jackson) and OVG Paediatric fellows.

Lastly, and critically, my whole family have been very supportive, and their patience, kindness and generosity have been unending. My parents never got to see me complete my first degree but to them I owe them everything. Thank you for the best start in life, raising me to be curious and your unerring encouragement. My husband Hugo and my children particularly have given up a lot to allow me to complete this, thank you with all my heart.

Declaration and Contribution

This thesis is based on data generated from the Oxford Vaccine Centre's COVID-19 Vaccine Trials and the result of a separate clinical study performed at the Oxford Vaccine Group. This was achievable because of the extensive efforts of multiple study teams, industry, academic and governmental partners as well as the generosity of the participants in these trials. Specific contributions are described below but every individual mentioned in the lists given in the appendix has played a role in delivering the COVID-19 vaccine trials. All data analyses and interpretations presented are my own and where necessary the contribution of others is specified.

Dr Katherine Emary

12th September 2022

Part 1: SARS-CoV-2

Complete lists for the Oxford COVID-19 Vaccine Trial Group and COG UK are included in Appendices.

Chapter 2: Methods

Conduct and supervision of the clinical studies: Oxford COVID-19 Vaccine Trial Group (UK sites), Emma Plested, Michelle Fuskova, Parvinder Aley, Hannah Robinson

Chapter 3-5 inclusive: Weekly surveillance and sequencing

Project initiation: Susanne Hodgson, Maheshi Ramasamy, **Katherine Emary**

Project management: Katherine Emary, Maheshi Ramasamy, Jessica Larwood, Liaquat Khan, Tanya Golubchik

Leadership and teams at Lighthouse Laboratories: Jodie Hay, Tony Cox, Farah Tahmasebi, Sarah Galloway

DHSC: Martin Shanahan, Amina Elmi, Harry Mayhew

Deloitte: Alex Hope

AstraZeneca: Claire Crook

Sequencing: David Bonsall, Jeremy Ratcliffe

Ampheus Project team

Oxford Viral Sequencing Group, Welcome Centre for Human Genetics, University of Oxford, UK: Christophe Fraser, David Buck, Angie Green, George MacIntyre-Cockett, Paolo Piazza, John A Todd, Amy Trebes, Luca Ferretti

Oxford Viral Sequencing Group, Big Data Institute, University of Oxford, UK: Laura Thomson

Leadership and teams at Lighthouse Laboratories: Jodie Hay, Tony Cox, Farah Tahmasebi, Sarah Galloway

COG UK: Sharon Peacock, Catherine Ludden, Beth Blane, Cristina Ariani, Leigh Jackson

Sanger: Jeffery Barrett

Statistical and data management advice: Merryn Voysey, Daniel Phillips, Gu-Lung Lin

Data team: Agnes Eordogh

Supervision: Maheshi Ramasamy, Andrew J Pollard, Tanya Golubchik

Chapter 6: ChAdOx1 nCoV-19 vaccination and the detection of virus and immunoglobulin in the stool of SARS-CoV-2 infected individuals

Sample processing: Lina Survilaite and her team at IHMA (Schaumburg, US)

Assay development: Emily Thornton, Dana Costigan

Sample processing: Emily Thornton, Dana Costigan, **Katherine Emary**

Project management: Jessica Bonnell, Lina Survilaite, **Katherine Emary,** Maheshi Ramasamy

Laboratory and sample management: Christina Dold, Elizabeth Jones

Statistical advice: Merryn Voysey

Data team: Agnes Eordogh

Supervision: Maheshi Ramasamy, Andrew J Pollard

Part 2: Bottlenecks

Conduct and supervision of the clinical study: Katherine Emary, Maria Moore, Hannah Robinson, Louise Willis, Naomi Douglas, Jamie Burbage, Rachel White, Claudio di Maso, Philomena Mweu, Juliana Ballaminut, Brama Hanumunthadu, Catherine Smith, Blanché Oguti, Maheshi Ramasamy, Brian Angus, Andrew J Pollard

Sample processing: Jennifer Hill, Elizabeth Jones, Marije Verheul, Lisa Stockdale, Spyridoula Marinou, Jennifer Clarke, Susanna Camara, Sonu Shrestha, Jaclyn Bowman, Charlie Brown-O'Sullivan, Christina Dold, **Katherine Emary**

Oxford University Hospitals NHS Foundation Trust laboratories: Marcus Morgan

PCR probe design and validation: Jennifer Hill, Marije Verheul, Jennifer Clarke

Project Management: David Smith, **Katherine Emary**, Jennifer Hill

Statistical advice: Xinxue Liu, Melanie Greenland (particularly regarding the probability calculations in Appendix 4.2)

Supervision: Maheshi Ramasamy, Andrew J Pollard, Jennifer Hill, Calman MacLennan, E Richard Moxon

Thesis formatting and design: adapted from CSD thesis template by Kayla Friedman and Malcolm Morgan used under CC BY 4.0. [1]

Illustrations: Referenced where relevant. Figures Figure 10.1, Figure 10.3, Figure 10.4, Figure 10.10 were created with BioRender.com.

Contents

Contents xiv

List of Tables xix

List of Figures xxii

List of Appendices xxv

List of Abbreviations and Acronyms xxvi

Part 1: SARS-CoV-2 28

1 Introduction 29
1.1 Virology 29
1.1.1 Taxonomy 29
1.1.2 Structure 30
1.1.3 Proteins 31
1.1.4 Cellular entry 32
1.1.5 Replication, and transcription 32
1.2 Epidemiology 34
1.3 Molecular epidemiology and viral evolution 36
1.3.1 Nomenclature 39
1.3.2 Pre-Alpha SARS-CoV-2 variants 40
1.3.3 Alpha (B.1.1.7) variant of concern 41
1.3.4 Significance of variants in the context of vaccination 43
1.3.5 Genomic surveillance 43
1.4 Transmission 44
1.4.1 Routes 44
1.4.2 Viral kinetics 45
1.4.3 Impact of symptoms on transmission 47
1.5 Diagnostics and Testing 48
1.5.1 PCR 48
1.5.2 Lateral flow testing 49
1.5.3 Antibody testing 50
1.5.4 Surveillance (wastewater) 50
1.6 Disease 50
1.6.1 Symptoms 50
1.6.2 Disease spectrum 51
1.7 Immune response to infection 53
1.7.1 Innate response 53
1.7.2 Humoral response 54
1.7.3 Cellular response 57
1.7.4 Mucosal response 58
1.7.5 Correlates of protection 59
1.7.6 Immune determinants of disease phenotype 62
1.7.7 Reinfection 64
1.8 Vaccination 65

1.8.1 Background to the development of SARS-CoV-2 vaccines 65
1.8.2 Vaccine platforms 77
1.8.3 UK Vaccination Programme 82
1.9 Non-Pharmaceutical Interventions for prevention of infection and disease 83
1.9.1 Masks 83
1.9.2 Handwashing 84
1.9.3 Social Distancing 84
1.10 Summary 84
1.11 Specific Aims 85

2 Methods 88
2.1 COV001 89
2.1.1 Study design 90
2.1.2 Study procedures 92
2.1.3 Ethics statement 93
2.1.4 Funding 93
2.2 COV002 94
2.2.1 Study design 94
2.2.2 Study procedures 94
2.2.3 Home swabbing pathway 101
2.2.4 Clinical oversight 101
2.2.5 Ethics statement 102
2.2.6 Funding 102
2.3 Symptomatic pathway 102
2.3.1 Procedures 102
2.4 Sample processing for the COV001 and COV002 trials 104
2.4.1 SARS-CoV-2 NAAT testing for symptomatic participants 104
2.4.2 Stool processing 104
2.4.3 SARS-CoV-2 sequencing 104
2.5 Data capture 105
2.6 Statistics 105

3 Longitudinal participant sampling for SARS-CoV-2 infection in a ChAdOx1 nCoV-19 vaccine trial 106
3.1 Introduction 106
3.2 Methods 108
3.2.1 Weekly self-swabbing 109
3.2.2 Matching and data capture 110
3.2.3 Legal safeguards for those repeatedly positive 111
3.2.4 Assessment of false positives 111
3.2.5 Statistical analysis 111
3.3 Results 113
3.3.1 Demographics of COV002 participants 113
3.3.2 Swab sources and results 116
3.3.3 Adherence to weekly swabbing 121
3.3.4 Ct values and gene targets 126
3.3.5 Assessment of false positives 127
3.3.6 Participants who tested positive 129
3.3.7 Symptoms and severity of disease in trial participants 133
3.4 Discussion 136

4 The impact of ChAdOx1 nCoV-19 vaccination on SARS-CoV-2 symptoms and viral load 144
4.1 Introduction 144
4.2 Methods 146
4.2.1 Trial methods and data collection 146
4.2.2 Immunology 147
4.2.3 Statistical analysis 148
4.3 Results 149
4.3.1 Analysis dataset for impact of vaccination on symptom and viral parameters 150
4.3.2 Cases 151
4.3.3 Symptom status and vaccination 152
4.3.4 Cycle thresholds, symptom status and vaccination 155
4.3.5 Predictors of Ct value 162
4.3.6 Asymptomatic infections with low Ct values 163
4.3.7 Duration of swab positivity, symptom status and vaccination 165
4.3.8 Predictors of NAAT positive duration 170
4.4 Discussion 170

5 ChAdOx-1 nCoV-19 vaccination and the B.1.1.7 variant of SARS-CoV-2 179
5.1 Introduction 179
5.2 Methods 185
5.2.1 Participants 185
5.2.2 Ct values and duration 186
5.2.3 S gene target failure 186
5.2.4 Ascertainment of sequences 186
5.2.5 Genome reconstruction & Phylogeny reconstruction 187
5.2.6 Minor variant calling 188
5.2.7 Statistical analysis 189
5.3 Results 190
5.3.1 Lineage, geography, and molecular clock 190
5.3.2 Demographics by sequencing result 200
5.3.3 Lineage, vaccine, and symptom status 202
5.3.4 Viral load 203
5.3.5 Duration 207
5.3.6 Efficacy 209
5.3.7 Consensus genome analysis 213
5.3.8 Intrahost viral diversity 224
5.4 Discussion 236

6 ChAdOx1 nCoV-19 vaccination and the detection of virus and immunoglobulin in the stool of SARS-CoV-2 infected individuals 248
6.1 Introduction 248
6.2 Methods 253
6.2.1 Participants 254
6.2.2 Sample collection 254
6.2.3 Sample processing 256
6.2.4 Sample size 258
6.2.5 Statistical analyses 262
6.3 Results 263
6.3.1 Participant characteristics 263

6.3.2 SARS-CoV-2 RNA was detectable in stool from participants of all vaccine status groups 264
6.3.3 Symptomatic infection was associated with stool sample SARS-CoV-2 PCR positivity 267
6.3.4 Gastrointestinal symptoms were common but not related to stool PCR result 269
6.3.5 Vaccination status was not associated with stool PCR result 270
6.3.6 Stool SARS-CoV-2 viral load increased with time from episode start to sampling 271
6.3.7 Anti-spike Immunoglobulin G and A was detectable in stool 273
6.4 Discussion 279

7 Concluding remarks and future directions 287
7.1 Summary of principal findings 288
7.1.1 Longitudinal sampling of vaccine trial participants 288
7.1.2 Application of longitudinal sampling to investigate symptom status, viral load and duration of PCR positivity and vaccination 288
7.1.3 Viral parameters and vaccine efficacy by SARS-CoV-2 variant 289
7.1.4 Genomic analyses by vaccination status 290
7.1.5 SARS-CoV-2 infection and the gastrointestinal tract 290
7.2 Limitations 290
7.3 Further work 293
7.4 Implications and future directions 294
7.5 Final words 296

8 References 297

9 Appendices 333

Appendix 1 Oxford Vaccine Trial Group, and other COVID-19 vaccine trial acknowledgements 334

Appendix 2 COG-UK consortium acknowledgements 354

Appendix 3 SARS-CoV-2 NAAT assays used at study sites for the symptomatic pathway (COV001 & COV002) 364

Appendix 4 Part 2: *Salmonella* Typhi 366

10 Bottlenecks 367
10.1 Introduction ("Bottlenecks") 367
10.1.1 Early evidence for population bottlenecks in experimental infection .. 370
10.1.2 Recent experimental work 372
10.1.3 Clinical evidence 375
10.1.4 Experimental models to investigate population dynamics 376
10.1.5 Natural *Salmonella* Typhi infection 378
10.1.6 Epidemiology & Burden 378
10.1.7 Infectious Dose & Transmission 380
10.1.8 Disease 381
10.1.9 Diagnostics 381
10.1.10 The Salmonella Typhi CHIM as a model for exploring population dynamics within a human host 382
10.1.11 Aims for Part 2 of this thesis 387

10.2 Methods (“Bottlenecks”) 388
10.2.1 Study Design 388
10.2.2 Laboratory and other methods 391
10.2.3 Data capture 399
10.2.4 Statistical analysis 399
10.2.5 Approvals 399
10.2.6 Funding 400
10.3 Results (“Bottlenecks”) 400
10.3.1 Recruitment & Participants 400
10.3.2 Challenge dose 401
10.3.3 Diagnosis of typhoid fever 401
10.3.4 Safety 403
10.3.5 PCR results from diagnosed participants 403
10.3.6 Blood Quantitation results 406
10.3.7 Stool PCR results 408
10.4 Discussion (“Bottlenecks”) 408
Further work 415

Appendix 4.1 Full Inclusion & Exclusion criteria for the ‘Bottlenecks’ Study 416
10.4.1 Inclusion Criteria 416
10.4.2 Exclusion Criteria 417

Appendix 4.2 Probability calculations for detection of strains 422

List of Tables

Table 1.1: Nomenclature of common SARS-CoV-2 variants. 40

Table 1.2: Defining mutations of B.1.1.7 (Alpha) Variant of Concern 42

Table 1.3: Associations of specific spike mutations with potential functional significance 42

Table 1.4: SARS-CoV-2 vaccines in Phase 3 and 4 clinical trials. 73

Table 2.1: COV001 Study Objectives 89

Table 2.2: Final table of groups for COV001. 92

Table 2.3: Study Objectives for COV002 94

Table 2.4: Study groups in COV002. 96

Table 3.1: Demographics of COV002 participants. 114

Table 3.2: Age distribution of COV002 participants 115

Table 3.3: Source of SARS-CoV-2 NAAT positive swabs. 117

Table 3.4: PCR results by Processing laboratory 118

Table 3.5: Linear model for predictors of the percentage of self-swabs returned. 125

Table 3.6: Linear model for predictors of the percentage of self-swabs returned in participants whilst blinded. 125

Table 3.7: Multiple logistic regression investigating those who did not return any weekly swabs.126

Table 3.8: Demographic details of those who tested positive from any swab source .. 130

Table 3.9: Generalised linear model for binary data assessing predictors for testing positive for SARS-CoV-2 131

Table 3.10: Total number of positive swabs for participants who tested positive at least once. 131

Table 3.11: Vaccination status at first positive swab for those who tested positive..... 132

Table 3.12: Demographics of all participants who returned at least one positive swab by symptom status. 134

Table 3.13: COVID-19 severity in COV002 participants who tested positive at least once during the trial. 136

Table 4.1: Demographics for included participants that tested positive 151

Table 4.2: Demographics of participants by symptom status. 153

Table 4.3: Symptom status by vaccination status at the time of first positive swab 154

Table 4.4: WHO progression score for symptomatic participants..... 155

Table 4.5: Minimum and first minimum Ct values for first episode of infection..... 156

Table 4.6: Multiple linear regression of variables associated with infection minimum Ct value. 162

Table 4.7: Comparison of demographic factors of asymptomatic participants with low Ct values 164

Table 4.8: Number of NAAT positive swabs per participant from any source. 166

Table 4.9: Multiple linear regression investigating factors associated with duration of NAAT positivity. 170

Table 5.1: Sequencing and protocols used to obtain sequences from COV002 participants. 190

Table 5.2: SARS-CoV-2 lineages identified from consensus sequences from COV002 participants. 192

Table 5.3: Details of participants by sequencing result. 200

Table 5.4: Comparison of symptom status by lineage and vaccine status..... 203

Table 5.5: Vaccine efficacy against B.1.177 and B.1.1.7 using sequencing data. 212

Table 5.6: Vaccine efficacy using SGTF in combination with sequencing data as a proxy for B.1.1.7. 213

Table 5.7: Median amino acid substitutions in regions of the genome by vaccination status 216

Table 5.8: Median nucleotide substitutions in regions of the genome by vaccination status 217

Table 5.9: Comparison of high frequency amino acid substitutions occurring in control and vaccinated participants. 221

Table 5.10: Number of unique iSNVs and median iSNV count by vaccination status and stratified by viral load status. 232

Table 5.11: Most frequent iSNVs observed for B.1.177 and B.1.1.7. 235

Table 6.1: Studies investigating stool shedding of SARS-CoV-2 RNA from infected participants. 259

Table 6.2: Sample sizes required for the detection of a range of differences between groups 260

Table 6.3: Characteristics of participants who were SARS-CoV-2 PCR positive during the stool sampling timeframe..... 264

Table 6.4: Demographics of participants who provided a stool sample. 265

Table 6.5: Dose groups in vaccinated participants. 266

Table 6.6: Symptom status and stool PCR result..... 267

Table 6.7: Generalised linear model for binary data assessing predictors for having detectable SARS-CoV-2 in stool. 269

Table 6.8: Gastrointestinal symptoms in SARS-CoV-2 infected participants..... 270

Table 6.9: Gastrointestinal symptoms and stool PCR results 270

Table 6.10: Stool PCR result and vaccination status 271

Table 6.11: Multiple linear regression model of predictors of stool viral load. 273

Table 10.1: Summary of study procedures 390

Table 10.2: Primer sequences 397

Table 10.3: Challenge participant demographics 401

Table 10.4: Challenge dose and ratio of strains within challenge inoculum for each participant 401

Table 10.5: Relationship between number of bacteria in the bloodstream and blood sample volume 422

Table 10.6: Probability of detecting both strains when present in equal proportions in a range of bacterial concentrations and volume scenarios. 424

Table 10.7: Probability of detecting both strains when present in a 60:40 ratio in a range of bacterial concentrations and volume scenarios. 425

Table 10.8: Probability of detecting both strains when present in an 80:20 ratio in a range of bacterial concentrations and volume scenarios. 425

List of Figures

Figure 1.1: Coronavirus family. 30

Figure 1.2: Schematic presentation of the SARS-CoV-2 genome structure. 31

Figure 1.3: SARS-CoV-2 replication. 33

Figure 1.4: Daily new cases in the UK 35

Figure 1.5: Diagram showing SARS-CoV-2 phylogeny by clade and PANGO lineage. 37

Figure 1.6: Normalised frequencies of sequences by Clade 38

Figure 1.7: SARS-CoV-2 viral load dynamics and period of infectiousness. 46

Figure 1.8: Schematic showing IgG/IgM/IgA/Neutralising Ab response over time 55

Figure 3.1: Enrolment and blind status of participants by age group 116

Figure 3.2: Daily counts for the UK 119

Figure 3.3: NAAT results from swabs taken at initial symptomatic visits. 120

Figure 3.4: Number of participants and percentage of swabs returned by week 121

Figure 3.5: Percentage of swabs returned by COV002 participants 122

Figure 3.6: Overall adherence with weekly self-swabbing in blinded participants by age group and sex. 123

Figure 3.7: Percentage of swabs returned by age group over time. 124

Figure 3.8: Gene targets and Ct values of weekly swabs over time. 127

Figure 3.9: Assessment of possible false positives. 128

Figure 3.10: Plot of potential false positives over time 129

Figure 3.11: Comparison of adherence between those who tested positive versus those who did not. 133

Figure 3.12: Proportion of symptom status by age group. 135

Figure 4.1: Participants who tested positive at least once during the study period. 150

Figure 4.2: Cases of infection by week by vaccine status. 152

Figure 4.3: Symptom status of infective episode by vaccination status. 154

Figure 4.4: Infection minimum Ct value, symptom status and vaccination status. 157

Figure 4.5: Number of positive self-swabs by vaccination and symptom status group. 159

Figure 4.6: First minimum Ct value and symptom and vaccination status..... 161

Figure 4.7: Neutralisation titres from asymptomatic vaccinees. 165

Figure 4.8: Duration of NAAT positive swabs, vaccination and symptom status..... 167

Figure 4.9: Duration of positive swabs by symptom and vaccination status..... 168

Figure 4.10: Number of swabs by vaccination and symptom status 169

Figure 5.1: Maximum likelihood consensus phylogenetic tree of SARS-CoV-2 genomes 193

Figure 5.2: Distribution of lineages across time from study sites. 194

Figure 5.3: Time-scaled consensus phylogenetic tree 195

Figure 5.4: Root to tip distance plotted for B.1.177 and B.1.1.7 sequences..... 196

Figure 5.5: Phylogenetic divergence of samples relative to the common ancestor..... 197

Figure 5.6: B.1.1.7 sequences show increased deviation from molecular clock. 199

Figure 5.7: Time distribution of different sequencing result groups. 201

Figure 5.8: Infection minimum Ct value by lineage result. 204

Figure 5.9: Infection minimum Ct value by sequencing result and vaccination status. 206

Figure 5.10: Duration of positive swabs for symptomatic participants..... 208

Figure 5.11: Consort diagram for participants in the COV002 trial included in efficacy analyses. 209

Figure 5.12: Lineage outcomes for NAAT positive participants..... 210

Figure 5.13: Maximum likelihood consensus phylogenetic tree by vaccination status 214

Figure 5.14: Total calculated nucleotide substitutions 219

Figure 5.15: Total substitution rate adjusted for genome coverage by Ct group..... 220

Figure 5.16: Amino acid substitutions by vaccination status in B.1.177 sequences. ... 222

Figure 5.17: Amino acid substitutions by vaccination status in B.1.1.7 sequences. 223

Figure 5.18: Relationship between Ct value and unique mapped reads. 225

Figure 5.19: Non-linear relationship between viral load (unique mapped reads) and iSNV counts. 226

Figure 5.20: Relationship between mean allele frequency and unique mapped reads . 227

Figure 5.21: Number of iSNVs per individual in group for high viral load samples. .. 228

Figure 5.22: Ct values and iSNV counts and episode onset to sampling interval. 230

Figure 5.23: Relationship between calendar date and iSNV count per sequence..... 231

Figure 5.24: Median MAF and cMAF by lineage and vaccination status 233

Figure 5.25: iSNVs by gene..... 234

Figure 6.1: Viral load in tissues of rhesus macaques challenged with SARS-CoV-2 .. 252

Figure 6.2: Demonstration of change in sample size required to detect differences between groups..... 261

Figure 6.3: Stool SARS-CoV-2 PCR result, symptom and vaccination status..... 268

Figure 6.4: Stool SARS-CoV-2 viral load and vaccination status..... 272

Figure 6.5: Stool SARS-CoV-2 viral load and interval between episode start and stool collection..... 272

Figure 6.6: Stool anti-spike IgG and IgA 274

Figure 6.7: Stool anti-spike IgG and IgA by vaccination status and stool PCR result . 275

Figure 6.8: Stool anti-spike IgG/IgA ratios 276

Figure 6.9: Interval between episode start date and IgG and IgA results by vaccination status and stool PCR result. 277

Figure 6.10: Stool viral load and stool IgG and IgA results. 278

Figure 6.11: Post vaccination serum anti-spike IgG and anti-spike IgG in stool at time of infection. 279

Figure 10.1: The effect of bottlenecks on genetic diversity. 369

Figure 10.2: The experimental design of a WITS experiment 373

Figure 10.3: Ideal features for an experimental model to explore population dynamics.377

Figure 10.4: Pathogenesis of Enteric fever 379

Figure 10.5: Recruitment flow chart for Bottlenecks study..... 400

Figure 10.6: Blood and stool culture results 402

Figure 10.7: PCR results from blood samples for diagnosed participants..... 405

Figure 10.8: Bacterial dynamics over time per participant..... 407

Figure 10.9: Stool culture results 408

Figure 10.10: Sources of experimental and theoretical information from different within-host levels 414

List of Appendices

Appendix 1 Oxford Vaccine Trial Group, and other COVID-19 vaccine trial acknowledgements 334

Appendix 2 COG-UK consortium acknowledgements 354

Appendix 3 SARS-CoV-2 NAAT assays used at study sites for the symptomatic pathway (COV001 & COV002) 364

Appendix 4 Part 2: *Salmonella* Typhi 366

Appendix 4.1 Full Inclusion & Exclusion criteria for the 'Bottlenecks' Study 416

Appendix 4.2 Probability calculations for detection of strains 422

List of Abbreviations and Acronyms

Abs 260	Absorption at 260 nm
BAL	Bronchoalveolar lavage
CAP	Community acquired pneumonia
CFU	Colony forming unit
CHIM	Controlled human infection model
COG-UK	UK COVID-19 genomics consortium
COP	Correlate(s) of protection
CRF	Case report form
Ct	Cycle threshold
CTD	C terminal domain
CTL	Cytotoxic T-Lymphocytes
DAMPS	Damage-associated molecular patterns
DHSC	Department of Health and Social Care
DPP4	Dipeptidyl peptidase 4
ELISpot assay	Enzyme linked immunospot assay
GCE	Genome copy equivalents
HADS	Hospital Anxiety and Depression Scale
HAP	Hospital acquired pneumonia
HCW	Healthcare worker
ICTV	International Committee on Taxonomy of Viruses
ID_{50}	Infectious dose required to infect 50% of subjects
IFR	Infection Fatality Ratio
ILC	Innate Lymphoid Cells
ISG	Interferon stimulated gene
LD_{50}	Lethal dose for 50% of subjects
MAF	Mean allele frequency
MHRA	Medicines and Healthcare products Regulatory Authority
NAAT	Nucleic acid amplification test
NHP	Non-human primate
NTD	N-terminal domain
PAMPs	Pathogen associated molecular patterns
PANGO	Phylogenetic Assignment of Names Global Outbreak LINeages
PBMC	Peripheral blood mononuclear cells
PBS-T	PBS plus 0.05% Tween™-20
PRRs	Pattern recognition receptors
PS80	Polysorbate 80
PCR	Polymerase Chain Reaction
qPCR	quantitative Polymerase Chain Reaction
RBD	Receptor binding domain
RdRP	RNA dependent RNA polymerase
RDT	Rapid diagnostic test
RTC	Replication and transcription complex
RT-PCR	Real Time polymerase chain reaction
RXN	Reaction

SNV	Single nucleotide variant
STAMP	Sequence tag-based analysis of microbial populations
TCID	Median tissue culture infectious dose
UTR	Untranslated region
VLP	Viral-like particle
VOC	Variant of concern
vp	Viral particles
VUI	Variant under investigation
WGS	Whole genome sequencing
WHO	World Health Organization
WITS	Wild-type isogenic tagged strains
WT	Wild type

Part 1: SARS-CoV-2

1 Introduction

In this introductory chapter I will discuss the background information on SARS-CoV-2 and vaccination as it relates to this thesis while reserving specific precursory comments relevant to each chapter for the individual chapter introductions. Unless otherwise specified the focus of this introduction is on SARS-CoV-2 prior to the availability of COVID-19 vaccines, up to and including the emergence of the Alpha variant (B.1.1.7) as this is the context for the period during which data collection was undertaken.

1.1 VIROLOGY

Severe acute respiratory syndrome-related coronavirus (SARS-CoV-2) is an enveloped single-stranded, positive sense RNA virus first identified in December 2019. [2] It has subsequently been shown that it can infect many different vertebrate host species. [3,4] The origin of this virus is most likely a result of zoonotic transmission to humans, from an unidentified animal reservoir or intermediate host with striking epidemiological similarities to the emergence of SARS. [5,6]

1.1.1 Taxonomy

The *Coronaviridae* are a family of viruses, consisting of two subfamilies, one of which, *Orthocoronavirinae* is further subclassified into four genera (Figure 1.1, after ICTV 2018). SARS-CoV-2 is a Betacoronavirus. It is related to two other zoonotic coronaviruses both of which have also caused outbreaks of human disease in the past two decades, *Severe acute respiratory syndrome-related coronavirus* (SARS-CoV) and *Middle East respiratory syndrome-related coronavirus* (MERS-CoV). [7-9]

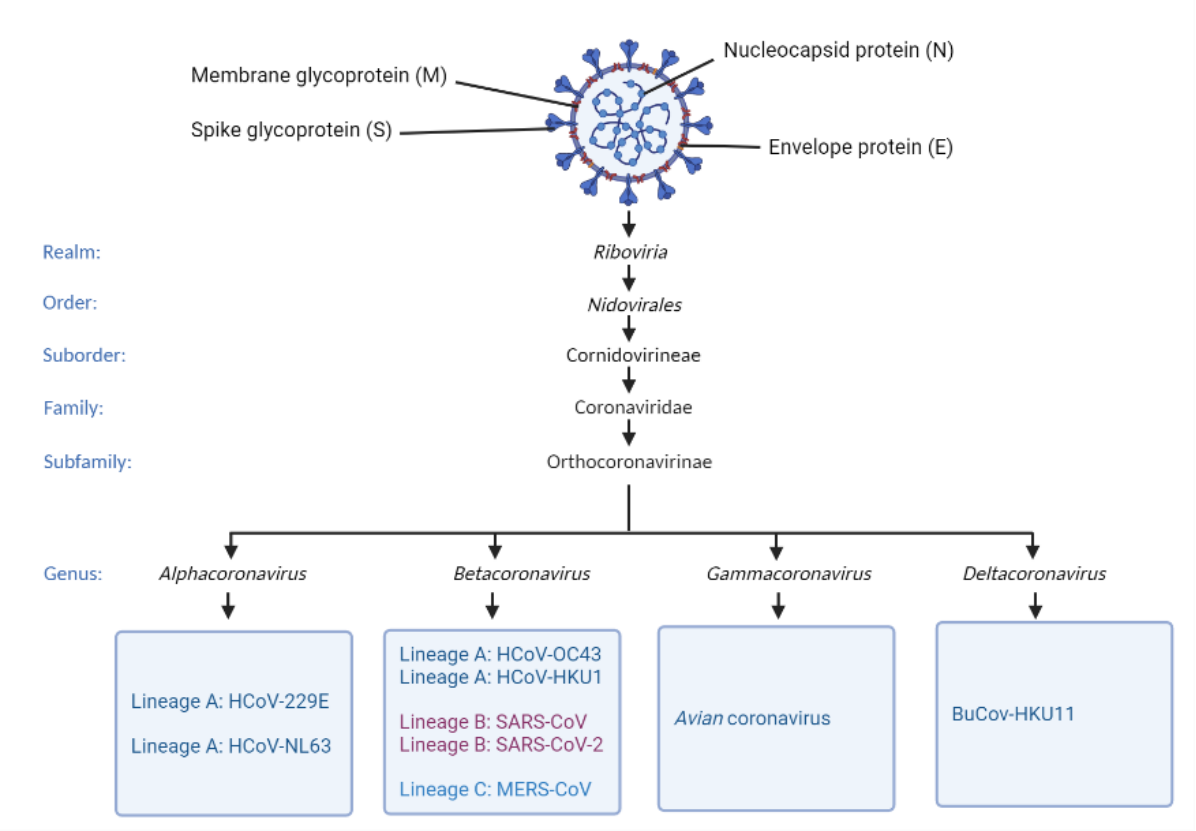


Figure 1.1: Coronavirus family. Figure adapted from Tang et al. (2020). [10] Reproduced under CC BY 4.0. [1]

In contrast to SARS-CoV-2, SARS-CoV is less readily spread between individuals and causes predominantly a lower respiratory tract infection, whilst MERS appears less transmissible but causes more severe human disease.

1.1.2 Structure

The SARS-CoV-2 genome was first sequenced from the bronchoalveolar lavage fluid sampled from a 41-year-old male patient who presented with a severe acute respiratory syndrome in Wuhan, China. [2] The genome is comparatively large for an RNA virus at 29-30 kb in size, and is arranged from the 5'UTR (untranslated region) with two large open reading frames encoding the non-structural proteins followed by genes for the major structural proteins: Spike,

Envelope, Membrane and Nucleocapsid (Figure 1.2). [2] The SARS-CoV-2 genome is most closely related to a number of bat derived coronaviruses and is more distant from SARS-CoV and MERS-CoV, sharing approximately 79% and 50% sequence identity respectively. [11]

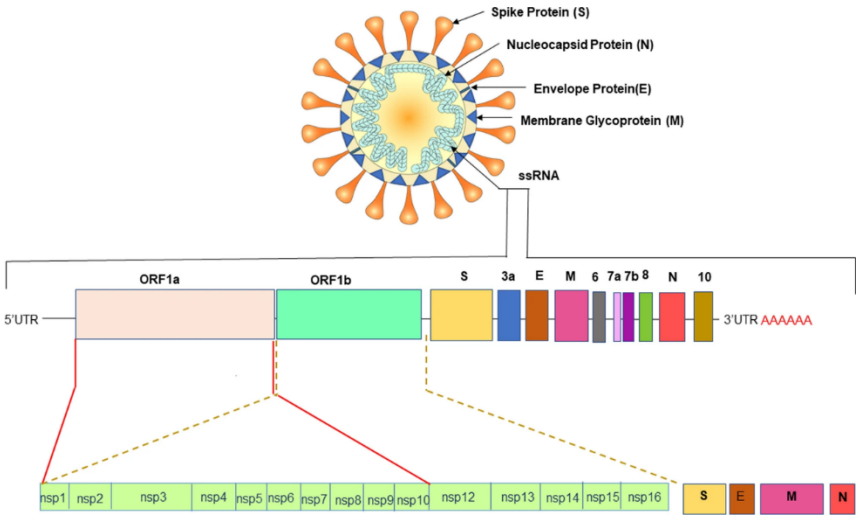


Figure 1.2: Schematic presentation of the SARS-CoV-2 genome structure. SARS-CoV-2 has a spherical structure. The virus has an outer lipid envelope, covered with spike glycoprotein. The SARS-CoV-2 represents a typical Betacoronavirus genome organisation. The full-length RNA genome comprises of approximately 29,903 nucleotides and has a replicase complex (comprised of ORF1a and ORF1b) at the 5′UTR. The ORF1a encodes for nsp1–nsp10, while ORF1b encodes for nsp1–nsp16. Four genes that encode for the Structural proteins: Spike gene, Envelope gene, Membrane gene, Nucleocapsid gene and a poly (A) tail at the 3′UTR. The accessory genes are distributed in between the structural genes. Figure and legend from Rastogi et al. (2020). [12] Reproduction licensed under CC BY 4.0. [1]

1.1.3 Proteins

1.1.3.1 Spike

The spike (S) protein is a trimeric glycoprotein which exists on the viral surface and begets the virus its characteristic appearance on electron microscopy. SARS-CoV-2 binds to human ACE-2 as a receptor, similarly to SARS-CoV, but in a tighter manner, to achieve cellular entry via membrane fusion. [2] For this reason whole spike protein, and specific regions of spike, were obvious vaccine targets, particularly considering the success of this strategy in the development of other coronavirus vaccines. [13]

1.1.3.2 The Envelope, Nucleocapsid and Membrane proteins

The envelope proteins are small polypeptides present in the viral membrane, which may play a role in pathogenesis but are also found in the endoplasmic reticulum and Golgi-complex where they play a role in virion assembly. [14] The nucleocapsid protein has a role in packaging viral RNA in the virions and interacts with the other structural proteins during various stages of replication. The membrane glycoprotein sits within the membrane and is likely to be important in the process of budding during replication as well as in the shape of the envelope. [15,16]

1.1.4 Cellular entry

Initially the S protein is synthesised as a precursor protein which is then cleaved into subunits S1 and S2. The S1 subunit contains the receptor-binding domain (RBD), N terminal domain (NTD) and C-terminal domains (CTD1 and 2). It is the RBD that is responsible for cell receptor recognition and whilst the role of the NTD is not fully understood, it may be related to the complex positional requirements for membrane fusion. The RBD has distinct 'up' and 'down' conformations, it is only in the former that receptor binding occurs. [17] In the pre-fusion position the S1 subunit binds to ACE-2 and thereby produces conformational change resulting in its dissociation. [18,19] Receptor binding also results in cleavage of the second subunit, S2 by transmembrane serine protease 2 (TMPRSS2). This causes the projection of the fusion protein from S2 into the host cell membrane; membrane fusion is achieved as the host and viral membranes are pulled together via this interaction. [15]

1.1.5 Replication, and transcription

Fusion of the SARS-CoV-2 viral particle with the cellular membrane uncoats the virus and leads to the translation of genomic RNA (Figure 1.3). ORF1a and ORF1b translation produces

polyproteins pp1a and pp1b which are then cleaved by self-activating proteases that undergo processing to form the non-structural proteins which make up the viral replication and transcription complex (RTC).

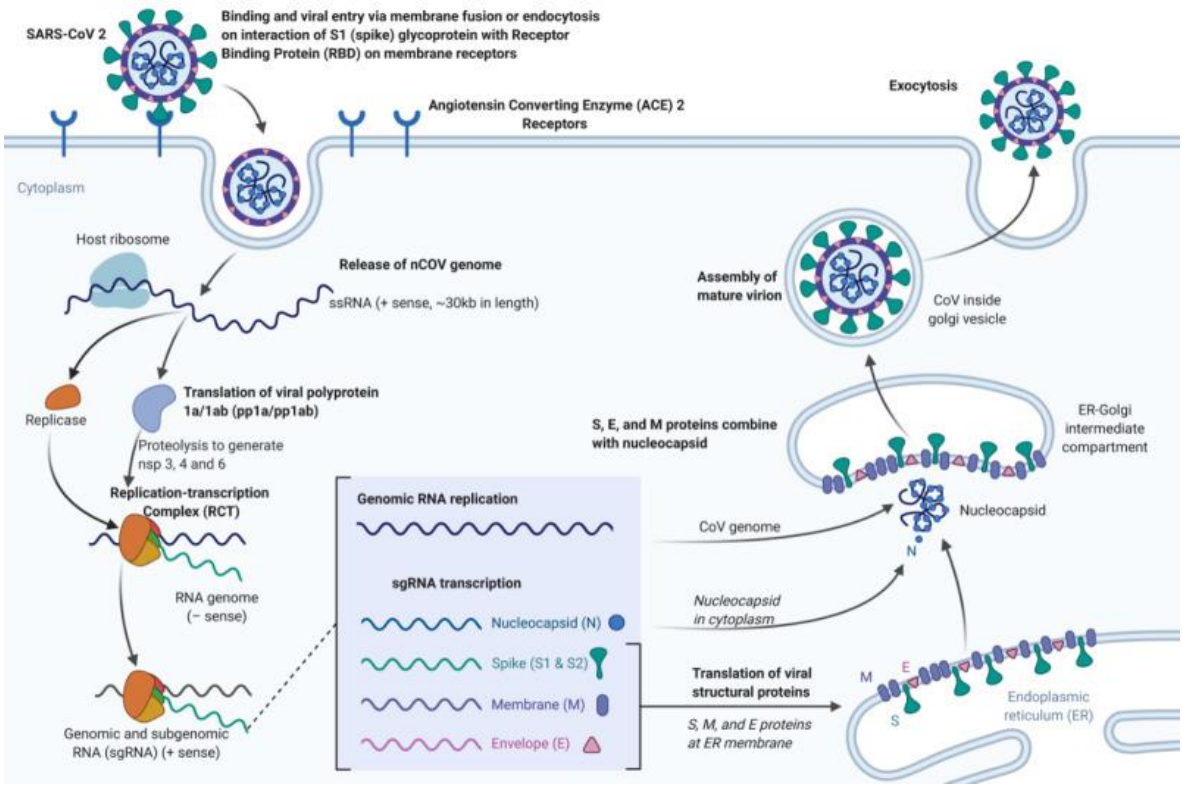


Figure 1.3: SARS-CoV-2 replication. Figure from Cascella et al. (2022) [20]. Figure contributed by Rohan Bir Singh, MD and reproduced under CC BY 4.0. [1]

The role of the RTC is to synthesise and proof-read RNA, a characteristic not typical of most viruses. [21] The RTC is made up of the RNA dependent RNA polymerase (RdRP) and three accessory subunits. The RdRP is a key enzyme in the replication of SARS-CoV-2 and the target of inhibition by remdesivir. [22] Double membrane vesicles are membrane bound structures which permit sgRNA replication for structural and accessory proteins within the cell. Once structural proteins have been translated, they are trafficked to the endoplasmic reticulum and are then combined with the genomic RNA to form virions. The host cell can then release virions by exocytosis. [16,23]

1.2 EPIDEMIOLOGY

By 8th September 2022, there had been over 607 million infections worldwide and more than 6.5 million deaths recorded. [24] These figures are likely to underestimate case numbers and deaths due to the different ways nation states record figures and attribute and record deaths, as well as differential access to testing both by geography, timing in the pandemic and political factors. Excess mortality estimates however suggest deaths 2-4 times that of official mortality numbers. [25]

Since 1st March 2020 until the time of writing in early September 2022, the UK has had over 23.5 million new confirmed cases (Figure 1.4). Cases initially were concentrated in two main waves (March-May 2020 and November 2020-February 2021). High levels of transmission were sustained from June 2021, with establishment of the Delta variant in the UK, until a large wave caused by the Omicron variant occurred in December 2021 which persisted for the first half of 2022. By September 2022 the UK has recorded over 205,000 deaths, disproportionately affecting those in older age groups, those in the most deprived areas and those living within the care sector. [26]

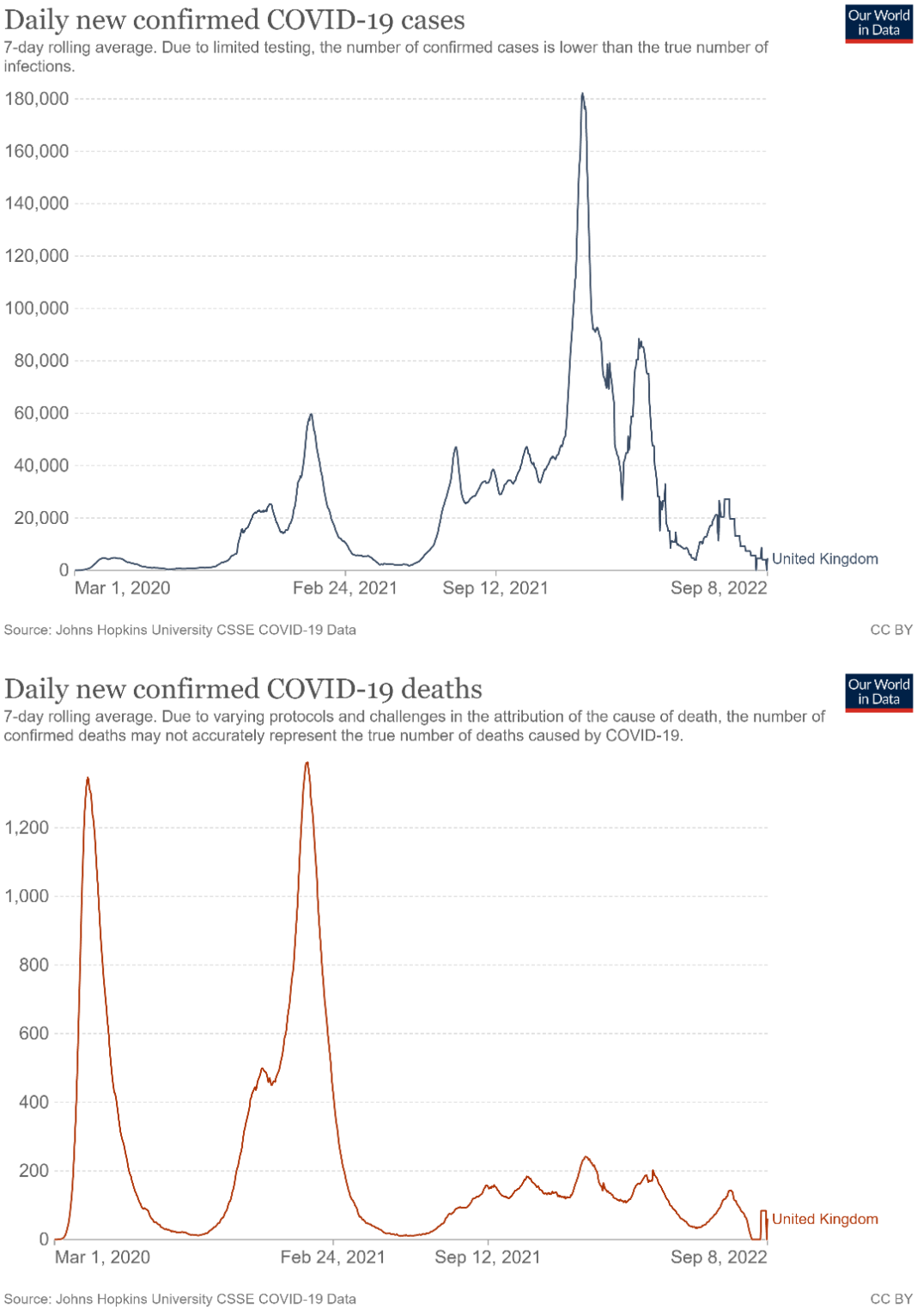


Figure 1.4: Daily new cases in the UK from 1st March 2020 to September 2022 (top panel) and daily confirmed deaths (lower panel). Figures reproduced from Our World in Data [27] under CC BY 4.0. [1]

Many respiratory viruses such as RSV and influenza demonstrate a seasonal pattern of transmission, with peaks typically during the winter months in each hemisphere. In the first two years of the pandemic SARS-CoV-2 has not demonstrated clear seasonality. There have been inconsistent findings of the relationship between case numbers and factors such as temperature and humidity. [28,29] This may be due to differing methodology and early studies were often confounded by influences such as population density and public health policy. Non-pharmaceutical interventions such as lockdowns and school closures appear to have had more effect on transmission, with meteorological factors playing a lesser role. [30] However, as population susceptibility is modulated by previous infection and vaccination the epidemiology of SARS-CoV-2 may change over time.

1.3 MOLECULAR EPIDEMIOLOGY AND VIRAL EVOLUTION

Mutations occur during viral replication and whilst many will not persist or exert a phenotypic change, others will contribute to the evolution of the virus. The initial phase of the pandemic was characterised by relatively few evolutionary changes, in keeping with estimates of the SARS-CoV-2 molecular clock being approximately 7×10^{-4} to 1.1×10^{-3} substitutions per site per year. [31,32] The D614G amino acid substitution was identified in January 2020, appears to have been positively selected for [33] and which has subsequently become ubiquitous. However, in October and December 2020 two new variants B.1.351 (later known as Beta) and B.1.1.7 (later known as Alpha) respectively were described, both of which contain numerous non-synonymous and synonymous mutations across the genome. [34,35] Since then there have been a number of variants described (Figure 1.5), the most significant of which (Delta, Omicron) have been the successive dominant variants responsible for UK case numbers (Figure 1.6).

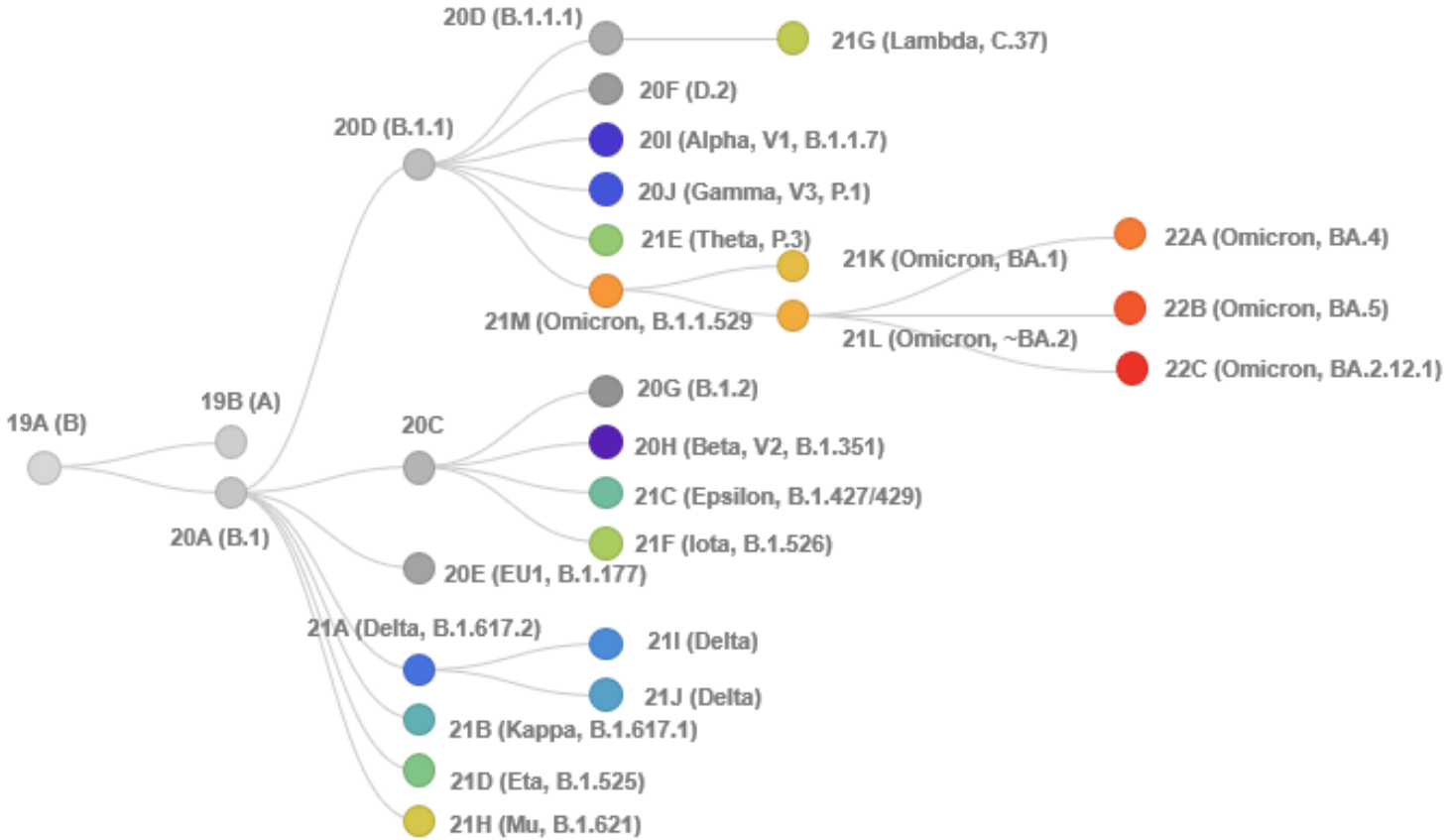


Figure 1.5: Diagram showing SARS-CoV-2 phylogeny by clade and PANGO lineage., from https://nextstrain.org [36], reproduced under CC by 4.0 license. [1]

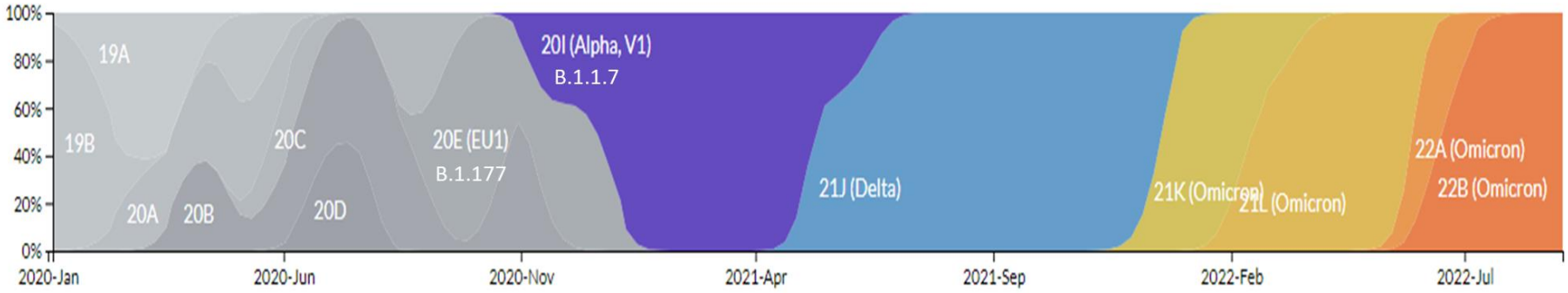


Figure 1.6: Normalised frequencies of sequences by Clade from the UK until September 2022. Screenshot from https://nextstrain.org [36], reproduced under CC by 4.0 license. [1]

A common theory as to why variants arise is during natural evolution of the virus in a chronically infected immunocompromised host. [37,38] Chronicity is thought to be required because within-host viral evolution during the short period of infection in an immunocompetent host is minimal. [39] Equally, the transmission bottleneck is tight, hence the number of viral particles typically involved in transmission is small, meaning that a mutation must be included in the dominant viral population to be reliably transmitted. [39,40]

Multiple mutations in the Spike protein, the primary immunogen for natural and vaccine-mediated immunity, have been identified. These have been associated with changes in ACE2 binding affinity, [41] viral transmissibility, [42] viral loads, [43] disease severity, [44] and immune evasion. [45,46] The consequences of mutations in non-Spike regions of the genome are still being elucidated. The R203K/G204R mutation in the Nucleocapsid protein observed in B.1.1.7 has been shown to be associated with a replication advantage *in vitro* and increased viral load and disease severity in a hamster model. [47] Agerer et al. demonstrated that even single nonsynonymous mutations can impact CTL IFN-γ responses in mutant peptide expanded and restimulated PBMCs (peripheral blood mononuclear cells). [48]

When a mutation, or constellation of mutations is suspected to have adverse implications for diagnostics, transmissibility, disease severity, or immune escape it is designated a variant under investigation (VUI; or variant of interest (VOI) in US) or, if confirmed, a variant of concern (VOC).

1.3.1 Nomenclature

As a novel pathogen the system to describe observed variants from sequenced isolates had to be generated and inevitably has changed during the pandemic. The most used systems of SARS-CoV-2 nomenclature are: PANGO (Phylogenetic Assignment of Names Global

Outbreak LINeages) which uses a letter followed by a number, where each dot means descendent of, to a maximum of three hierarchical levels, for example B.1.1.7. [49] The Nextstrain nomenclature, similarly to influenza virus nomenclature begins with a year then a letter for example 20I and this system has been revised as the conventions have evolved with the pandemic. [50] Subsequently the WHO announced the intention to use Greek letters for variants of interest and concern to streamline communication. [51]

Table 1.1 shows the relationship between these three nomenclature systems in use during the pandemic. Nations where variants are first described may not reflect where the variant originated and associations or nomenclature including regions or countries should be avoided. [52] A clear naming system for variants of concern allows all stakeholders, the media, and the public to communicate effectively and such naming conventions have never been more essential than during this pandemic.

Table 1.1: Nomenclature of common SARS-CoV-2 variants. *This variant was not a variant of concern therefore has no designated Greek letter. Alpha was first identified in the UK, Beta and Omicron in South Africa, Gamma in Brazil, and Delta in India.

Greek letter	Pango	Nextstrain
*****	B.1.177	20E
Alpha	B.1.1.7	20I
Beta	B.1.351	20H
Gamma	P.1	20J
Delta	B.1.617.2	21A
Omicron	B.1.1.529	21K

1.3.2 Pre-Alpha SARS-CoV-2 variants

The beginning of the pandemic in the UK was characterised by multiple independent introductions of virus into the country, and following the introduction of lockdown in March 2020 was then associated with extinction of many of these by June 2020. [53] When travel resumed to and from the UK in the summer of 2020, B.1.177 (20E) was introduced, likely from Europe and became dominant in autumn 2020 before B.1.1.7 emerged. [54]

1.3.3 Alpha (B.1.1.7) variant of concern

The Alpha (B.1.1.7) variant of concern was identified in December 2020, having emerged in approximately September 2020. [35] It was particularly notable for the large number of mutations, especially non-synonymous mutations and was phylogenetically distinct from virus that had been sequenced previously. These lineage-defining mutations of the Alpha (B.1.1.7) VOC are shown in Table 1.2.

The deletion of two amino acids at positions 69 and 70 (69-70del) causes a reproducible failure of signal on the S gene target on the Thermo Fisher TaqPath three gene assay, referred to as S gene target failure (SGTF). [55] This assay was widely used in the UK government Lighthouse laboratories for SARS-CoV-2 diagnosis of Pillar 2 (community population) samples. Whilst these deletions are not unique to B.1.1.7, the SGTF result was used as a proxy marker for this variant from November 2020 to January 2021. SGTF demonstrated rapid expansion forming 3% of positive results the week commencing 12th October, to 95% of infections during the week beginning 30th November 2020 to 99.9% by 11th January 2021 from all Pillar 2 testing. [56] B.1.1.7 remained the dominant SARS-CoV-2 variant in the UK until the advent of the Delta (B.1.617.2) variant in late March 2021 which became dominant until the end of the testing period in the studies discussed in this thesis. Several observations suggested that individual spike mutations present in B.1.1.7 may cause viral phenotypic changes in terms of infectivity, tropism, receptor interaction and susceptibility to neutralisation. The key mutations and evidence of impact on viral phenotype are summarised in

Table 1.3.

Table 1.2: Defining mutations of B.1.1.7 (Alpha) Variant of Concern

Non-synonymous mutations				
Spike	**Nucleocapsid**	**ORF1a**	**ORF1b**	**ORF8**
H69- (NTD)	D3L	T1001I	P314L	Q27stop
V70- (NTD)	R203K	A1708D		R52I
Y144- (NTD)	G204R	I2230T		Y73C
N501Y (RBD)	S235F	S3675-		
A570D		G3676-		
D614G		F3677-		
P681H				
S982A				
D1118H				
Synonymous mutations				
C913T, C3037T, C5986T, C14676T, C15279T, T16176C				

Table 1.3: Associations of specific spike mutations with potential functional significance

Spike Mutation	**Evidence of functional significance**	**Reference**
69-70del	Twofold increased infectivity compared with wild-type spike within a pseudovirus model	38
N501Y	Enhanced viral replication in primary human airway tissue model	42
	Enhanced spike-receptor interaction via increased binding affinity	41
	Weakened antibody binding	57
D614G	Evidence of enhanced replication through increased virion infectivity from in vitro cell culture, hamster infection model and primary human airway tissue model	58
	Increased infectivity in human lung cells	59
	Higher viral loads in patients	33

The B.1.1.7 variant has now been associated with reduced neutralisation by convalescent and vaccine induced antibodies but without evidence of widespread immune evasion. [57] Compared with pre-B.1.1.7 lineages it has been shown to have increased transmissibility and viral loads by proxy, [60-63] and SGTF has been associated with increased severity of disease in studies using different methodologies. [44,64,65]

1.3.4 Significance of variants in the context of vaccination

Whilst the molecular clock of SARS-CoV-2 is relatively slow, individual variants can swiftly become dominant within a population during periods of high transmission. The presence of multiple mutations in variants, particularly within regions such as the RBD of the spike protein raises the concern that both the immunity conferred by natural protection and from vaccination may be compromised, especially because most vaccine platforms use the spike protein as the immunogen. In addition to pre-clinical data, evidence has accumulated that suggests convalescent and vaccine induced antibodies may have reduced neutralising ability against successive SARS-CoV-2 variants. [57,66] This became evident with the Delta and Omicron variants. [67,68] Whilst vaccine effectiveness against severe outcomes such as hospitalisation and death appear maintained across the major variants, rapid waning of protection following two doses of ChAdOx1 nCoV-19 or mRNA vaccines against both infection and severe outcomes was evident in older long-term care residents during 2021. [69] Protection against all outcomes was shown to be restored with third (booster) doses of vaccine, but rapid wane observed after boosting suggests an annual boosting strategy may be insufficient. [70]

A related concern was whether vaccination could in fact give rise to variants, as vaccine induced immunity may select for variants that can adequately cause infection in the face of vaccination. This is still unknown.

1.3.5 Genomic surveillance

The UK COVID-19 Genomics consortium (COG-UK) initiated large-scale whole genomic sequencing (WGS) of SARS-CoV-2 in the UK during the pandemic. Globally these efforts have been mirrored by an enormous number of sequences being made available, however inevitably sequences from areas without the infrastructure to support large-scale WGS are

poorly represented. These genomes have provided a unique, near real-time perspective into the molecular epidemiology of the virus which has provided additional knowledge to integrate with clinical information as the basis for policy decision making.

1.4 TRANSMISSION

1.4.1 Routes

SARS-CoV-2 is highly contagious. Like other respiratory viruses, SARS-CoV-2 can theoretically be transmitted via several routes.

Spread by respiratory particles can be differentiated into droplets or aerosols where droplets are defined as ≥5-10 μm. [71] However this definition is somewhat arbitrary as an infected individual may produce infectious particles along the size spectrum. Smaller aerosolised particles can remain suspended in the air for longer and are more readily inhaled, the smallest of which may be able to reach beyond the upper respiratory tract into the lungs. Attempts to recover viable virus from air samples have not been consistently successful but there remains supporting evidence that aerosol or airborne transmission is highly likely to be the principle means of transmission. [72] Aerosol transmission has resulted in the infection of individuals who have not had close contact with index cases and numerous "super-spreader" indoor events for example in restaurants, on transport and at choir practice. [73-75]

Droplets generated by an infected individual talking, coughing, or singing can directly inoculate the mucous membranes of another individual in close proximity, with gravity exerting a greater effect on these larger particles and therefore meaning that these will fall to surfaces relatively quickly. In a study of viral RNA contamination in hospital rooms, the most likely surfaces to be contaminated were the floor, air vent, bed rail and bedside locker despite

twice daily cleaning. [76] In a laboratory setting virus can persist on environmental surfaces for up to 72 hours. [77] In theory surface contamination could permit individuals to self-inoculate by transferring fomites from their hands to their mucous membranes. However, there is no evidence to support fomite transmission as a principle means of transmission compared with evidence for SARS-CoV-2 RNA environmental recovery and many have questioned both this and the relative importance that infection control measures have placed on surface decontamination. [78,79]

SARS-CoV-2 RNA is readily shed in stool, [80] and live virus has been isolated[81] therefore faeco-oral transmission is plausible. There are descriptions in the literature of likely instances of this type of transmission, though aerosolization of virus on toilet flushing may also contribute. However, community-based studies of household contacts and other study types have concluded that the risk of transmission via stool or wastewater is likely to be low. [82,83]

In chapter 6 I will assess the impact of vaccination on SARS-CoV-2 shedding in stool.

1.4.2 Viral kinetics

The viral kinetics of SARS-CoV-2 are still being understood, with the advent of each new SARS-CoV-2 variant posing the question about whether fundamentals such as incubation, shedding and duration of infectiousness are different. Understanding these aspects of infection is critical for formulating effective and proportionate interventions to prevent spread, at an individual and population level and in the evaluation of the impact of vaccination.

In a systematic review which included papers published up to June 2020, a meta-analysis of the 43 papers that included information on duration of viral RNA shedding from at least one bodily fluid, showed a mean shedding duration of 17.0 days (95% CI 15.5-18.6 days) in the upper respiratory tract, which increased with age. [80] The viral load was found to be highest

around the time of symptom onset and was higher in the first week of symptoms compared with later in the infection (Figure 1.7). A modelling study using data from a French inpatient cohort found peak viral load 1.1 days prior to the onset of symptoms, and evidence of a biphasic decline from the peak. [84] Five out of six studies that compared severe vs non severe illness showed a longer duration of shedding with more severe illness. [80]

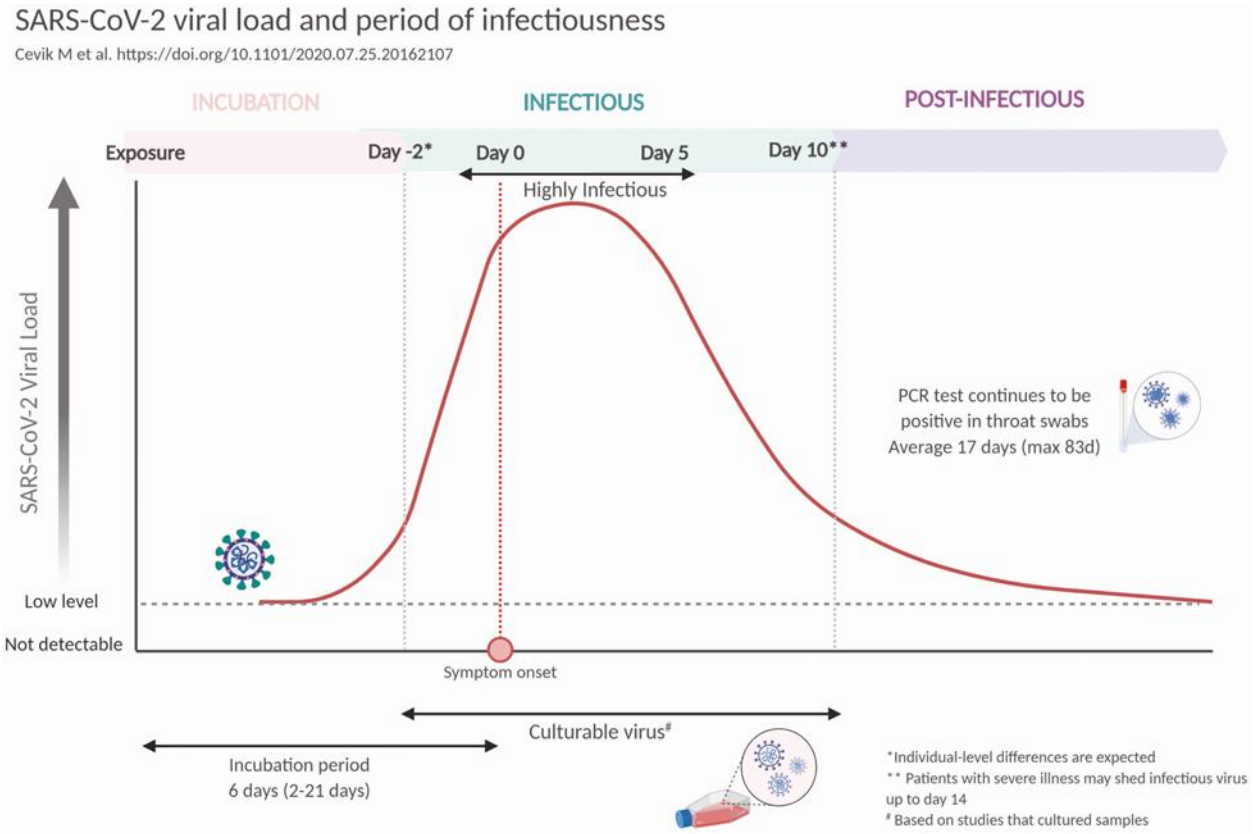


Figure 1.7: SARS-CoV-2 viral load dynamics and period of infectiousness. Incubation period (time from exposure to symptom onset) of 6 days (2–21 days), peak viral load levels documented from day 0 (symptom onset) to day 5, infectious period starts before symptom onset up to 10 days (this may be extended in patients with severe illness), and RNA shedding continues for a prolonged period of time but culturable virus has been identified up to day 9 of illness. (This figure was created by the authors on Biorender, https://biorender.com based on available literature about SARS-CoV-2 viral load dynamics.) Abbreviations: max, maximum; PCR, polymerase chain reaction; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2. Figure and legend from Cevik et al. (2021) [85] with permission from Oxford University Press (license number: 5626310103294).

The relationship of symptoms to viral load and shedding duration is less clear. In the same systematic review, there were conflicting studies on viral loads between those with and without symptoms; two found lower levels, whilst four others showed similar levels between these two

groups. Five studies found that RNA clearance in those who were asymptomatic was more rapid. [80] A study of hospital outpatients published after this systematic review, found that 25% of asymptomatic patients had low viral loads, significantly more than the 11% of symptomatic patients with low viral loads. However, 75% of asymptomatic patients therefore had mean to high viral loads, the median of which was slightly more than for those who were symptomatic. [86] This group also demonstrated a clear relationship between the cycle threshold (Ct) value of these samples and the fold increase of viral RNA levels following viral culture thus supporting the basis for relating Ct value and potential for infectivity. [86] The clearance of viable virus, identified by culture in upper respiratory samples was 3-12 days from the beginning of symptoms in one of the early case series from Germany. [87] This has more recently been reproduced in controlled human infection models of SARS-CoV-2 using virus from the 20A clade of B.1. [88] Due to the safety requirements and difficulty in evaluating the presence of SARS-CoV-2 replication in cell culture many larger studies do not proceed to culture their samples. Therefore, the Ct value is often used as a proxy for transmission potential, with the acknowledgement that this is imperfect.

In chapters 4 and 5 I will present the results of the longitudinal upper respiratory tract sampling in trial participants considering viral loads and duration of positive PCR testing in individuals in relation to vaccine status.

1.4.3 Impact of symptoms on transmission

Individuals can transmit SARS-CoV-2 infection both when they have symptoms of infection and when they do not, either when pre-symptomatic or when they are infected but never develop symptoms. A decision-analytic modelling study investigated the relative contribution of different symptomatic states to transmission. [89] Using evidence-based assumptions for the incubation and infectiousness periods and adjusting for these, this simple model suggested that

approximately 50% of transmission could result from asymptomatic individuals. Real-world evidence supports that a significant amount of transmission comes from individuals that do not have symptoms such as the findings from the *Diamond Princess* Cruise ship outbreak where individuals who shared a cabin with an asymptomatic but infected individual had an attack rate of 63% compared with 81% if the cabinmate was symptomatic when infected. [90] In a study of an outbreak in a nursing facility in the US, 27 out of 48 residents who tested positive were asymptomatic at the time of testing, 24 subsequently developed symptoms and in samples from 17 of these, viable virus was identified in culture. [91] Asymptomatic transmission presents additional barriers to disease control when infectious individuals may not be readily identifiable and may continue interactions with susceptible individuals. Identification of asymptomatic cases using RDTs has been a key component of the UK government's containment strategy though use of such screening, even when free and voluntary has been associated with inequality for example in the large-scale community testing in Liverpool using lateral flow tests. [92] Uptake was lower within areas of deprivation, amongst men and within certain ethnic groups, yet these were also the groups that were more likely to receive positive tests.

1.5 Diagnostics and Testing

1.5.1 PCR

The most used method for diagnosis of SARS-CoV-2 infection has been RT-PCR using samples from the upper respiratory tract, commonly nasal and oropharyngeal specimens but can also be performed on other samples such as saliva, BAL fluid and stool. Commercially available assays typically employ more than one genetic target from: spike (S), nucleocapsid (N), ORF1ab, envelope (env) and RNA-dependent RNA polymerase (RdRp). The use of

multiple gene targets was vindicated with the finding that viral mutations may result in primer failure such as the Δ69-70 deletion present in the Alpha (B.1.1.7) and Omicron variants causing so called S-gene target failure. [55,93]

RT-PCR assays produce a cycle threshold (Ct) result, when the fluorescence generated by the reaction exceeds a given threshold above background and is therefore considered positive. As a real-time assay it is semi-quantitative, with a non-linear inverse relationship between Ct value and viral load. [94] Ct values cannot be compared between different assays.

The caveat with the use of PCR as a diagnostic test, although highly sensitive and specific, [95,96] is that the presence of viral RNA does not necessarily indicate presence of culturable virus with infective potential. RNA can be detected for many weeks after infection in healthy individuals in a range of clinical specimens. [80] Whilst there are reports of isolating virus later the probability of doing so decreases 8-9 days after infection even when high viral loads persist. [87,88,97] The likelihood of culturing live virus also drops with increasing Ct value. [86,98] Therefore samples with a high Ct value collected late in the infectious course are unlikely to have the capacity to infect others, however ascertaining absolute Ct cut offs is complicated by sample type and lack of standardisation both for PCR testing and viral culture. [97]

1.5.2 Lateral flow testing

Rapid point of care testing using viral antigen immunoassays (RDTs) have been widely used in a variety of diagnostic situations during the SARS-CoV-2 pandemic. These tests, in general, are highly specific and variably sensitive, the extent of which depends on test operator, symptom status, time after symptom start and viral load. [99,100] Antigen testing may have a higher positive predictive value for positive viral culture compared with RT-PCR testing. [101]

1.5.3 Antibody testing

Antibody response to natural infection has been established with early increases in IgA, IgM, and IgG to different antigens (S, N) but the levels of antibodies vary in time to detection and durability. The usefulness of serology as a diagnostic test is limited due to the availability of pathogen based molecular testing, sensitivity of available tests, the timing of antibody response in the context of symptoms, antibody kinetics over time and difficulties in interpretation in terms of determining current infection. However, at a population level serological testing may be useful for estimating prevalence and response to vaccination. [102] Wide variation of results across different settings have been found; [103] these may relate not just to geographic and temporal differences but also to range of methodological and procedural factors. [104]

1.5.4 Surveillance (wastewater)

SARS-CoV-2 virus in wastewater, identified by PCR and sequencing techniques has been used to track infection rates and as epidemiological surveillance of viral variants. [105-107] Wastewater based surveillance can detect SARS-CoV-2 significantly prior to recognition of cases and wastewater signal associates well with community cases suggesting it may be a cost-effective, non-invasive way to monitor prevalence and serve as an early warning indicator of changes in infection trend in some settings. [108]

1.6 DISEASE

1.6.1 Symptoms

There is a clinical spectrum of SARS-CoV-2 infection from asymptomatic infection through to severe COVID-19 disease. Most symptomatic disease is mild and lasts 1-3 weeks and initially was characterised by non-specific symptoms such as fever, persistent cough, shortness

of breath, sputum production, headache, myalgia, and fatigue. [109-112] More specific symptoms of anosmia and ageusia were rapidly recognised as common findings. [113] Involvement of other systems such as the gastrointestinal tract and skin rashes are also well described. [112,114,115] As the pandemic has progressed data from large scale population monitoring suggest that increasing levels of immunity and more recent variants (Delta, Omicron) may be associated with more prominent upper respiratory tract symptoms (rhinorrhoea, sore throat, sneezing). [116,117]

Severe disease is characterised by deterioration after approximately one week of symptoms, with shortness of breath and evidence of respiratory failure. The disease may then progress to acute respiratory distress syndrome with bilateral pulmonary infiltrates visible on thoracic imaging. Multi-organ failure, particularly renal failure may then occur. [118] Complications include thrombo-embolism, [119] coagulopathy, [120] cardiovascular, [121] respiratory, [122] neuropsychiatric [123] and post-COVID syndrome. [124]

1.6.2 Disease spectrum

Asymptomatic SARS-CoV-2 infection, and the ability of humans to transmit whilst asymptomatic was evident from investigations of early outbreaks in confined spaces. [90,125-127] Amongst the 229 passengers aboard the *Diamond Princess* cruise ship in January 2020 who participated in a cross-sectional retrospective survey, 21% of those who tested positive (14/66) reported no symptoms though these could represent pre-symptomatic cases in the absence of longitudinal follow up and are subject to the caveats of retrospective surveys. [90] In a similar outbreak study, the crew of U.S.S. *Theodore Roosevelt* aircraft carrier which had a younger, healthy, and predominantly male population showed a rate of asymptomatic infection of 43% (572/1331). [126] Neither of these studies identified any demographic or predictive factors associated with the absence of symptoms. Other estimates of asymptomatic rates vary widely

though this is likely to be influenced by several host factors, the difficulty of defining and capturing symptom data, as well as whether the study design permits differentiation between pre-symptomatic and asymptomatic infections. Studies using different methodologies and in varied populations have shown asymptomatic rates ranging from 6.25% (3/48) in residents of a skilled nursing facility in the US, 42.5% (95% CI: 31.5-54.6%) in a population survey conducted in Italy and 90.2% amongst marine recruits in the US. [91,125,127]

Similarly, the data regarding the proportion of individuals that are asymptomatic at the time of their positive SARS-CoV-2 PCR but who subsequently go on to develop symptoms (i.e. pre-symptomatic) are inconsistent. On the USS Theodore Roosevelt 30.5% (406/1271) individuals were pre-symptomatic at the time of their time their first positive test, conversely in a skilled nursing facility outbreak, 56% (27/48) residents had no symptoms at the time of their first positive test, however 89% (24/27) of these went on to develop symptoms. [91,126].

A systematic review and meta-analysis involving 170 of the identified 390 studies published up to April 2021 found that approximately a third of infections were asymptomatic using two different analyses, the first including all studies with adequate follow up time (35.1%, 95% CI 30.7 to 39.9%) and the second with a stricter inclusion of those in whom the methodology could adequately delineate pre-symptomatic cases (36.6%, 95% CI 31.8 to 42.4%). [128] In this second group of studies pre-symptomatic infections were found to occur in 31.8% (95% prediction interval 5.6 to 78.7%). Overall, at the time of testing 42.8% (95% prediction interval 5.2 to 91.1%) of cases were silent and testing was performed in the absence of symptoms. This meta-analysis found that asymptomatic infection was less common in the elderly, in residents of long-term care facilities and in those with comorbidities. [128] These findings are broadly in keeping with an earlier systematic review which included a different dataset of papers published until November 2020. This review combined findings from 19

studies that had PCR testing and longitudinal follow up (median 14 days, IQR: 14-15.8 days) and found that a median of 42.5% (IQR 29.6 to 77.8%) of cases are asymptomatic throughout infection. [129] This review also calculated that a median of 72.3% (IQR 56.7 to 89.7%) of those who were asymptomatic at the time of their positive test remained asymptomatic.

Whilst asymptomatic cases clearly account for a proportion of infections, most symptomatic cases are mild to moderate in severity and occur within the community. Severe disease and hospitalisation occur in a significant proportion of unvaccinated individuals, figures which are highly dependent on the demographic setting, host risk factors and transmission rates. Risk factors for severe disease and death have been identified. [130] A log-linear increase in IFR has been demonstrated with increasing age [131,132]. Male sex, pre-morbid medical conditions including obesity and non-White ethnicity are also associated with severe disease and poorer outcomes. [133-137]

Rates of estimated mortality vary widely with multiple factors affecting these figures: methods of calculation, mortality reporting, testing availability, population demographics and access to healthcare. [138] An overall estimate of Infection fatality ratios (IFR), "the proportion of those infected who will go on to die from that infection", [132] using serological surveys range from 0.14 to 1.79% for different income settings in 2020. [118,132]

1.7 IMMUNE RESPONSE TO INFECTION

1.7.1 Innate response

The innate immune system is the first part of the immune system to respond to pathogens. It is comprised of physical barriers, chemical defence, and cells such as natural killer (NK) cells and macrophages. It is triggered by many steps of SARS-CoV-2 life cycle but especially within

the intracellular compartment during replication. [139] Pattern recognition receptors (PRRs) on innate immune cells recognise signatures of potential threat called pathogen-associated molecular patterns (PAMPs) and damage-associated molecular patterns (DAMPs) which trigger inflammatory pathways and can induce cell death. [139] Innate lymphoid cells (ILCs) are a type of T lymphocyte that do not have antigen-specific receptors but do possess PRRs. [140] ILCs are found throughout tissues and may play a role in mucosal defence and homeostasis. Patients with COVID-19 infection show reductions in multiple cell lines, for example patients with moderate/severe disease show reduced circulating ILCs on admission and circulating NK cells have also been found to be reduced in hospitalised patients. [141,142]

Investigations of pathophysiology of severe COVID-19 disease suggest that an uncontrolled inflammatory response mediated by cytokines such as IL-6, IL-10 and TNF-α may play a role. [143,144] High levels of cytokines have been observed in hospitalised COVID-19 patients and associations of some cytokines with more severe disease. [145] Early increases of cytokine levels have been associated with poorer clinical outcomes. [146] A "core COVID-19 signature" of inflammatory cytokines has been observed in moderate and severe COVID-19 disease, and a further association was found between severe disease and cytokines linked to cytokine release syndrome which started after day 10 of symptoms. [146] These cytokine responses include key chemokines which recruit neutrophils and monocytes to infected areas which may in themselves contribute to more severe disease. [147] Dysregulation of the innate immune response and downstream effectors is likely to be related to the pathophysiology of severe disease in some individuals. [145,147]

1.7.2 Humoral response

In most cases of natural, symptomatic SARS-CoV-2 infection, anti-spike and anti-nucleocapsid antibodies are elicited and the primary goal of most vaccinations is to stimulate a humoral

response. Spike-specific antibodies develop within 1-3 weeks after infection and peak before declining following convalescence (Figure 1.8). [148,149] IgM is typically detectable first, but can co-occur with IgG and IgA, and usually declines earlier than IgG and IgA. [148,150]

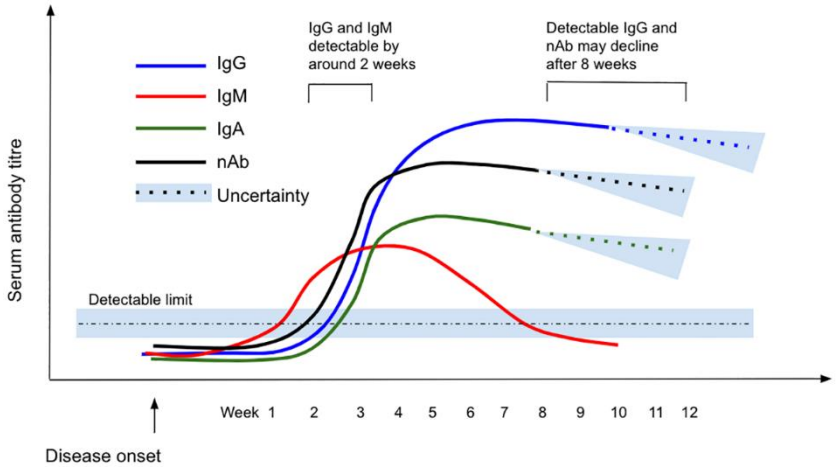


Figure 1.8: Schematic showing IgG/IgM/IgA/Neutralising Ab response over time from disease onset. Note that the y-axis is illustrative only and therefore no scale is given. Figure and legend from Post et al., 2021. [151] Reproduced under CC BY 4.0. [1]

Antibody magnitude and kinetics may be mediated by several factors including disease severity and viral load. Earlier IgM responses were seen in milder disease and higher titres of IgG may be related to more severe disease though this association is only discernible two weeks after infection. [148,150] In a large survey study individuals that showed a classical IgG seroconversion profile had lower Ct (cycle threshold) values (therefore a higher viral load), more reported symptoms and more classic symptoms than those who either failed to seroconvert or those whose levels peaked around the time of their positive test and so could have been cases of reinfection. [152]

Early studies performed involving hospitalised patients suggested seroconversion was ubiquitous, [148] however subsequent data have shown not all individuals produce detectable levels of IgG. [152-154] Lack of measurable anti-spike IgG is more common if disease is

asymptomatic or mild, if individuals are from some ethnic groups or on immunosuppressive medication and may also be associated with increasing age. [149,152,153]

The rate of natural infection induced antibody wane might be biphasic with a shorter half-life initially, with the half-life becoming longer with increasing time since infection. [155] A range of models derived from a large dataset support this reduction of decline over time though relationship between rate of decline and peak levels have not been consistent between studies. [152,156] Longitudinal studies show the persistence of anti-spike IgG antibodies out to at least 14 months post symptom onset with 16/22 of these participants maintaining neutralising antibodies out to this timepoint. [155]

Similar factors to those noted for anti-spike IgG were observed to influence peak levels and trajectories of anti-nucleocapsid IgG however the antibody trajectories of the two antigenic targets are quite different with anti-nucleocapsid antibodies waning more rapidly. [156] Large cohort studies estimate anti-spike antibody half-life to be 184 days [152] in contrast to 85 days for anti-nucleocapsid antibodies. [156] These findings have implications both for seroprevalence studies and for understanding serological contribution to protection against infection and which antigenic targets are associated with serological persistence.

Neutralising antibodies are a subset of the total antibody population that can prevent disease by blocking infection or eliminate pathogens by antibody-mediated effector functions. Serum neutralising activity is often present in the first week after symptom onset and has been shown in a viral inhibition assay using an HIV viral particle pseudo-typed with spike, to be detectable prior to the ability to detect IgG by ELISA. [149] Early production of neutralising antibodies may be critical in viral control and failure to mount such a response correlates with an increased risk of death. [157] Peak neutralising IgG titres have been found to be higher with more severe disease, and more likely to be absent in those who had asymptomatic disease. [154]

Monoclonal preparations of neutralising antibodies have been used as treatments in severe COVID-19, with the RECOVERY trial finding a combination of casirivimab and imdevimab (REGEN-COV) reduced 28-day mortality in those who had no evidence of an antibody response at baseline [158] though further research is required to verify the efficacy of these on a range of clinical outcomes. [159]

1.7.3 Cellular response

Cellular responses to viral infection are important through direct CD8+ mediated cytotoxicity of infected cells, production of cytokines and co-stimulation of B cells. The cellular response to SARS-CoV-2 remains incompletely understood. It is likely to play a crucial role in protection but may also be involved in the pathophysiology of severe disease with prolonged dysfunctional inflammation[146] and CD8+ T cell expression of PD-1 and Tim-3 particularly in severe disease, suggesting either exhaustion or prolonged activation. [160,161] Lymphopenia is common in COVID-19 patients, leading to an elevated neutrophil/lymphocyte ratio during infection, and a reduction in CD4+ T cells. [143,162]

Circulating CD4+ and CD8+ cells have been found in high proportions of SARS-CoV-2 convalescent patients, with responses to both spike and non-spike targets. [163,164] In a study of T cell memory from convalescent patients evaluating response to peptide pools using ex vivo IFN-γ ELISpot assays, those with severe disease had higher and broader overall responses compared with mild cases. [165] ChAdOx1 nCoV-19 and mRNA vaccination induces broad CD4+ and CD8+ T cell response against SARS-CoV-2 spike peptides. [166,167] The breadth of T cell targets, including those beyond spike may have implications for next generation vaccination strategies as spike-based vaccines do not include these epitopes. T cells can also play a role in long-term immunity, with evidence of SARS-CoV memory evident in T cells

from SARS patients 17 years after infection as well as showing cross-reactivity against SARS-CoV-2 N peptides. [164]

Reduced neutralising ability of disease or vaccine-induced antibodies in the face of heterologous variants of concerns has been widely reported however the data regarding cellular immunity is more diverse. Spike-specific CD4+ T cells from convalescent individuals and mRNA vaccinees had similar activation-induced marker (CD69 and CD137) expression to wild type and VOC peptide pools. [167] However, CD8+ IFNγ responses from convalescent individuals and inactivated vaccine vaccinees were found to be reduced to spike and non-spike epitopes containing mutations present in VOCs compared with wild type. [168] These findings suggest the cellular response may be more resilient to evasion. [169]

1.7.4 Mucosal response

A wealth of knowledge has accumulated about immune responses measurable in the peripheral blood both in response to infection and vaccination, far less is known about the response at the mucosa. Mucosal surfaces are the sites at which the virus is first encountered and from where viral shedding can occur. Animal studies suggested that vaccines may not block mucosal shedding [170,171] and whilst vaccination has been shown to provide excellent protection against severe disease and the worst clinical outcomes, vaccines do not provide absolute protection against infection suggesting that the immune responses necessary to prevent infection in different anatomical compartments may be discrete.

In a study of cytokines from paired serum/nasal washes in symptomatic and asymptomatic COVID-19 patients, nasal washes from symptomatic patients had higher levels of a range of cytokines compared with asymptomatic patients, whilst asymptomatic patients had higher levels of IL-1β and G-CSF. [172] Results regarding the correlation between nasal fluid or salivary

IgG and IgA are not aligned, though differing participant groups and methods make comparisons challenging. In another study of infected, symptomatic healthcare workers (HCWs) serum and nasal IgG and IgA values were correlated however the relationship between the two sample types for IgA was less consistent than for IgG. [173] Comparisons of serum and saliva IgG and IgA have produced similar findings. [174] IgA was also detected in mucosal samples from seronegative participants. [173] However in a further study of nasopharyngeal and serum samples in COVID-19 patients with moderate to severe disease, spike specific IgG and IgA responses were poorly correlated which this group interpreted as possible compartmentalisation of the immune response between tissues. [175] In some studies, high levels of mucosal spike or RBD specific IgA have been found in samples from controls [174] which was also reported in a study of stool IgA. [176] Salivary and serum IgA levels were found to decline with time following natural infection. [174] This might explain susceptibility to upper respiratory tract infection and viral shedding following vaccination however there remains much to understand about the mucosal immune response and relationship to vaccination routes to inform vaccination approaches.

1.7.5 Correlates of protection

Plotkin and Gilbert define a correlate of protection (COP) as "An immune marker statistically correlated with vaccine efficacy (equivalently predictive of vaccine efficacy) that may or may not be a mechanistic causal agent of protection". [177] As with other pathogens, for SARS-CoV-2 the definition of protection by study endpoint is important, as the COP is likely to vary; for example, the definition of efficacy in question may be protection from upper airway infection, severe disease, or death, and the COP may vary with vaccine platform and with host characteristics such as age.

Both animal and human data provided evidence that natural infection may confer protection against subsequent infection. In NHPs that were re-challenged 35 days after initial challenge, there was a significant reduction in median viral load after re-challenge, accompanied by anamestic antibody, neutralising and cellular responses. [178] Studies of reinfections occurring in large cohorts such as SIREN, which were conducted early in the pandemic showed that previous exposure to SARS-CoV-2 provided 84% risk reduction for reinfection. [179] Risk reduction for symptomatic infections was shown for up to seven months post initial infection. [179] Other sizeable prospective studies in healthcare workers confirmed a reduced risk of infection in the presence of anti-spike antibodies. [180]

Early findings from an outbreak in a small cohort on a US fishing ship showed that neutralising anti-spike antibodies may correlate with protection against SARS-CoV-2. None of the three crew members who had pre-existing neutralising antibodies developed symptoms or showed virological or immunological evidence of infection during an outbreak. In contrast, a statistically significant higher proportion (103/117) of those who did not have neutralising antibodies became infected. [181] Animal studies showed that passive transfer of convalescent purified IgG could protect NHPs from subsequent challenge in a dose dependent manner. [182]

Two studies of COVID-19 vaccine trial participants (ChAdOx1 nCoV-19 and mRNA-1273) have shown that the risk of symptomatic infection decreased with increasing levels of anti-spike and anti-RBD antibodies and that vaccine efficacy increased with increasing antibody and neutralisation titres. [183,184] In a study of seven vaccine trials, using normalised neutralisation titres to allow comparison between the different assays used between trials, there was a strong non-linear association between protective efficacy and mean neutralisation level. [185] However, at the level of an individual, exposure outcomes vary widely whereby people with high antibody titres can still develop infection. [183] It is therefore unlikely that an absolute

correlate based on a neutralising antibody titre, i.e. an antibody titre above which protection is assured, can be defined.

Whilst neutralising antibodies may be sufficient to protect, they do not account for all protection. It is likely that there are co-correlates of protection and that cellular responses contribute to infection and vaccine mediated protective efficacy. Passive transfer of neutralising antibodies can protect naïve macaques from infection following SARS-CoV-2 challenge in a dose dependent way, however this protective effect can in part be reduced by CD8+ depletion. [182] Animals that had undergone CD8+ depletion demonstrated increased viral replication after re-challenge compared with those macaques that were in the sham group who did not undergo CD8+ depletion. Cellular immunity may help confer early protection as S-reactive CD4+ and CD8+ T cells are detectable by day 10 after vaccination, [186] and protection from infection is seen rapidly following a single dose vaccination of mRNA or ChAdOx1 nCoV-19 vaccinations when there are only modest levels of neutralising antibodies. [13,187-189] Equally, binding and neutralising antibodies wane quickly following vaccination or infection, whilst T cell responses appear more robust, are polyfunctional and able to proliferate 6 months after second mRNA vaccination shown in a recent pre-print. [190] This cellular immunity may be responsible for the sustained protection against severe outcomes observed. [191]

Immune correlates of protection are further complicated by the emergence of viral variants. Antibodies from vaccinees have been shown to have reduced neutralising ability in the face of spike mutations, [192,193] whilst T cell responses are similar in magnitude to ancestral and variant strains are similar due to viral T cell epitopes being conserved. [192]

1.7.6 Immune determinants of disease phenotype

Ageing, male sex and comorbidities have been identified as host risk factors for severe disease, [194] however the immunological signatures of those who have poorer outcomes have been less readily established. Some associations with severe disease appear in common with other severe respiratory diseases whilst others appear to be pathogen related. [195,196]

Severe disease is characterised by early inflammation ('cytokine storm'), followed by a slower adaptive response which is then excessive. Using metatranscriptomic sequencing, Zhou et al demonstrated a greater increase in expression of proinflammatory genes, particularly chemokines and their receptors (especially neutrophil chemoattractants) and correspondingly neutrophil response in the BAL fluid of COVID-19 patients compared with community acquired pneumonia (CAP) patients or healthy controls. [197] In a small retrospective case series comparing severe (signs of respiratory distress) and moderate (symptomatic with evidence of pneumonia) COVID-19 disease higher serum levels of IL-2R, IL-6, IL-10 and TNF-alpha were seen in severe disease. [198] Others have shown increased IFN-γ, IL-2 and GM-CSF following in vitro stimulation of T cells and differentially expressed gene analysis shows hyperinflammatory CD4+ cells in severe disease in common with other patients with severe hospital acquired pneumonia (HAP). [196] Conflicting findings have been reported regarding interferon responses. Reduced IFN responses (and persistent high viral loads) have been linked to severe COVID-19 and neutralising autoantibodies against IFN have been detected in individuals with life-threatening COVID-19. [199,200] Others have however found increased IFN-1, TNF/IL-1β responses and interferon stimulated genes (ISGs) using a variety of methods in plasma, PBMCs and BAL fluid [195,197,198] Kim & Shin propose a 'delayed but exaggerated' type 1 IFN response, causing lung immunopathology and severe pneumonia may underpin these apparent contradictory observations. [201]

The overall humoral response to SARS-CoV-2 infection is better characterised due to the relative ease of measuring antibody titres, however findings are not wholly consistent, and the range of neutralisation assays, populations and study designs can make it difficult to compare results between studies. As highlighted in section 1.7.2, asymptomatic or mild disease may not be associated with IgG seroconversion measurable in commercial assays though can still be associated with neutralising antibodies suggesting that neutralisation may relate to anti-Spike IgM/A in absence of demonstrable IgG. [149,202] Some studies have found that those with severe disease appear to IgG seroconvert earlier, [202] whilst others have observed the opposite with IgG seroconversion occurring ~1 week later in severe compared with mild cases. [203] Others have shown that the timing of neutralising antibody response, either to be detectable or to reach a peak was not shown to be different in severe compared with mild disease. [149] These observed variations in the kinetics of neutralising antibody titres may well relate to the interval after infection with which different disease severities present and the difficulty of performing these types of observational cohort studies at the beginning of the pandemic. However, the timing of the response for a given individual may be key; early neutralising antibody generation within 14 days of disease onset was seen in those who recovered from COVID-19. [204] More consistent is the finding that severe disease results in a higher peak IgG titre than those with mild disease. [202,203,515] Neutralising antibody responses have been shown to correlate with anti-spike IgG [149,205] and in some, but not all studies, peak neutralising antibody response has been shown to be greater in those with more severe disease. [149,206]

Lymphocytopenia is a prominent feature of COVID-19 and is more profound with increasing disease severity [145,195,198,207] Mild disease is associated with a higher proportion of CD8+ T cell responses [165] and higher frequency of memory B cells. [199,207] The observation of CD4+ T cell exhaustion (PD-1 expression) has been observed in multiple studies with higher

expression of PD-1 in severe COVID-19 patients compared with both mild disease and other severe respiratory syndromes. [196,199] Severe COVID-19 (as well as other severe respiratory syndromes) may also be characterised by an impairment in antigen presentation particularly in myeloid cells. [196]

In keeping with other respiratory viruses, individuals who have an immunodeficiency or are immunosuppressed from illnesses or treatment (such as those with cancers, chronic inflammatory conditions, or organ transplant recipients) are at higher risk of severe outcomes of SARS-CoV-2 infection. [208-211] Immunosuppression, particularly secondary to treatments involving B cell depletion, may reduce antibody response to vaccination. [212-216] T cell vaccine responses do not differ in those who are immunocompromised and immunocompetent following COVID-19 vaccination though magnitude of response may be impaired by specific agents such as tacrolimus or tofacitinib therapy. [216-218] Immunosuppressed individuals may also have reduced vaccine effectiveness to infection and symptomatic COVID-19 with ChAdOx1 nCoV-19 or BNT162b2 vaccines. [215,219] Vaccine effectiveness against more severe outcomes such as hospitalisation after mRNA vaccination was found to be high for immunocompromised groups but still lower than for immunocompetent individuals. [220]

1.7.7 Reinfection

Large cohorts conducted in 2020 and early 2021 demonstrated prior infection providing protection from reinfection for up to seven months. [179,180] SARS-CoV-2 reinfections were relatively infrequent until late 2021. [221,222]

UK data suggests re-infection rates became relatively common in 2022. In week 13 of 2022 (April), the last week for which these data are available, showed that probable reinfections accounted for 11.6% of all new infections in the UK that week [223] despite high levels of

immunity both from vaccination and previous infection in the community. Reinfection may be due to wane of protective immune responses as well as the impact of emerging variants. *In vitro* work has demonstrated the reduction in neutralising ability of both convalescent and vaccinee sera in the face of variants of concern, [224] though at present real-world data regarding variant infections suggest that whilst vaccines remain broadly protective, it is likely to be reduced if there is significant mismatch between vaccinating and infecting strain and that protection wanes over time, and often rapidly post vaccination. [67]

The data from the UK correspond with early epidemiological survey reports from South Africa that reinfection is more likely with the Omicron variant than with either Beta or Delta variants. [225] Vaccination levels in South Africa are lower than the UK but it may be that the ability of the Omicron variant to evade immunity may be common to natural and vaccine mediated protection.

1.8 VACCINATION

1.8.1 Background to the development of SARS-CoV-2 vaccines

Although nearly two decades had elapsed since the emergence of SARS, when the COVID-19 pandemic began there were no licensed human vaccines against any members of the coronavirus family however data from studies of other coronaviridae provided key information facilitating SARS-CoV-2 vaccine development.

1.8.1.1 SARS vaccine development

In southern China in November 2002 multiple cases of atypical pneumonia with evidence of human-to-human transmission were recognised and the causative pathogen was identified as SARS. [226] Over 8000 probable cases have been reported, with regional spread plus sporadic

other cases due to importation by travellers. [227] SARS spike glycoprotein mediates human host cell entry via the ACE2 receptor [228] and human monoclonal antibodies have been shown in vitro to block the spike protein-receptor interaction, suggesting that this may be a potential vaccine target. [229]

Animal studies of SARS infection provided evidence that neutralising antibodies may be protective of lower respiratory tract disease. Studies on golden Syrian hamsters showed that following SARS re-challenge, high neutralising antibody titres protected against viral replication in the lower respiratory tract and were associated with a significant reduction in the viral titre from the nasal turbinates. [230] Similarly, BALB/c mice could be more effectively protected from viral replication in the lower respiratory tract compared with the nasal turbinates after intranasal SARS challenge by transfusion of hyperimmune serum. [231] The ferret model also demonstrated an increase in neutralising antibody titre following re-challenge, and no detectable viral RNA in the rechallenged animals. [232] SARS challenge studies in different species of non-human primates (cynomolgous monkeys, African green monkeys, rhesus macaques) showed that NHPs can be infected and manifest minimal or no clinical illness, but there was variability between species and between studies. Most NHPs generated a neutralizing antibody response to challenge, which correlated to viral titres in the respiratory tract. [233] Animal data however also raised concern regarding the potential for enhanced disease following vaccination or challenge after convalescent antibody infusion (see section 1.8.1.3 below).

Data from patients also showed that neutralising antibodies may be associated with clinical outcome. Convalescent serum from patients showed high levels of neutralizing IgG antibodies which remained detectable for at least 5 months after initial infection. [234] There is also evidence from a small number of patients that there may be an association between titres of neutralizing

anti-S and anti-N antibodies between survivors and patients who died from SARS. [235] Passive immunisation using convalescent plasma was attempted in infected patients during the SARS outbreak but not in randomised controlled trials. [236,237] A more recent meta-analysis of 8 observational studies showed that mortality was reduced in plasma recipients, particularly if administered early, however these studies were at moderate to high risk of bias due to the way in which patients were chosen for therapy by physicians and based on plasma availability. [238]

Vaccination strategies for SARS tested in preclinical studies have included either whole virus (inactivated or attenuated) or platforms such as viral vectors, recombinant proteins, or DNA-based approaches to deliver portions of the spike protein, full-length spike protein or the nucleocapsid protein as the immunogen. Antibody and T cell responses can be elicited against nucleocapsid, [239] however the spike protein appeared to be the most important in terms of stimulating an immune response and conferring protective immunity. [240] A hamster study used recombinant parainfluenza virus vector constructs to express the SARS structural proteins individually or in combination to assess the relative contribution of each to protection. [240] The study showed that the constructs that included the S protein resulted in protection of the lower respiratory tract. Moreover, animals immunised with only M, E and/or N in the absence of S did not have detectable neutralising antibodies suggesting that spike is the superior target. [240] Mice vaccinated with recombinant RBD subunit vaccine produced a durable S-specific antibody response and provided protection from subsequent challenge. [241] This type of response was also seen in NHPs: African green monkeys mucosally vaccinated with a whole spike-expressing recombinant parainfluenza virus mounted a neutralising response had no detectable SARS in upper or lower airways following SARS challenge. [242]

To date only two phase 1 studies of SARS vaccines have been performed. Both the inactivated and DNA vaccine (which encoded the full-length spike protein for SARS) report

participants with neutralising antibodies after 2 (inactivated vaccine) [243] or 3 doses (DNA-based vaccine, using a plasmid expressing spike protein) [244]. The DNA-based vaccine was also shown to elicit CD4+ T cell responses in 8 out of 10 participants. [244] No further phase 1 studies have been conducted, nor have either of these progressed further to Phase 2 studies. This is related to reduction in perceived threat, safety concerns (see 1.8.1.3) and competing funding priorities such as MERS-CoV which emerged in 2012. [9]

1.8.1.2 MERS-CoV vaccine development

MERS-CoV was first isolated from a patient from Saudi Arabia who died from acute severe respiratory disease. [9] Like both SARS coronaviruses, MERS-CoV is also believed to be of zoonotic origin as it is closely related to bat coronaviruses, with dromedary camels identified as intermediate hosts. [245] Over 2600 cases have been reported between April 2012 and December 2022. [246] Most cases have occurred in the Middle East/Eastern Mediterranean region, with a large healthcare related outbreak in Korea. [246,247] MERS-CoV appears less transmissible from human-to-human outside healthcare settings but carries a higher case fatality rate than SARS-CoV or SARS-CoV-2 at approximately 35%. In small case series, patients have been shown to have detectable anti-spike IgG and these responses appear to be long-lasting out to at least six years. [248-250]

The surface spike glycoprotein of MERS-CoV uses human dipeptidyl peptidase 4 (DPP4) as the host receptor which mediates viral fusion with the host cell. Vaccine design approaches have included fragments of spike such as the RBD or full-length spike protein. Recombinant trimeric RBD was shown to block MERS-CoV pseudovirus entry into cells *in vitro* and when given with an alum adjuvant in a human DPP4-transgenic mouse model, generated neutralising antibodies and protected mice from lethal MERS-CoV challenge. [251] An MVA-MERS-S (full-length spike) vaccine reduced viral excretion in seronegative camels, but though neutralising

antibodies were detectable, it did not block infection following MERS-CoV wild type challenge suggesting that this vaccine may not protect from upper respiratory tract infection in some animal models. [252] A consensus spike glycoprotein DNA vaccine also generated neutralisation against live MERS-CoV in NHPs and protected the animals from clinical, radiographic, and histological changes following MERS-CoV challenge. [253]

A DNA vaccine (GLS-5300) and two MERS viral vectored vaccines (ChAdOx1 and MVA), all of which include full-length MERS-CoV spike have been evaluated in Phase 1 studies and demonstrated safety and tolerability. [13,254-256] Following a single dose of the ChAdOx1-MERS vaccine 92% of participants had anti-spike IgG 56 days post vaccination, which persisted to one year in 68% of those tested. [13] These findings were replicated in the phase 1b study conducted in Saudi Arabia. [254] Neutralisation tests using pseudotyped lentiviruses expressing MERS spike protein from different strains showed >70% of samples were positive at day 28 against each strain, though live virus neutralisation assays elicited neutralising antibodies in only 44% of the highest dose group. [13] T cell responses peaked at 14 days post vaccination and were maintained at one year. After two doses of MVA-MERS-S most vaccinated participants seroconverted at days 42 and 56 days post prime with ELISA OD values correlating with plaque reduction neutralisation assay (PRNT) results suggesting that many of these antibodies were functional neutralising antibodies. These titres waned over time and by day 180 positive ELISA responses were only seen in 14% of participants. [255] Similarly 92% of participants who received three doses of GLS-5300 seroconverted with 58% having detectable responses at week 60. Evidence of neutralising antibodies was also seen in a lower percentage (39% at week 24, 10 weeks post third dose). [256] As for SARS-CoV, clinical development has not progressed further though this may change with increased focus on coronaviruses and the discovery of platforms such as mRNA vaccines.

1.8.1.3 Safety

From the outset of SARS-CoV-2 vaccine development there were concerns about safety regarding the risk of vaccine-associated enhanced disease.

Two mechanisms were the focus of concern, firstly, antibody dependent enhancement of disease whereby anti-spike immune serum non-neutralising antibodies facilitate viral entry into host cells. [257] This was observed in an animal vaccine against feline infectious peritonitis virus (FIPV), a related coronavirus, was developed using a recombinant vaccinia virus which expressed the FIPV spike protein. [258] In trials in kittens this led to enhanced disease and early death; adverse effects which were antibody mediated. [258,259] In humans, antibody-dependent enhancement is observed in dengue reinfections and has challenged dengue vaccine development. [260]

Secondly, vaccines can be associated with an eosinophilic immunopathology. This mechanism is proposed to explain the lung response described in infants given an alum-adjuvanted inactivated RSV vaccine, though it is not fully understood. During a trial in the 1960s, 80% of vaccinated children who subsequently were naturally infected with RSV were hospitalised and two children died. [261]

In SARS-CoV vaccine development, viral vectored vaccines encoding nucleocapsid protein have been shown to cause severe lung inflammation in aged mice after SARS challenge whilst anti-spike IgG infusion were also associated with excess inflammation and increased lung injury following SARS challenge in macaques. [262,263] Virus-like particle (VLP), double-inactivated whole virus and recombinant DNA encoded S protein vaccines have been associated with neutralising antibodies and variable protection from challenge but also immunopathology in small animal models including aged mice. [264,265] The eosinophilic

infiltration found in the lungs is thought to be caused by a TH_2-skewed immune response which results in immune complex deposition. [264] These concerns led to caution in transitioning to clinical phase studies for SARS. [266]

The development of MERS-CoV vaccines have also been associated with safety concerns from small animal models. Whilst neutralising antibodies have been shown to be protective and correlate with viral load, infection does not always elicit sufficient titres of neutralising antibodies and/or titres may wane rapidly. [267] Rabbits exhibited enhanced lung pathology on histology when reinfected with MERS-CoV, and passive transfer experiments showed that this effect was likely mediated by non-neutralising antibodies. [267] As for SARS, mice challenged after receiving an inactivated vaccine showed eosinophilic infiltrates, not seen in those mice that had only received adjuvant alone. [268]

No safety concerns were raised from phase 1 trials of MERS vaccine candidates however, (as for SARS) these were not conducted at times of circulating MERS-CoV. [13,254-256] Vaccine enhanced disease was not observed in preclinical SARS-CoV-2 testing of viral vectored, mRNA or inactivated vaccines; TH1 biased responses were observed in mice, Syrian hamsters and NHPs with no evidence of antibody-dependent enhancement or antibody enhanced disease. [170,171,269]

1.8.1.4 SARS-CoV-2 vaccine development

Development of a vaccine against SARS-CoV-2 by multiple groups started with the characterisation of the viral genome in early January 2020. [2] At the time of writing in September 2022 there were 171 vaccines in clinical development and 198 vaccine candidates in preclinical phase studies. [270] Cognizant of the public health imperative for a vaccine and the potential risks highlighted by previous coronavirus attempts, animal studies were initiated, and as these results

accrued, human volunteers were recruited to allow rapid progression to clinical phases of development. Vaccines currently in Phase 3 and 4 studies are shown in Table 1.4. A full review of COVID-19 vaccine technologies, development and studies during the pandemic is beyond the scope of this chapter but for context I will give a brief overview of COVID-19 vaccine strategies that have resulted in products in use with particular focus on ChAdOx1 nCoV-19.

Table 1.4: SARS-CoV-2 vaccines in Phase 3 and 4 clinical trials. Data from WHO COVID vaccine tracker and landscape, last updated 6th September 2022. [270] DNA= DNA based, IAV=inactivated virus, LAV= live attenuated virus, PS= protein subunit, RNA= RNA based, VV(nr)= viral vector (non-replicating), VV®= viral vector (replicating), VLP = virus like particle, ID= intradermal, IM= intramuscular, IN= intranasal, IH= inhaled

Vaccine platform	Type	Route	Developer	Phase
DNA	INO-4800+electroporation	ID	Inovio Pharmaceuticals + International Vaccine Institute + Advaccine	3
DNA	nCov vaccine	ID	Zydus Cadila	3
IAV	CoronaVac; inactivated SARS-CoV-2 vaccine (vero cell)	IM	Sinovac Research and Development Co., Ltd	4
IAV	Inactivated SARS-CoV-2 vaccine (Vero cell)	IM	Sinopharm	4
IAV	Inactivated SARS-CoV-2 vaccine (Vero cell), vaccine name BBIBP-CorV	IM	Sinopharm	4
IAV	BBV152 vaccine	IM	Bharat Biotech International Limited	3
IAV	Inactivated SARS-CoV-2 vaccine (Vero cell)	IM	Shenzhen Kangtai Biological Products Co., Ltd.	3
IAV	VLA2001	IM	Valneva, National Institute for Health Research, UK	3
IAV	TURKOVAC, inactivated virus	IM	Erciyes University and the Health Institutes of Turkey (TUSEB)	3
IAV	Inactivated Whole Virion Concentrated Purified Vaccine (CoviVac)	IM	Chumakov Federal Scientific Center for	3
IAV	Omicron COVID-19 inactivated Vaccine (Vero Cell)	IM	China National Biotec Group Company Limited	3
IAV	UNAIR Inactivated COVID-19 Vaccine	IM	Airlangga University, Indonesia	3
IAV	Inactivated COVID-19 Vaccine	IM	KM Biologics Co., Ltd.	3
IAV	QazCovid-in® - COVID-19 inactivated vaccine	IM	Research Institute for Biological Safety Problems, Rep of Kazakhstan	3
LAV	COVI-VAC	IN	Codagenix/Serum Institute of India	3

Vaccine platform	Type	Route	Developer	Phase
PS	SARS-CoV-2 rS/Matrix M1-Adjuvant (Full length recombinant SARS CoV-2	IM	Novavax	3
PS	Recombinant SARS-CoV-2 vaccine (CHO Cell)	IM	Anhui Zhifei Longcom Biopharmaceutical+ Institute of Microbiology	3
PS	VAT00008: SARS-CoV-2 S protein with adjuvant	IM	Sanofi Pasteur + GSK	3
PS	CpG 1018/Alum-adjuvanted Recombinant SARS-CoV-2 Trimeric S-protein Subunit Vaccine (SCB-2019)	IM	Clover Biopharmaceuticals Inc./Dynavax	3
PS	MVC-COV1901 (Spike-2P protein + adjuvant CpG 1018)	IM	Medigen Vaccine Biologics + Dynavax + National Institute of Allergy and	4
PS	COVAX-19® Recombinant spike protein + adjuvant	IM	Vaxine Pty Ltd./CinnaGen Co.	3
PS	FINLAY-FR-2 anti-SARS-CoV-2 Vaccine (RBD chemically conjugated to tetanus toxoid plus adjuvant)	IM	Instituto Finlay de Vacunas	3
PS	EpiVacCorona (EpiVacCorona vaccine based on peptide antigens for the prevention of COVID-19)	IM	Federal Budgetary Research Institution State Research Center of Virology and Biotechnology "Vector"	3
PS	RBD (baculovirus production expressed in Sf9 cells) Recombinant SARS-CoV-2 vaccine (Sf9 Cell)	IM	West China Hospital + Sichuan University	3
PS	UB-612 (Multitope peptide based S1-RBD-protein based vaccine)	IM	Vaxxinity	3
PS	CIGB-66 (RBD+aluminium hydroxide)	IM	Center for Genetic Engineering and Biotechnology (CIGB)	3
PS	BECOV2	IM	Biological E. Limited	3
PS	Recombinant Sars-CoV-2 Spike protein, Aluminum adjuvanted (Nanocovax)	IM	Nanogen Pharmaceutical Biotechnology	3
PS	Recombinant protein vaccine S-268019 (using Baculovirus expression vector system)	IM	Shionogi	3
PS	GBP510, a recombinant surface protein vaccine with adjuvant AS03 (aluminium hydroxide)	IM	SK Bioscience Co., Ltd. and CEPI	3

Vaccine platform	Type	Route	Developer	Phase
PS	Razi Cov Pars, recombinant spike protein	IM and IN	Razi Vaccine and Serum Research Institute	3
PS	ReCOV: Recombinant two-component spike and RBD protein COVID-19 vaccine (CHO cell).	IM	Jiangsu Rec-Biotechnology	3
PS	Recombinant SARS-CoV-2 Fusion Protein Vaccine (V-01)	IM	Livzon Pharmaceutical	3
PS	RBD protein recombinant SARS-CoV-2 vaccine (Noora Vaccine)	IM	Bagheiat-allah University of Medical Sciences/AmitisGen	3
PS	Recombinant protein RBD fusion dimer adjuvanted vaccine (COVID-19 Vaccine Hipra)	IM	Laboratorios Hipra, S.A.	3
PS	SCTV01C. A Bivalent Recombinant Trimeric S Protein vaccine against SARS-CoV-2 Variants	IM	Sinocelltech Ltd.	3
PS	SARS-CoV-2 Protein Subunit Recombinant Vaccine adjuvanted With Alum+CpG 1018	IM	PT Bio Farma	3
RNA	mRNA-1273	IM	Moderna + National Institute of Allergy and Infectious Diseases (NIAID)	4
RNA	BNT162b2 (3 LNP-mRNAs), also known as "Comirnaty"	IM	Pfizer/BioNTech + Fosun Pharma	4
RNA	mRNA-1273.351.	IM	Moderna + National Institute of Allergy and Infectious Diseases (NIAID)	4
RNA	CVnCoV Vaccine	IM	CureVac AG	3
RNA	SARS-CoV-2 mRNA vaccine (ARCoV)	IM	Academy of Military Science (AMS), Walvax Biotechnology and Suzhou Abogen Biosciences	3
RNA	ARCT-154 mRNA Vaccine	IM	Arcturus Therapeutics, Inc.	3
RNA	Coronavirus mRNA vaccine (LVRNA009)	IM	CanSino Biologics Inc.	3
RNA	mRNA-1273.214 (Booster)	IM	ModernaTX	3
VV (nr)	ChAdOx1-S - (AZD1222)	IM	AstraZeneca + University of Oxford	4
VV (nr)	Recombinant novel coronavirus vaccine (Adenovirus type 5 vector)	IM	CanSino Biological Inc./Beijing Institute of Biotechnology	4

Vaccine platform	Type	Route	Developer	Phase
VV (nr)	Recombinant COVID-19 vaccine (adenovirus type 5 vector) for Inhalation (Ad5-nCoV-IH)	IH	CanSino Biological Inc./Beijing Institute of Biotechnology	4
VV (nr)	Convidecia Vaccine (Ad5-nCoV). Bivalent Recombinant COVID-19 Vaccine (Adenovirus Type 5 Vector)	IM	CanSino Biological Inc./Beijing Institute of Biotechnology	3
VV (nr)	Ad26.COV2.S	IM	Janssen Pharmaceutical	4
VV (nr)	Gam-COVID-Vac Adeno-based (rAd26-S+rAd5-S)	IM	Gamaleya Research Institute; Health Ministry of the Russian Federation	3
VV (nr)	BBV154, Adenoviral vector COVID-19 vaccine	IN	Bharat Biotech International Limited	3
VV(r)	DelNS1-2019-nCoV-RBD-OPT1 (Intranasal flu-based-RBD)	IN	University of Hong Kong, Xiamen University and Beijing Wantai Biological Pharmacy	3
VLP	Coronavirus-Like Particle COVID-19 (CoVLP)	IM	Medicago Inc.	3
VLP	ABNCoV2 capsid virus-like particle (cVLP) +/- adjuvant MF59	IM	Radboud University	3

1.8.2 Vaccine platforms

1.8.2.1 Viral vectored vaccines

Vaccines based on a viral vector platform have been constructed for a variety of pathogens including Ebola, HIV and Zika virus. Viral vectors are viruses that are non-pathogenic to humans such as chimpanzee adenovirus and modified vaccinia virus Ankara to avoid host immune response attacking the vector. The viral vector is engineered to include the genes for a selected antigen(s). The vector is taken up by host cells and delivers the desired genetic code within the cell. This allows a protein antigen to be manufactured by the host cells and expressed on the cell surface thereby triggering an immune response. Whilst there had been multiple clinical trials of vaccines using this type of platform, at the time the COVID-19 epidemic reached Europe only the Ebola Zaire Vaccine (rVSVΔG-ZEBPV-GP) had received European licensure. [271] Equally, no viral vector vaccine had been manufactured on scales sufficient for a population, presenting considerable practical barriers to be overcome should this type of platform prove suitable.

Two viral vectored COVID-19 vaccines have since been licensed in the UK, ChAdOx1 nCoV-19 (AZD1222, AstraZeneca) and Janssen's COVID-19 vaccine Ad26.COV2-S.

1.8.2.2 ChAdOx1 nCoV-19

ChAdOx1 is a simian adenovirus vector constructed from isolate Y25 and rendered non-replicating by the deletion of the transcriptional unit E1. [272] The E3 subunit is also deleted to allow a greater size of genetic insert. In the case of the COVID-19 vaccine this insert is a codon-optimised full-length spike protein of SARS-CoV-2 with a tissue plasminogen activator leader sequence. This vector had been used in multiple clinical trials at the University of Oxford.

Mouse data revealed specific neutralising antibodies as well as a T-cell response following vaccination with ChAdOx1 nCoV-19, with a prominent T_H1-type response, providing preliminary reassurance that vaccine induced immunopathology was less likely to be elicited. [170] Non-human primate data indicated a beneficial neutralising antibody response to a second dose of vaccine but there was no difference between one and two-dose groups for spike-specific T cell responses whilst a pig model showed significantly higher T cell response as well as antibodies following a second dose. [170,273] One and two dose vaccine groups were compared with controls (who received a ChAdOx1-GFP construct), all groups were challenged with 4×10^5 $TCID_{50}$ ml^{-1} live SARS-CoV-2 virus involving inoculation intraocularly, intranasally, orally and into the trachea. Vaccinated animals had better clinical scores and lower viral loads in BAL fluid. This was supported by the finding from histology that no vaccinated monkeys had lung pathology whilst three of six controls were shown to have interstitial pneumonia. [170]

The data from animal models provided the evidence to vaccinate 1067 healthy adults in a randomised controlled phase 1/2 study in the UK (COV001) which used the MenACWY vaccine as an active comparator, associated with mild side effects post vaccination, to enhance the trial blinding. [274] It was originally designed as a single dose study given the logistically favourable option of one dose delivery if a vaccine must be deployed to whole populations. A group of ten unrandomized individuals did received two doses with a 28-day dosing interval. The dose of 5 x 10^{10} viral particles was chosen to correspond with the highest dose in the ChAdOx1 MERS study which was shown to be immunogenic and well tolerated despite being reactogenic. [13] There were no serious adverse reactions to vaccination and the adverse event profile was in keeping with other viral vector vaccines.

Due to the favourable results from the Phase 1/2, the phase 2/3 studies in the UK (COV002), Brazil (COV003) and South Africa (COV005) commenced including adults of all

ages and with well controlled comorbidities. The primary endpoint was protection against primary symptomatic COVID-19 disease and an interim analysis showed an efficacy of at least 64·1% (95% CI 50·5–73·9) following a single dose, and an overall efficacy of 70·4% (95·8% CI 54·8–80·6) for two doses. [275] Efficacy was lower for any NAAT positive result, and further reduced with confidence intervals that spanned zero for infections of asymptomatic or unknown symptom status. [275] A later analysis showed that a longer interval between the initial two doses was associated with a higher efficacy. [276]

Early real-world data post-roll out confirmed effectiveness against infection after a single dose against pre-Alpha variants and the B.1.1.7 (Alpha) VOC, and the use of an extended dose interval is likely to have contributed to high levels of effectiveness. [188,277,278] Protection from moderate to severe disease and death by ChAdOx1 nCoV-19 is maintained against the Delta variant in different study designs. [279,280] In spite of evidence that antibodies elicited by ChAdOx1 nCoV-19 have reduced neutralising ability of VOCs compared with an ancestral variant [224] effectiveness of ChAdOx1 nCoV-19 against multiple endpoints appears preserved against some VOCs and may, in part, be explained by preserved spike specific CD4+ and CD8+ T cells. [279]

In 2021 following roll out of ChAdOx1 nCoV-19 in multiple national vaccination campaigns, a rare, potentially fatal complication of ChAdOx1 nCoV-19 was observed that had not been seen in clinical trial populations. Individuals presented with thrombosis or thrombocytopenia following ChAdOx1 nCoV-19 vaccination in a syndrome resembling autoimmune heparin-induced thrombocytopenia which may be mediated by platelet factor-4 antibodies. [281]

Other viral vectored COVID-19 vaccines have also demonstrated efficacy, the single dose Ad26.COV2.S vectored Janssen vaccine against infection, severe disease, and death[282] and the heterologous rAd26 and rAd5 (Gam-COVID-Vac) vaccine against symptomatic disease. [283]

1.8.2.2.1 ChAdOx1 nCoV-19-induced immune responses

The COV001 study of ChAdOx1 nCoV-19 showed peak anti-spike IgG antibodies (predominantly IgG1 and IgG3) at day 28 post one dose sustained at day 56, with ten-fold boosting seen after the second dose given at 28 or 56 days. [274,284] A range of different neutralisation assays confirmed that neutralising antibodies are elicited as well as a range of non-neutralising functional antibodies. [284] Anti-spike IgM and IgA titres were slightly elevated post first ChAdOx1 nCoV-19 dose but at lower titres than seen in convalescent individuals. [284] Antibody responses to ChAdOx1 nCoV-19 after one or two doses were significantly lower than BNT162b2. [285] Subsequent studies showed antibody titres were higher in trial participants who had had a longer interval between first and second ChAdOx1 nCoV-19 vaccine doses. [276,286] Following second dose antibodies wane but by day 180 remain above the baseline of seronegative participants. [287] Boosting of both antibody (binding and neutralising) and T cell responses occur following a third ChAdOx1 nCoV-19 dose with a >36-week interval between second and third doses. [286]

IFN-γ ELIspot responses peaked at 14 days after one dose and were not increased in two-dose groups with intervals of day 28 or day 56 but maintained out to day 84. IFN-γ and IL-2 levels increased following PBMC stimulation with SARS-CoV-2 spike peptides. [166] These findings confirmed the T-helper cell type 1 (T_H1)-skewed response seen in animal models, providing additional reassurance regarding safety concerns about antibody-dependent enhancement. [166,284] Cellular responses elicited by ChAdOx1 nCoV-19 were found to be significantly higher after a single dose compared with one dose of BNT162b2 however a

heterologous schedule of ChAdOx1 nCoV-19 with a BNT162b2 second dose or two doses of BNT162b2 gave the highest responses after the second dose. [285]

Vaccination in those that have been shown to have serological evidence of prior SARS-CoV-2 infection has been shown to have an increased humoral and cellular response in older adults and in a recent pre-print was associated with broader T cell responses. [190,288]

1.8.2.3 mRNA vaccines

Two different mRNA-based vaccines, BNT162b2 and mRNA1273 both use a lipid nanoparticle to deliver RNA encoding a prefusion stabilised full-length SARS-CoV-2 spike protein into cells. Stabilisation of the spike protein by the introduction of additional prolines in the S2 subunit maintains it in its pre-fusion configuration thought to be optimal for the generation of antibodies that could block fusion. [15] Both have demonstrated excellent immunogenicity with antibody titres significantly higher than those seen with two doses of ChAdOx1 nCoV-19, [285] and efficacies of 95% (95% credible interval, 90.3 to 97.6) for BNT162b2 and 94.1% for mRNA-1273 (95% CI, 89.3 to 96.8%) [187,289] and comparable real-world effectiveness. [179,290]

1.8.2.4 Inactivated vaccines

Inactivated vaccines are made by growth of viral isolates in cell culture, and then chemical inactivation using for example β-propiolactone. [291,292] Typically, inactivated vaccines are combined with an adjuvant such as alum to optimise immunogenicity but because the whole virion is presented to the host immune system there is the theoretical benefit of exposure to multiple antigens and therefore potentially eliciting a broader immune response. From a manufacturing perspective at the outset of the pandemic, expertise existed about manufacture of such a product at scale placing these vaccine candidates in a logistically favourable position.

Using clinical isolates as starting stock, inactivated vaccine candidates (e.g. CoronaVac initially known as PiCoVacc (Sinovac Biotech), BBiBP-CorV (Sinopharm)) were found to be immunogenic in a range of animal species and protective against SARS-CoV-2 intratracheal challenge in non-human primates though these studies focused on lower respiratory tract pathology and did not measure upper airway protection or effect on viral shedding. [291,292] No safety concerns were raised by the animal studies.

Phase 1/2 studies confirmed immunogenicity and tolerability in humans, [293] and phase 3 trials showed protective efficacy against primary symptomatic disease ranging from 50.7% (95%CI 36·0-62·0) to 83·5% (95% CI 65·4–92·1) in different settings. [294,295] There is evidence of relatively rapid waning of antibody after inactivated vaccines, but immune memory is established, and boosting can be achieved. [296,297]

1.8.2.5 Protein subunit vaccines

Subunit vaccines contain purified proteins, derived either from cell cultures or from recombination, often in combination with an adjuvant to stimulate an immune response. Recently the NVX-CoV23/3 vaccine (Novavax), recombinant nanoparticle stable prefusion spike protein with Matrix-M adjuvant, has received MHRA authorisation for use in the UK following efficacy results of 89.7% against symptomatic infection in phase 3 trials. [298]

1.8.3 UK Vaccination Programme

The COVID-19 vaccination programme began on 8th December 2020, at first targeting high risk groups such as those in residential care, older adults, and frontline healthcare workers, then progressing to the whole adult population in phases determined by age and risk of severe disease. [299] Subsequently in mid-2021 vaccine programme roll out was extended to adolescents and at-risk younger children. In September 2021 the UK began third dose ‘booster’

vaccinations which was then accelerated to include all adults who had received two prior doses due to ongoing high levels of cases and due to the advent of the Omicron variant.

1.8.3.1 Impact of SARS-CoV-2 Vaccination

Modelling data using officially reported deaths and excess mortality figures suggest that globally, COVID-19 vaccination prevented 14.4 million (95% credible interval 13.7-15.9) deaths due to COVID-19. [300] In the UK, real world data has shown COVID-19 vaccination to be associated with reduced risk of death, hospitalisation, symptomatic disease, and transmission. [70,188,191,301]

1.9 NON-PHARMACEUTICAL INTERVENTIONS FOR PREVENTION OF INFECTION AND DISEASE

1.9.1 Masks

Guidance on the use of masks has been available since early in the pandemic [302] though accumulated evidence of limited quality from other respiratory viruses has not been conclusive. [303] Masks are available in a range of different types from homemade fabric face coverings, single use surgical masks, to medical grade respirators with a varying ability to filter aerosols, droplets, or particles. Case-control studies and modelling data have shown face mask adoption as an effective non-pharmaceutical intervention to reduce the risk of SARS-CoV-2 infection by preventing the transmission of large respiratory droplets between individuals. [304,305] High quality randomised controlled trials of mask wearing in relation to respiratory virus transmission pre-pandemic have been difficult to deliver due to the low compliance seen within intervention arms and, with respect to COVID-19, also due to the variability of public health interventions present at a given time in a particular epidemiological setting. [303,306] An updated Cochrane review of the impact of physical interventions including studies carried out within

the pandemic is awaited. [303] Pooled data from six studies showed a reduced incidence of COVID-19 with mask wearing and high-level data of country wide practices suggests mandated mask use may be associated with lower mortality. [307,308]

1.9.2 Handwashing

Many public health messages around the world have emphasised the importance of hand sanitation in the prevention of SARS-CoV-2 transmission. Prior to COVID-19, a systematic review and meta-analysis found a modest relative reduction in respiratory illnesses in groups randomised to hand hygiene measures but there was high heterogeneity in the included studies. [303] Three studies included in a meta-analysis of physical interventions on COVID-19 showed a non-statistically significant reduction in incidence in the handwashing group but with substantial risk of bias in the studies. [307]

1.9.3 Social Distancing

Social distancing for entire populations was an early measure introduced by many countries during the pandemic. As for many other interventions this intervention was supported by limited pre-existing evidence. [309] The impact of social distancing measures is closely linked with other factors including, but not limited to, mask-wearing, setting, and level of ventilation. Studies performed during the pandemic have shown social distancing was associated with a reduction on SARS-CoV-2 transmission and reduced mortality in countries or states where lockdowns had been implemented. [307]

1.10 SUMMARY

The recognition of multiple cases of severe respiratory disease in late 2019 evolved into the pandemic that became the dominant news story globally for much of the next two years. Much

of the world's population has been affected by the SARS-CoV-2 virus directly, or indirectly due to efforts to contain the virus.

Scientific knowledge and literature have expanded very quickly, with now a large volume of data collected on most aspects of this infection but at its onset scientists were operating with an unknown pathogen and disease. Rapid identification of multiple efficacious vaccines and their swift deployment has been a huge success; modelling estimates that 14.4 million (95% credible interval 13.7-15.9) deaths have been averted. [300] However unequal access, leaving low-income countries without adequate vaccination coverage means that many lives that could have been saved, were not.

Nearly three years on from its emergence, SARS-CoV-2 continues to evolve, and despite the knowledge and experience attained, the future remains uncertain as to what will happen next, particularly as social contact has resumed but the experience of the population's immunity has been atypical in recent years. In the UK many of the same questions that were asked at vaccine roll out remain pertinent: what type of protection does vaccination afford, how long does it last and will viral variants be able to evade pre-existing immunity?

1.11 Specific Aims

In this thesis, I will discuss several sub-studies conducted during the COVID-19 vaccine trials initiated in 2020. The overarching aim of these vaccine studies was to develop a COVID-19 vaccine that was protective against symptomatic disease suitable for population deployment.

The sub-studies that are presented here sought to explore basic questions about the SARS-CoV-2 viral parameters and the impact of ChAdOx1 nCoV-19 vaccination within the trial population.

The aims of the first study involving regular and symptomatic participant sampling included estimation of asymptomatic SARS-CoV-2 cases and evaluate the relationship between symptom status, viral parameters, and vaccination status. Part of this study also involved sequencing of viral isolates.

The aims of this project are to address these specific hypotheses:

- Weekly self-swabbing for SARS-CoV-2 PCR testing in trial participants is feasible as a novel approach to longitudinal surveillance within a pandemic (Chapter 3).
- Trial participants will not be highly adherent to weekly self-swabbing and adherence will decline over time (Chapter 3).
- Asymptomatic infection will occur in a minority of participants (Chapter 4)
- ChAdOx1 nCoV-19 vaccination will be associated with an increased proportion of asymptomatic cases (Chapter 4).
- ChAdOx1 nCoV-19 vaccination will result in reduced viral load and reduced duration of viral shedding in infected participants (Chapter 4).
- B.1.1.7 (Alpha) infections will have a higher viral load compared with pre-Alpha variants (Chapter 5).
- ChAdOx1 nCoV-19 vaccination may be associated with reduced efficacy against infections caused by the B.1.1.7 (Alpha) variant (Chapter 5).
- SARS-CoV-2 sequences obtained from controls may have fewer nucleotide substitutions compared to those from ChAdOx1 nCoV-19 vaccinees due to pre-existing immune responses (Chapter 5).
- SARS-CoV-2 sequences from ChAdOx1 nCoV-19 vaccinees may show a greater number of nucleotide, and potentially amino acid, substitutions compared with sequences from control participants (Chapter 5).

- Minor allele frequencies will be higher in sequences from vaccinated participants compared with controls (Chapter 5).
- Sequences from vaccinated participants will have a higher number of iSNVs compared with those from controls (Chapter 5).

The second study involved sampling stool from SARS-CoV-2 infected trial participants.

The specific aims of this project are to investigate the hypotheses that:

- Prior vaccination with ChAdOx1 nCoV-19 will increase SARS-CoV-2 shedding in stool in infected participants (Chapter 6).
- ChAdOx1 nCoV-19 will decrease stool SARS-CoV-2 viral load (Chapter 6).
- Prior vaccination with ChAdOx1 nCoV-19 will increase the titres of spike specific IgA measured in stool in participants who go on to develop SARS-CoV-2 infection (Chapter 6).

2 Methods

This methods section contains the information regarding the COVID-19 vaccine trials from which my samples and data were generated. These studies were originally designed, and protocols written by Pedro Folegatti, Daniel Jenkin, Sarah Gilbert, Andrew Pollard, Adrian Hill, Hannah Robinson, and Merryn Voysey but have had input from many other team members during the clinical delivery of the trials. Early iterations of the protocols of these studies have already been published as supplementary material to Voysey et al. (2000) [275] but I will summarise the key elements of these studies to provide the context for the chapters that follow.

Multiple formal substantial and non-substantial amendments have taken place during the conduct of these trials, both in response to new information regarding the vaccine and in response to the changing nature of the pandemic. For example, both the COV001 and COV002 studies were conceived as one dose studies, because single dose administration would have been preferable, particularly in low and middle-income settings in the context of a pandemic. However, early immunogenicity data from the COV001 study revealing the benefit of a second dose resulted in all participants in both trials being offered more than one dose of their allocated intervention (active vaccine or control vaccine). In addition, there have been many changes to the practical aspects of the trial such as the management of suspected or confirmed COVID-19 cases in response to frequent changes of government guidance. For clarity, I have taken the decision to only highlight the chronology and changes in trial methodology that are most important, or which have direct bearing to those sub-projects which are included in this thesis. Methods pertaining to specific sub-projects are detailed in the relevant chapters.

CLINICAL STUDIES

2.1 COV001

COV001 was a single-blinded, controlled, phase 1/2 study assessing the efficacy, safety, and immunogenicity of the intramuscular ChAdOx1 nCoV-19 coronavirus vaccine using a licensed intramuscular meningococcal conjugate vaccine (MenACWY) as a comparator. An early version of the protocol of this study was published as a supplement to the original paper reporting the preliminary results. [274] The primary and secondary objectives are shown in Table 2.1.

Table 2.1: COV001 Study Objectives

Objectives	**Outcome Measures**
Primary Objective	
To assess efficacy of ChAdOx1 nCoV-19 against COVID-19	a) Virologically confirmed (PCR or other nuclei acid amplification test positive) symptomatic cases of COVID-19
Co-primary Objective	
To assess the safety of the candidate vaccine ChAdOx1 nCoV-19	a) Occurrence of serious adverse events (SAEs) throughout the study until a cut-off date of 1st July 2021 or 6 months post late vaccination visit, whichever is latest
Secondary Objectives	
To assess the safety, tolerability and reactogenicity profile of the candidate vaccine ChAdOx1 nCoV-19	a) Occurrence of solicited local reactogenicity signs and symptoms for 7 days following vaccination b) Occurrence of solicited systemic reactogenicity signs and symptoms for 7 days following vaccination c) Occurrence of unsolicited adverse events (AEs) for 28 days following vaccination; (7 days following vaccination for groups 1c,1d,5a and 5b) d) Change from baseline for safety laboratory measures e) Occurrence of SAE of special interest: disease enhancement episodes

Objectives	Outcome Measures
To assess efficacy of ChAdOx1 nCoV-19 against COVID-19	a) Hospital admissions associated with COVID-19 b) Intensive care unit (ICU) admissions associated with COVID-19 c) Deaths associated with COVID-19 d) Severe COVID-19 disease (defined according to clinical severity scales) e) Seroconversion against non-Spike antigens
To assess cellular and humoral immunogenicity of ChAdOx1 nCoV-19	a) Interferon-gamma (IFN-γ) enzyme-linked immunospot (ELISpot) responses to SARS-CoV-2 spike protein b) Quantify antibodies against SARS-CoV-2 spike protein (seroconversion rates)

2.1.1 Study design

Between 23 April 2020 and 21 May 2020, 1077 healthy adults aged 18-55 were recruited at five study sites within the UK (Oxford, Southampton, Bristol, Imperial, and St George's in London). Participants underwent detailed screening by medical history, examination and screening blood and urine testing. Participants with a history of COVID-19 symptoms since 1 February 2020, household contact with individuals vulnerable to severe COVID-19 disease, in high-exposure occupations and who had received a diagnosis of COVID-19 were excluded. Full inclusion and exclusion criteria have been published. [274] Once antibody testing became available high-exposure healthcare workers were eligible if they had negative serology at screening (Substantial Amendment 3, 21 April 2020). Each site performed their own serological test which variably included N and/or S antigen testing. Serology was not performed on all participants prior to enrolment.

At the initiation of the study, participants were divided into 3 groups, with participants in the groups 1, 2 and 4 being randomised 1:1 to receive either vaccine or vaccine comparator. Group 3 had no randomisation and was unblinded as to allocation. Group 4 was added shortly after study initiation to increase the sample size.

Analysis of interim immunogenicity results showed an improvement of neutralising antibody titres in the prime-boost regimen compared with prime alone. [274] During subsequent iterations to the protocol (second dose) boost doses were offered to participants, first to a subset of group 2 (SA08 protocol version 8.0, 22nd June 2020), then to all participants in groups 2 and 4 (SA09 protocol version 9.0, 30th July 2020). Second doses occurred at variable intervals from first doses due to group allocation and vaccine availability.

Following national rollout of vaccination, further groups were constructed from original groups to assess the impact of a delayed second dose, (second dose) boosting with vaccine at half dose and third dose boosting (Table 2.2).

Table 2.2: Final table of groups for COV001. MenACWY = Meningococcal Group A, C, W-135, Y conjugate vaccine, ChAdOx1 nCoV-19 = replication-deficient simian adenovirus vector expressing the spike (S) protein of SARS-CoV-2, vp = viral particles.

Group	Allocation
1a	ChAdOx1 nCoV-19 5 x 1010 vp
1b	MenACWY
1c (from 1a)	Delayed homologous ChAdOx1 nCoV-19 boost
1d (from 1b)	Two-dose ChAdOx1 nCoV-19
2a	ChAdOx1 nCoV-19 5 x 1010 vp
2b	MenACWY
2c (from 2a)	Homologous Prime-Boost ChAdOx1 nCoV-19 5 x 1010 vp (8-week interval)
2d (from 2a)	Homologous Prime-Boost ChAdOx1 nCoV-19 5x1010 vp (prime) and 2.5x1010 vp (boost) (8-week interval)
2e (from 2b)	Two-dose MenACWY (8-week interval)
2f (from 2a)	Homologous Prime-Boost ChAdOx1 nCoV-19 5x1010 vp (prime) and 3.5-6.5x1010 vp (boost) (4-week minimum interval)
2g (from 2b	Two dose MenACWY (4-week minimum interval)
3a	Homologous Prime-Boost ChAdOx1 nCoV-19 5x1010 vp
3b (from 3a)	Delayed three dose ChAdOx1 nCoV-19 5 x 1010 vp
4a	ChAdOx1 nCoV-19 5 x 1010 vp
4b	MenACWY
4c (from 4a)	Homologous Prime-Boost ChAdOx1 nCoV-19 5x1010 vp (prime) and 3.5-6.5x1010 vp (boost) (4-week minimum interval)
4d (from 4b)	Two dose MenACWY (4-week minimum interval)
5a (from 2c, 2f,4c)	Delayed three dose ChAdOx1 nCoV-19 5 x 1010 vp
5b (from 2e, 2g, 4d)	Two dose MenACWY, two dose ChAdOx1 nCoV-19 5 x 1010 vp

2.1.2 Study procedures

Participants were monitored via electronic diary reporting of their symptoms post prime and (second dose) boost vaccinations and at clinic visits which varied in intensity according to group. Additionally, participants were asked to contact the study team if they experienced COVID-19 symptoms (fever $\geq$37.8°C, cough, shortness of breath, ageusia or anosmia). Symptomatic participants entered the Symptomatic Pathway for assessment of SARS-CoV-2 infection (see section named Symptomatic pathway).

Follow up visits for participants occurred (at variable intervals depending on group allocation) and involved the collection of safety information, blood samples as appropriate to group and review of interim medical history. All participants were followed up for six months following their last vaccine, all groups other than 1c, 1d and 5a had an optional visit at one year after their last vaccination. All participants in COV001 were compensated for their participation as previously described. [275]

2.1.3 Ethics statement

All participants provided written consent to take part in the study and re-sought when required following amendments. The study was approved by the Medicines and Healthcare products Regulatory Authority (ref: 21584/0424/001-0001) and the South-Central Berkshire Research Ethics Committee (ref: 20/SC/0145). This study is registered with ClinicalTrials.gov (NCT04324606) and with the ISRCTN registry (ISRCTN15281137). The study was performed in accordance with the principles of the Declaration of Helsinki and Good Clinical Practice guidelines.

2.1.4 Funding

UK Research and Innovation, Coalition for Epidemic Preparedness Innovations, National Institute for Health Research (NIHR), NIHR Oxford Biomedical Research Centre, Thames Valley, and South Midland's NIHR Clinical Research Network, and the German Center for Infection Research (DZIF), Partner site Gießen-Marburg-Langen. Full acknowledgements and role of funding source detailed in published paper. [274]

2.2 COV002

2.2.1 Study design

COV002 was a single-blind, randomised controlled phase 2/3 trial assessing the efficacy of the intramuscular ChAdOx1 nCoV-19 vaccine using a licensed intramuscular MenACWY vaccine as a comparator. Version 14.0 of the protocol (dated 9th November 2020) was published as supplementary appendix to the interim efficacy analysis which combined data from the four randomised controlled trials testing ChAdOx1 nCoV-19 efficacy. [275]

The primary and secondary objectives are shown in Table 2.3.

Table 2.3: Study Objectives for COV002

Objectives	
Primary	To assess efficacy of the candidate ChAdOx1 nCoV-19 against COVID-19 in adults aged 18 years and older
Co-Primary	To assess the safety of the candidate vaccine ChAdOx1 nCoV-19 in adults
Secondary	To assess the safety, tolerability and reactogenicity profile of the candidate vaccine ChAdOx1 nCoV-19
	To assess efficacy of the candidate ChAdOx1 nCoV-19 against severe and non-severe COVID-19
	To assess humoral immunogenicity of ChAdOx1 nCoV-19
	To assess cellular immunity of ChAdOx1 nCoV-19 in older adults (groups 1, 2, 7 and 8 only)
	To assess the safety and immunogenicity of a booster dose of ChAdOx1 nCoV-19 in older adults aged 56 years or older (two-dose schedules for groups 1, 2, 7 and 8 only)

2.2.2 Study procedures

Briefly, 10,812 adults aged 18 years and older were recruited at 19 sites across England, Wales, and Scotland between 31st May and 24th November 2020. Co-morbidities were permitted if mild or moderate and well controlled. Recruitment of individuals with high exposure risk to

SARS-CoV-2 were prioritised. All participants underwent a screening visit, medical history and if required by group, screening blood tests.

Participants were allocated into several study groups, which, for the most part was related to age and intensity of study visits (see Table 2.4). Other groups included participants who had received a ChAdOx1 viral vectored vaccine (for non-Covid vaccine trials) previously and individuals who were HIV positive. Participants were randomised to receive either the ChAdOx1 nCoV-19 vaccine or the licensed MenACWY vaccine according to the specific ratio for their allocated group.

Table 2.4: Study groups in COV002. Abs 260 = Absorption at 280 nm, PS80=Polysorbate 80. Table courtesy of Dr Robert Shaw.

Group	**Age**	**N**	**Intervention**	**Randomisation (ChAdOx1 nCoV-19: Comparator vaccine)**
1a1	56-69	30	Single dose ChAdOx1-nCoV-19 5x10e10vp (Abs 260)	3:1
1a2	56-69	10	Single dose MenACWY	
1a3	56-69	From 1a1	ChAdOx1-nCoV-19 5x10x^{10}vp followed by 0.5 ml (3.5 – 6.5 x 10^{10} vp)	
1a4	56-69	From 1a2	Two dose MenACWY	
1b1	56-69	30	ChAdOx1 nCoV-19 5x10e10vp (Abs 260) prime and 2.2 x10e10vp (qPCR) boost (4-6 weeks apart)	3:1
1b2	56-69	10	Two-dose MenACWY (4-6 weeks apart)	
2a1	>=70	50	Single dose ChAdOx1-nCoV-19 5x10e10vp (Abs 260)	5:1
2a2	>=70	10	Single dose MenACWY	
2a3	>=70	From 2a1	Two-dose ChAdOx1 nCoV-19 5x10e10 vp (Abs 260) prime and 0.5ml (3.5 – 6.5 x10e10 vp, Abs 260, corrected for PS80 boost)	
2a4	>=70	From 2a2	Two dose MenACWY	
2b1	>=70	50	Two-dose ChAdOx1 nCoV-19 5x10e10vp (Abs 260) prime and 2.2 x10e10vp (qPCR) boost (4-6 weeks apart)	5:1
2b2	>=70	10	Two-dose MenACWY (4-6 weeks apart)	
4a1	18-55	1775	Single dose ChAdOx1-nCoV-19 5x10e10vp (Abs 260)	1:1
4a2	18-55	1775	Single dose MenACWY	
4b1	18-55	50	Two-dose ChAdOx1 nCoV-19 5x10e10vp (Abs 260) prime and 2.2 x10e10vp (qPCR) boost (4-6 weeks apart)	1:1
4b2	18-55	50	Two-dose MenACWY (4-6 weeks apart)	

Group	Age	N	Intervention	Randomisation (ChAdOx1 nCoV-19: Comparator vaccine)
4c1	18-55	1725 From 4a1	Two-dose ChAdOx1 nCoV-19 5x10e10vp (Abs 260) prime and 0.5mL (3.5 – 6.5 × 10e10 vp, Abs 260, corrected for PS80) OR ChAdOx1 nCoV-19 5x10e10vp (qPCR), (4-12 weeks apart, +2 weeks)	
4c2	18-55	1725 From 4a2	Two-dose MenACWY, (4-12 weeks apart, +2 weeks)	
5a1	18-55	50	ChAdOx1 nCoV-19 5x10e10vp (Abs 260)	1:1
5a2	18-55	50	Single dose MenACWY	
5a3	18-55	From 5a1	Two-dose ChAdOx1 nCoV-19 5x10e10vp (Abs 260) prime and 0.5mL (3.5 – 6.5 × 10e10 vp, Abs 260, corrected for PS80)	
5a4	18-55	From 5a2	Two-dose MenACWY	
5b1	18-55	25	ChAdOx1 nCoV-19 5x10e10vp (qPCR)	1:1
5b2	18-55	25	Single dose MenACWY	
5c1	18-55	25	ChAdOx1 nCoV-19 5x10e10vp (qPCR)	1:1
5c2	18-55	25	Single dose MenACWY	
5d1	18-55	50	Two dose ChAdOx1 nCoV-19 0.5mL (3.5 – 6.5 × 10e10 vp Abs 260, corrected for PS80)	5:1
5d2	18-55	10	Two-dose MenACWY (4-6 weeks apart)	
5e	18-55	Up to 15	Two-dose ChAdOx1 nCoV-19 vaccine (Covishield 0.9 x 10e11 vp/ml), 0.5 ml (4-6 weeks apart)	Open label
5f	18-55	Up to 15	Two-dose ChAdOx1 nCoV-19 vaccine (Covishield 0.9 x 10e11 vp/ml), 0.25ml prime and 0.5 ml boost (4-6 weeks apart)	Open label
6a1	18-55	3000	ChAdOx1 nCoV-19 5x10e10vp (qPCR)	1:1
6a2	18-55	3000	Single dose MenACWY	

Group	Age	N	Intervention	Randomisation (ChAdOx1 nCoV-19: Comparator vaccine)
6b1	18-55	3000 From 6a1	Two-dose ChAdOx1 nCoV-19 5x10e10vp (qPCR) prime and 0.5mL (3.5 – 6.5 × 1010 vp Abs 260, corrected for PS80) OR ChAdOx1 nCoV-19 5x10e10vp (qPCR), (4-12 weeks apart, +2 weeks)	1:1
6b2	18-55	3000 From 6a2	Two-dose MenACWY, (4-12 weeks apart, +2 weeks)	
7a1	56-69	30	Single dose ChAdOx1-nCoV-19 5x10e10vp (qPCR)	3:1
7a2	56-69	10	Single dose MenACWY	
7b1	56-69	30	Two-dose ChAdOx1 nCoV-19 5x10e10vp (qPCR) (4-6 weeks apart)	3:1
7b2	56-69	10	Two-dose MenACWY (4-6 weeks apart)	
8a1	>=70	50	Single dose ChAdOx1-nCoV-19 5x10e10vp (qPCR)	5:1
8a2	>=70	10	Single dose MenACWY	
8b1	>=70	50	Two-dose ChAdOx1 nCoV-19 vaccine, 5x10e10 vp (qPCR) prime and 0.5mL (3.5 – 6.5 × 10e10 vp, Abs 260, corrected for PS80) boost OR 5x10e10vp (qPCR) boost (4-6 weeks apart)	5:1
8b2	>=70	10	Two-dose MenACWY (4-6 weeks apart)	
9a1	56-69	500	Two dose ChAdOx1 nCoV-19 0.5mL (3.5 – 6.5 × 10e10 vp, Abs 260, corrected for PS80), (4-6 weeks apart)	1:1
9a2	56-69	500	Two-dose MenACWY (4-6 weeks apart)	
10a1	>=70	500	Two dose ChAdOx1 nCoV-19 0.5mL (3.5 – 6.5 × 10e10 vp, Abs 260, corrected for PS80), (4-6 weeks apart)	1:1
10a2	>=70	500	Two-dose MenACWY (4-6 weeks apart)	

Group	Age	N	Intervention	Randomisation (ChAdOx1 nCoV-19: Comparator vaccine)
11	18-55	60	Two dose ChAdOx1 nCoV-19 0.5mL (3.5 – 6.5 × 10e10 vp, Abs 260, corrected for PS80), (4-6 weeks apart) in individuals who have previously received a ChAdOx1 vectored vaccination (non-Covid)	Open label
12	18-55	60	Two dose ChAdOx1 nCoV-19 0.5mL (3.5 – 6.5 × 10e10 vp, Abs 260, corrected for PS80), (4-6 weeks apart) in HIV positive individuals	Open label

Follow-up visits were conducted after vaccination according to the schedule for each study group. All study groups were followed up for a total of 12 months from their last (within trial) vaccination visit. Participants in groups 1-3, 5, 7, 8, 11 and 12 were asked to complete 7 days of solicited adverse event diary entry following vaccination and 28 days post vaccination for unsolicited adverse events. A subset of participants in groups 4, 6, 9 and 10 (approximately 3000) were asked to record solicited and unsolicited adverse events for 7 days after vaccination. During follow up visits, participants were assessed for adverse events, interim medical history, and examination where necessary and blood testing as per schedule appropriate for their group. Participants in groups 1, 2, 5, 7, 8 and 12 received compensation for their participation and those in groups 4, 6, 9, 10 and 11 did not, as previously described. [275]

The ChAdOx1 nCoV-19 vaccines used in the trial were manufactured by different organisations which employed different methods of viral vector quantification (spectrophotometry and quantitative PCR). The first participants vaccinated in the trial received an initial batch of ChAdOx1 nCoV-19 from one manufacturer. When the second batch underwent quality control it was noticed that quantification by different methods produced different results. Following discussions with MHRA, the lower value (5×10^{10} vp by spectrophotometer, 2.2×10^{10} vp by qPCR) was chosen for use in the trial, as this reflected the spectrophotometer readings for the batch used in the COV001 study, so it was known to be safe and likely to be immunogenic. However, when this second batch was administered to participants, the reactogenicity observed was less than that seen in the COV001 trial. It was later found that an excipient (PS80) in the vaccine resulted in a falsely high spectrophotometry reading and the qPCR was thence taken as the more accurate of the two. This was amended in protocol version 6.0, dated 5 June 2020.

This resulted in participants aged 18-55 receiving different dosage regimens, thereby producing a LD/SD (low dose/standard dose) and a SD/SD (standard dose/standard dose) cohort. All participants aged 56 and older received two standard doses.

As immunogenicity data accrued, the original intention of a single dose study for younger cohorts (18-55 years) was amended to allow for second doses (version 9.0 20 July 2020). Participants aged 56 and older were recruited into two dose groups from their enrolment.

As for COV001, COV002 participants were asked to contact the study team if they experienced COVID-19 symptoms (primary symptoms defined as: fever $\geq$37.8°C, cough, shortness of breath, ageusia or anosmia). Symptomatic participants entered the Symptomatic Pathway for assessment of SARS-CoV-2 infection (see section Symptomatic pathway).

2.2.3 Home swabbing pathway

Additionally, COV002 participants were asked to perform weekly self-administered nose and throat swabs and return by post during their study participation until 30th June 2021. These were tested using PCR for SARS-CoV-2 (detailed methods in Chapter 3) which were performed in collaboration with the UK Department of Health and Social Care (DHSC).

2.2.4 Clinical oversight

All participants in both trials had 24-hour contact information of the on-call physician for their study site for the duration of their involvement in the trial. Serious Adverse Events were recorded for the duration of both trials covered in this thesis.

2.2.5 Ethics statement

All participants provided written consent to take part in the study. The study was approved by the Medicines and Healthcare products Regulatory Authority (Ref: 21584/0428/001-0001) and the Research Ethics Committee (Ref: 20/SC/0179). This study is registered with ClinicalTrials.gov (NCT04400838) and with the ISRCTN registry (ISRCTN90906759). The study was performed in accordance with the principles of the Declaration of Helsinki and Good Clinical Practice guidelines.

2.2.6 Funding

UK Research and Innovation, National Institutes for Health Research (NIHR), Coalition for Epidemic Preparedness Innovations, Bill & Melinda Gates Foundation, NIHR Oxford Biomedical Research Centre, Thames Valley, and South Midland's NIHR Clinical Research Network, and AstraZeneca. Full acknowledgements and role of funding source in Voysey et al. (2021). [275]

2.3 SYMPTOMATIC PATHWAY

This pathway was used for symptomatic participants from both the COV001 and COV002 studies.

2.3.1 Procedures

For the COV001 and COV002 a shared symptomatic pathway was in place with clinical procedures common to both trials. An early version of the clinical study plan outlining these procedures was published as supplementary material to the original COV001 paper. [274]

Participants were asked to contact their study sites if they experienced symptoms of COVID-19 (fever ≥37.8°C, cough, shortness of breath, ageusia or anosmia) or if they were admitted to hospital for any reason. They were reminded weekly by email or text message by their study site to do this.

If a participant contacted their study site and met the criteria for assessment they were invited for a clinical visit and underwent NAAT testing by nasopharyngeal swab, clinical assessment and blood sampling for safety and immunology.

Initially all symptomatic participants were asked to record their symptoms in an electronic diary, this element was discontinued on 22nd Feb 2021. Follow up visits after the first symptomatic visit varied as the trial continued. Initially, when the viral kinetics of SARS-CoV-2 were less well understood, all participants had oral and nasopharyngeal testing at symptom onset and after 7 days. After 8th August 2020, participants who tested negative at their original symptomatic pathway visit performed a self-administered swab at home within days 3-5 after their original visit. If participants tested positive on either their first symptomatic pathway visit or home swab on day 3-5 then a follow up 7 days after their positive test was arranged. From 22nd February 2021 the visit at day 7 after a positive test was converted to a phone visit.

If participants were NAAT tested outside of the trial for any reason, they were asked to let their study site know of the result. If they communicated this to the study site, the results, and type of test were recorded in the electronic case report form (CRF). Participants were asked to provide evidence of the test (screenshot or email of result) where possible.

All cases of COVID-19 were assessed by two members of a clinical review endpoint committee who were blinded as to vaccine allocation. These clinicians reviewed the NAAT

result for each positive case in the context of clinical history and adverse events and then assigned individuals with a symptom status and WHO clinical progression score. [310]

2.4 SAMPLE PROCESSING FOR THE COV001 AND COV002 TRIALS

2.4.1 SARS-CoV-2 NAAT testing for symptomatic participants

At the time these studies were initiated, several different NAAT assays were available and to ensure rapid turnaround of NAAT results from symptomatic visits, testing was performed locally at each study site. NHS laboratories often had several assays validated for use due to concerns about reagent and test availability during the pandemic. NAAT assays used at each study site are shown in Appendices (Appendix 3).

2.4.2 Stool processing

Between 19th August to 10th November 2020 date participants who tested positive by NAAT for SARS-CoV-2 from an upper airway sample were asked to provide a stool sample. Samples were analysed by PCR and ELISA for anti-spike IgG and IgA. Please see chapter 6 for detailed methods.

2.4.3 SARS-CoV-2 sequencing

Swabs from the symptomatic pathway or weekly self-swabs which tested positive were identified where possible and sequenced by COG consortium partners. Please see chapter 5 for detailed methods.

2.5 DATA CAPTURE

Study data for COV001 and COV002 were collected and managed using REDCap electronic data capture tools hosted at University of Oxford. [311,312] REDCap (Research Electronic Data Capture) is a secure, web-based software platform designed to support data capture for research studies, providing 1) an intuitive interface for validated data capture; 2) audit trails for tracking data manipulation and export procedures; 3) automated export procedures for seamless data downloads to common statistical packages; and 4) procedures for data integration and interoperability with external sources.

2.6 STATISTICS

Analyses were performed using R (version 4.1.2), individual statistical tests are described in specific chapters. [313]

3 Longitudinal participant sampling for SARS-CoV-2 infection in a ChAdOx1 nCoV-19 vaccine trial

3.1 INTRODUCTION

SARS-CoV-2 was formally recognised in China in December 2019 [145] whilst the first cases of infection in the UK were reported in late January 2020. [314] SARS-CoV-2 vaccine trials began shortly after in the UK in April and May 2020. [274,275] The choice, definition and measurement of endpoints associated with efficacy within vaccine clinical trials are key. For some pathogens, particularly for low incidence but serious diseases such as those caused by *Neisseria meningitidis*, a randomised controlled trial to prove vaccine efficacy would be unfeasibly large, and so the endpoint may be attainment of an established immunological surrogate of protection. [315] Alternatively, for more common infections such as influenza and pneumococcal disease vaccine efficacy may be assessed by probable disease based on symptoms or signs (such as pneumonia), detection of the pathogen, medical attendance, or mortality. [316,317]

Several COVID-19 vaccine trials used differing definitions of primary endpoints to assess efficacy from symptomatic COVID-19. [275,282,283,289,298] COVID-19 vaccine trials were initiated and designed rapidly. At that time there was a limited knowledge base about this emerging pathogen as well as uncertainty about how many cases were likely to occur within study populations, particularly in the face of public health restrictions designed to limit transmission. This presented difficulties in making an informed decision as to which endpoints should be chosen, how endpoints should be defined in terms of symptom profile and severity, and whether an endpoint involving infection in the absence of symptoms should be included within a clinical trial to allow assessment of the impact of vaccination on SARS-CoV-2.

Evidence from SARS-CoV infections in 2003 suggested that peak viral load may occur in the second week of infection, viral shedding may be prolonged, and that higher viral loads were associated with increasing illness severity. [80] Serological studies following SARS-CoV exposure also revealed the ability of coronaviruses to cause asymptomatic infection. [318] Early SARS-CoV-2 evidence from mild cases in a German hospital confirmed a similar but not identical pattern to SARS-CoV; nose and throat samples were PCR positive at the onset of symptoms, were highest during the first week before declining, and detectable viral RNA could persist in the upper respiratory tract for many days after onset. [87] The potential for individuals to be infected with SARS-CoV-2 without symptoms was established relatively early in the pandemic from outbreak assessments, but the relative contribution of asymptomatic disease to transmission was not understood. [90,125-127]

One means of assessing a pathogen with the potential for asymptomatic infection as well as gaining information on viral dynamics is by regular longitudinal sampling of a trial population. A meta-analysis of SARS-CoV-2 viral dynamics demonstrated that the mean duration of RNA shedding was 17.0 days (95% CI 15·5–18·6) [80] suggesting that weekly testing may be frequent enough to capture multiple samples per individual if they became infected. The weekly swabbing sub-study of COV002 sought to gain weekly upper airway SARS-CoV-2 PCR tests from trial participants during their trial participation to look for asymptomatic infection and capture information about duration of infection. It was unknown whether trial participants would adhere to trial requests to send swabs regularly, whether this could be sustained for a prolonged study period, and whether changes in access to other testing, prevalence of SARS-CoV-2 infection or public health guidance for individuals testing positive (and their contacts) would influence these. It was also unknown at trial initiation if self-performed unsupervised testing would provide suitable samples, whether transit via the regular

postal system without temperature control would impact negatively on sample quality and if genomic sequences could be obtained from these samples after PCR processing in high-throughput settings.

In this chapter the data from symptomatic and weekly self-swabs from the COV002 vaccine trial will be explored, with the following specific objectives:

- to describe the demographics of COV002 participants and the subset of those who returned at least one SARS-CoV-2 NAAT positive swab.
- to describe the sources and temporal pattern of SARS-CoV-2 NAAT positive swabs.
- to assess which weekly self-swabs may represent false positive results.
- to describe SARS-CoV-2 infection rates within the trial.
- to investigate adherence to weekly self-swabbing within the COV002 vaccine trial.
- to explore participant factors associated with infection.

3.2 METHODS

The study design and generic methods for COV002 are given in chapter 2, including the testing of participants with symptoms. Symptomatic visits were performed at study sites and NAAT testing performed locally. Specific details relating to weekly swabbing pathway are described below.

My role was to set up the weekly self-swab pathway, project manage its delivery, liaise with DHSC and Deloitte Consultancy to ensure extensions to the programme, stock delivery to sites and provide support to participants at the Oxford sites and external sites on swabbing issues, maintain the data agreement with NHS Digital and manage the relationships with the Lighthouse Laboratories.

I have collated the data from the trial and from the Lighthouse laboratories and performed the analyses contained in this thesis.

3.2.1 Weekly self-swabbing

All COV002 participants were asked to perform a home self-swab test once per week starting one week after their prime vaccination. DHSC provided home test kits to the COV002 clinical trial which were distributed directly to participants at study visits by sites. Initially participants were asked to do this for 8 weeks, but as feasibility and security of kit supply was established, the duration of self-swabbing was extended during the trial and participants were asked to swab until the end of their participation in the trial or until 30th June 2021, whichever came sooner for a given individual.

Participants were given standardised instructions and asked to swab their nose and throat. Packaging was provided within the kits to allow the swabs to be returned using Royal Mail. Participants had to register their personal details and the kit barcode on a government online portal, which allowed the results from these tests to be delivered directly to participants by email or telephone. Isolation advice was provided directly to participants by government services (NHS Test and Trace in England, Test, Trace, Protect in Wales, and NHS Scotland Test and Protect).

Tests were processed by DHSC Lighthouse laboratories, principally Glasgow, Alderley Park (Manchester) and Milton Keynes which all used the same RT-PCR assay described below. Cycle threshold (Ct) data was not available for other processing laboratories as these laboratories used alternative PCR assays.

RT-PCR for three SARS-CoV-2 target genes (Nucleocapsid (N), Spike (S) and ORF1ab (ORF) used the Thermo Fisher TaqPath RT-PCR COVID-19 assay (Thermo Fisher Scientific,

Waltham, MA, USA), analysed using UgenTec Fast Finder 3.300.5 (TaqMan 2019-nCoV Assay Kit V2 UK NHS ABI 7500 v2.1, UgenTec, Cambridge, MA, USA). The Assay Plugin contains an Assay specific algorithm and decision mechanism that allows conversion of the qualitative amplification Assay PCR raw data from the ABI 7500 Fast into test results with minimal manual intervention. Samples were called positive in the presence of at least single N gene and/or ORF1ab but may be accompanied with S gene (1, 2 or 3 gene positives). S gene was not considered a reliable single gene positive (as of mid-May 2020).

The minimum Ct value for an individual swab was taken as the minimum Ct value from any positive gene target.

3.2.2 Matching and data capture

A data agreement between NHS Digital and the University of Oxford allowed data matching between barcoded swabs and trial participants by the study data team. Positive weekly swab results were uploaded to the REDCap® trial database. [311,312] This did not occur in real time. Participants were not contacted by study sites with their weekly swabbing results, but all participants received standardised weekly reminders by email or text message to remind them to return their home kits each week.

The NHS swab dataset was then matched by code written by the trial statistics team in R statistical software (version 4.1.2, R Foundation for Statistical Computing, Vienna, Austria) [313] to the trial dataset. Matching was defined by at least one of six combinations whereby at least two personal identifiers matched (such as name, NHS number or date of birth).

3.2.3 Legal safeguards for those repeatedly positive

Initially, participants who tested positive on serial weekly self-swabs were repeatedly asked to self-isolate in keeping with original UK national coronavirus legislation. A subsequent change in legislation allowed participants to continue to provide weekly swabs but not to have to repeat their periods of isolation for subsequent positive tests within 90 days [319](legislation since revoked).

3.2.4 Assessment of false positives

False positives were assessed using criteria adapted from Wei et al. [152] and Walker et al. [320].

Wei et al. defined 'strong evidence for a true-positive PCR result' as those that had Ct values $\leq$32 and $\geq$2 genes detected. Swabs not meeting these criteria were further refined by excluding swabs if they were from an individual who returned other positive swabs (either within 90 days or with fewer than three negative consecutive swabs between positives) or they were associated with known symptoms.

Walker et al. defined 'low quality' positives as single gene positives with a Ct value>34 and no evidence of symptoms. Swabs meeting these criteria were further refined by excluding swabs if they were from an individual who returned other positive swabs as defined above.

3.2.5 Statistical analysis

For swab-based analyses all swabs received and matched to study participants during their enrolment in the trial were included.

For participant-based analyses, participants who had not reached 21 days after their initial control or ChAdOx1 nCoV-19 vaccination by the time of their first positive swab were

excluded from analyses involving vaccination status. Potential reinfections were defined as instances where there were three consecutive negative swabs or a duration of >90 days between positive swabs. For participant and infection-based analyses involving participants with potential multiple infections by this definition, only swabs from the first episode were included, swabs from subsequent infections were excluded.

Demographic univariate analyses for sex, the listed comorbidities, any co-morbidity, smoking, and alcohol consumption were performed using Pearson's Chi-squared test, age and BMI using Wilcoxon rank sum test and ethnicity using Fisher's exact test.

For assessment of adherence for individual participants the number of weeks a participant was eligible to send swabs was calculated; starting one week after prime vaccination and their eligibility ended at the end of their individual study participation or 30th June 2021, whichever came first. Adherence was calculated by the number of swabs a participant had returned divided by the number of eligible weeks. Blinded swab adherence for individual participants was calculated in the same way for those swabs returned during the period when the participant was blinded. Where data are analysed by calendar week swabs received during that week were divided by eligible population.

Predictors of the percentage of returned self-swabs were analysed using a multiple linear regression model. Factors associated with not returning any swabs and testing positive for SARS-CoV-2 from any testing source were analysed using generalised linear models for binary data. Model assumptions were tested and only models that met the assumptions were retained. The percentage of swabs returned for those who had at least one positive swab was compared with those who never tested positive using Wilcoxon rank sum test.

3.3 RESULTS

3.3.1 Demographics of COV002 participants

There were 10,812 participants enrolled across all 19 sites in the UK into the COV002 trial between 28th May 2020 and 8th December 2020. The demographics for the participants are shown in Table 3.1. Over half of the participants were female, two thirds were healthcare workers, and the majority were white and in employment. The age distribution is shown in Table 3.2, most participants were between the ages of 25 and 55. Co-morbidities were infrequent, the most common of which were cardiovascular disease, asthma, and gastrointestinal disease.

Participants received their first vaccinations between 28th May 2020 and 8th December 2020 and their booster doses (where applicable) between 27th June 2020 and 11th March 2021.

Table 3.1: Demographics of COV002 participants. BMI was missing for 2 participants.

Characteristic	Participants	n (%)
Female sex, n (%)	10,812	6,441 (60%)
Age at screening, years (IQR)	10,812	44 (33, 54)
Ethnicity, n (%)	10,812	
White		9,941 (92%)
Asian		553 (5.1%)
Mixed		180 (1.7%)
Black		49 (0.5%)
Other		59 (0.5%)
Arab		26 (0.2%)
Prefer not to disclose		4 (<0.1%)
Occupation, n (%)	10,812	
Health and social care setting workers		6,998 (65%)
Other occupation		1,947 (18%)
Retired		1,438 (13%)
Student		120 (1.1%)
Hospitality/retail workers in non-essential shops and branches		62 (0.6%)
Public safety workers (e.g. police, fire services, security)		65 (0.6%)
Retail workers in essential shops and branches (e.g. food, chemists, banks)		47 (0.4%)
Drivers and transport workers		48 (0.4%)
Unemployed		41 (0.4%)
Construction workers and labourers		33 (0.3%)
Religious workers		10 (<0.1%)
Cleaning and domestic workers		3 (<0.1%)
Cardiovascular disease, n (%)	10,812	1,311 (12%)
Asthma, n (%)	10,812	1,111 (10%)
Chronic obstructive airways disease, n (%)	10,812	23 (0.2%)
Bronchiectasis, n (%)	10,812	13 (0.1%)
Other respiratory disease, n (%)	10,812	216 (2.0%)
Diabetes mellitus, n (%)	10,812	220 (2.0%)
Renal disease, n (%)	10,812	207 (1.9%)
History of stroke, n (%)	10,812	70 (0.6%)
Gastrointestinal disease, n (%)	10,812	1,288 (12%)
Liver disease, n (%)	10,812	174 (1.6%)
Any co-morbidity, n (%)	10,812	3,649 (34%)
BMI, n (%)	10,810	25.4 (22.9, 28.9)
Smoker, n (%)	10,812	563 (5.2%)
Consumes alcohol, n (%)	10,812	9,317 (86%)

Table 3.2: Age distribution of COV002 participants

Characteristic	Participants n	n (%)	UK population (2011 census data)[321]
Age group	10,812		
Under 25		585 (5.4%)	5267401 (9.4%)
25-34		2,397 (22%)	7520524 (17.1%)
35-44		2,558 (24%)	7831250 (17.8%
45-54		2,595 (24%)	7702220 (17.5%)
55-64		1,141 (11%)	6561077 (14.9%)
65-74		1,142 (11%)	4852833 (11.0%)
75-84		383 (3.5%)	3115552 (7.1%)
85 and over		11 (0.1%)	1254688 (2.8%)

The number of participants enrolled in the COV002 trial remained high until the end of the swabbing period (Figure 3.1A). Older participants were enrolled later than younger participants (Figure 3.1C). The roll of UK national vaccination programme prompted unblinding which led to a rapid drop in blinded participation across age groups (Figure 3.1B). Due to the high proportion of healthcare workers who were prioritised for vaccination unblinding occurred across all age groups (Figure 3.1D). Those who were not eligible for vaccination for occupational or medical criteria were typically unblinded when invited for vaccination by the UK national programme which was stratified by age so the proportion of younger participants who remained blinded rose over time (Figure 3.1D).

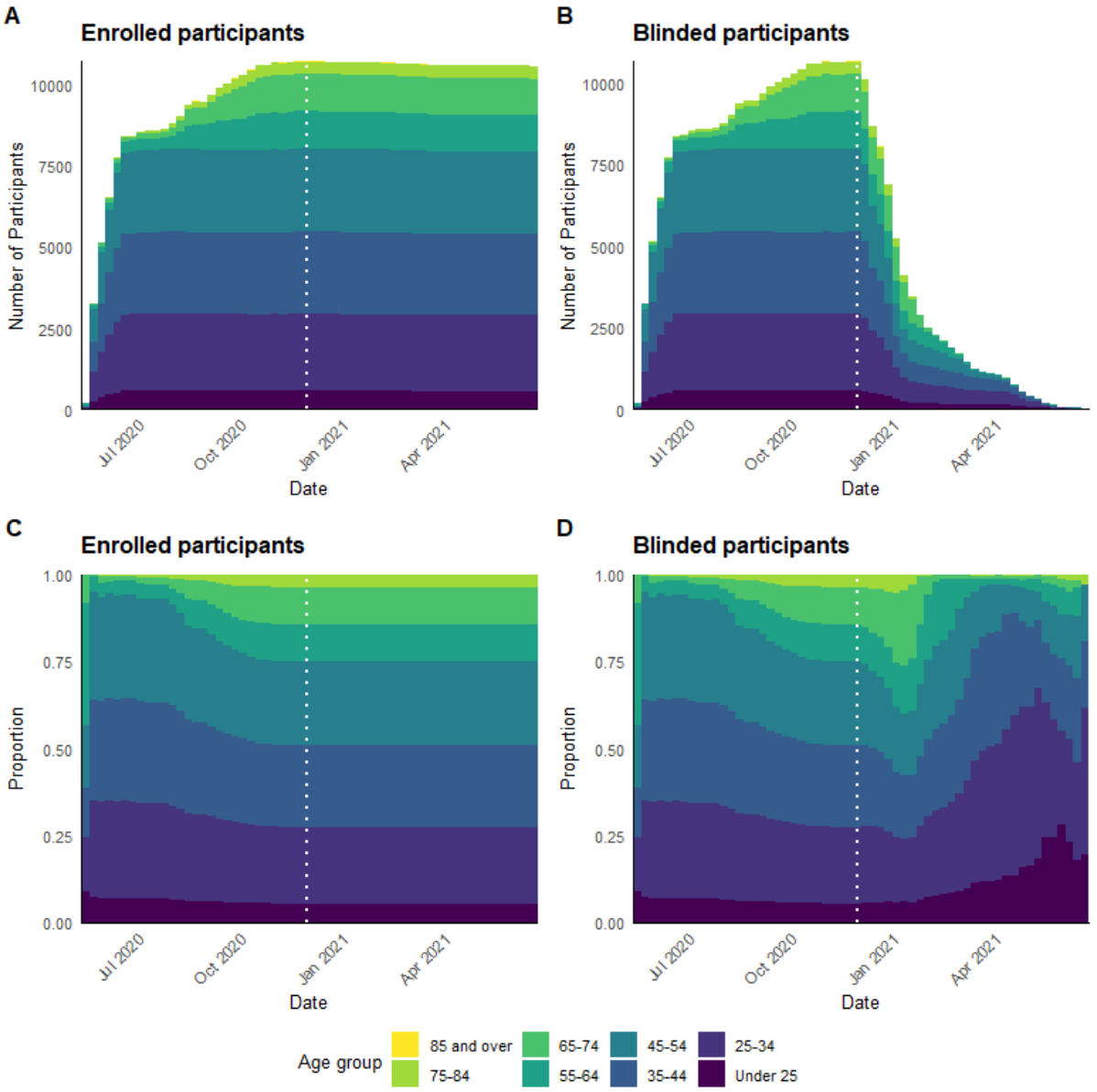


Figure 3.1: Enrolment and blind status of participants by age group from enrolment to 30th June 2021. Figures A and B present absolute numbers, figures C and D present proportion of total number (enrolled or blinded) by week. Dotted white line represents week when UK national vaccination programme began (8th December 2020).

3.3.2 Swab sources and results

Between 30th May 2020 and 30th June 2021, a total of 358,802 swabs were included within the trial analyses, of which 99.9% (358,440) were processed either at trial sites or at the Lighthouse laboratories and for which verification of the result was therefore available. Of the

10,812 participants enrolled in COV002 10,500 participants had a NAAT result available from at least one upper airway swab.

During this period 1130 (10.45%) participants tested positive for SARS-CoV-2 and returned 2251 (0.63%) positive swabs in total. Weekly swabs formed the main source of positive swabs in the trial with initial symptomatic visits providing most of the other positive results (Table 3.3).

Most weekly swabs were negative because these typically were sent by asymptomatic individuals as surveillance. However, 37% of swabs taken from participants who attended initial symptomatic visits and therefore had trial-defined swab criteria symptoms (any one or more of: cough, fever $\geq$37.8°C, shortness of breath, anosmia, ageusia) were SARS-CoV-2 positive on NAAT tests performed by trial staff.

Table 3.3: Source of SARS-CoV-2 NAAT positive swabs. s3-5: swab taken 3-5 days after initial symptomatic visit (only performed if initial symptomatic visit swab was negative), s7: swab taken 7 days after initial symptomatic visit (until August 2020, taken if initial symptomatic visit swab was negative).

		Swab NAAT Result		
	Overall n (%) N = 358,802	**Negative n (%) N = 350,000**	**Positive n (%) N = 2,251**	**Indeterminate n (%) N = 6,551**
Source of swab				
Weekly self-swab	356,707	348,550 (98%)	1,623 (0.5%)	6,534 (1.8%)
Initial symptomatic visit	1,153	717 (62%)	430 (37%)	6 (0.5%)
s3-s5 home self-swab	448	434 (97%)	4 (0.9%)	10 (2.2%)
s3-s5 visit	40	40 (100%)	0 (0%)	0 (0%)
s7 visit	92	92 (100%)	0 (0%)	0 (0%)
Workplace routine test	90	66 (73%)	24 (27%)	0 (0%)
Workplace test due to symptoms	127	67 (53%)	60 (47%)	0 (0%)
Other e.g. NHS testing	145	34 (23%)	110 (76%)	1 (0.7%)

Weekly self-swabs (n=356,707) were processed by the Glasgow (n=189762, 53.2%), Milton Keynes (94328, 26.4%), and Alderley Park (66869, 18.7%) Lighthouse laboratories. The remaining 1.6% (5748) swabs were processed in alternative Pillar 2 laboratories. There

was no difference in the rate of positives between different laboratories. There was a higher rate of indeterminate results at other laboratories which may have been related to different assays and QC procedures (Table 3.4).

Table 3.4: PCR results by Processing laboratory

	Glasgow N=189762	**Milton Keynes N=94328**	**Alderley Park N=66869**	**Other N=5748**	**Total n=356,707**
Positive	1056 (0.6%)	279 (0.3%)	264 (0.4%)	24 (0.4%)	1623 (0.5%)
Negative	185044 (98%)	93122 (99%)	65061 (97%)	5323 (93%)	348,550 (98%)
Indeterminate	3662 (1.9%)	927 (1%)	1544 (2.3%)	401 (7%)	6534 (1.8%)

Positive results from weekly self-swabs over time mirrored national case data (Figure 3.2).

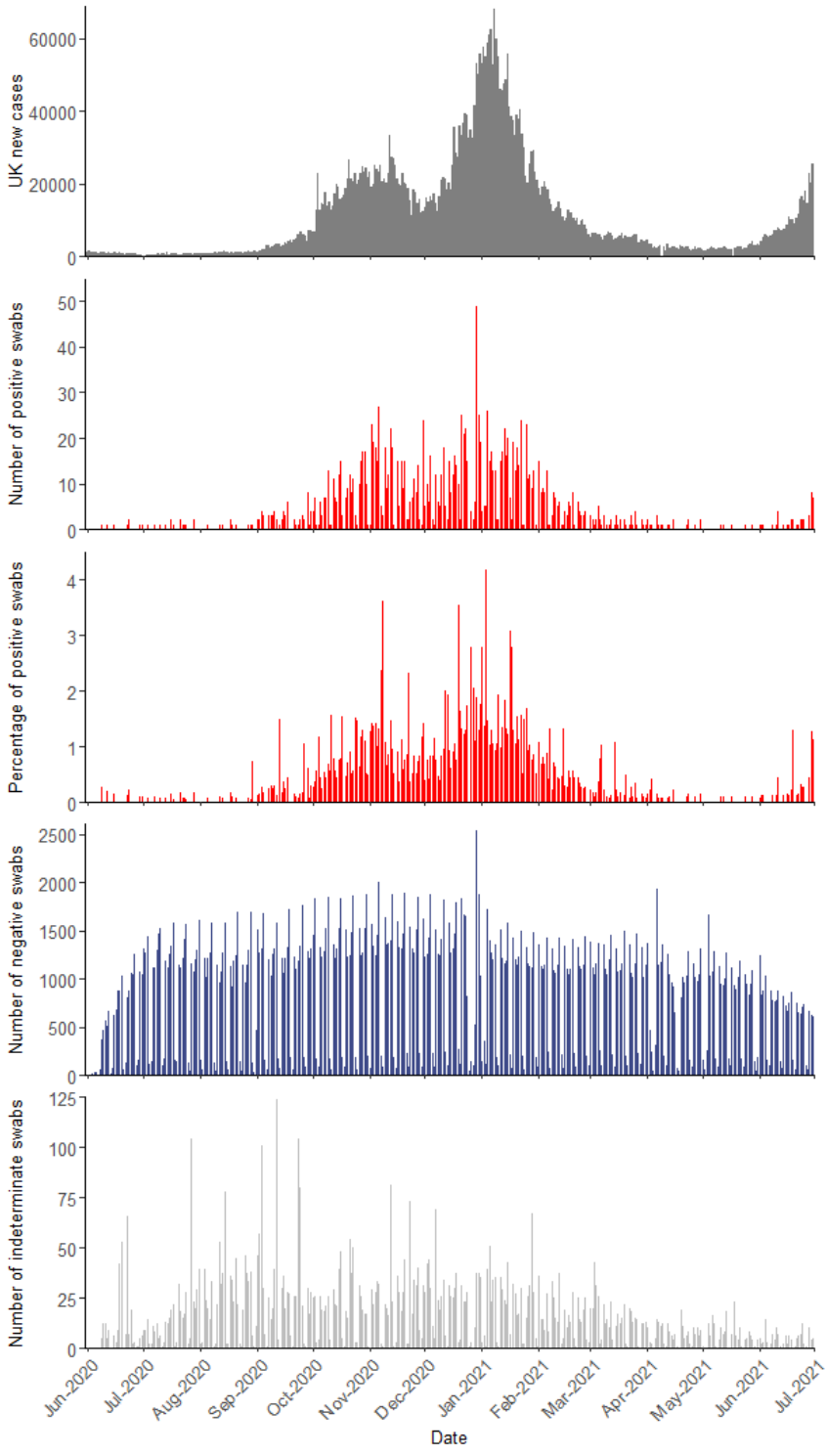


Figure 3.2: Daily counts for the UK (averaged over seven days), positive swabs received in the trial, percentage of swabs returned that were positive, negative, and indeterminate results. Trial swab numbers are presented without smoothing. Note participants were asked not to send swabs at the weekend or bank holidays due to lack of postal collections. The raw data therefore shows fewer swabs sent on 1-2 days each week and following bank holidays. Swab processing took place daily throughout.

Positive NAAT results from swabs taken at initial symptomatic visits also mirrored UK data with more positive swabs taken during period of high UK case numbers (Figure 3.3). Negative swabs from initial symptomatic visits were more common when UK rates were low from June to September 2020. Symptomatic visits were less common in the last four months of this study period.

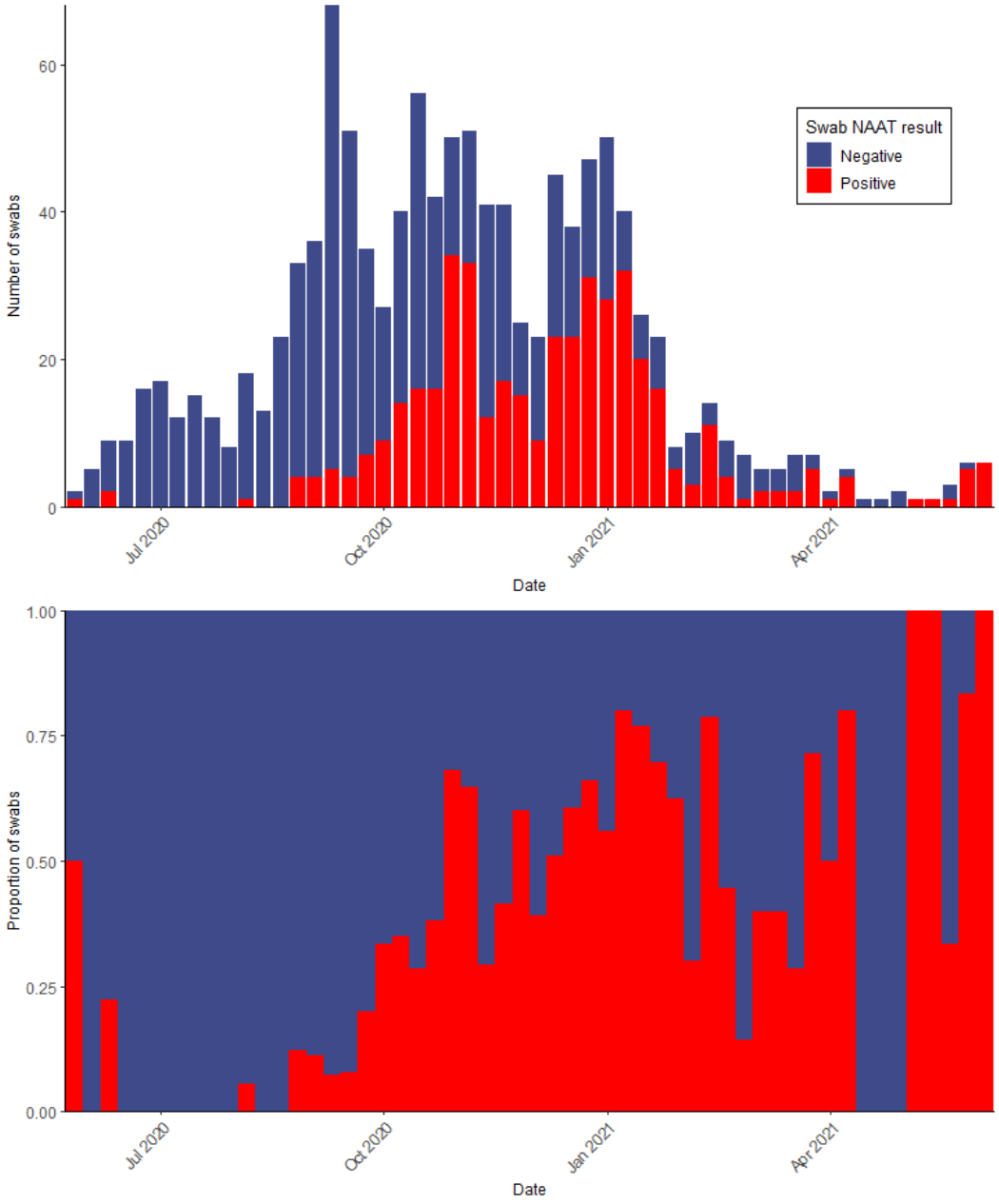


Figure 3.3: NAAT results from swabs taken at initial symptomatic visits. Data aggregated by calendar week. Indeterminate results not shown.

3.3.3 Adherence to weekly swabbing

Swab return rates remained high during the trial period (Figure 3.4) with rapid increase in returns during the trial recruitment phase. The number of swabs returned declined over Christmas 2020 and then from January 2021, which coincided with the national roll out of the vaccine programme, unblinding, and the end of the swabbing period.

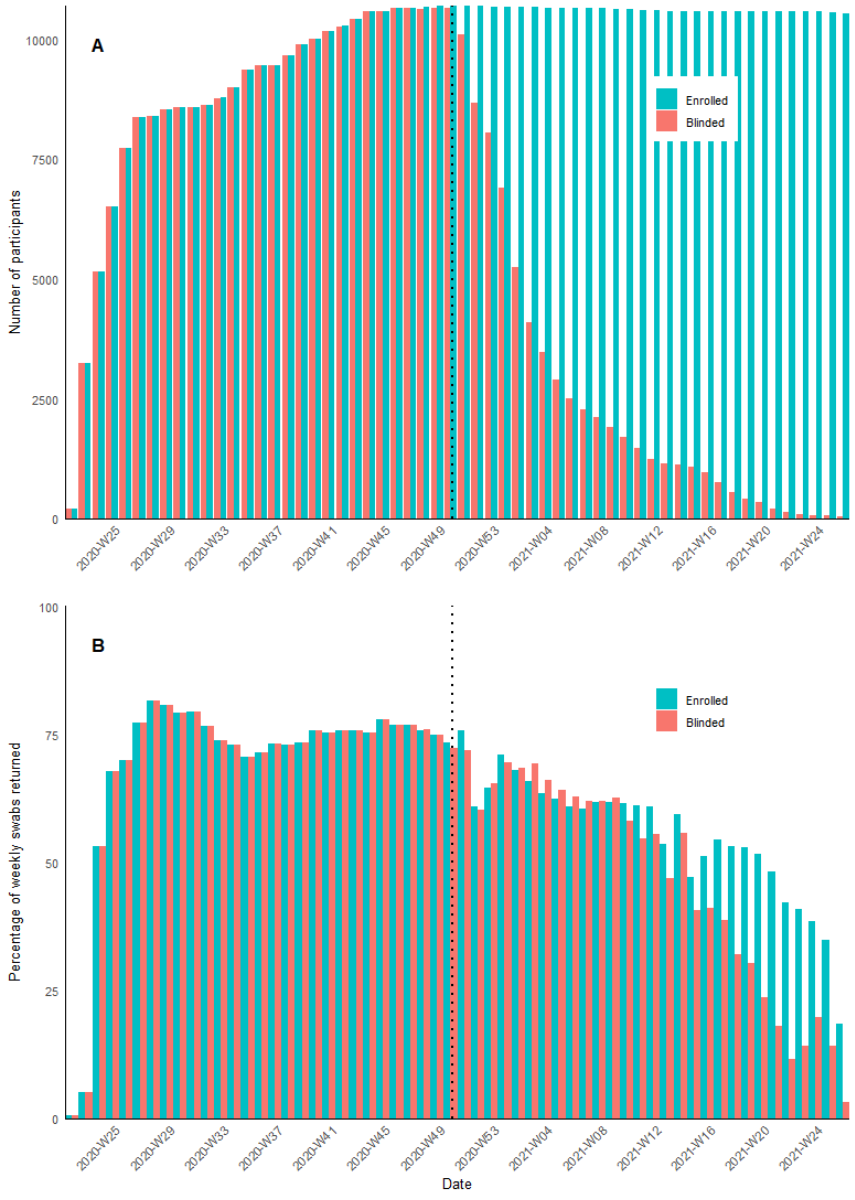


Figure 3.4: Number of participants and percentage of swabs returned by week by participants in COV002 trial. Absolute numbers, in enrolled and blinded participants, shown in A, percentage of weekly swabs returned, in enrolled and blinded participants shown in B. Dotted black line represents week when UK national vaccination programme began (8th December 2020).

Individual adherence to weekly swabbing overall was high (Figure 3.5). Median percentage of swabs returned for enrolled participants over the swabbing period was 76.8% (IQR: 44.7 to 92.8), and higher at 89.1% (IQR: 64.1 to 98) for blinded participants. There were 336 participants (3.1%) who did not return any weekly swabs. A small minority of participants (13/10812, 0.12%) returned more than 110% of tests that they were eligible to return over the whole weekly self-swabbing period, for blinded participants this figure was 0.54 % (59/10812.)

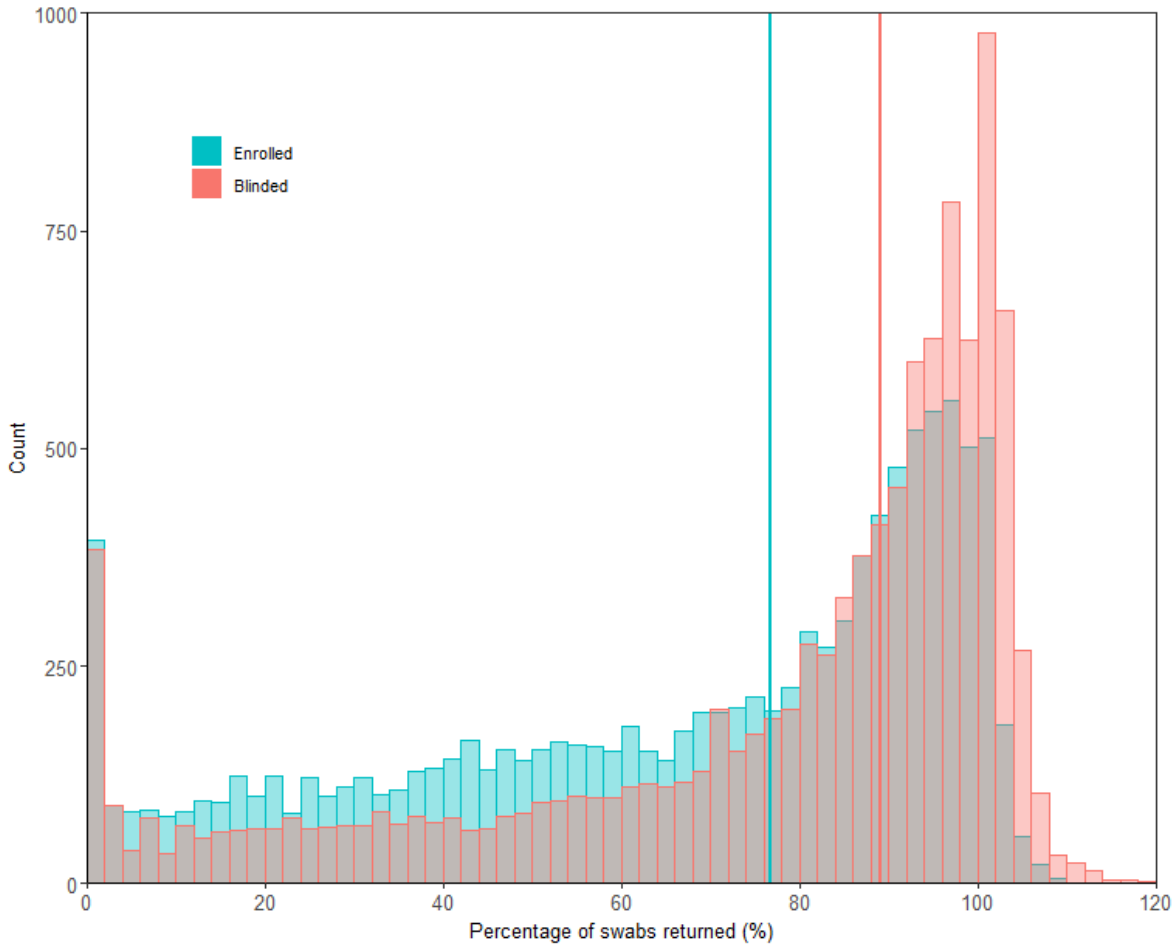


Figure 3.5: Percentage of swabs returned by COV002 participants for entire period of enrolment (turquoise) or for period when a participant was blinded to arm allocation (red). Lines represent medians for each group. Participants who returned >120% of expected swabs were omitted from this plot (enrolled n=10, blinded n=9).

When adherence was considered for blinded participants by age group (as blinded participants are considered in the analyses in chapters 4 and 5), adherence was lowest in the under 25-year-old age group (median: 70%, IQR 27 to 86%) and went up with increasing age (Figure 3.6).

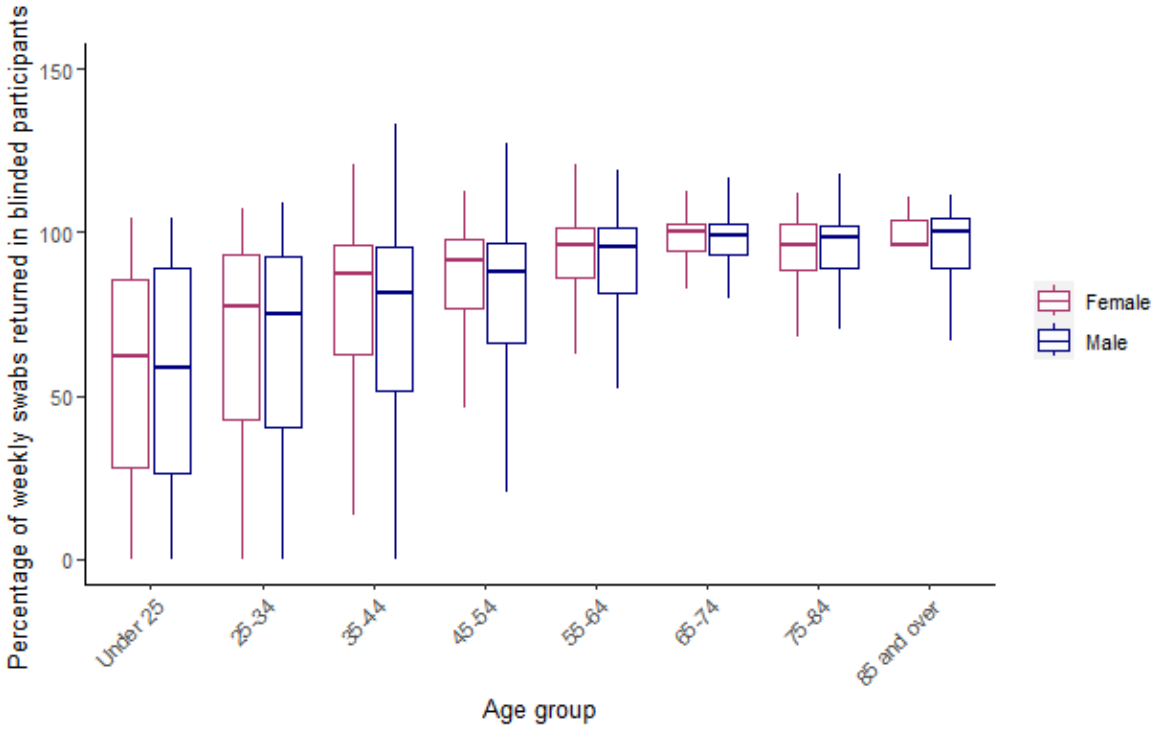


Figure 3.6: Overall adherence with weekly self-swabbing in blinded participants by age group and sex. The midlines of the boxes show medians, and the lower and upper bounds of the boxes correspond to the 25th and 75th percentiles. The whiskers show the most extreme value no further than 1.5 x IQR above or below the 75th or 25th percentile.

When adherence was examined by age group over time, reduction in adherence over time appeared most marked in the youngest age category and become progressively less marked with increasing age (Figure 3.7).

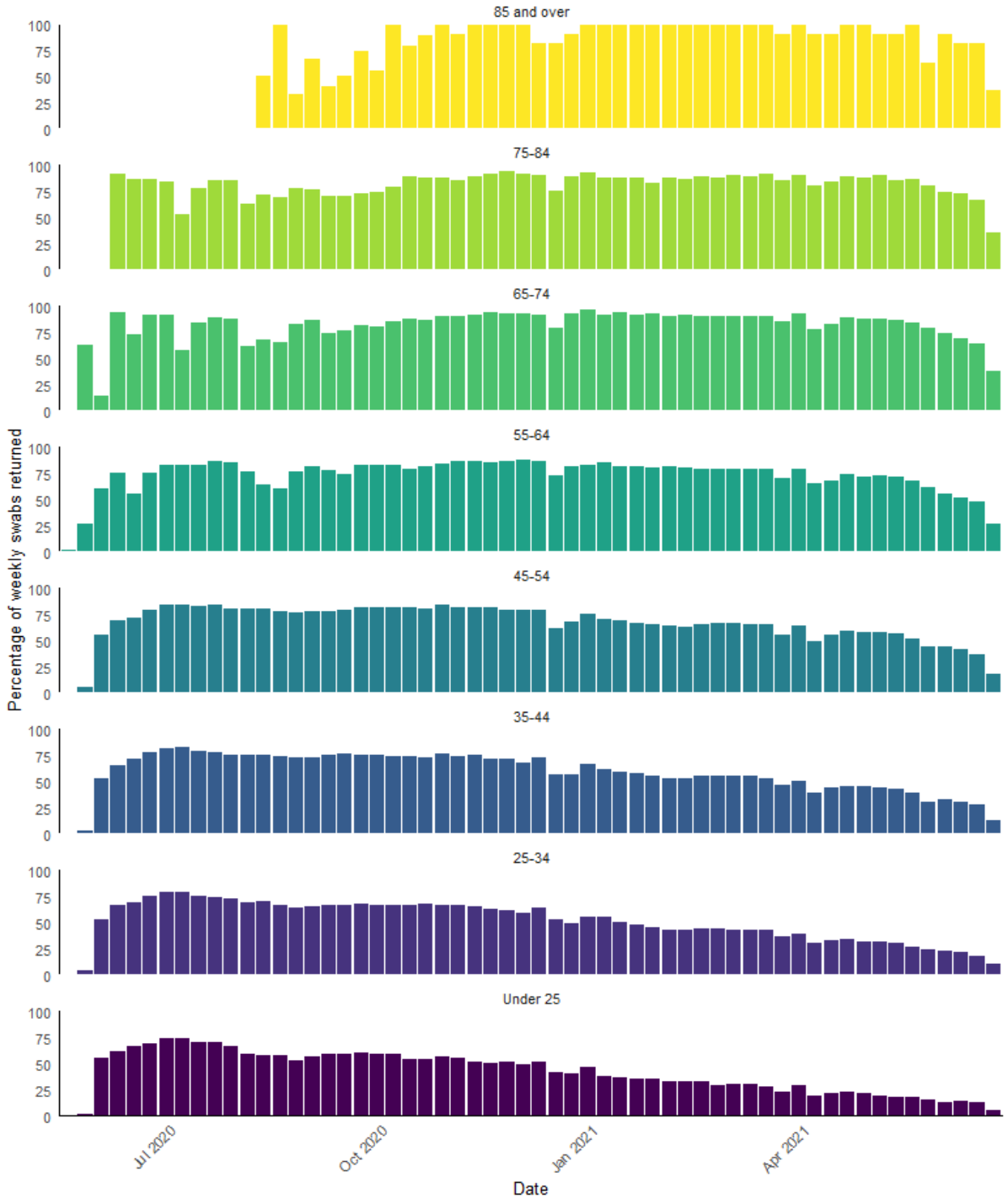


Figure 3.7: Percentage of swabs returned by age group over time. All enrolled participants included.

Multiple linear regression was performed to investigate the factors associated with a higher rate of self-swab returns for individuals (Table 3.5). The percentage of swabs that a participant

returned was significantly lower in those who were male (-3.46% , 95% CI: -4.56, -2.36) and in those who were health or social care workers (-5.32%, 95% CI: -6.57, -4.08). The percentage of swabs returned was higher in those who were older (0.78% per year, 95% CI: 0.74, 0.82). The interaction between sex and age was tested and this was non-significant (p=0.06). The median percentage of swab returns in the 75-to-84-year age group was 95.15% (IQR: 86.4-99.6) compared with 42.1% (IQR: 18.2-67.7) in the under 25-year-old age group. However, this model only explained 17% of the variance so there are other factors related to adherence which are not considered. The data met the assumptions of homogeneity of variance and linearity, and the residuals were approximately normally distributed. When the models were performed for blinded participants, sex and age remained significant predictors whilst being a healthcare worker did not (Table 3.6 and Figure 3.6).

Table 3.5: Linear model for predictors of the percentage of self-swabs returned (Adj. R^2=0.18, F(3,10808)=777.7, $p<2.2x10^{-16}$).

Variables		**Adj. coefficient**	**95% CI**		**p value**
Sex	Male	-3.5	-4.6	-2.4	**<0.001**
Age (years)	Per year	0.78	0.74	0.82	**<0.001**
Health care worker	Yes	-5.3	-6.6	-4.1	**<0.001**

Table 3.6: Linear model for predictors of the percentage of self-swabs returned in participants whilst blinded (Adj. R^2=0.12, F(3,10808)=481.5, $p<2.2x10^{-16}$).

Variables		**Adj. coefficient**	**95% CI**		**p value**
Sex	Male	-4.4	-5.5	-3.3	**<0.001**
Age (years)	Per year	0.68	0.64	0.72	**<0.001**
Health care worker	Yes	0.13	-1.12	1.4	0.8

Correspondingly, in a generalised linear model for binary data investigating those who did not return any weekly swabs, these participants were likely to be younger (adjusted OR 0.98 (95% CI: 0.97-0.99) and male (adjusted OR: 1.62 (95% CI: 1.29-2.02) but occupation in health or social care was not significant (Table 3.7).

Table 3.7: Multiple logistic regression investigating those who did not return any weekly swabs.

Variables		**Adj. OR**	**95% CI**	**p value**
Age (years)	Per year	0.98	0.97, 0.99	**<0.001**
Sex	Male	1.62	1.29, 2.02	**<0.001**
Health care workers	Yes	1.24	0.96, 1.62	0.0981

3.3.4 Ct values and gene targets

Of the 1130 participants who tested positive at least once, 992 (87.8%) returned at least one positive weekly self-swab. Ct values were available for 1568 of the 1623 (96.6%) positive swabs from weekly self-swabs from these participants that tested positive at least once. The relationship between Ct values, gene targets and date of swab collection are shown in Figure 3.8. Positives with S gene target failure consistent with the SARS-CoV-2 B.1.1.7 (Alpha) variant (shown in light green in Figure 3.8) were dominant from December 2020 to April/May 2021. Gene targets and Ct values for symptomatic visit swabs were not available due to the different commercial NAAT assays used at trial sites (details in Appendix 3).

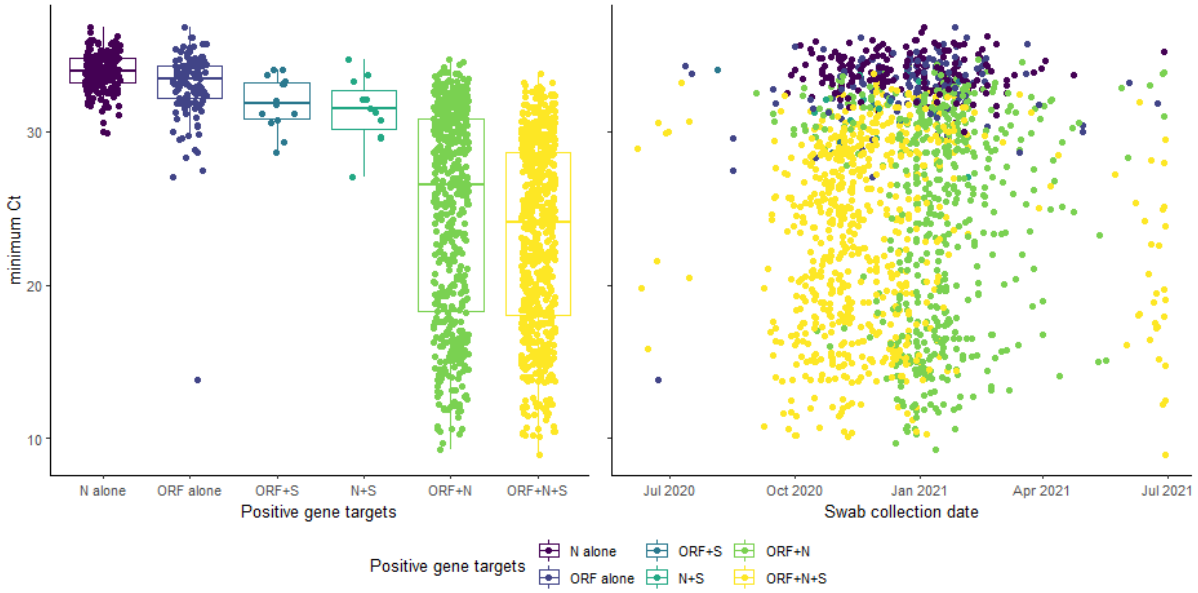


Figure 3.8: Gene targets and Ct values of weekly swabs over time. Each point represents an individual swab result. N=1568 (96.6%) out of total 1623 positive results. The midlines of the boxes show medians, and the lower and upper bounds of the boxes correspond to the 25th and 75th percentiles. The whiskers show the most extreme value no further than 1.5 x IQR above or below the 75th or 25th percentile. Ct values were not available when swabs were processed at alternative laboratories or for some samples at the beginning of the sampling period as laboratories were increasing capacity.

3.3.5 Assessment of false positives

Longitudinal swabbing of asymptomatic individuals may give rise to false positive results which may have implications for the assessment of infection and therefore vaccine efficacy within a trial. To try and assess which positive tests in the trial may have been falsely positive, two different criteria using Ct values and symptom status were employed adapted from other methods (Figure 3.9) resulting in 165 and 72 swabs (10.1% and 4.4% of all positive swabs respectively) that were identified as potential false positives. Potential false positives occurred throughout the study period, with higher numbers during the periods of higher rates of positive tests overall in the trial and nationally (Figure 3.10). These estimates could be refined further by measuring nucleocapsid antibody status from serum taken at intervals during the trial.

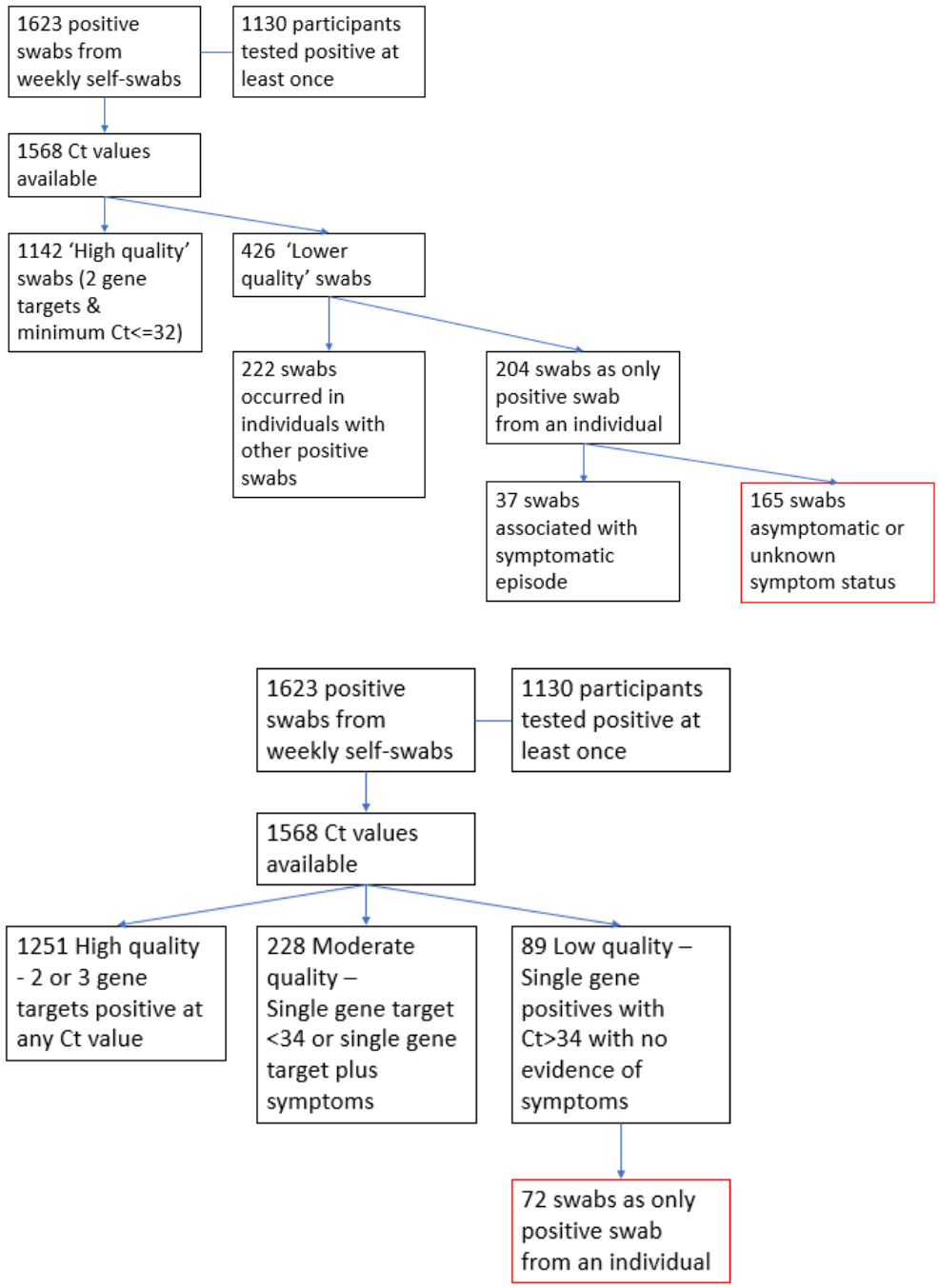


Figure 3.9: Assessment of possible false positives. Criteria adapted from Wei et al. (2021) [152] for upper panel, criteria adapted from Walker et al. (2021) [320] for lower panel.

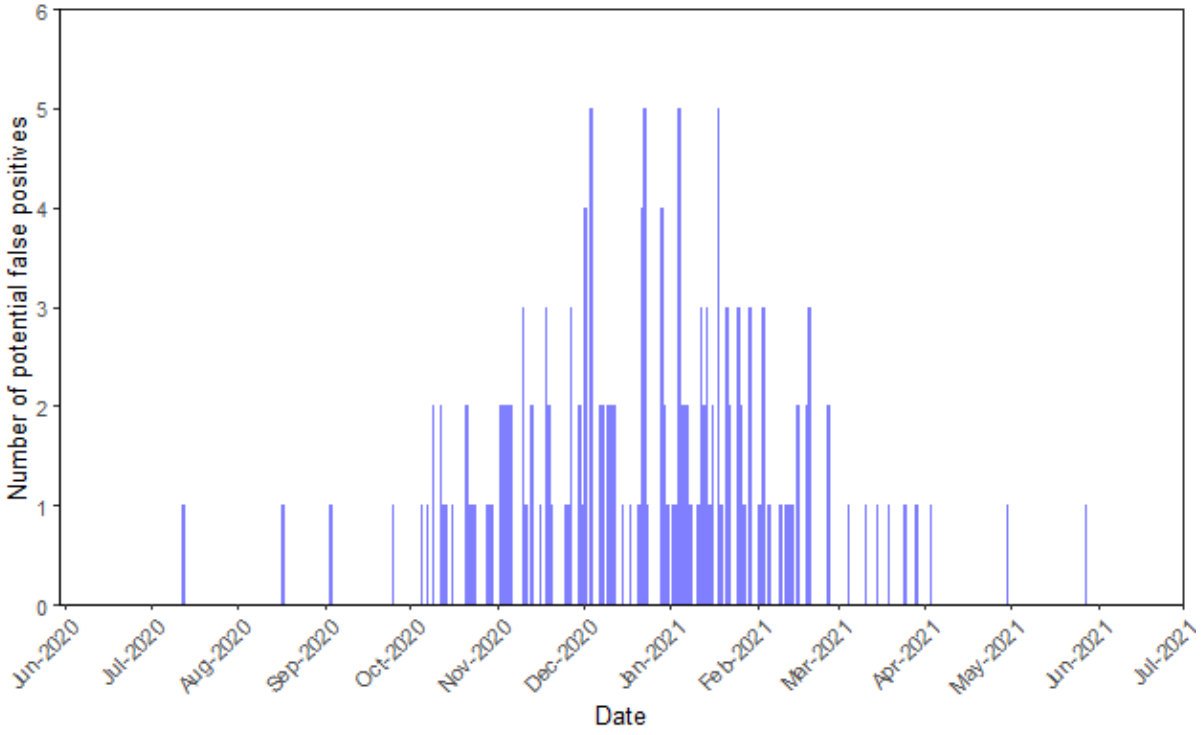


Figure 3.10: Plot of potential false positives over time

3.3.6 Participants who tested positive

Between 30th May 2020 and 30th June 2021, over 10% of participants (n=1130) had at least one positive swab for SARS-CoV-2 from any testing source, both within and outside the trial. There were few significant demographic differences between those who tested positive compared with those who never tested positive before 30th June 2021 (Table 3.8).

Table 3.8: Demographic details of those who tested positive from any swab source during the weekly swabbing period. BMI values were missing for two participants. Significant p values from univariate analyses are shown in bold. The group that did not return a positive swab also includes those that did not return a swab from any source (n=312). Ethnicity not shown or analysed due to small group sizes. P values from univariate analyses are shown. The following statistical tests were used: Pearson's Chi-squared test: sex, the listed comorbidities, any co-morbidity, smoking, alcohol consumption); Wilcoxon rank sum test: age, BMI, time in study.

Characteristic	**N**	**Tested positive from any source at least once**		
		No n = 9,682	**Yes n = 1,130**	**p-value**
Female sex, n (%)	10,812	5,771 (60%)	670 (59%)	0.8
Age at screening (years) Median (IQR)	10,812	44 (34, 55)	41 (31, 52)	**<0.001**
Age group, n (%)	10,812			
Under 25		506 (5.2%)	79 (7.0%)	
25-34		2,104 (22%)	293 (26%)	
35-44		2,270 (23%)	288 (25%)	
45-54		2,330 (24%)	265 (23%)	
55-64		1,035 (11%)	106 (9.4%)	
65-74		1,068 (11%)	74 (6.5%)	
75-84		358 (3.7%)	25 (2.2%)	
85 and over		11 (0.1%)	0 (0%)	
Healthcare worker, n (%)	10,812	6188 (64%)	810(72%)	**<0.001**
Cardiovascular disease, n (%)	10,812	1,187 (12%)	124 (11%)	0.2
Asthma, n (%)	10,812	986 (10%)	125 (11%)	0.4
Chronic obstructive airways disease, n (%)	10,812	21 (0.2%)	2 (0.2%)	>0.9
Bronchiectasis, n (%)	10,812	12 (0.1%)	1 (<0.1%)	>0.9
Other respiratory disease, n (%)	10,812	192 (2.0%)	24 (2.1%)	0.7
Diabetes mellitus, n (%)	10,812	197 (2.0%)	23 (2.0%)	>0.9
Renal disease, n (%)	10,812	187 (1.9%)	20 (1.8%)	0.8
History of stroke, n (%)	10,812	65 (0.7%)	5 (0.4%)	0.4
Gastrointestinal disease, n (%)	10,812	1,152 (12%)	136 (12%)	0.9
Liver disease, n (%)	10,812	160 (1.7%)	14 (1.2%)	0.3
Any co-morbidity, n (%)	10,812	3,273 (34%)	376 (33%)	0.7
BMI, Median (IQR)	10,810	25.4 (22.9, 28.8)	25.7 (23.3, 29.4)	**0.015**
Smoker, n (%)	10,812	501 (5.2%)	62 (5.5%)	0.7
Consumes alcohol, n (%)	10,812	8,345 (86%)	972 (86%)	0.9
Time in study (weeks), Median (IQR)	10,812	54 (51,56)	55(53,56)	**<0.001**

In a generalized linear model for binary data participants who tested positive were younger (OR (per 10-year increase in age): 0.91 (95%: 0.87-0.96), p<0.001) and had a higher BMI (OR

(per 5-point increase in BMI): 1.1 (95% 1.03-1.16%), p<0.05). Participants were more likely to test positive if they participated in the study for longer (OR: 1.1 (95%: 1.05-1.15), p<0.01), whilst sex and being a healthcare worker were not significant (Table 3.9).

Table 3.9: Generalised linear model for binary data assessing predictors for testing positive for SARS-CoV-2 amongst COV002 trial participants. p values with significance <0.05 are shown in bold.

Variables		**Estimate (Adjusted OR)**	**95% CI**	**p value**
Sex	Male	1.133	0.99-1.29	0.0595
Age (years)	Per 10 years	0.914	0.87-0.96	**0.0006**
Healthcare worker	Yes	1.155	0.99-1.35	0.0674
BMI	Per five-point increase	1.096	1.03-1.16	**0.0017**
Time in study	Per 4 weeks	1.093	1.05-1.15	**0.0018**

The median number of positive swabs per participant was 1, with a range of 1-14 (Table 3.10). There were 336 (3.1%) participants who did not return any weekly swabs and of these, 24 (0.22%) had at least one swab from another source (e.g. symptomatic pathway, workplace testing).

Table 3.10: Total number of positive swabs for participants who tested positive at least once.

Number of positives from any source, n (%)										
1	2	3	4	5	6	7	8	9	10	14
561 (50)	260 (23)	171 (15)	88 (7.9)	28 (2.5)	8 (0.7)	6 (0.5)	3 (0.3)	3 (0.3)	1 (<0.1)	1 (<0.1)

Vaccination status at the time of their first positive PCR result for participants who tested positive at least once is shown in Table 3.11.

Table 3.11: Vaccination status at first positive swab for those who tested positive. Participants were excluded from analyses if 21 days had not elapsed between their first control or ChAdOx1 nCoV-19 vaccination and their first positive swab.

Two doses control vaccine	One dose control	One dose ChAdOx1 nCoV-19	Two doses of ChAdOx1 nCoV-19	Excluded
532 (47%)	108 (9.5%)	112 (9.9%)	354 (31%)	24 (2.1%)

To assess if there was any difference in the percentage of swabs returned by participants in those who tested positive from any source during the self-swab period compared with those who never tested positive, the percentage of swabs returned between these two groups were compared. Overall participants who tested positive during this period returned a lower median percentage of weekly swabs than participants who never tested positive (Never tested positive: 77.8% (IQR: 45.1-93.0%), Tested positive at least once: 69.0% (IQR: 43.9-89.4%), Wilcoxon rank sum, $p<0.001$, Figure 3.11). As participants were more likely to test positive if they were younger, the percentage of swabs returned by age group was compared and in most age groups there was no significant difference in the median percentage of swabs returned. It was only in the 75- to 84-year-old group where those who never tested positive sent significantly more swabs than those who had at least one positive swab (Wilcoxon rank sum, $p<0.05$).

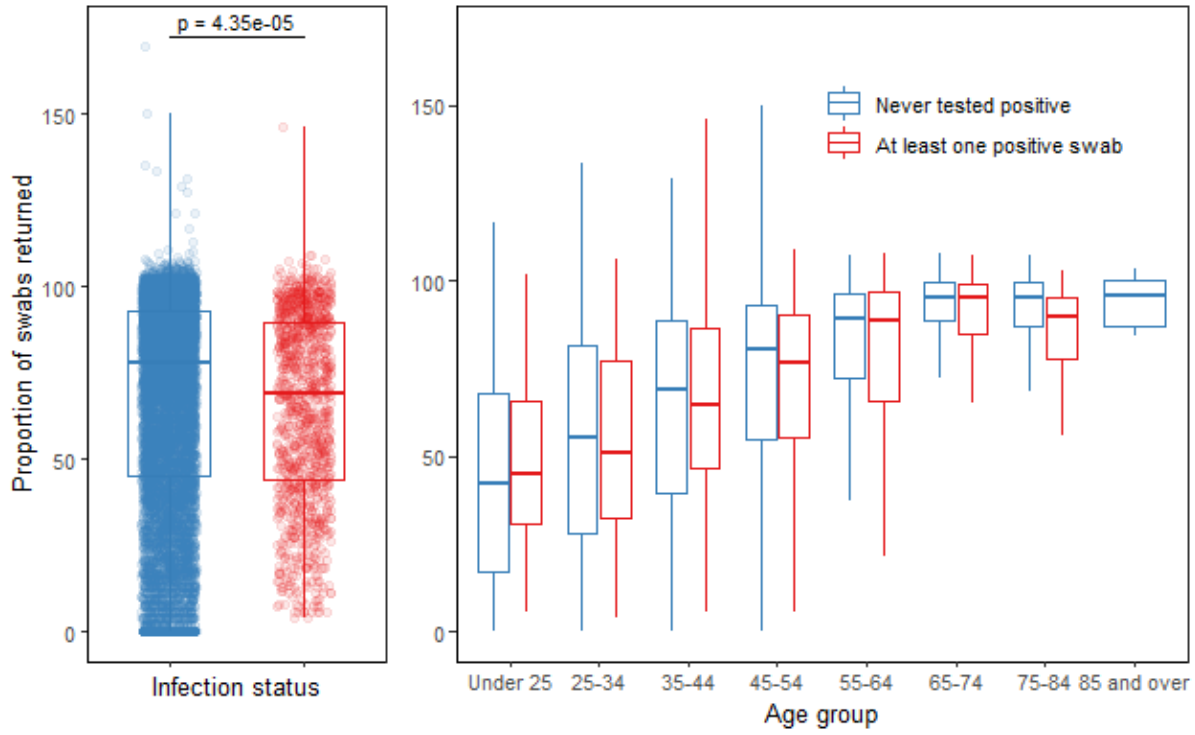


Figure 3.11: Comparison of adherence between those who tested positive versus those who did not. Comparison of the percentage of swabs returned by those who returned at least one positive swab in the trial compared with those who never tested positive and by age group. The midlines of the boxes show medians, and the lower and upper bounds of the boxes correspond to the 25th and 75th percentiles. The whiskers show the most extreme value no further than 1.5 x IQR above or below the 75th or 25th percentile. In the left panel each point represents one participant. Medians compared with Wilcoxon sum rank test.

3.3.7 Symptoms and severity of disease in trial participants

The demographics of participants who tested positive by symptom status are shown in Table 3.12. There was a significant difference in age between the symptom groups (Kruskal Wallis rank sum test p<0.001). Wilcoxon signed rank pairwise tests were carried out for the three pairs of groups, participants who were symptomatic were significantly younger (median age 39, IQR: 30-49) than those who were asymptomatic (median age 45, IQR: 34-56) or unknown symptom status (median age 45, IQR: 35-55; (p<0.001 for both, adjusted using the Bonferroni

correction). There was no difference in age between the asymptomatic and unknown groups.

The proportion of symptomatic cases by age group are shown in Figure 3.12.

Table 3.12: Demographics of all participants who returned at least one positive swab by symptom status. Symptomatic defined as any recorded symptom deemed to be associated with positive SARS-CoV-2 swab by blinded, independent endpoint committee. Age and BMI were assessed by Kruskal Wallis rank sum testing, categorical variables by Pearson's chi squared testing apart from ethnicity which was tested using Fisher's exact test. P values <0.05 are shown in bold.

Characteristic	***Symptomatic*** *n=669*	***Asymptomatic*** *n=191*	***Unknown*** *n=270*	***p-value***
Female sex, n (%)	405 (61%)	112 (59%)	153 (57%)	0.5
Age at screening (years) Median (IQR)	39 (30, 49)	45 (34, 56)	45 (35, 55)	**<0.001**
Age group, n (%)				
Under 25	63 (9.4%)	10 (5.2%)	6 (2.2%)	
25-34	196 (29%)	38 (20%)	59 (22%)	
35-44	176 (26%)	44 (23%)	68 (25%)	
45-54	151 (23%)	37 (25%)	67 (25%)	
55-64	53 (7.9%)	21 (11%)	32 (12%)	
65-74	21 (3.1%)	24 (13%)	29 (11%)	
75-84	9 (1.3%)	7 (3.7%)	9 (3.3%)	
85 and over	0	0	0	
Ethnicity, n (%)				0.2
White	611 (91%)	176 (92%)	254 (94%)	
Asian	39 (5.8%)	10 (5.2%)	10 (3.7%)	
Mixed	11 (1.6%)	1 (0.5%)	1 (0.4%)	
Black	3 (0.4%)	0	4 (1.5%)	
Arab	3 (0.4%)	2 (1%)	0	
Other	2 (0.3%)	2 (1%)	1 (0.4%)	
Any co-morbidity, n (%)	223 (33%)	71 (37%)	82 (30%)	0.3
BMI, Median (IQR)	25.7 (23.3, 29.8)	26.3 (23.9, 30.0)	25.1 (23.0, 27.9)	**0.011**
BMI>25	383 (57%)	119 (62%)	138 (51%)	0.051
BMI>30	162 (24%)	46 (24%)	41 (15%)	**0.008**
Healthcare worker, n (%)	517 (77%)	122 (64%)	171 (63%)	**<0.001**

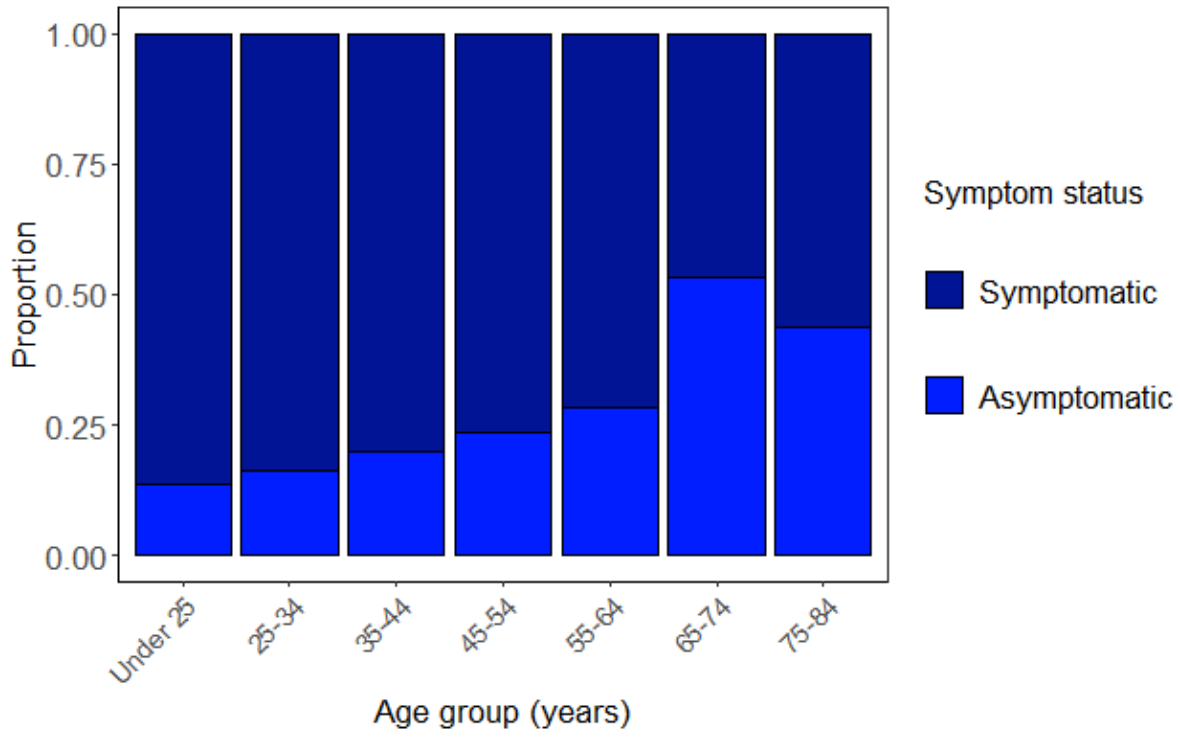


Figure 3.12: Proportion of symptom status by age group. Unknown symptom status not shown.

There was also a significant difference in proportion of healthcare workers between the symptom status groups, with 77% of those who were symptomatic being healthcare workers compared with 64% and 63% in the asymptomatic and unknown symptom groups respectively (Pearson chi-squared test, $p<0.001$). There was no difference in the proportion of female participants, ethnicity, or presence of any co-morbidity between the symptom groups. There was a significant difference in the median BMI between symptom groups, but this was driven by a lower BMI in the unknown group than either of the known symptom groups, there was no difference between symptomatic and asymptomatic groups.

Participants in the trial typically had mild COVID-19, there were only 7 cases with a WHO progression score of 4 or above (Table 3.13).

Table 3.13: COVID-19 severity in COV002 participants who tested positive at least once during the trial. A score of 1 denotes asymptomatic infection. Unknown participants were those for whom no symptom data was available and therefore a WHO progression score could not be assigned.

	WHO progression score										
	1	2	3	4	5	6	7	8	9	10	Unknown
Participants, n	191	246	416	3	2	1	0	0	0	1	270
%	17%	22%	37%	0.3%	0.2%	0.09%				0.09%	24%

3.4 DISCUSSION

This chapter describes the swabbing results of participants from the COV002 vaccine trial which captured the SARS-CoV-2 NAAT data which supported the endpoint assessments for participants in this trial. These data form the basis for the analyses in the following two chapters.

Our trial population was dominated by white adults, with a female majority, aged 25-55 years old. This study was like many others in its demographic make-up: a cross-sectional study of 178 adult vaccine trials in the US found under-representation of ethnic groups and older adults and an overrepresentation of female participants compared with US census data. [322] Participants employed in health and social care roles were specifically targeted due to high occupational COVID-19 exposure, thereby maximising the chance of achieving primary endpoints for the trial. However, HCWs may have had different exposures to disease and usage of non-pharmacological interventions e.g. PPE during the pandemic. Equally health and social care workers are more likely to be female. [323] Further work is needed to understand why some groups remain underrepresented in trials and what the barriers are to access and inclusion.

While the study population was not demographically representative of the wider UK population, the weekly trial SARS-CoV-2 positive PCR events tracked concurrent national case numbers. The distribution of negative and positive swabs from symptomatic visits over

time shows that from October 2020, the proportion of symptomatic visits that resulted in a positive swab was greater than in the preceding months where almost all symptomatic visits were in participants who tested negative and therefore were likely to have an alternative upper respiratory tract viral infection. There were fewer national restrictions in place until November 2020 and these have been shown to reduce the incidence of other respiratory viruses. [324,325]. This may explain why during periods of high COVID-19 transmission the symptomatic pathway was dominated by SARS-CoV-2 positive swabs as other viruses which could have caused symptoms meeting the criteria for swabbing were less prevalent. Ideally the swabs taken from symptomatic participants would have undergone additional molecular testing for a wider panel of respiratory pathogens to understand what other pathogens were circulating despite public health guidelines to prevent the spread of respiratory infections however unfortunately this was not able to be pursued during this study. Weekly self-swabbing took place for the first thirteen months of the trial. Access to this alternative means of diagnosis (as well as increased availability of workplace/public NHS testing during 2020) may have impacted on the likelihood of participants presenting for a symptomatic visit. Swab returns (Figure 3.4) and attendance for symptomatic visits (data not shown) fell over the course of the study. Preliminary ChAdOx1 nCoV-19 efficacy results were published in November 2020 and unblinding for many participants started in December 2020, hence participants may have perceived symptomatic visits to be less important thereafter as well as developing fatigue with study procedures over time.

The data show that trial participants were motivated and returned a high percentage of weekly swabs, and that adherence was predicted by being female and increasing age. Literature on large scale longitudinal sampling often considers maintaining response rates to survey studies over much longer time periods with greater intervals, often of a year or more between

each study intervention whilst our study demanded frequent engagement at a particularly challenging time personally and professionally for many participants. A reduction in response rate over time is often seen in registry or cohort studies, [326] which may be due to the enthusiasm and motivation for participation seen at enrolment waning over time. Financial incentives have been shown to increase the number of returned surveys [327] but our participants did not receive any additional incentive to send weekly swabs above that which they were given for their study participation, which varied between study groups. A survey study in eight high-income countries, which controlled for socioeconomic factors and employment status, demonstrated a larger proportion of women perceived SARS-CoV-2 to be a very serious health problem and that women were more likely to comply with public health rules such as social distancing [328] compatible with our finding that the percentage of self-swab returns were higher in women. Increasing age is a well-publicised association with increased risk of severe COVID-19 disease [329] which may have prompted older participants to be more motivated to engage in regular surveillance. Younger age has also been associated with lower compliance with COVID-19 preventative behaviours [330] which may be analogous to self-swab returns in our study. However, there may be more prosaic reasons for older participants to have had a higher compliance if they were more likely to be retired and have more time to adhere to this aspect of the study.

The SARS-CoV-2 pandemic in the UK provided a unique research context, especially for healthy volunteers given that coronavirus was dominant in the media and impacted on daily life with public health restrictions on movements and contacts. This meant there were constant reminders from sources unconnected to the trial about infection which may have prompted higher adherence to weekly swabbing. One study (TrackCOVID) of general adult population in the San Francisco Bay area also showed high retention at 5 months for monthly

nasopharyngeal swabs and other samples to be 86.6% [331] and in a community cohort study which sought faecal samples those who opted to take part in sampling were more likely to be older in keeping with our data. [83] Returning weekly self-swabs may also have had additional benefits such as the provision of reassurance and access to testing during periods where it was more difficult to access, rather than simply the provision of trial information. However, the risk of testing positive was not inconsiderable during the study period discussed in this thesis; 10% of individuals recorded at least one positive which would have resulted, during this time in mandatory self-isolation and isolation of contacts. Those who never swabbed may have viewed swabbing when they did not have symptoms as a risk, as 24/336 (7%) of those who never sent a weekly self-swab did present for symptomatic testing suggesting that symptoms were their personal threshold for testing. The high levels of adherence observed highlight how individuals enrolled in interventional studies are motivated to take part in research and can also provide epidemiological data. However, these participants are self-selecting in that they have presented for participation in the trial and are therefore motivated to pursue the recruitment and undertake trial procedures.

In this trial population younger age, having a higher BMI and having a longer duration in the study were associated with higher odds of returning at least one positive swab. In a primary care cohort in the pre-vaccination era where diagnostic swabs were mainly done in those with symptoms, researchers found that obesity was associated with increased risk of infection, [332] but they also found male sex to be associated with infection which was not found in our study population. This may be because of the large number of female healthcare workers in our study group were occupationally at risk of infection which may have masked the effect of being male. In the same primary care cohort, [332] adults aged between 40-64 and over 75 were at greater risk which differs to the findings in our study, but again this may be due to occupational exposure

in our youngest groups. Other factors such as chronic kidney disease, being black, living in a deprived or urban area were found also to be associated with a positive test which we could not look at in our population. Many of these primary care findings were also echoed in a Biobank study of adults 49-83 that found that being over 75 was associated with increasing risk. [333] Our study population was not large enough, nor designed to explore risk factors to infection, but it is interesting that increasing BMI was observed as associated with infection. Obesity is recognised as a major risk factor for progression to severe disease, but vulnerability to infection with an increased BMI has specific implications for the assessment of vaccines. There are theoretical reasons why obesity may cause a dysregulated immune response or altered immunodynamics (for example faster antibody wane) to vaccination, [334] though SARS-CoV-2 vaccine efficacy in trials does not appear to be different in obese participants. [335]

Symptomatic participants were younger than those in the asymptomatic or unknown categories. This contrasts with other findings in different populations that symptoms are more likely with increasing age. [91,128,336,337] When inclusive of the whole study population a higher proportion of symptomatic cases were also healthcare workers. These observations may be explained by the different force of infection between healthcare workers (who were more likely to be younger) compared with those who were older and not healthcare workers. Younger people may have been more likely to have had more social contact (in addition to occupational exposure) during this period of the pandemic compared with older individuals and therefore be more likely to be exposed. This study had few infections in the oldest age groups, likely reflective of the public health guidelines in place in the UK at this time advising those in older age groups to minimise social interactions. Alternatively, the finding that symptomatic participants were younger may reflect the lower adherence to weekly swabbing by the younger age groups. Lower adherence to surveillance may mean that fewer asymptomatic infections in

these individuals were identified and therefore symptomatic disease appears more common in the younger groups.

The finding that being a healthcare worker was associated with a lower percentage of returned swabs during the whole weekly swabbing period but not during whilst blinded may be due to multiple confounders. Firstly, healthcare workers may more typically have been attending their workplace during the pandemic meaning they may have had reduced availability to return swabs, especially in the period in early 2021 when rates of infection were high. Occupational focus on COVID-19 may have resulted in use of PPE and a changed perception of risk and personal behaviours amongst healthcare workers. Other healthcare workers may have been required by their NHS employer to participate in local SARS-CoV-2 surveillance initiatives thereby reducing the individual incentive to return weekly swabs, or simply place excess burden on those individuals. Lastly, healthcare workers were included in priority group two for vaccination which means that they were likely to have been unblinded early and found out that they had either been vaccinated in the trial or were vaccinated shortly after unblinding which may have reduced the motivation for sending weekly swabs.

Longitudinal sampling of asymptomatic vaccine trial participants presents its own challenges. It requires active engagement of participants to provide weekly samples, it is logistically demanding and expensive to run at scale and has the potential to return false positives or impact on other higher priority trial endpoints. However, randomisation of participants offers well matched groups within the same epidemiological context which can be unattainable in population level surveys. At the start of the pandemic, it was unknown if swabs taken by untrained individuals provided with instructions would provide results equivalent to those taken or supervised by a trained healthcare professional. We did not assess the self-swab

technique of participants however subsequent research has shown that self-testing of the upper airway is comparable in performance to testing done by a healthcare worker. [338-342]

Ct values from self-swab samples with two or more positive gene targets were generally lower than samples with a single positive gene target but there remained a wide range of values. This is in keeping with the results of a large ONS study, conducted over a partially matched timeframe to this study period. The ONS study also tested individuals agnostic of symptoms, and similar to our findings, those with a single positive gene target had higher Ct values. [320]

There are limitations to the data presented here. Some of the reported positive weekly swabbing results will be false positives, though the specificity of the assay used in the Lighthouse laboratories is very high, and therefore the estimated false positive rate is thought to be extremely low. [320] False positive PCR results are difficult to discern from true positive swabs with low viral loads that may represent the end (or less likely, the beginning, given the asymmetrical trajectory) of an infection where viral burden is low. I chose two criteria from studies where the samples were processed in the same laboratories as those used in this study to try to estimate the swabs within the population dataset for which there was lower evidence of being positive, and therefore which may represent false positive results. [152,320] Estimates from this data set using adapted criteria show that up to 10% of the positive swabs were the only positive swab returned by an individual, had a low viral load, and were not known to be associated with symptoms. Using alternative adapted criteria this is reduced to 4.4% when 2 or more positive gene targets are used as a preliminary filter, but it is unlikely that these all represent false positive results especially given high levels of transmission. When these were plotted over time they occurred throughout the study period, with more seen during periods of greater transmission at a national (and trial) level. This demonstrates uncertainty, asymptomatic positives are more likely when there is transmission in the community, but equally at times of

high transmission there will be more positives on a 96-well RT-PCR plate and therefore more chance of sample contamination. Serological identification of nucleocapsid antibody seroconversion between study visits could help discern these further. This would require visits to be appropriately timed with respect to the infection and not too temporally distant from potential infection as nucleocapsid antibody wanes more quickly than other targets. However not all asymptomatic infections seroconvert and whilst trajectories are comparable, asymptomatic infected individuals who do seroconvert have been shown to have lower initial antibody titres. [156] As these swabs were not associated with symptoms, they did not contribute to the primary endpoint of vaccine efficacy in this trial, however further work continues as to what impact such results may have on the overall assessment of vaccine efficacy.

It is also plausible that there are false negatives within the dataset that are hard to identify, but swabs resulted as 'indeterminate' but with temporally related symptomatic positive NAAT test may represent false negatives. A small number of trial allocated swabs were returned by individuals outside of the trial (presumably contacts of trial participants). These were identified by a mismatch between study-held participant personal identifiers and demographic details used to register the swabs on the government online portal and were therefore excluded. It is also possible that some swabs included in the analysis were labelled as belonging to a trial participant but were actually performed on a different individual (again likely to be a household member).

This dataset represents the largest longitudinal study of weekly SARS-CoV-2 NAAT testing in vaccine trial participants. The demographic factors associated with high rates of swab return and adherence to study procedures may be used to inform future studies. In the following two chapters I will present further analyses based on swabbing data of those who became infected with SARS-CoV-2 during the trial.

4 The impact of ChAdOx1 nCoV-19 vaccination on SARS-CoV-2 symptoms and viral load

4.1 INTRODUCTION

The WHO outlined in the target product profile for COVID-19 vaccines in April 2020 that the aim of vaccines should be to provide "clear demonstration of efficacy (on population basis) ideally with ~50% point estimate". [343] Efficacy can, however, be measured via a range of phase 3 trial endpoints for example: microbiologically confirmed infection, clinical endpoints such as symptomatic or severe disease, or transmission, all of which at a population level may be impactful. [344] Ideally, a vaccine that primarily protects against symptomatic disease may also attenuate any infection that may occur despite vaccination, and in doing so may reduce potential for transmission either by reducing the period or magnitude of infectious viral shedding.

Original pre-clinical data suggested that SARS-CoV-2 vaccines may not prevent shedding and therefore transmission, even if protection against disease is achieved. [170,171] Macaques in the highest dose group inoculated with an inactivated vaccine showed no detectable virus in the lungs or pharynx 7 days after SARS-CoV-2 challenge. [291] However other animal vaccine-challenge studies, including of ChAdOx1 nCoV-19 and BNT162b vaccines did show detectable virus in the upper respiratory tract in vaccinated animals, despite reduced or absent virus in lower respiratory tract samples following challenge. [170,171] Investigations using other species, such as ferrets did however show a reduction in nasal shedding after ChAdOx1 nCoV-

19 vaccination. [345] Such animal studies are inevitably comprised of small numbers of animals and protocol differences, in terms of dose and assays may make comparisons difficult. The differences observed in terms of nasal viral shedding may be related to the routes by which the animals received the challenge inoculum; those that were administered virus into the nose and/or orally had detectable virus despite vaccination. As upper airway exposure is the route of inoculation in most human infections this has implications for the ability of SARS-CoV-2 vaccines to interrupt transmission, which may be due to the differential immune response required to protect the lungs from severe disease compared with protection of the nasal mucosa from infection.

Asymptomatic or pauci-symptomatic infections were recognised early in the pandemic. [90,125-127] Pre-symptomatic or asymptomatic infectious individuals pose significant challenges for disease control as interventions such as isolation of cases and contract tracing rely on identification of cases prompted by symptom onset. [86,91] Consequently, when evaluating vaccine efficacy within a trial setting it is important to understand what impact vaccination may have on the proportion of individuals that are infected but asymptomatic and may still have the capacity to transmit, especially as this type of outcome is more difficult to study outside of a trial setting. Conversely, if a vaccination is effective at reducing disease severity such that vaccinees are more likely to be asymptomatic, a vaccine may counter-intuitively not reduce the risk of transmission unless vaccination has additional effects on viral burden or duration of infectious shedding. Understanding the effects of vaccination on asymptomatic infection is important to support public health decision making such as informing testing and isolation strategy.

A meta-analysis of shedding studies performed in the pre-vaccination and pre-variant era suggested that SARS-CoV-2 RNA remains detectable by PCR in the upper respiratory tract for

more than two weeks, with the viral load peak around the time of symptom onset, though the duration of shedding of culturable virus beyond day 9 was not reported. [80] The duration of shedding and viral load may be longer and higher in those with more severe disease, and there are conflicting results on the relationship between viral load and asymptomatic versus symptomatic disease. [80,346,347] [86,127] At the time this study was initiated the impact of vaccination on viral load or shedding duration was not known.

The objectives of this sub study were:

- To calculate what proportion of trial participants may have had asymptomatic SARS-CoV-2 infection.
- To compare the proportion of symptomatic infection in controls compared with ChAdOx1 nCoV-19 vaccinees.
- To compare the severity of symptoms in controls compared with ChAdOx1 nCoV-19 vaccinees.
- To compare the Ct values as a proxy for viral load in controls compared with ChAdOx1 nCoV-19 vaccinees and explore factors that predict Ct values.
- To compare the duration of NAAT positive swabs in controls compared with ChAdOx1 nCoV-19 vaccinees and explore factors that predict duration of NAAT positive swabs.

4.2 METHODS

4.2.1 Trial methods and data collection

Data collection and procedural methods are as described in chapters 2 & 3. Briefly, participants underwent symptomatic visits if they contacted their study site with any one or more of the

following symptoms: persistent cough, fever ≥37.8°C, shortness of breath, ageusia or anosmia. If participants contacted the site with symptoms but did not meet criteria for swabbing, these were recorded. All participants were asked to return weekly nose and throat self-swabs for PCR testing. Participants were reminded weekly by email or text to contact their study site if they had symptoms and to continue to send weekly swabs.

All cases of COVID-19 were assessed by two members of a clinical review committee who were blinded as to vaccine allocation. These clinicians reviewed the NAAT result for each positive case in the context of clinical history and adverse events and then assigned individuals with a symptom status and WHO clinical progression score. [310] Sources of information for the assessment of symptom status include electronic symptom diaries, safety reports, ad hoc contacts with participants, investigator comments, medication histories, as well as evidence from study visits. If an episode of infection was designated as symptomatic this was either primary (featured at least one of the primary symptoms listed above) or without primary symptoms (symptoms present that were attributed to SARS-CoV-2 infection but none of the primary symptoms). Asymptomatic infections were designated when there was explicit documentation saying that the participant was asymptomatic. Participants were not contacted about any positive results from the weekly swabbing pathway, so information regarding this was on an ad hoc basis depending on whether the participant contacted their study site. Episodes were designated as unknown if there was no information in a participant's CRF with respect to symptoms.

4.2.2 Immunology

Live and pseudoneutralisation assays were performed on samples taken 28 days after boosting. Details of these assays have been published in Feng et al. (2021). [348] Normalised values (NF50)

of the live neutralisation assay were used for these analyses, the LLOQ of the assay is 8.6. The LLOQ for the pseudoneutralisation assay is 40.

4.2.3 Statistical analysis

Only participants who returned at least one positive swab were included in these analyses.

At the time of their first positive swab, participants were categorised as either:

1. excluded from analyses if less than 22 days had passed since their first vaccine of either MenACWY or ChAdOx1 nCoV-19,
2. one dose recipients (MenACWY or ChAdOx1 nCoV-19) if more than 21 days had passed since their first vaccine but a second dose had not been received, or
3. controls or vaccinated individuals if both vaccine doses had been received more than 14 days prior.

All ChAdOx1 nCoV-19 dose groups (low dose/low dose (LD/LD), low dose/standard dose (LD/SD), standard dose/standard dose, (SD/SD)) were considered vaccinated.

Potential reinfections were defined as instances whereby there were three negative swabs or a duration of >90 days between positive swabs. Where potential reinfections were identified only swabs from the first episode were included in analyses, swabs from subsequent infections were excluded.

For the analysis of demographics and baseline characteristics, categorical variables were compared using Pearson Chi-square tests, except when cell size was $\leq$5 where a Fisher's exact test was used. Continuous variables were assessed using Wilcoxon rank sum tests.

Ct values were only available for the weekly self-swabs and not for swabs from symptomatic visits. An 'infection minimum Ct value' was generated for each participant who tested positive as an inverse proxy for viral load. The 'infection minimum Ct value' was the lowest Ct value for any gene target from all swabs obtained during the participant's infection episode. 'First minimum Ct value' was the lowest Ct value was taken for any gene target from their first PCR positive self-swab. Ct values were compared using Wilcoxon rank sum tests. Where multiple comparisons were made between groups Bonferroni corrections were applied.

To calculate duration of NAAT positive swabs, all swabs that a participant had recorded in their study record were included from any source. As participants were requested for a visit if they were experiencing any of the swab criteria symptoms and therefore may have had an additional swab, participants were considered within symptom categories. Duration was defined as time in weeks from first NAAT positive swab to last NAAT positive swab of the first infection episode. Duration of NAAT positivity between groups was compared using a Wilcoxon rank sum test.

Multiple linear regressions were performed for Ct values and duration of NAAT positivity. Model assumptions were tested and only models that met the assumptions were retained.

4.3 RESULTS

Of the 10812 participants enrolled in the COV002 trial, 1130 participants returned at least one positive swab from any source between 30th May 2020 and 30th June 2021. The overall demographics of trial participants and those who tested positive are described in chapter 3.

4.3.1 Analysis dataset for impact of vaccination on symptom and viral parameters

The analysis dataset for the following analyses is shown in Figure 4.1. Participants were included if they had received both doses of study intervention (MenACWY or ChAdOx1 nCoV-19) more than 14 days before their first positive SARS-CoV-2 NAAT swab was collected and had not yet been unblinded to study allocation at the time of swab collection. This resulted in 616 participants, 397 from the control group and 219 who were fully vaccinated with ChAdOx1 nCoV-19. Of the ChAdOx1 nCoV-19 vaccinees, there were 7 participants who received two low dose vaccines, 54 who received a low dose followed by a standard dose and 158 who received two standard doses.

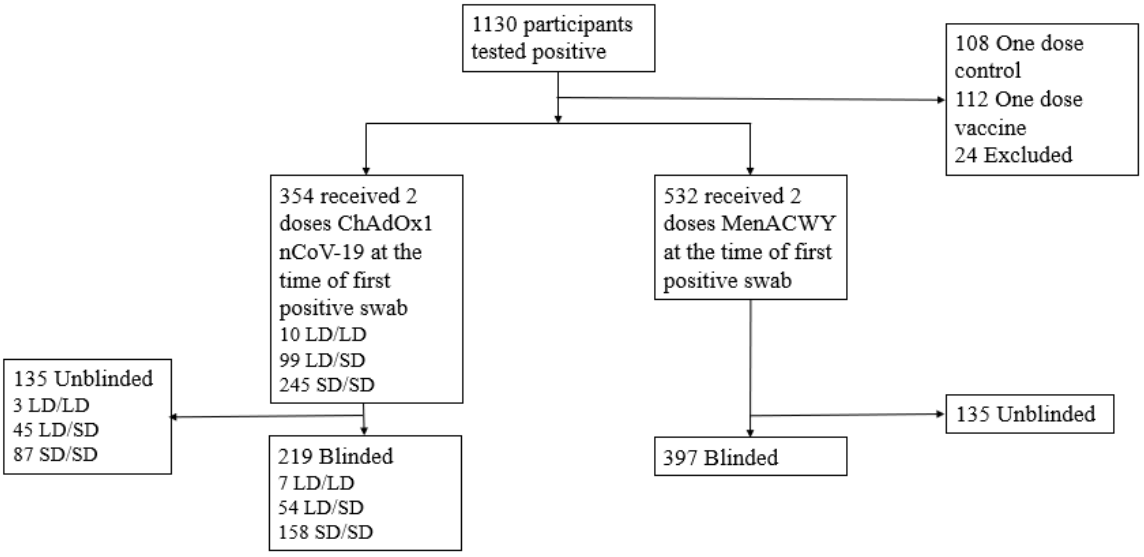


Figure 4.1: Participants who tested positive at least once during the study period. This dataset is used for symptom, viral load, and duration analyses. LD = low dose, SD =standard dose. Those excluded as one dose vaccine or control groups also includes participants who had their first NAAT positive test <15 days after their second dose.

The demographics for these participants are shown in Table 4.1. Vaccinated participants were slightly older (Wilcoxon rank sum test, $p<0.05$) and there were fewer healthcare workers in the

vaccinated group (Pearson's chi squared test, $p<0.01$). There was no difference in the sex, BMI, ethnicity, or the presence of any comorbidity between the control and vaccinated groups.

Table 4.1: Demographics for included participants that tested positive between 28th May 2020 and 30th June 2021. Age and BMI were assessed by Wilcoxon rank sum testing, categorical variables by Pearson's chi squared testing apart from ethnicity which was tested using Fisher's exact test.

Characteristic	**Control n=397**	**Vaccinated n=219**	**p value**
Female sex, n (%)	234 (59%)	132 (60%)	0.7
Age at screening (IQR)	41 (30, 52)	43 (34, 52)	**0.047**
Under 25	29 (7.3%)	11 (5.0%)	
25-34	110 (28%)	46 (21%)	
35-44	99 (25%)	65 (30%)	
45-54	93 (23%)	53 (24%)	
55-64	39 (9.8%)	16 (7.3%)	
65-74	24 (6.0%)	19 (8.7%)	
75-84	3 (0.8%)	9 (4.1%)	
85 and over	0	0	
Ethnicity, n (%)			0.2
White	367 (92%)	204 (93%)	
Asian	24 (6.0%)	8 (3.7%)	
Mixed	2 (0.5%)	5 (2.3%)	
Black	2 (0.5%)	1 (0.5%)	
Other	2 (0.5%)	1 (0.5%)	
Healthcare worker	278 (70%)	129 (59%)	**0.005**
Any co-morbidity, n (%)	118 (30%)	69 (32%)	0.6
BMI, n (%)	25.6 (23.3, 29.4)	25.6 (23.2, 29.0)	0.8

4.3.2 Cases

Infections in both vaccination groups occurred across the duration of the study period (Figure 4.2).

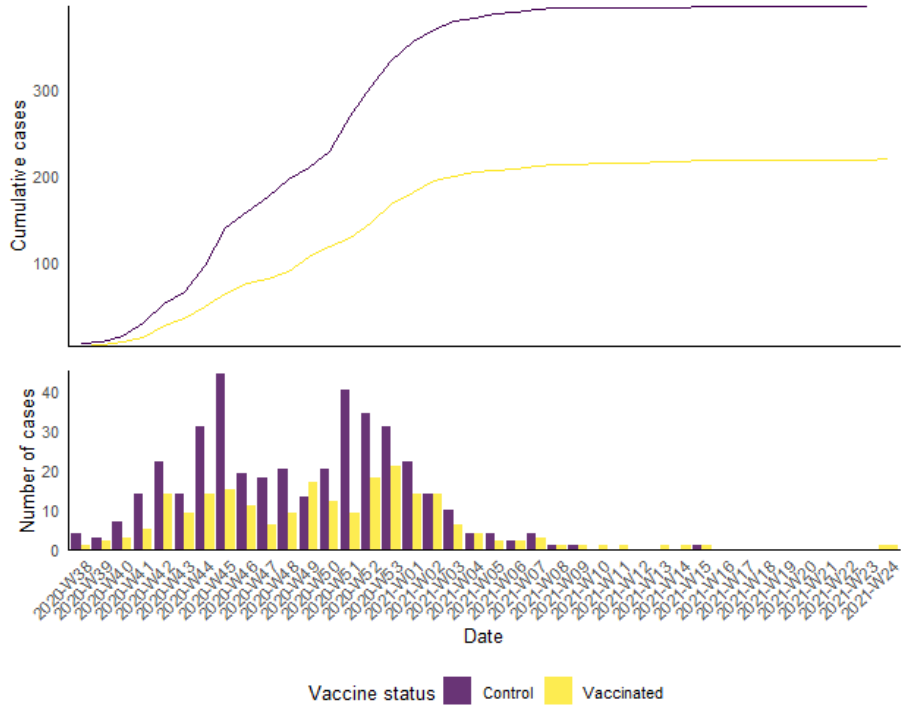


Figure 4.2: Cases of infection by week by vaccine status.

4.3.3 Symptom status and vaccination

The demographics of the different symptom status groups are shown in Table 4.2.

There was a significant difference in median age between the symptom groups (Kruskal Wallis rank sum test $p<0.001$). Wilcoxon signed rank pairwise tests were carried out for the three pairs of groups, participants who were symptomatic were significantly younger (median age 40, IQR: 30-50) than those who were asymptomatic (median age 44, IQR: 33-54) or unknown symptom status (median age 46, IQR: 35-56; $p<0.05$ for both, adjusted using the Bonferroni correction). There was no difference in age between the asymptomatic and unknown groups. There was also a significant difference in proportion of healthcare workers between the symptom status groups, with 72% of those who were symptomatic being

healthcare workers and 63% and 54% in the asymptomatic and unknown symptom groups respectively (Pearson chi-squared test, p<0.001), this statistical result though was due to the difference between symptomatic and unknown groups. There was no significant difference in proportion of healthcare workers between those who were symptomatic and asymptomatic. Similarly, the BMI of those in the unknown group was significantly lower than for the symptomatic group but not between those who were symptomatic and asymptomatic.

Table 4.2: Demographics of participants by symptom status. Age and BMI were assessed by Kruskal Wallis rank sum testing, categorical variables by Pearson's chi squared testing apart from ethnicity which was tested using Fisher's exact test. P values <0.05 are shown in bold.

Characteristic	**Symptomatic n=356**	**Asymptomatic n=103**	**Unknown n=157**	**p value**
Female sex, n (%)	209 (59%)	60 (58%)	97 (62%)	0.8
Age at screening (IQR)	40 (30-50)	44 (33-54)	46 (35-56)	**<0.001**
Under 25	34 (9.6%)	5 (4.9%)	1 (0.6%)	
25-34	101 (28%)	22 (21%)	33 (21%)	
35-44	98 (28%)	26 (25%)	40 (25%)	
45-54	85 (24%)	26 (25%)	35 (22%)	
55-64	21 (5.9%)	12 (12%)	22 (14%)	
65-74	14 (3.9%)	9 (8.7%)	20 (13%)	
75-84	3 (0.8%)	3 (2.9%)	6 (3.8%)	
85 and over	0	0	0	
Ethnicity, n (%)				>0.9
White	327 (92%)	97 (94%)	147 (94%)	
Asian	21 (5.9%)	4 (3.9%)	7 (4.5%)	
Mixed	5 (1.4%)	1 (1%)	1 (0.6%)	
Black	2 (0.6%)	0	1 (0.6%)	
Other	1 (0.3%)	1 (1%)	1 (0.6%)	
Any co-morbidity, n (%)	107 (30%)	32 (31%)	48 (31%)	>0.9
BMI, n (%)	25.8 (23.3, 30.3)	25.8 (23.4, 29.8)	24.9 (23.2, 27.2)	**0.041**
Healthcare worker	257 (72%)	65 (63%)	85 (54%)	**<0.001**

Episode symptom status was related to vaccination status (Figure 4.3 and Table 4.3). Infections in control participants were more likely to be symptomatic than infections in vaccinated participants (66% (95% CI: 61-71) vs 43% (95% CI: 36-50), Pearson Chi square

test, p<0.001) and for those who were symptomatic, control participants were more likely to experience a primary symptom (89% (95% CI: 84-92) vs 78% (95% CI: 68-86), Pearson Chi square test, p<0.05).

Table 4.3: Symptom status by vaccination status at the time of first positive swab

Characteristic	**Control, n = 397**	**Vaccinated, n = 219**
Symptom status		
Symptomatic, n (%)	261 (66%)	95 (43%)
Primary symptom experienced, n (%)	231 (89%)	74 (78%)
Asymptomatic, n (%)	52 (13%)	51 (23%)
Unknown, n (%)	84 (21%)	73 (33%)

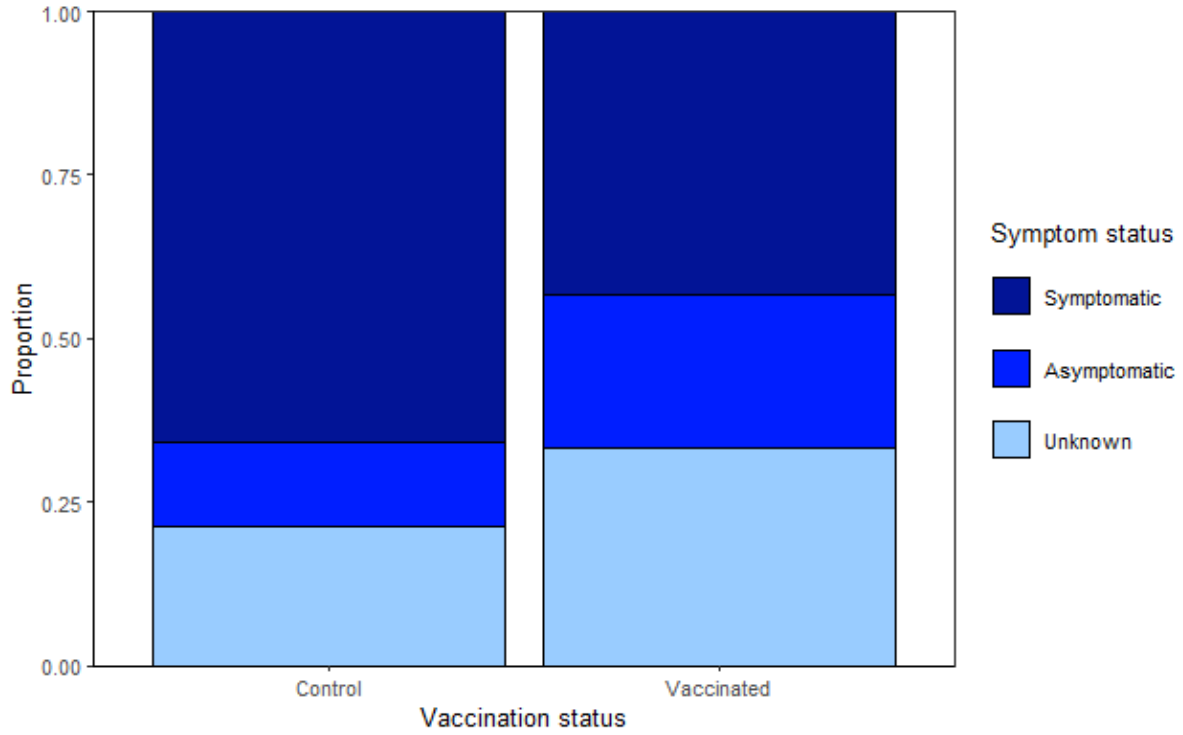


Figure 4.3: Symptom status of infective episode by vaccination status.

A larger percentage of the vaccinated group were asymptomatic compared with the control group. There was also a greater proportion of participants in the unknown symptom status category in the vaccinated than the control group.

There was also a significant difference in the WHO progression score between the control and vaccinated groups (Table 4.4) demonstrating those who developed symptomatic infection and had been vaccinated had less severe disease (Pearson chi-squared test, $p<0.01$).

Table 4.4: WHO progression score for symptomatic participants. Note a WHO progression score of 1 denotes asymptomatic infection and were not included.

Characteristic	**Control, n = 261**	**Vaccinated, n = 95**
WHO progression score		
2	60 (23%)	40 (42%)
3	198 (76%)	55 (58%)
5	1 (0.4%)	0
6	1 (0.4%)	0
10	1 (0.4%)	0

4.3.4 Cycle thresholds, symptom status and vaccination

Using the same dataset, the relationship between vaccination status, symptom status and cycle threshold value (as a proxy for viral load) was explored. From these 616 participants, 1252 positive swabs were returned, Ct values were available from 1163 swabs (92.9%) from 547 (88.8%) participants.

For both control and vaccinated participants, infection minimum Ct values were significantly lower for those with known symptomatic infection compared with those with asymptomatic or unknown infection (Table 4.5, Figure 4.4). There was no difference in infection minimum Ct value for those with asymptomatic and unknown symptom status for either vaccine group (Figure 4.4, top panel).

Table 4.5: Minimum and first minimum Ct values for first episode of infection. The 'infection minimum Ct value' was the lowest Ct value for any gene target from all swabs obtained during the participant's infection episode. 'First minimum Ct value' was the lowest Ct value was taken for any gene target from their first PCR positive self-swab. Ct values were not available for 69 participants, 53 in the control group (47 symptomatic, 5 asymptomatic, 1 unknown) and 16 (13 symptomatic, 3 asymptomatic) in the vaccinated group.

	Infection minimum Ct value, median (IQR)		**First minimum Ct value, median (IQR)**	
	Control n = 344	**Vaccinated n = 203**	**Control n = 344**	**Vaccinated N=203**
Symptomatic	17.95 (14.99-25.43)	21.07 (16.21-29.46)	18.95 (15.26-26.44)	21.73 (16.61-30.25)
Asymptomatic	30.13 (20.60-33.38)	32.60 (29.52-34.13)	30.63 (21.55-33.38)	32.60 (29.52-34.13)
Unknown	29.10 (20.16-32.90)	30.74 (25.61-33.76)	29.10 (21.83-32.90)	31.17 (26.55-34.07)

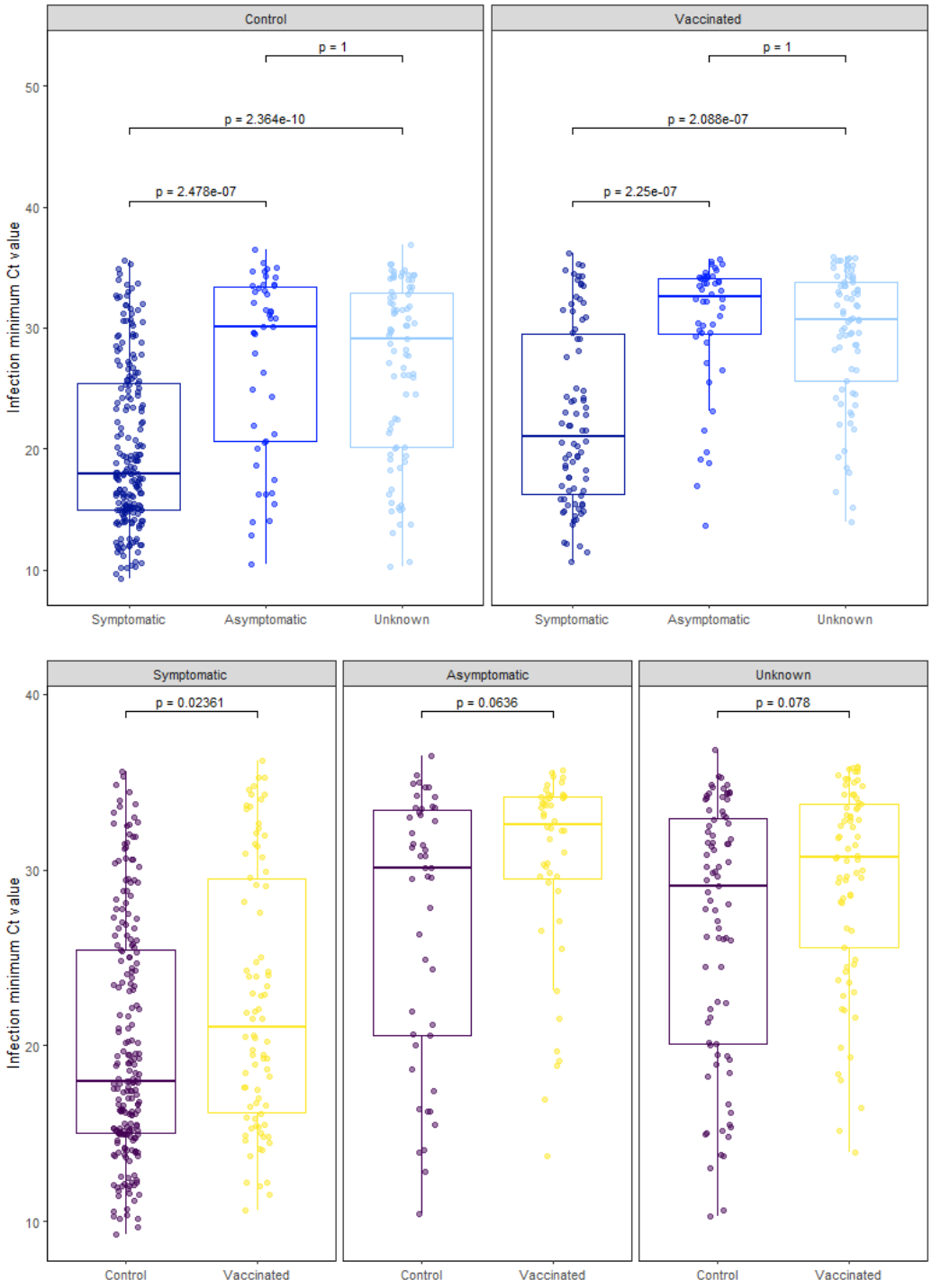


Figure 4.4: Infection minimum Ct value, symptom status and vaccination status. Each point represents the 'infection minimum Ct value' which was the lowest Ct value for any gene target from all swabs obtained during the participant's infection episode. Top panel comparisons within vaccination group, lower panel comparisons within symptom status group. The midlines of the boxes show medians, and the lower and upper bounds of the boxes correspond to the 25th and 75th percentiles. The whiskers show the most extreme value no further than 1.5 x IQR above or below the 75th or 25th percentile. Comparisons of infection minimum Ct between groups by Wilcoxon rank sum tests. Infection minimum Ct values were missing for 69 participants. 53 in the control group (47 symptomatic, 5 asymptomatic, 1 unknown) and 16 (13 symptomatic, 3 asymptomatic) in the vaccinated group.

When considered within symptom status group, the infection minimum Ct values were lower (therefore viral load was higher) for control participants compared with vaccinated participants for those who were symptomatic (median 17.95 (IQR: 14.99-25.43) vs 21.06 (IQR: 16.2-22.55), Wilcoxon rank sum, adj. $p<0.05$, Figure 4.4, lower panel) and whilst the trend was similar for those with asymptomatic or unknown infection it was not significant.

Participants who returned more swabs for analysis may have been more likely to have a lower infection minimum Ct value due to more sampling opportunity. To assess whether the results regarding vaccination may have been confounded due to different total numbers of swabs available for control versus vaccinated participants the numbers of positive swabs available for each vaccination group were calculated.

Symptomatic control participants did have more positive weekly swabs for analysis compared with symptomatic vaccinated participants (median of 1 (IQR: 1-3) versus median of 2 (IQR: 1-2), Wilcoxon rank sum test $p<0.0001$, Figure 4.5), thereby providing more data points for analyses of Ct values for this group. There was no difference between the number of swabs for controls versus vaccinated participants if they had an asymptomatic infection or an episode of unknown symptom status.

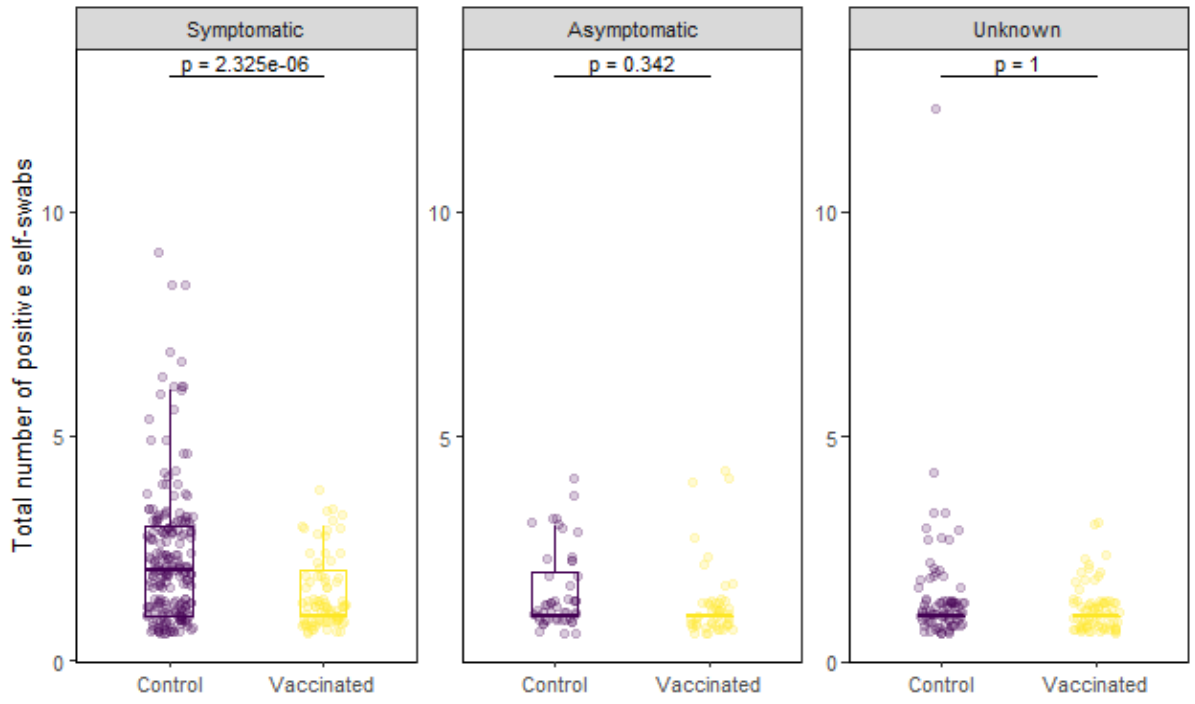


Figure 4.5: Number of positive self-swabs by vaccination and symptom status group. The midlines of the boxes show medians, and the lower and upper bounds of the boxes correspond to the 25th and 75th percentiles. The whiskers show the most extreme value no further than 1.5 x IQR above or below the 75th or 25th percentile. Comparison by Wilcoxon rank sum test.

Therefore, the analysis was repeated using the Ct value from the first positive weekly swab from a given participant ('first minimum Ct value'). The first positive swab was chosen due to the asymmetrical dynamics of SARS-CoV-2 infection whereby peak viral load is attained early in the infective process, [349] with a slower decline; hence, early swabs may be more likely to approximate an individual's highest viral load.

First minimum Ct values were similar to the infection minimum Ct for each symptom/vaccine status group (Table 4.5). First minimum Ct values were lower for those with symptomatic infections compared with asymptomatic infections for both controls (symptomatic control median Ct: 18.95, IQR: 15.26-26.44, vs asymptomatic control median Ct: 30.63, IQR: 21.55-33.38, Wilcoxon rank sum, adj. p<0.0001) and vaccinees (symptomatic

vaccinated median Ct: 21.73, IQR: 16.61-30.25, vs asymptomatic vaccinated median Ct: 32.6, IQR: 29.52-34.13, Wilcoxon rank sum, adj. $p<0.0001$). This was also the case for when symptomatic and unknown symptom status infections were compared within the control and vaccinated groups (Figure 4.6). First minimum Ct values were lower for control participants compared with vaccinated participants if infections were symptomatic (symptomatic controls median Ct: 18.95, IQR: 15.26-26.44, vs symptomatic vaccinated median Ct: 21.73, IQR: 16.61-30.25, Wilcoxon rank sum, adj. $p<0.05$) or of unknown symptom status (unknown symptoms controls median Ct: 29.10, IQR: 21.83-32.90, vs unknown symptoms vaccinated median Ct: 31.17, IQR: 26.55-34.07, Wilcoxon rank sum, adj. $p<0.05$). There was no difference in first minimum Ct values between controls and vaccinees in participants with asymptomatic infection.

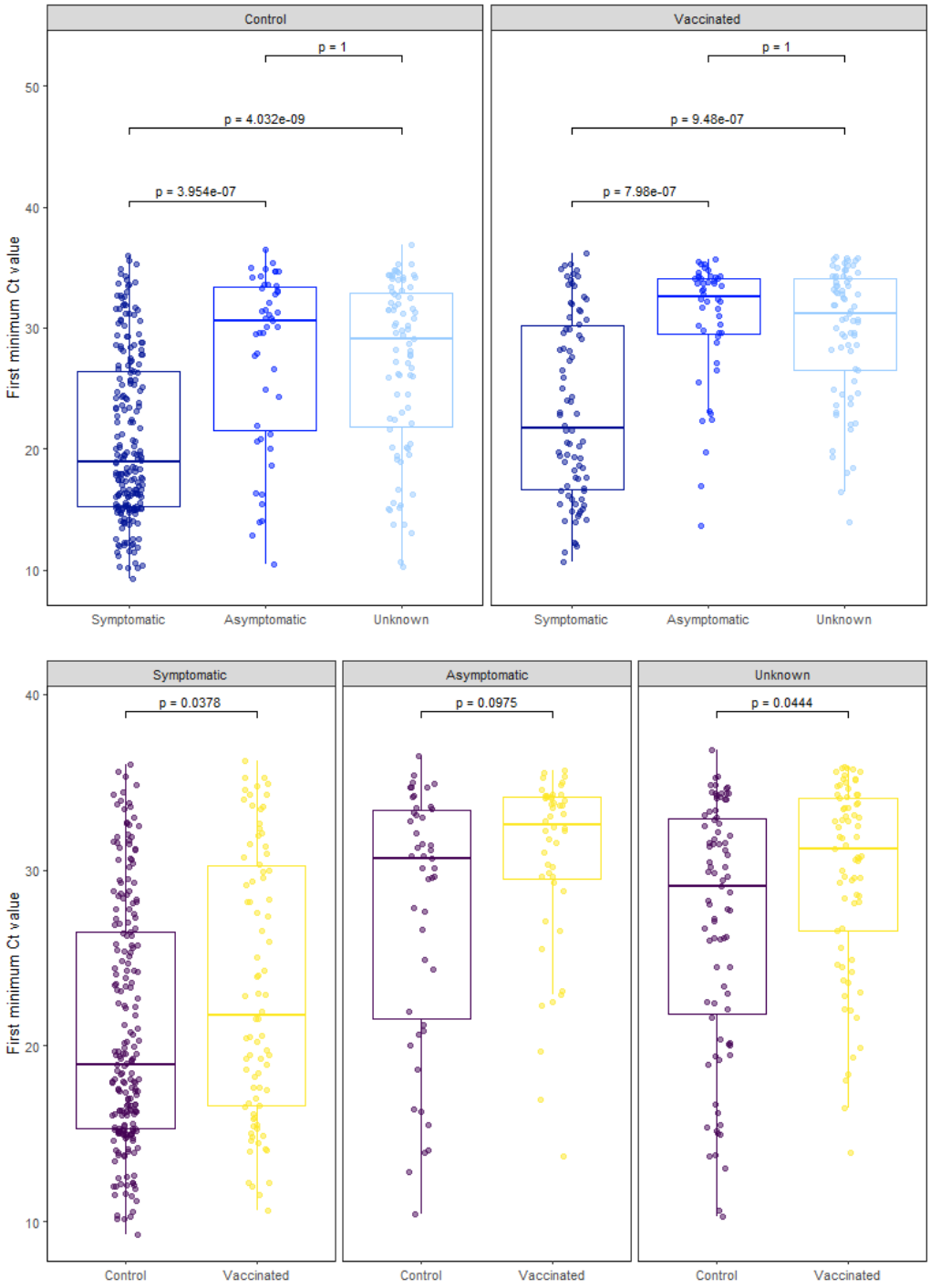


Figure 4.6: First minimum Ct value and symptom and vaccination status. Each point represents the 'first minimum Ct value' which was the lowest Ct value for any gene target from the first positive swab obtained during the participant's infection episode. Top panel comparisons within vaccination group, lower panel comparisons within symptom status group. The midlines of the boxes show medians, and the lower and upper bounds of the boxes correspond to the 25th and 75th percentiles. The whiskers show the most extreme value no further than 1.5 x IQR above or below the 75th or 25th percentile. Comparisons of groups by Wilcoxon rank sum tests. First minimum Ct values were missing for 69 participants. 53 in the control group (47 symptomatic, 5 asymptomatic, 1 unknown) and 16 (13 symptomatic, 3 asymptomatic) in the vaccinated group.

4.3.5 Predictors of Ct value

Multiple linear regression was performed to investigate the predictors of the infection minimum Ct value attained for participants (Table 4.6). Infection minimum Ct values attained for patients were no different between males and females or between those who with or without primary symptoms (cough, fever, shortness of breath, anosmia, ageusia) so were removed from the final model. It was found that age significantly predicted infection minimum Ct value (adjusted coefficient 0.05, 95% CI: 0.01-0.09, p<0.001). It also showed that symptom status predicted infection minimum Ct value: asymptomatic episodes had higher Ct values by 6.13 points (and thus lower viral loads) compared with symptomatic episodes (95% CI: 4.62-7.65, p<0.0001), the same was true for unknown symptom status compared with symptomatic episodes (adjusted coefficient 5.12, 95% CI: 3.82-6.43, p<0.0001). Vaccination also predicted infection minimum Ct value, with vaccinated participants having a 1.65-point increase in their infection minimum Ct value (95% CI: 0.50-2.08, p<0.01). Having more positive swabs for inclusion in the analysis also predicted infection minimum Ct value, reducing the value by 1.88 for each additional swab (95% CI: -2.34- -1.44, p<0.0001).

Table 4.6: Multiple linear regression of variables associated with infection minimum Ct value. Adj. R^2=0.34, F(5,541)=58.24, p=<$2.2x10^{-16}$

Variables		**Adj. coefficient**	**95% CI**		**p value**
Age (years)	Per year	0.05	0.01	0.09	**<0.001**
Symptom status					
Symptomatic		Reference			
Asymptomatic		6.13	4.62	7.65	**<0.0001**
Unknown		5.12	3.82	6.43	**<0.0001**
Vaccination status					
Control		Reference			
Vaccinated		1.65	0.50	2.80	**<0.01**
Number of positive self-swabs	Per swab	-1.88	-2.34	-1.44	**<0.0001**

4.3.6 Asymptomatic infections with low Ct values

Participants who had asymptomatic infection had a wide range of Ct values including low values indicating a high viral load (Figure 4.4, Figure 4.6). These individuals with high viral loads despite vaccination may make a greater contribution to transmission but would not be identified by symptom-based testing approaches hence identifying the risk factors for these individuals or immune profiles may be valuable.

There were 103 asymptomatic infections in total in this dataset, of which 51 were vaccinated and 52 were controls. Ct values were available for 95/103 participants. There were no significant demographic differences when participants with Ct values in the lowest quartile were compared with those with Ct values above the median, or when these groups were compared by vaccine status (Table 4.7).

For the vaccinated asymptomatic participants (n=51), 28-day post boost vaccination live microneutralisation and pseudoneutralisation titres were available for 29 and 41 participants respectively. ELISA data were available for only two participants so was not analysed. Ct values were not available for 3 participants. Although high titres for both microneutralisation assays were common in participants with higher Ct values there was no correlation between either titre and the infection minimum Ct values and high titres were also observed in participants with first quartile Ct values in the live microneutralisation assay (Figure 4.7).

Table 4.7: Comparison of demographic factors of asymptomatic participants with low Ct values (high viral load). Comparison of participants with lowest quartile infection minimum Ct values compared with those with above median infection minimum Ct values for all participants and by vaccination status. Statistical comparisons of continuous variables by Kruskal Wallis rank sum test, Sex by Pearson's chi squared for all participants, Sex by Fisher's exact test for controls and vaccinated participants.

	All participants				**Control**				**Vaccinated**			
Characteristic	**n**	**Lowest quartile infection minimum Ct** n=24	**Above median infection minimum Ct** n=49	**p value**	**n**	**Lowest quartile infection minimum Ct** n=12	**Above median infection minimum Ct** n=25	**p value**	**n**	**Lowest quartile infection minimum Ct** n=12	**Above median infection minimum Ct** n=24	**p value**
Age, median (IQR)	73	48 (31,56)	45 (38,53)	0.8	37	46 (30, 52)	46 (38, 54)	0.4	36	49 (31, 72)	45 (39, 51)	0.7
Female Sex, n (%)	73	15 (62%)	28 (57%)	0.7	37	7 (58%)	16 (64%)	>0.9	36	8 (67%)	12 (50%)	0.5
Prime boost interval, median (IQR)	73	57 (49, 95)	49 (22, 69)	0.07	37	64 (44, 121)	56 (31, 69)	0.4	36	57 (52, 74)	30 (17, 70)	0.09
Time since boost, median (IQR)	73	71 (47, 89)	81 (60, 127)	0.12	37	75 (34, 86)	75 (57, 99)	0.5	36	62 (49, 93)	89 (70, 129)	0.2

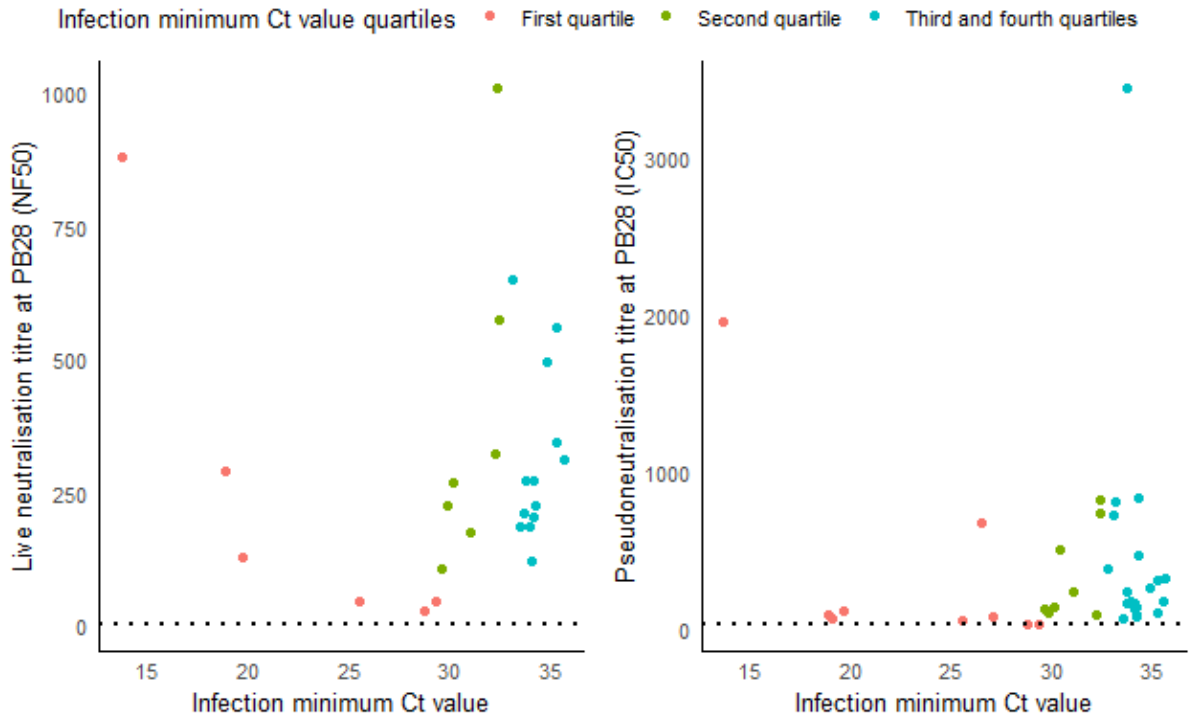


Figure 4.7: Neutralisation titres from asymptomatic vaccinees. Dot plots demonstrating infection minimum Ct value and live neutralisation and pseudoneutralisation titres from day 28 post boost in vaccinated asymptomatic participants (total n=51). Dotted lines on each plot represent the LLOQ for the assay. Infection minimum Ct data were not available for 3 participants, NF50 data were not available for 22 participants and IC50 data were not available for 10 participants.

4.3.7 Duration of swab positivity, symptom status and vaccination

From the 616 participants included in this analysis, 20563 swabs were included from any source (1252 positive (6%), 18843 negative (92%) and 468 indeterminate (2%)). The number of positive swabs per participant is shown in Table 4.8. Just over half of the participants (325, 52.7%) had more than one swab included in the analysis.

Table 4.8: Number of NAAT positive swabs per participant from any source. Number of participants, n=616.

Number of NAAT positive swabs from any source, n (%)											
1	2	3	4	5	6	7	8	9	10	14	Total
291 (47)	132 (21)	108 (18)	54 (8.8)	15 (2.4)	4 (0.6)	5 (0.8)	2 (0.3)	3 (0.5)	1 (0.2)	1 (0.2)	1252

To explore whether symptom status or vaccination status was related to duration of NAAT positivity the time in weeks from first to last positive swab were used as a proxy and analysed. When considered within vaccination status group, symptomatic episodes were associated with longer duration of positivity in weeks compared with those who were asymptomatic for both controls (symptomatic controls median duration: 2, IQR: 1-3, vs asymptomatic controls median duration: 1, IQR:1-1, Wilcoxon rank sum adj. $p<0.0001$) and vaccinees (symptomatic vaccinees median duration: 1 (IQR: 1-2) vs asymptomatic vaccinees median duration: 1 (IQR:1-1), Wilcoxon rank sum adj. $p<0.05$). Comparisons between symptomatic and unknown symptom status groups produced similar findings (Figure 4.8).

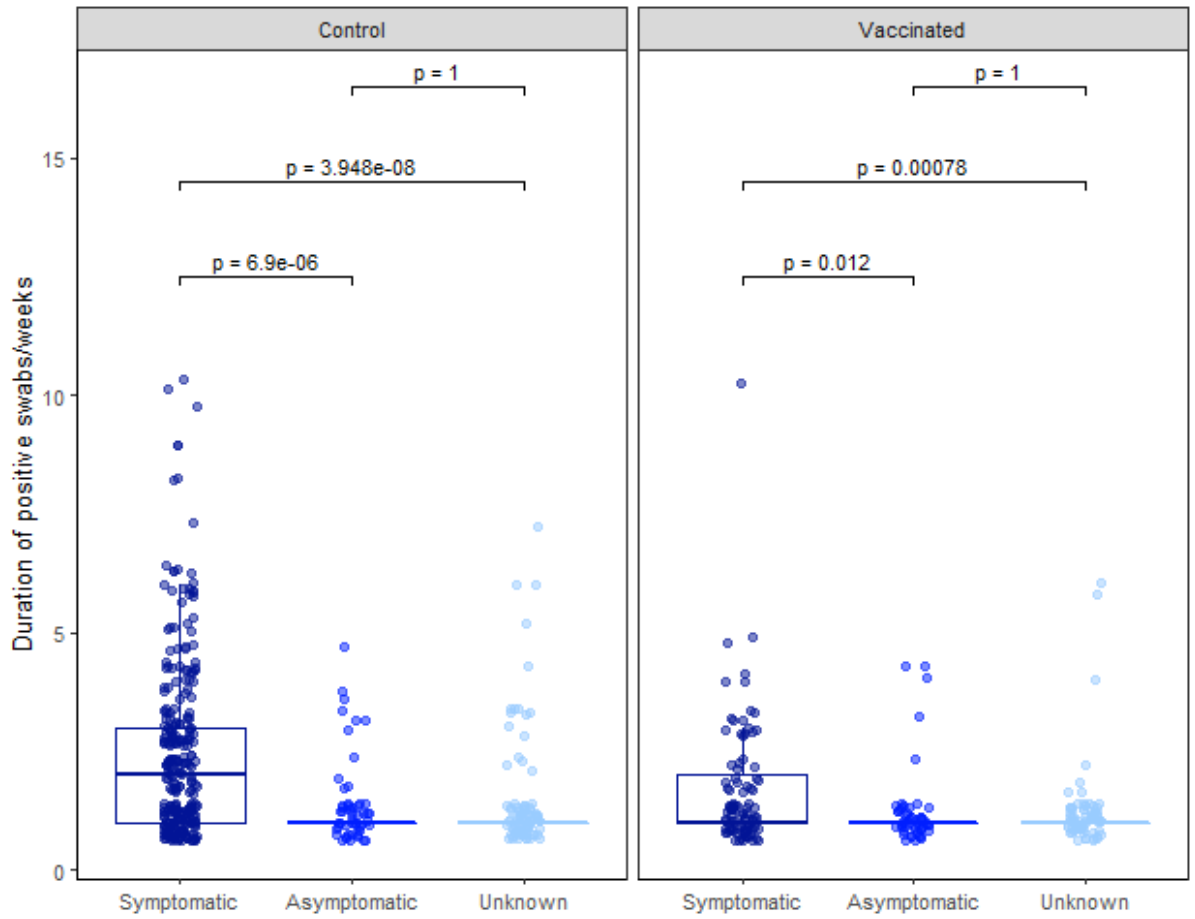


Figure 4.8: Duration of NAAT positive swabs, vaccination and symptom status. Each point represents one participant. The midlines of the boxes show medians, and the lower and upper bounds of the boxes correspond to the 25th and 75th percentiles. The whiskers show the most extreme value no further than 1.5 x IQR above or below the 75th or 25th percentile. Comparisons of groups by Wilcoxon rank sum tests.

For symptomatic participants, those who were vaccinated had a statistically significantly shorter duration of positive swabs compared with control participants (vaccinated, median of 1 week, IQR: 1-2, compared with controls, median 2 weeks, IQR:1-3, Wilcoxon rank sum test p<0.0001). There was no difference between vaccine status groups for participants who had asymptomatic or unknown symptom status infections (

Figure 4.9).

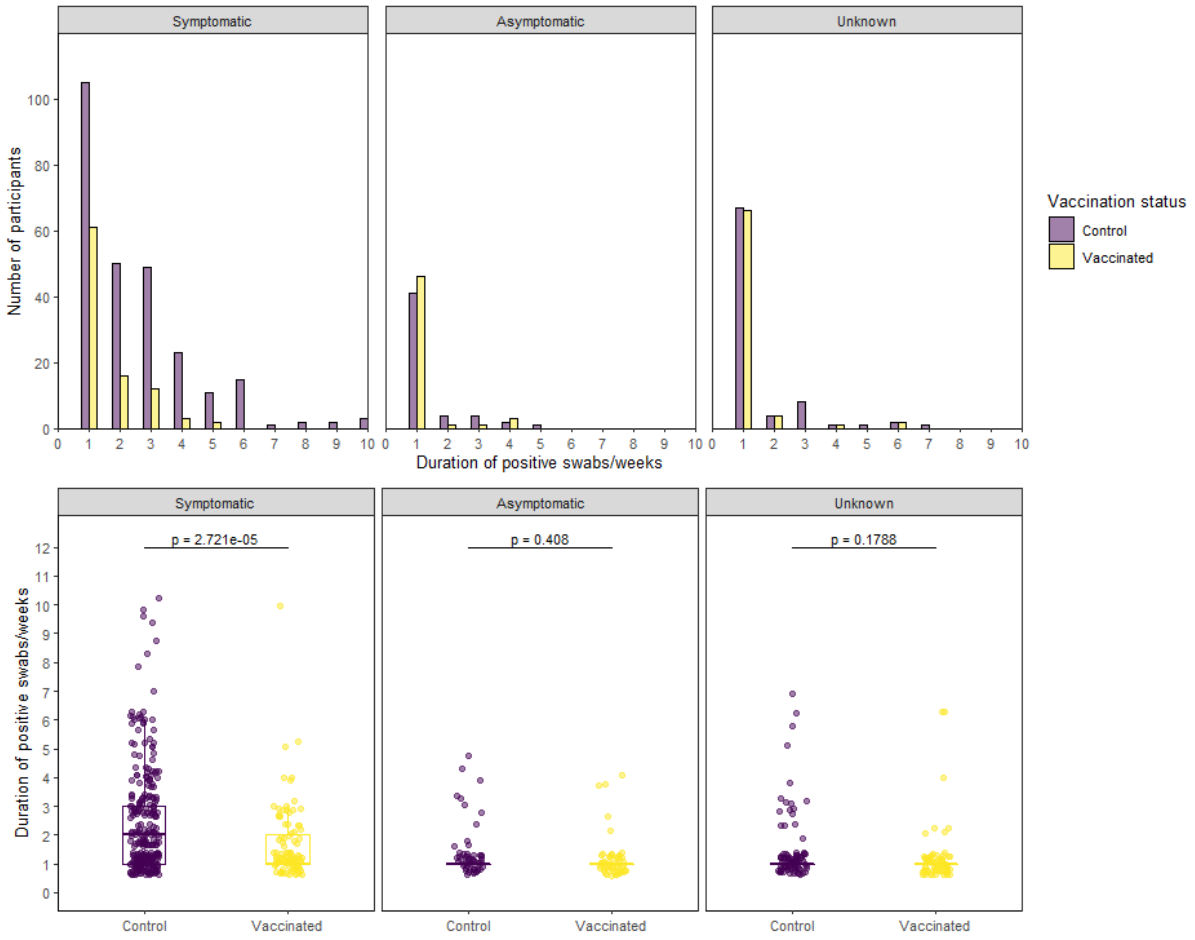


Figure 4.9: Duration of positive swabs by symptom and vaccination status. The midlines of the boxes show medians, and the lower and upper bounds of the boxes correspond to the 25th and 75th percentiles. The whiskers show the most extreme value no further than 1.5 x IQR above or below the 75th or 25th percentile. Comparisons of duration of positivity between groups by Wilcoxon rank sum tests.

If symptomatic control participants returned a greater number of total swabs (and thereby had more chance of testing positive multiple times) than vaccinated participants, this could explain why their duration of positive swabs was significantly longer in this group. To assess whether participants in different vaccine and symptom status groups returned a differing number of swabs the total number of swabs returned and the total number of positive swabs per group were calculated.

For swabs from any source there was no difference in the total number of swabs sent between vaccination groups by symptom status. Symptomatic control participants did however return more positive swabs (controls, median = 3, IQR: 2-4) compared with symptomatic vaccinated participants (vaccinated, median = 2, IQR:1-3, Wilcoxon rank sum test, p<0.0001, Figure 4.10).

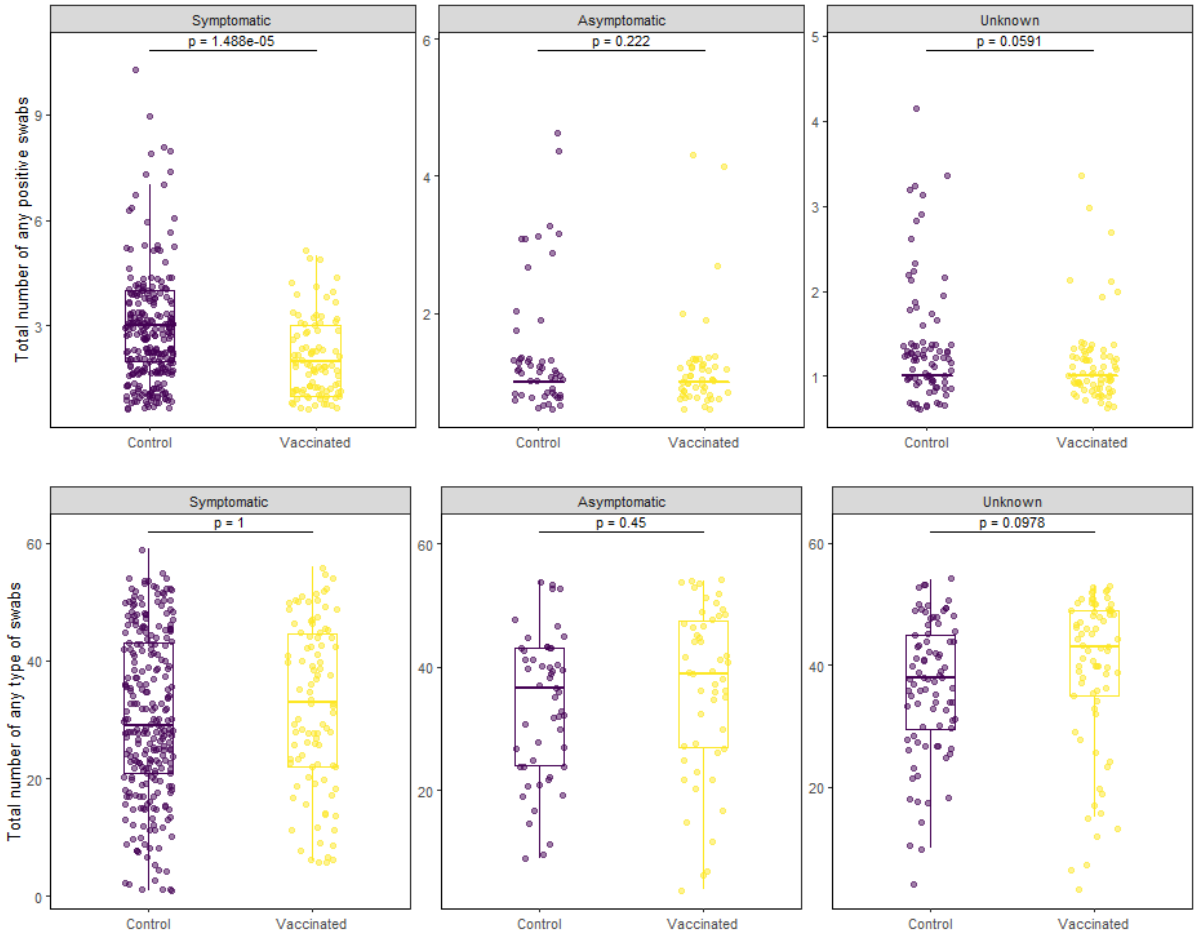


Figure 4.10: Number of swabs by vaccination and symptom status. Number of positive swabs from any source by vaccination and symptom status (top panel), total number of swabs from any source by vaccination and symptom status (lower panel). The midlines of the boxes show medians, and the lower and upper bounds of the boxes correspond to the 25th and 75th percentiles. The whiskers show the most extreme value no further than 1.5 x IQR above or below the 75th or 25th percentile. Comparisons of number of swabs between groups by Wilcoxon rank sum tests.

4.3.8 Predictors of NAAT positive duration

Multiple linear regression was performed to see which variables were associated with duration of positivity (Table 4.9). Only vaccination status and infection minimum Ct value were associated with duration of NAAT positivity; sex and age were removed from subsequent models. Symptom status and primary symptoms were retained as these variables approached significance in the original models. Duration of NAAT positivity decreased by approximately 5 days (-0.72 weeks) for asymptomatic individuals compared with those who were symptomatic. For each 10 unit increase in cycle threshold (i.e. lower viral load) NAAT positive duration reduced by approximately 3 days. Vaccinated participants had a 3.7 day (0.53 weeks) shorter duration of NAAT positivity compared with controls.

Table 4.9: Multiple linear regression investigating factors associated with duration of NAAT positivity.

Variables	**Adj. coefficient**	**95% CI**		**p value**
Symptom status				
Symptomatic	Reference			
Asymptomatic	-0.72	-1.09	-0.35	**<0.001**
Unknown	-0.75	-1.07	-0.44	**<0.001**
Infection minimum Ct	-0.04	-0.06	-0.03	**<0.001**
Vaccination status				
Controls	Reference			
Vaccinated	-0.53	-0.79	-0.27	**<0.001**

4.4 DISCUSSION

ChAdOx1 nCoV-19 vaccination was shown to have efficacy in protecting individuals from symptomatic COVID-19 in an interim analysis [275] however, additional knowledge can be gained from analysis of the SARS-CoV-2 infections that occurred in both trial allocation arms. In this chapter, I have demonstrated control participants who had SARS-CoV-2 infections were more likely to be symptomatic, experience primary symptoms, and have more severe disease

(defined by WHO clinical progression score) compared with ChAdOx1 nCoV-19 vaccinees who experienced infection. I have also shown that vaccinated participants had higher Ct values (therefore lower viral loads) and were NAAT positive for a shorter duration from the upper airway. The implication of these findings is that ChAdOx1 nCoV-19 confers protection from symptoms in mild disease and may reduce transmission by reducing peak viral load and duration of shedding.

In this study blinded participants with symptomatic infection were younger than those who were asymptomatic or unknown, mirroring that which was seen for all participants who tested positive in the study. As discussed in Chapter 3, the finding of symptomatic disease being associated with younger participants does not accord with the existing literature on risk of symptomatic disease but may relate to increased social contact of younger individuals and the risk of high exposure. It is unclear as to whether there is a dose-response relationship between SARS-CoV-2 inoculum and symptoms. For other respiratory pathogens such as influenza challenge studies have not found a consistent relationship between inoculum and symptom profile. [350,351] In the SARS-CoV-2 challenge model for naïve participants the lowest dose of 10 $TCID_{50}$ was sufficient to cause infection within the specified primary endpoint range (53%, aim: 50-70%) of participants so there was no further dose escalation however symptoms within this dose group were variable. [88] Real world data suggests that household contact with a positive case may be associated with increased symptom severity, as for other viral pathogens such as varicella and measles but this is unclear if this relates to magnitude or duration of infectious exposure. [352,353] Young individuals may have been more likely to be sharing accommodation and therefore exposed to higher inoculums resulting in symptomatic infection. Alternatively, this observation may be related to reduced adherence to weekly swabbing in the younger age groups. Reduced adherence in these groups will have led to an underestimation of

infections that were asymptomatic or pauci-symptomatic which may account for the age of participants to be younger in the symptomatic group.

Systematic reviews exploring asymptomatic SARS-CoV-2 infection covering studies conducted over similar time periods to the data included in this analysis have been conducted. They range in their estimates from 17% (95% CI 14% to 20%) [354] to approximately one third of infections being asymptomatic (31%, 95% CI 26% to 37%, [355]; 36.6%, 95% CI 31.8 to 42.4% [128]). A further systematic review found a higher rate of 42.5% (IQR 29.6 to 77.8%) [129] however there are methodological difficulties with many of the included studies. The limitations of these studies typically involve adequacy of follow up and bias which highlight the challenges of conducting large scale research of this nature, initiated rapidly within a pandemic setting. [356] Whilst the proportion of cases that are asymptomatic looks low in our dataset, it is likely that many of those that are classified as unknown may represent asymptomatic cases, but this is speculation. If the asymptomatic and unknown categories are combined as 'possible asymptomatic', this gives a proportion of 34% (95% CI 29.3-38.9) in the control group which, despite the limitations of this study accords well with the published literature. The proportion of infections that were asymptomatic in vaccinees was greater than in controls, suggesting the vaccine may convert infections from being symptomatic to asymptomatic or pauci-symptomatic. Indeed, real-world data in health care workers in the UK who had been offered BNT162b2 vaccination showed a similar reduction in symptomatic infection and increased proportion of asymptomatic infections in vaccinees. [179] Importantly, the secondary attack rate has been estimated to be lower from asymptomatic individuals in the pre-vaccination era [355] suggesting that if vaccination does indeed convert what would otherwise have been symptomatic cases to asymptomatic, it may also reduce transmission risk. This is supported in our data, in which asymptomatic cases has higher Ct values (lower viral load) than

symptomatic cases. Later studies of the B.1.1.7 (Alpha) variant have demonstrated that vaccination (with two doses of either BNT162b2 or ChAdOx1 nCoV-19) is associated with a reduction in transmission. [301]

Our study highlights the potential tension between maximising information derived from a unique large clinical trial on a novel pathogen and the fundamental aims of assessing vaccine efficacy when investigating human participants for subjective symptoms if these form part of the primary endpoint definition. It was essential not to prejudice the trial primary endpoint of primary symptomatic disease and for this reason whilst all participants were reminded weekly to contact their study site if they were symptomatic, participants with a positive weekly swab were not directly contacted to ascertain whether they were experiencing symptoms. Therefore, symptom status misclassification is likely to have occurred for some cases in this dataset, which is also noted in other observational studies. [356] It is also possible that those who received a positive swab result and contacted their study sites to inform them and were thus recorded as asymptomatic may have in fact been pre-symptomatic, particularly those who went on to have multiple positive swabs. [356] Longitudinal follow up of symptoms in the short-term following a positive result would have allowed more clarity on this. However, results from self-swabs took days for the result to be communicated to participants due to being transported by post therefore additional time had passed to allow for symptoms to develop before the result was received. A systematic review including data from 14 longitudinal studies published until November 2020 allowed distinction between asymptomatic and pre-symptomatic individuals suggested that 72.3% (IQR: 56.7%-89.7%) of individuals that were asymptomatic at the time of their positive PCR test remained symptom-free. [129] Therefore it may be assumed that those participants in this study who contacted their study site to say they were asymptomatic after receiving a positive result are likely to have remained truly asymptomatic.

Endpoint assessments only took place after the infective episode was complete so the evidence from the entire episode was appraised to assign a symptom status. Inevitably, there will have been some participants who did not contact their study site if they developed symptoms (or who had symptoms but not those that prompted a symptomatic visit) and therefore the unknown group may have included some symptomatic individuals however other indicators such as swabbing adherence rates suggest that compliance with other trial procedures was high. Cases of infection in the trial were typified by mild disease, due in part to the demography of the trial population who were mostly young and healthy however, even in this group ChAdOx1 nCoV-19 vaccination was associated with a reduction in severity. Differentiating between WHO progression score 2 and 3 is more subjective, than between higher scores which are distinguished by objective criteria (such as route of oxygen administration). However, assessors were blinded as to vaccine allocation, hence this finding is likely to be robust. Whilst the COV002 study in the UK had too few cases to demonstrate a reduction in more severe endpoints, later effectiveness studies have shown the protection efficacy of ChAdOx1 nCoV-19 vaccination against severity endpoints such as hospitalisation and death. [188]

In both vaccinated and control participants, symptomatic disease was associated with lower PCR cycle threshold values compared with asymptomatic disease, though there were a range of cycle threshold values associated with asymptomatic infection. Some studies conducted in different settings in the pre-vaccination era have found no difference in cycle thresholds with symptom status [91,125,349], whilst others have observed lower cycle thresholds in pre-symptomatic or symptomatic cases. [127,357] As in other studies, in our participants, being asymptomatic did not preclude low cycle threshold values (high viral loads) in some cases. [86,127] Our finding of a longer duration of positive swabs in symptomatic cases is in keeping

with other groups. [349] Regression analysis of duration of NAAT positivity showed that lower infection minimum cycle threshold value was associated with a longer duration of NAAT positivity showing that the trajectory of duration may relate to peak viral load.

Taken together these findings suggest that ChAdOx1 nCoV-19 vaccination has an impact on symptomatic disease and parameters such as Ct value and duration of NAAT positivity when associated with symptomatic disease but the same is not true for asymptomatic infection. This evidence in humans is analogous to the original non-human primate data which suggested that viral replication may still occur in the upper respiratory tract at the same time as the lower respiratory tract being protected. [170] Effectiveness data following vaccination roll out in the UK confirmed the finding in this study by demonstrating that vaccination with ChAdOx1 nCoV-19 had a greater effect in symptomatic infections and higher viral burden. [358]

There are several caveats around extrapolating from PCR Ct values to the risk of viral transmission. During the SARS-CoV-2 pandemic PCR cycle threshold values have been widely used as a surrogate for viral load, however this is a non-linear relationship and ideally requires reference to a human target. [94] The presence of RNA detectable by PCR does not equate to viable virus with infective potential, however good correlation between cycle thresholds and culture positivity has been established. [86,98,320] A further step is required between culturable virus and infection of a host in terms of route of transmission. Recent evidence from a COVID-19 human challenge study showed that the dose required to infect naïve individuals is as low as 10 $TCID_{50}$ implying that the presence of even small quantities of culturable virus could plausibly result in transmission. [88] Infectivity assays were beyond the scope of this trial given the number of participants and swabs but would have provided additional information about the implications of these Ct findings. However, a large-scale retrospective cohort study using contact tracing data in the UK showed vaccination with ChAdOx1 nCoV-19 or

BNT162b2 was associated with a reduction in transmission for the Alpha variant which was one of the two most dominant lineages during this study period. [301]

The strengths of this study are the large number of closely monitored participants who were randomised to receive an active control vaccine or the ChAdOx1 nCoV-19 and were highly adherent to sending regular self-swabs which were tested on the same assay. Unlike many early studies, this study includes many individuals who tested positive, and those with symptomatic infection are well characterised.

Limitations of this study include ascertainment of asymptomatic status, as previously discussed and the mild infection experienced by most participants which precluded exploration of more severe endpoints. The demographics of the trial population are a function of targeted recruitment, meaning that it is difficult to know if these findings could be extrapolated to other populations, particularly those at greatest risk of severe outcomes such as long-term care facility residents. However, the younger, relatively healthy cohort studied here are likely to be important contributors to community transmission. Younger participants, especially those less than 25 years old returned a lower percentage of self-swabs (Chapter 3) than older participants. This may mean that these participants may have been less likely to have had asymptomatic infections detected by self-swabbing and have fewer swabs to contribute to analyses. As previously discussed, this is likely to bias the data and make it appear that younger participants had fewer asymptomatic infections. Due to the viral kinetics of SARS-CoV-2 it is possible that asking participants to self-swab weekly may have been insufficient to capture all cases and more frequent swabbing would have provided additional data points to better define peak viral load, but this may have proved too burdensome for participants thereby reducing adherence. However, modelling from Kissler et al. [349] based on data from 2020 suggests that weekly swabbing should have captured most cases, with mean pre-peak viral load duration of 3.5 days

(95% CI 2.5, 4.5), and clearance phase of 7.8 days (95% CI 6.1, 9.7) in those who were asymptomatic at diagnosis, with a longer clearance phase for those who were symptomatic.

The use of the infection minimum Ct from a participant's entire episode may have introduced bias as symptomatic controls returned more positive swabs then their vaccinated counterparts and therefore had a greater chance of a swab being taken closest to peak viral load. However, the finding that symptomatic individuals had lower Ct values was still observed when only the first positive swab for a given participant was analysed. A large study of Ct values in randomly selected households showed that Ct values were lower in the first positive swab for an individual [320] suggesting this confirmatory approach was appropriate.

The epidemiological context during the period of this study was dynamic with large changes in the rates of infection in the UK, with early infections occurring more sporadically and later infections resulting from a pandemic wave of B.1.1.7 (Alpha) infection within the UK. Evidence suggests that at a population level during a growing epidemic, Ct values will be lower as individuals have more chance of being sampled early in their infection and get higher as a wave of infection subsides. [359] These data are additionally complicated by the emergence of the B.1.1.7 (Alpha) variant during late 2020 when infections overall were increasing, and concern arose that vaccine efficacy may be impacted. Unlike a later variant such as Delta, the B.1.1.7 (Alpha) variant infection increased at a time when there were other lineages prevalent thereby precluding assigning a lineage on time based epidemiological data. The next chapter examines these data from this perspective.

This was an ambitious project to include within a rapidly delivered clinical trial but has produced results which were subsequently confirmed at a population level after vaccine roll out. If such a project were to be repeated for a future respiratory virus, prior to full understanding of the viral dynamics of a novel pathogen there are several changes to the design

that would optimise the quality of the data collected. These would include, but are not limited to, additional intensive sampling following a positive result for all participants, automated symptom diaries and samples taken at sites at symptomatic visits for testing on the same assay as surveillance swabs.

5 ChAdOx-1 nCoV-19 vaccination and the B.1.1.7 variant of SARS-CoV-2

5.1 INTRODUCTION

In late 2020, a novel variant of SARS-CoV-2, named B.1.1.7 (subsequently named Alpha) was first identified in the UK and became the dominant lineage in the UK from December 2020 until April 2021. In contrast to the lineages circulating in Europe during 2020 which had relatively few mutations, [54] B.1.1.7 has 23 mutations across the genome relative to the Wu-1 reference genome. These include nine non-synonymous mutations affecting the spike protein which mediates host cell entry by the virus. [35]

The overall population mutation rate in SARS-CoV-2 is approximately 1-2 nucleotide mutations per month and typically infections within an immunocompetent individual are relatively short. [31,32] Given this rate, it would take approximately 6 months for the number of mutations in B.1.1.7 to arise by repeated transmission of intermediate variants. However, there is no evidence of any such B.1.1.7 precursors in the UK or global genomic data prior to September 2020, despite extensive genome sequencing and monitoring, (including groups that could be hypothesised to be at risk for such mutations), [63,387] suggesting that stepwise acquisition of mutations was not how this VOC evolved. One potential mechanism for the occurrence of this constellation of mutations is the accumulation of multiple mutations secondary to chronic infection in an immunocompromised individual. [38]

The most notable mutation present in B.1.1.7 is N501Y affecting the receptor binding domain (RBD), a mutation which independently emerged in other variants of concern, B.1.351

(Beta) and P.1 (Gamma) during 2020. [34,360] Mutations at N501 can increase the affinity of the RBD for ACE2, [41,57,361] with kinetic analyses showing that for the N501Y mutation this is mediated by a moderately increased association rate and a large reduction in dissociation rate. [41,57] A deletion in the N-terminal binding domain, H69-/V70-, also increases ACE2 binding efficiency, increases infectivity mediated via promotion of cleaved spike into virions and cell-to-cell fusion. [362] It also causes a reproducible drop-out in the S gene in the commonly used Thermo Fisher TaqPath three gene target diagnostic assay. [55] Non-spike substitutions may also contribute to variant phenotype such as the nucleocapsid mutations R203K/G204R, which may confer a replication advantage mediated by replication efficiency and thereby contribute to increased transmission and virulence. [47]

Evidence suggests that the B.1.1.7 lineage may be associated with a higher reproduction number compared with prior lineages [363,364], higher upper airway viral loads [63,365] and more severe illness, [44] although these associations have not been consistently replicated. [63,366,367]

In vitro reports of results from different neutralisation assays show that convalescent and mRNA and viral vector vaccinee sera demonstrated reduced neutralisation of B.1.1.7 (or pseudoviruses carrying some or all mutations present in B.1.1.7) compared with parental strains. [57,368-370] Other groups showed no difference in convalescent sera to neutralise B.1.1.7 compared with older strains. [371,372] It was unclear as to whether the observed reduction in neutralisation observed by some groups, in some assays, would translate into an impact on vaccine efficacy.

At the time of identification of B.1.1.7, several COVID-19 vaccines using spike protein as their immunogen had announced efficacy against symptomatic COVID-19 infection in phase

3 clinical trials. [187,275,289] Mutations affecting the spike protein caused concern regarding whether immunity induced by vaccination with ancestral spike protein would remain protective, particularly in light of the emerging *in vitro* evidence described above.

In addition to the question as to whether viral mutation may render vaccines less efficacious there is also the question of whether the nature of the dominant infecting viral variant may be influenced by vaccination status of the host. In general, during the period included in this study the genetic diversity of the virus was low [40,373] however it was unknown if vaccination with a spike containing product may be associated with infections by viruses containing a greater number of novel mutations (particularly of the spike region), or whether samples from vaccinated individuals may show phylogenetic clustering driven by shared mutations. Presence of any hypothetical vaccination-associated mutations would be observable either as close phylogenetic relatedness of samples from vaccinated individuals (clustering), or as unusually long phylogenetic divergence from the root, resulting from homoplasy (recurrent mutation along the phylogeny, or 'toggling'). Analysis of consensus sequences also permits review of whether there were any regions of the genome with higher rates of mutation (with specific interest in the spike protein) or specific mutations that may be associated with vaccination given the unique trial context allowing direct comparison between randomised groups.

As with other viral infections such as HCV and HIV, [374,375] an individual infected with SARS-CoV-2 at any time in their infection may harbour populations of genetically distinct viruses, detectable by the presence of sub-consensus ("minor") alleles relative to the reference genome. These minor allele variations (or intra-host single nucleotide variants, iSNV), when present in gene regions, can result in non-synonymous changes or deletions in a proportion of

the resulting protein product, with the potential to cause meaningful viral phenotypic changes such as altered transmissibility or resistance to therapeutics. [43,376] Others may cause synonymous mutations or deletions and not have any functional impact, though may still result in the loss of PCR binding sites and therefore the risk of molecular diagnostic assay failure exemplified by H69-/V70- deletion present in B.1.1.7 which produced SGTF. [55] The majority of spontaneous mutations will be deleterious to viral fitness and be swiftly lost. Even if functional, most iSNVs will be lost via stochasticity (genetic drift) during infection and transmission. Within a given host, iSNVs can be gained or lost during an infection. [40,377-380] Whilst iSNVs can be detected in most sequences, intrahost viral diversity of SARS-CoV-2 has been shown to be low in published literature, with each sequence typically harbouring fewer than five iSNVs. [40,378,379,381]

The analysis of iSNVs is complicated by the difficulty in distinguishing between genuine sub-consensus variation and sequencing error. [382] It is further exacerbated by the confounding influence of sample viral load [379,383,384] and methodological differences that make direct comparisons of datasets challenging and may account for some of the variance observed in studies. [382,385] As a compromise between sensitivity and specificity of iSNV detection, recent studies have generally used iSNVs at sites supported by at least 100 sequence reads in total (depth 100x), with a minimum of 5% minor allele frequency (MAF) – that is, a minimum of 20 reads supporting the iSNV. [40,383,386] This reduces false detections but is a much stricter detection threshold compared with consensus variants, which are generally called as majority alleles (>50% MAF) at sites supported by a total of at least 10 reads (depth 10x). Samples with a low viral load by definition have fewer starting template molecules for sequencing and typically result in lower sequencing depth and coverage, and therefore a greater variance of

MAFs (as detection of a minor allele may be disproportionately amplified). For this reason, minor variants can be almost impossible to detect reliably in low viral load samples. This can lead to confounding since viral load and generation of iSNVs both plausibly vary over the course of infection. For this reason, most iSNV studies involve samples with moderate to high viral load.

Intrahost diversity, when reliably inferred, can be used to look at transmission [40] and can also inform which genomic positions are most prone to mutation. The presence of intrahost viral diversity may be associated with positive selection, which may be exerted by vaccine-mediated or pre-existing immunity, or by therapeutics such as monoclonal antibodies. The transmission bottleneck in SARS-CoV-2 is thought to be narrow [40] and independent of lineage [386] meaning that typically an individual is infected with virus exhibiting little or no diversity. Therefore, it is most likely that VOCs first arise within a host. Ultimately, within host viral diversity may result in the emergence of new variants as iSNVs serve as a "mutational pool" which may contribute to viral evolution. [378]

SARS-CoV-2 vaccination may exert selective pressure on intrahost viral diversity which could be hypothesised to act in either direction. Vaccine mediated immunity may reduce viral diversity at the time of infection via the action of neutralising antibodies (in contrast to antivirals or monoclonals, which are introduced during an established infection); or alternatively, the presence of vaccine-mediated humoral and cellular immunity may exert positive selection during replication during infection within the host and promote viral diversity. The effects, if any, of vaccination on viral diversity may also be lineage-dependent and relate to the match between spike sequence used in the vaccine and that which the lineage expresses.

In this study, sequences from UK participants in a randomised controlled trial from 2020/2021 with two main viral lineages (B.1.177 and B.1.1.7) were compared by vaccination status, to compare viral, clinical, vaccine efficacy and genomic characteristics between lineages.

The objectives of this study were:

1. To describe the SARS-CoV-2 lineages and phylogeny identified from trial participants.
2. To compare Ct values (as a proxy for viral load) from infections caused by B.1.1.7 (Alpha) and a pre-Alpha variant (B.1.177) in control and ChAdOx1 nCoV-19 vaccinated groups.
3. To compare duration of NAAT positivity for infections caused by B.1.1.7 (Alpha) and a pre-Alpha variant (B.1.177) in control and ChAdOx1 nCoV-19 vaccinated groups.
4. To perform exploratory efficacy analyses for infections associated with different symptom profiles of ChAdOx1 nCoV-19 vaccination for the B.1.1.7 (Alpha) variant.
5. To perform exploratory analysis of SARS-CoV-2 consensus sequences to assess whether if vaccination status was associated with unusual phylogenetic structure or the presence of specific mutations.
6. Describe and compare the number of intrahost single nucleotide variants (iSNVs), and number of unique iSNV sites by group.
7. Explore iSNV density along the genome, accounting for gene length.
8. Identify any highly recurrent iSNV sites, which may represent mutational hotspots or sites of positive selection.

5.2 METHODS

This analysis was performed on the participants from the multi-centre Phase 2/3 COV002 COVID-19 vaccine trial where adult participants aged 18 and over were randomised to receive ChAdOx1 nCoV-19 vaccination or control vaccination (MenACWY vaccination). The study design and generic methods for COV002 are given in chapter 2. Participants were sampled if they experienced primary symptoms (fever ≥37.8°C, cough, shortness of breath, anosmia or ageusia) and asked to perform weekly self-swabs regardless of symptoms. Specific details relating to NAAT testing of symptomatic and weekly self-swabs are described in chapters 2, 3, and 4. The analyses in this chapter include participant data from recruitment in 2020 until 30th June 2021 or the end of their participation in the study, if this occurred prior to this date.

My role was the initiation, set up and management of sequencing from the trial, liaison with COG partners, the Sanger Centre and Lighthouse laboratories. I managed the lineage and trial data to complete the following analyses.

5.2.1 Participants

Participants who underwent randomisation were included in efficacy analyses. Participants were analysed according to the vaccine received. Baseline SARS-CoV-2 serology was not available for all participants so was not used as an exclusion criterion for efficacy analyses. All ChAdOx1 nCoV-19 dose groups (low dose/low dose (LD/LD), low dose/standard dose (LD/SD), standard dose/standard dose, (SD/SD)) were considered vaccinated. Cases were excluded as described in section 4.2.

5.2.2 Ct values and duration

Infection minimum Ct values were computed taking the minimum Ct value for any present positive gene target across all the self-swabs obtained for the first SARS-CoV-2 infection experienced by a participant. Duration was defined as time in weeks from first positive swab to last positive swab of the first infection.

5.2.3 S gene target failure

S gene target failure (SGTF) was defined as minimum ORF or N gene cycle threshold value of 32 or lower with a negative S gene result and/or B.1.1.7 identified on sequencing or the absence of data from sequencing. Non-SGTF was defined as a positive swab which included a positive S gene result and alternative lineage identified on sequencing. All other swabs that did not meet either of these criteria or did not have gene target cycle threshold data available were placed in the 'not known' category. Where multiple positive swabs were available for a participant, each swab was assigned an S gene target failure category and an overall infection episode was classed as SGTF, non-SGTF or 'not known'. Analyses performed with SGTF data were restricted to cases occurring after 1st November 2020 because before this time SGTF was a poor proxy for B.1.1.7.

5.2.4 Ascertainment of sequences

Sequences were obtained for positive samples from symptomatic visits at study sites which were NAAT tested for SARS-CoV-2 locally at study sites using MHRA derogated assays, or from weekly self-swabs which underwent PCR testing at Lighthouse laboratories (TaqPath RT-PCR COVID-19 assay, Thermo Fisher Scientific, Waltham, MA, USA).

NAAT positive samples were identified, extracted and sequenced by COG partners using the platform (Oxford Nanopore, Oxford Nanopore Technologies, Oxford, UK; Illumina, Illumina Inc., San Diego, CA, USA) and protocol (e.g. veSEQ, [388] ARTIC [389]) available at individual sites. Alternatively, when PCR positive samples were identified at Lighthouse laboratories, plates containing positive samples were initially sent to the Sanger or local COG partners for sequencing, later samples were consolidated onto plates and sent to the Oxford Viral Sequencing Group for sequencing using the quantitative veSEQ approach. [388]

Where participants had multiple sequences available, the first chronological sequence available where a lineage was identified was used. If the sequence lineage was B.1.1.7 but the corresponding swab PCR result showed the presence of the S gene (n=2), the sequence was excluded.

The following designations were used for groups of participants. The 'Other lineages' group includes all identified lineages other than B.1.177 and B.1.1.7, 'no sequence result' refers to those cases where a sequence was attempted for at least one swab but a sequence sufficient to assign a lineage was not obtained, 'not sequenced' refers to infections where no swab could not be retrieved for sequencing.

5.2.5 Genome reconstruction & Phylogeny reconstruction

Viral genomes were assembled by Oxford Viral Sequencing Group as previously described, [40] and variant frequencies were computed using shiver (tools/AnalysePileup.py), [390] with the default settings of no base alignment quality and maximum pileup depth of 1 000 000. Lineages were assigned by Pangolin version 2.1.7 (lineages version 2021–02–12) using the determined consensus genome for each sequenced sample. If the proportion of gaps was greater than 50%

(the recommended minimum for Pangolin), lineages were assigned on the basis of presence of lineage-defining mutations in the spike gene. [35]

Samples that could not be assigned a lineage by either method, due either to absence of reads covering the lineage-defining sites or to mixed allele calls consistent with contamination, were classified as unknown. Consensus sequences were aligned using MAFFT version 7.402. [391]

Phylogenetic reconstruction was performed on the alignment consisting of consensus sequences rooted with the Wuhan-Hu-1 reference sequence (RefSeq NC_045512), using IQ-TREE version 1.6.12, [392] with the generalised time reversible + FreeRate model and 1000 bootstrap replicates. Phylogenetic trees were visualised using the R package ggtree. [393] Mutation analysis was performed using Nextclade version 1.6.0 (2021-08-31). [394] Patristic root to tip values, molecular clock analyses and root placement were determined using TreeTime (open-source package for implementation in python for Maximum Likelihood Phylodynamic Analysis) [395] with assistance from Dr Golubchik.

5.2.6 Minor variant calling

Analysis of alternative or minor allele frequency was limited to sequences determined by Illumina short read sequencing (sequences derived from both veSEQ and ARTIC protocols used). [396] To limit false detections, genomes with at least 40% of the reference genome reconstructed at consensus level were used, and a conservative approach was employed whereby iSNVs were only included if they occurred at positions supported by at least 100 reads, with a minimum MAF of 5%. Cumulative minor allele frequency (cMAF) was calculated as the combined proportion of all alternative non-consensus bases at a given position.

5.2.7 Statistical analysis

Lineage, vaccination, and symptom status were compared with Fisher's exact tests. Comparison of primary symptoms was performed using Pearson's chi-square test.

Infection minimum Ct values and duration of NAAT positivity by lineage and vaccine received were compared using Wilcoxon rank sum tests, multiple tests were corrected for with Bonferroni correction.

Vaccine efficacy was calculated as 1-relative risk (ChAdOx1 nCoV-19 vs MenACWY groups)*100 and confidence intervals were calculated using the gsDesign package for R. [397]

Analyses of phylogenetic signals were performed in R using the phylogenetic trees constructed above and the metadata as described. The *D* statistic measures phylogenetic signal for a binary trait: "The *D* statistic is equal to 1 if the observed binary trait has a phylogenetically random distribution across the tips of the phylogeny and to 0 if the observed trait is as clumped as if it had evolved by Brownian motion under [the] threshold model". [398] The *D* statistic was calculated in R using the 'phylo.d' function of the 'caper' (Comparative Analyses of Phylogenetics and Evolution in R) package with 1000 permutations. [399]

Genome coverage, amino acid and nucleotide substitutions by whole genome and gene were performed by Wilcoxon rank sum testing. Comparison between the proportions of control and vaccine participants with specific amino acid substitutions was performed with Fisher's exact tests. Median MAF and cMAF values were compared using Wilcoxon rank sum tests.

5.3 RESULTS

5.3.1 Lineage, geography, and molecular clock

Between 30th May 2020 and 30th June 2021, 1130 participants tested positive for SARS-CoV-2 on at least one nasopharyngeal swab from the 10,812 participants who were originally recruited. From these 1130 participants, 2251 positive swabs in total were returned. Of these positive swabs, 2057 were processed from weekly self-swabs or initial symptomatic visits and therefore were potentially available for sequencing. Of these, at least 1494 (72.6%) had cycle threshold values of 30 or below (Ct values were not available for swabs from symptomatic visits). 1241 swabs (60% of the total taken from symptomatic visits or self-swabs, 83% of those with known Ct value of 30 or below) from 766 participants had sequencing attempted. The remaining 816 samples were not able to be retrieved for sequencing. Of those samples that had sequencing attempted, 628 (50.6%) failed to return a lineage either due to low viral load or inadequate genome coverage attained.

Table 5.1: Sequencing and protocols used to obtain sequences from COV002 participants.

		Protocol		**Total attempted (%)**	**Total lineage data attained (%)**
		ARTIC	veSEQ		
Platform	**Illumina**	175	819	994 (80)	425 (69)
	Oxford Nanopore	207	0	207 (17)	188 (31)
	Unspecified	40	0	40 (3)	0
	Total	422	819	1241	613

From 1241 samples processed, sequences were obtained from 994 (80%) using Illumina, with 425 (43%) having sufficient coverage to define lineage using this method (Table 5.1). Of

the 207 that were sequenced using Oxford Nanopore, 188 (91%) produced lineage level data. From both methods 613/1241 (49%) returned a lineage. Three sequences from three separate individuals demonstrated more than one viral lineage and were excluded from further analyses as these may have represented contamination, two of these individuals had lineage results from alternative samples which were included. Two sequences were identified as B.1.1.7 but had identified S genes on PCR and were excluded. A total of ten sequences from three individuals were excluded as these samples came from infection episodes after the initial episode.

Lineages were then matched to the dataset of blinded participants who had received two doses of either control or ChAdOx1 nCoV-19 vaccine (n=616, see chapter 4 for construction of this dataset) with NAAT positive swabs dated from 14th September 2020 to 19th June 2021. From this dataset 245 (40%) participants had a lineage result from at least one swab.

Of those that had a lineage result (n=245) the most common lineages from study participants were B.1.177 (n=129, 52.6%) and B.1.1.7 (85, 35%) which accounted for 87% of the dataset (Table 5.2). Other than B.1.1.7 (Alpha) there were two other variants of interest or concern identified in our dataset by sequencing, P.1 (Gamma, n=1) and B.1.525 (Eta, n=1). It is likely that the latest infection amongst the included participants was caused by the B.1.617 (Delta) variant due to S gene detection at PCR testing and the prevailing epidemiology in the UK in June 2021 (approximately 95% of sequences from 7-21 June 2021 were identified as Delta) [400] but the sample was not retrieved for sequencing for this participant. The consensus phylogeny is displayed in Figure 5.1.

Table 5.2: SARS-CoV-2 lineages identified from consensus sequences from COV002 participants. *B1 other lineages include 6 participants with B.1.258 (5 controls and 1 vaccinated), 4 participants with B.1 identified (3 control, 1 vaccinated), 3 controls with B.1.1.311, 4 participants with B.1.1.37 (2 controls and 2 vaccinated), 2 participants with B.1.160 (1 control and 1 vaccinated), 1 control each with B.1.1.15, B.1.1.170, B.1.1.98, B.1.1.218, B.1.1.236, B.1.1.307, B.1.2, B.1.235, B.1.260, B.36.17.

	Control	**Vaccinated**
Lineage	n=397	n=219
B.1.177	109	20
B.1.1.7	58	27
B.1 other lineages*	24	5
B.1.525	1	0
P.1	0	1
Sequence attempted, no lineage	93	83
Isolates not retrieved	112	83

For analysis the dataset was then restricted to individuals who had received both doses of study intervention at least 14 days previously and had not yet been unblinded to study intervention at the time of their first positive swab leaving a dataset of 616 participants with infections starting from 14th September 2020 until 19th June 2021. Cases of B.1.177 were evident from 14th September 2020 to 14th January 2021 whilst the earliest sequencing-confirmed B.1.1.7 case identified in this dataset was 25th November 2020 in Glasgow. The geographical distribution of cases across time is shown in Figure 5.2. Sequences from throughout the tree were seen across study sites (Figure 5.3). Cases of both B.1.177 and B.1.1.7 occurred across all study sites with nearly half of the cases of confirmed B.1.1.7 occurring at London and Northwick Park sites (49% (42/85) of all confirmed B.1.1.7 cases). B.1.1.7 became the dominant identified lineage in this dataset from January 2021 until April 2021 (Figure 5.2). Other lineages were sporadically identified during the period.

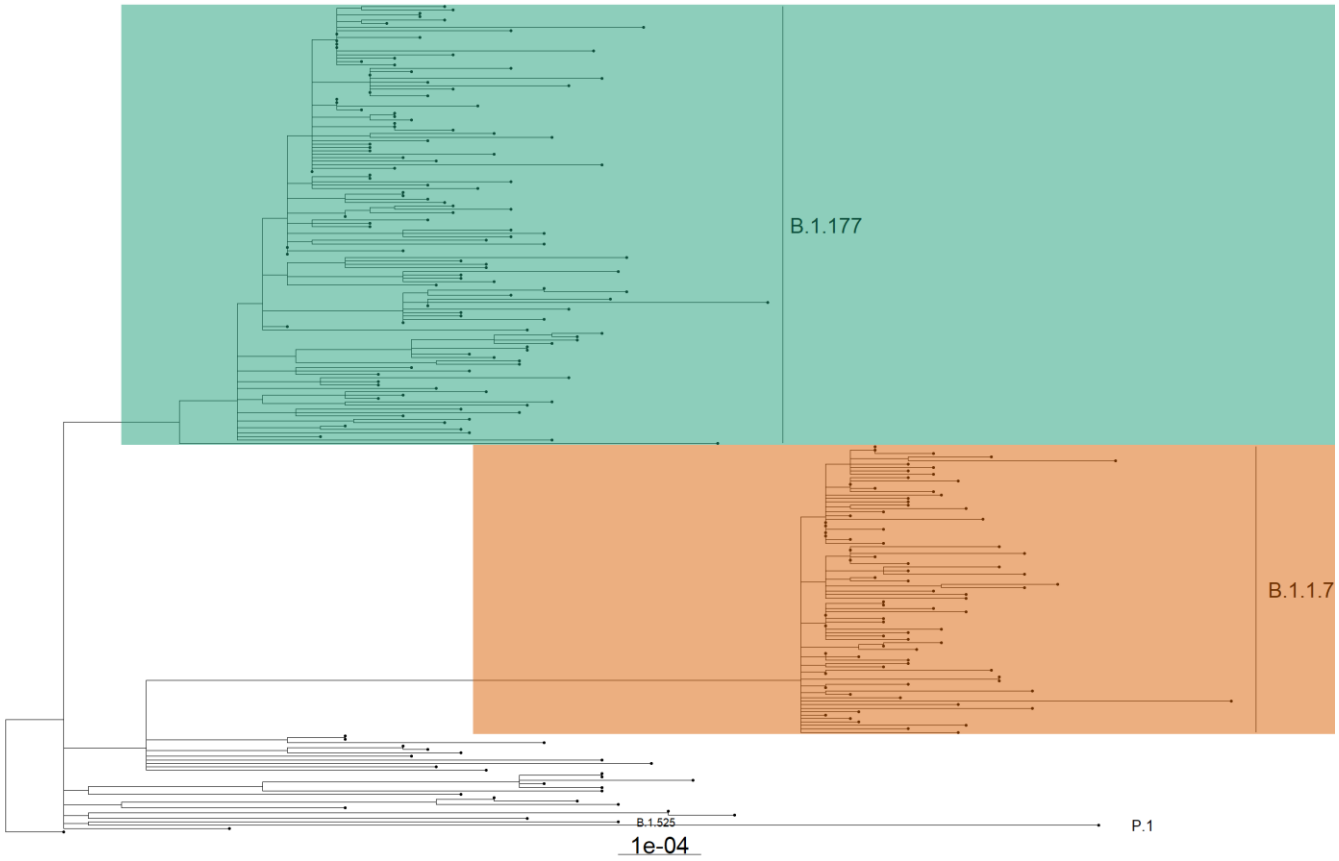


Figure 5.1: Maximum likelihood consensus phylogenetic tree of SARS-CoV-2 genomes from blinded COV002 participants (n=240). Rooted by Wuhan-1 NC_045512.2_ Only genomes with at least 40% coverages are included, three sequences were of low quality and were removed due to >50% ambiguity.

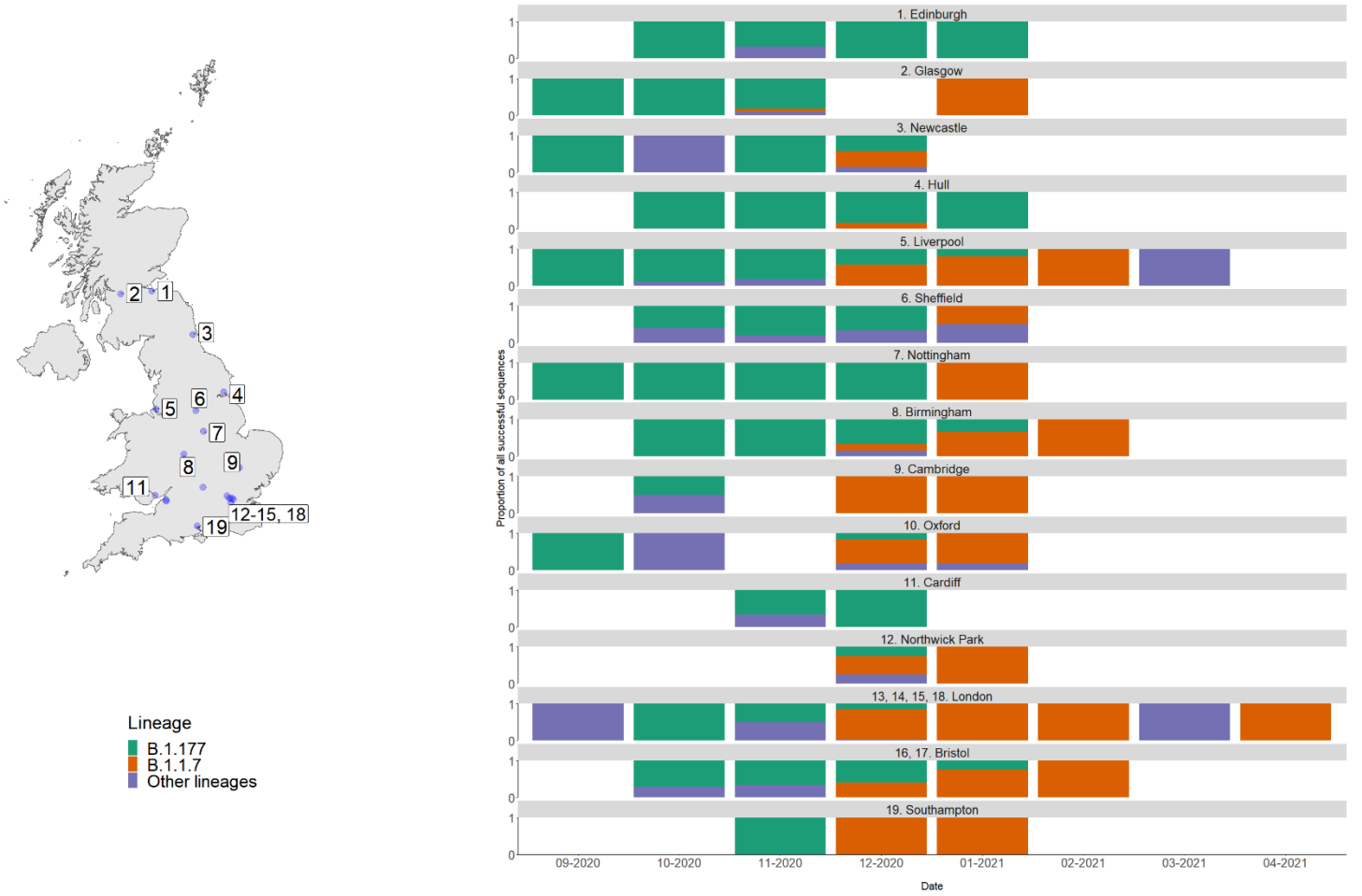


Figure 5.2: Distribution of lineages across time from study sites. Cases displayed as a proportion of sequences for that location for that calendar month. Four London locations (UCLH, Hammersmith/Imperial, Guy's and St George's) and two Bristol locations (North Bristol and Bristol) combined due to geographical proximity as 'London' and 'Bristol' respectively.

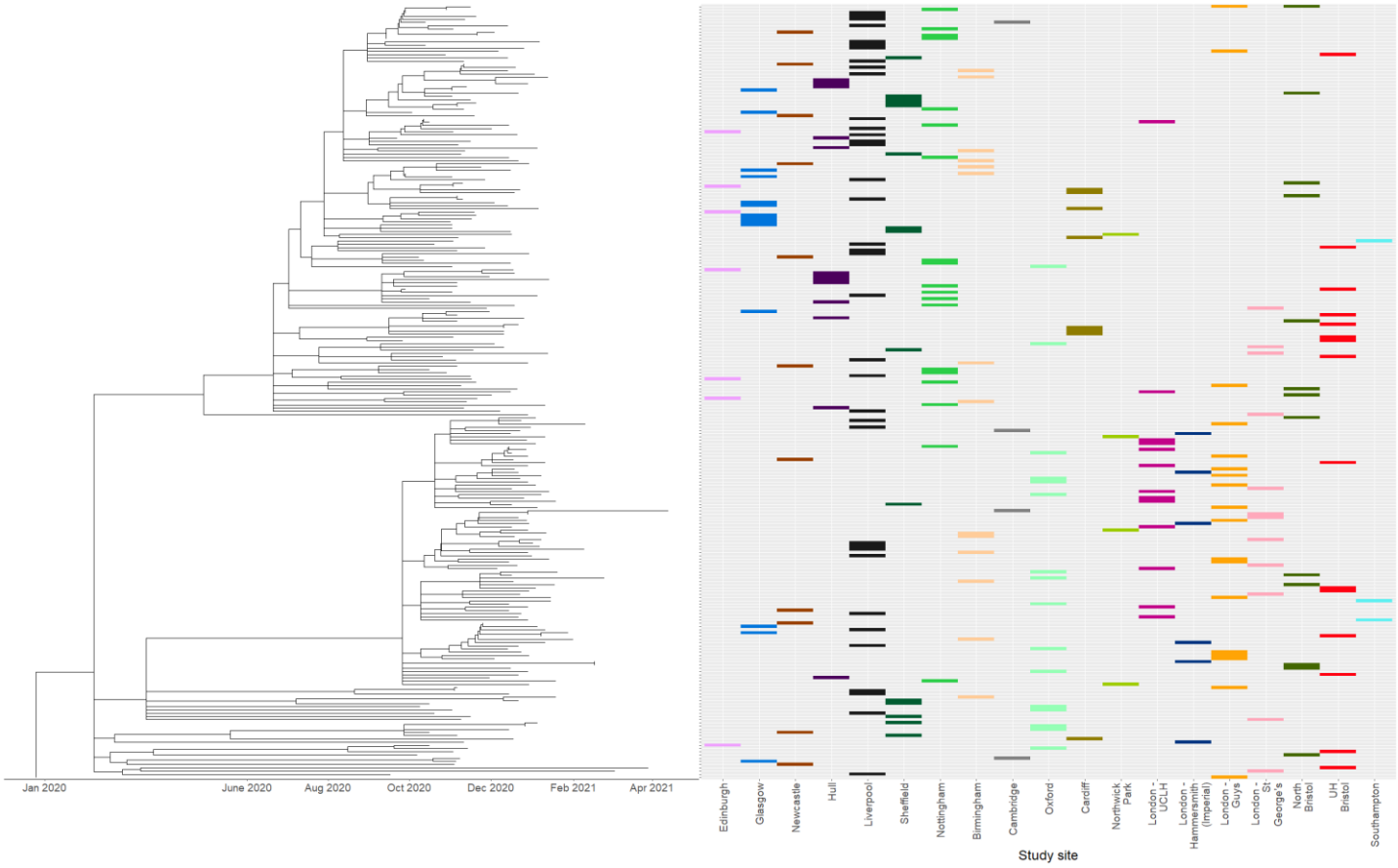


Figure 5.3: Time-scaled consensus phylogenetic tree on left of plot with locations of sequence collection on right. Consensus phylogeny coloured by study site. From blinded COV002 participants (n=240). Rooted by Wuhan-1 NC_045512.2_ Only genomes with at least 40% coverages are included, three sequences were of low quality and were removed due to >50% ambiguity.

Phylogenetic divergence (root to tip distance) was plotted over time for B.1.177 and B.1.1.7 (Figure 5.4). The evolutionary rate of B.1.177 was 4.8 x 10^{-4} nucleotide changes/per site/per year (equivalent to ~15 changes per genome per year) compared with B.1.1.7 with 3.8 x 10^{-4} nucleotide changes/per site/per year (equivalent to ~12 changes per genome per year).

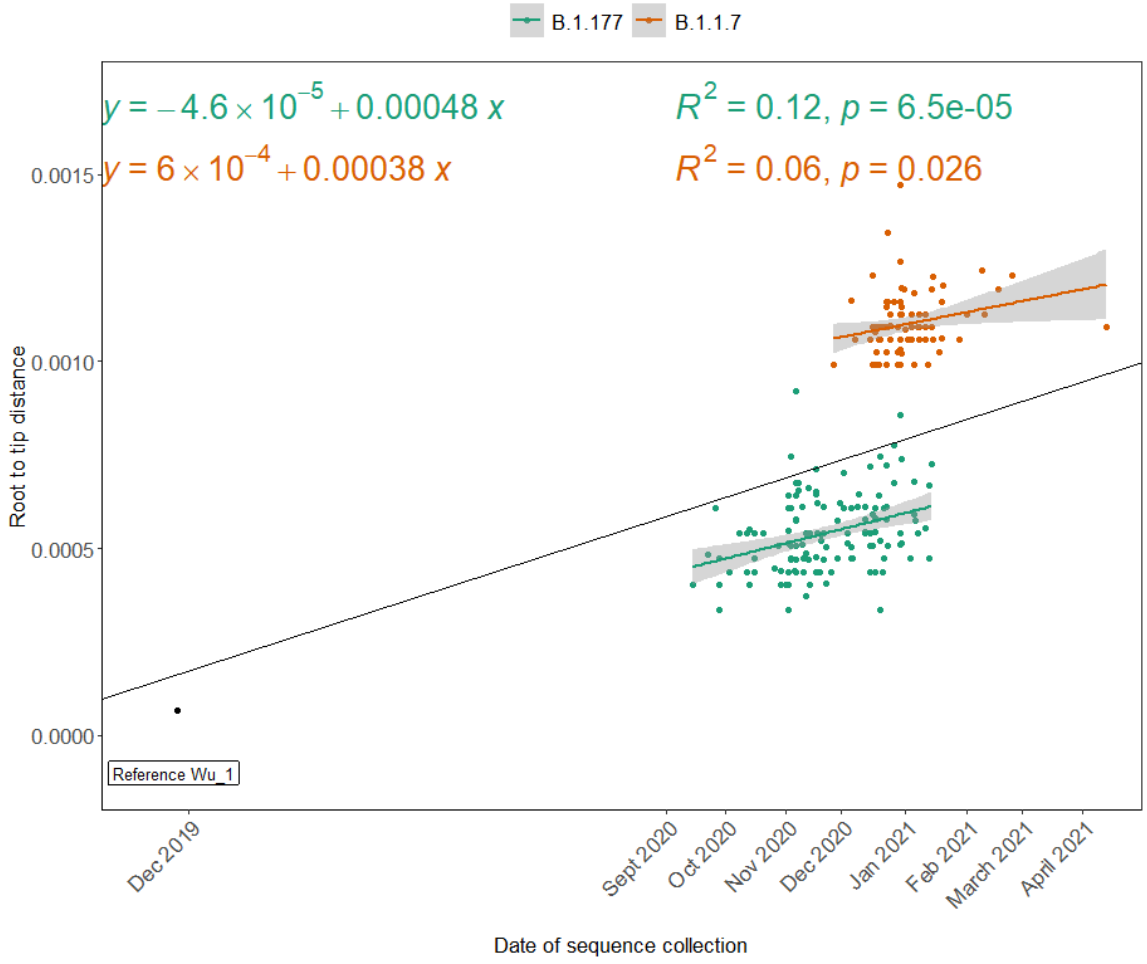


Figure 5.4: Root to tip distance plotted for B.1.177 and B.1.1.7 sequences. B.1.177, n=128, root to tip distance values missing for 1 participant, B.1.1.7 n=82, root to tip distances missing for 2 participants. Black line represents overall molecular clock as calculated by TreeTime (see methods).

A linear model was performed to compare these rates and found that the estimated rate for B.1.1.7 was significantly faster than for B.1.177. The overall model was significant (adj.

R^2 = 0.89, (F(1,208) =1612, p<0.001, β coefficient = 0.00056 (95% CI: 0.00053, 0.00059), p<0.001), though the difference was small in magnitude.

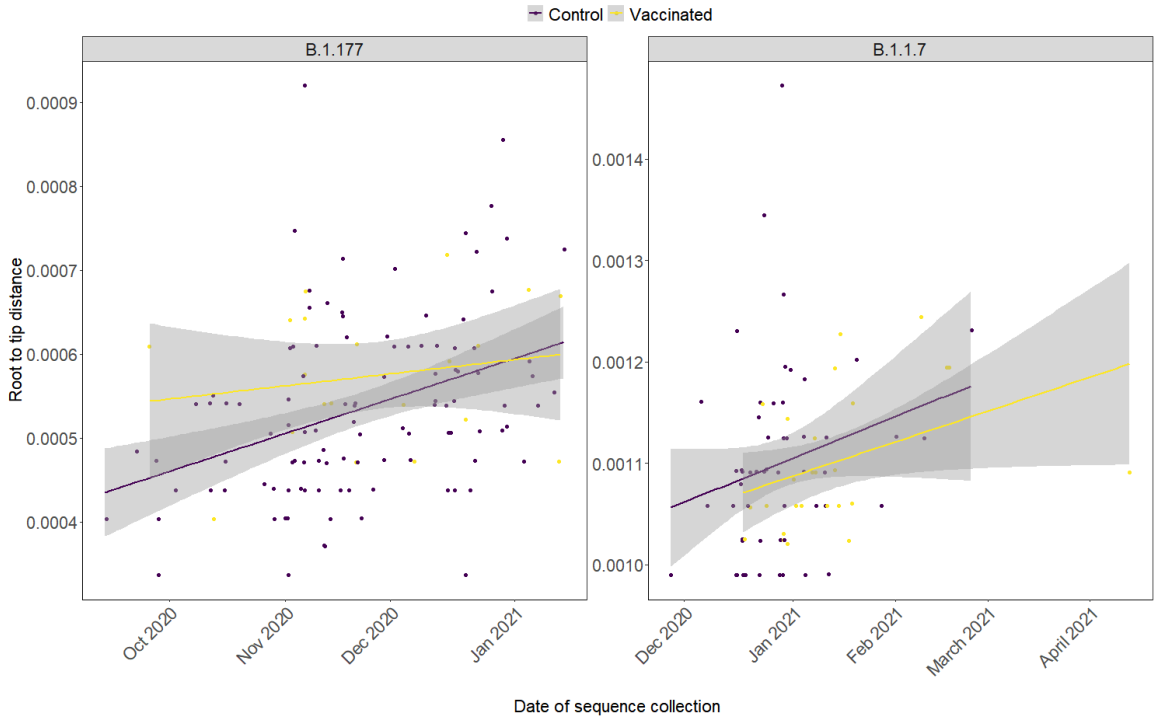


Figure 5.5: Phylogenetic divergence of samples relative to the common ancestor (root to tip distance), by lineage and vaccination status.

Multiple linear regression was used to test if vaccination status predicted phylogenetic divergence (root to tip value) separately by lineage for B.1.177 and B.1.1.7 sequences (Figure 5.5). For B.1.177 the overall regression was statistically significant (Adj. R^2= 0.42, F(2, 234)=60.24, p<0.001), and as expected sequence collection date significantly predicted root to tip value (β=5.05x10^{-6}, p<0.001) however vaccination status did not (β=5.4x10^{-3}, p=0.7). There was no evidence of an interaction between vaccination status and sequence date. For B.1.1.7 a regression model including vaccination status was not significant. When vaccination status was removed, a model for sequence collection date and root to tip value

was significant (Adj. R^2=0.04, F(1, 80)=5.122, p<0.05) and as anticipated, sequence collection date significantly predicted root to tip value ($\beta=1.01x10^{-6}$, p<0.05). A higher deviation from the molecular clock was evident for B.1.1.7 (Figure 5.6) as reported in the original observation of B.1.1.7. [35,387]

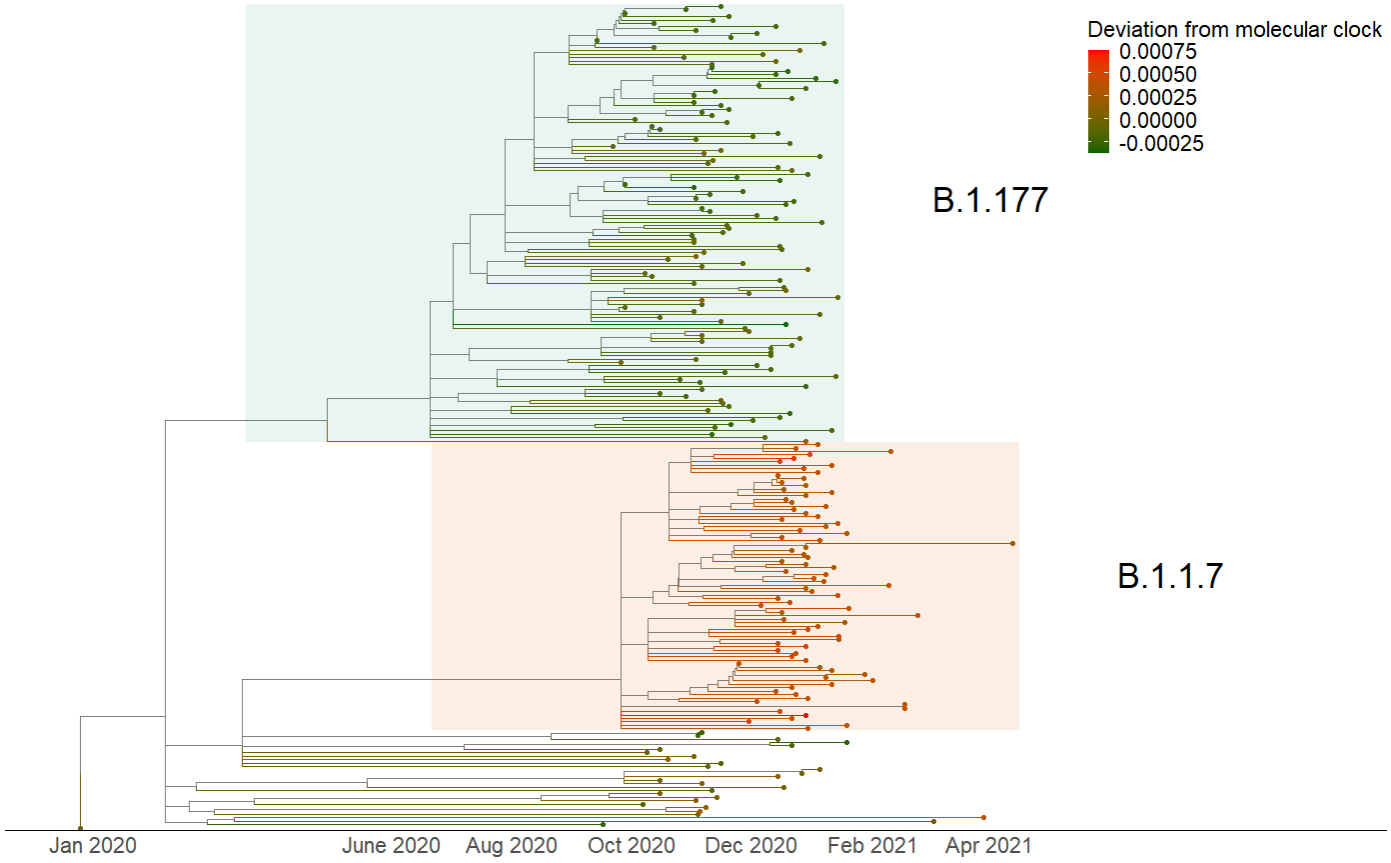


Figure 5.6: B.1.1.7 sequences show increased deviation from molecular clock. Time scaled phylogenetic tree from blinded COV002 participants (n=240). Rooted by Wuhan-1 NC_045512.2_ Only genomes with at least 40% coverages are included, three sequences were of low quality and were removed due to >50% ambiguity. Rooted by Wuhan-1 NC_045512.2. N=240.

5.3.2 Demographics by sequencing result

Participants in each of the sequencing result groups were similar in terms of sex, ages, ethnicity, and comorbidity (Table 5.3).

Table 5.3: Details of participants by sequencing result. HCW= healthcare worker, LD = low dose, SD = standard dose

	n	**B.1.177** n=129	**B.1.1.7** n=85	**Other lineages** n=31	**Sequence attempted** n=176	**Not retrieved** n=195
Female Sex, n (%)	616	80 (62%)	49 (58%)	17 (55%)	101 (57%)	119 (61%)
Age, median (IQR)	616	43 (30-51)	40 (31-49)	44 (30-53)	44 (35-56)	40 (30-51)
White ethnicity, n (%)	616	121 (94%)	80 (94%)	26 (84%)	165 (94%)	179 (92%)
HCW, n (%)	616	107 (83%)	49 (58%)	16 (52%)	97 (55%)	138 (71%)
Any comorbidity, n (%)	616	41 (32%)	29 (34%)	12 (39%)	55 (31%)	50 (26%)
Controls, n (%)	397	109 (84%)	58 (68%)	25 (81%)	93 (53%)	112 (58%)
Vaccinees, n (%)	219	20 (16%)	27 (32%)	6 (19%)	83 (47%)	83 (43%)
LD/LD	7	0	3	0	3	1
LD/SD	54	7	7	1	19	20
SD/SD	158	14	17	5	61	61
Time since second dose, weeks, median (IQR)	616	10.7 (7-13.7)	17.3 (15.0-19.7)	9.9 (5.9-11.6)	12.4 (8.6-17.0)	9.7 (7-14.4)
Dose interval, weeks, median (IQR)	616	12 (9-13)	10 (9-12)	10 (8-13)	10 (6-13)	11 (8-13)

Most B.1.177 infections occurred in healthcare workers (107/129, 83%) whilst only 58% (49/85) of B.1.1.7 infections were in healthcare workers likely to reflect the increased rates of transmission within the community as B.1.1.7 became dominant. B.1.1.7 infections occurred a median of 17.3 weeks (15-19.7 weeks) after second vaccine dose compared with a median of 10.7 weeks (7-13.7 weeks) for B.1.177 infections. The interval between first and second vaccine doses was similar for all lineage groups.

When the time difference of infections caused by different lineage results were compared by vaccination status, there was no difference in dates of the first positive swab for any lineage group by Wilcoxon rank sum testing except for infections caused by B.1.1.7 where infections in vaccinees were significantly later than for control participants ($p<0.01$, Figure 5.7).

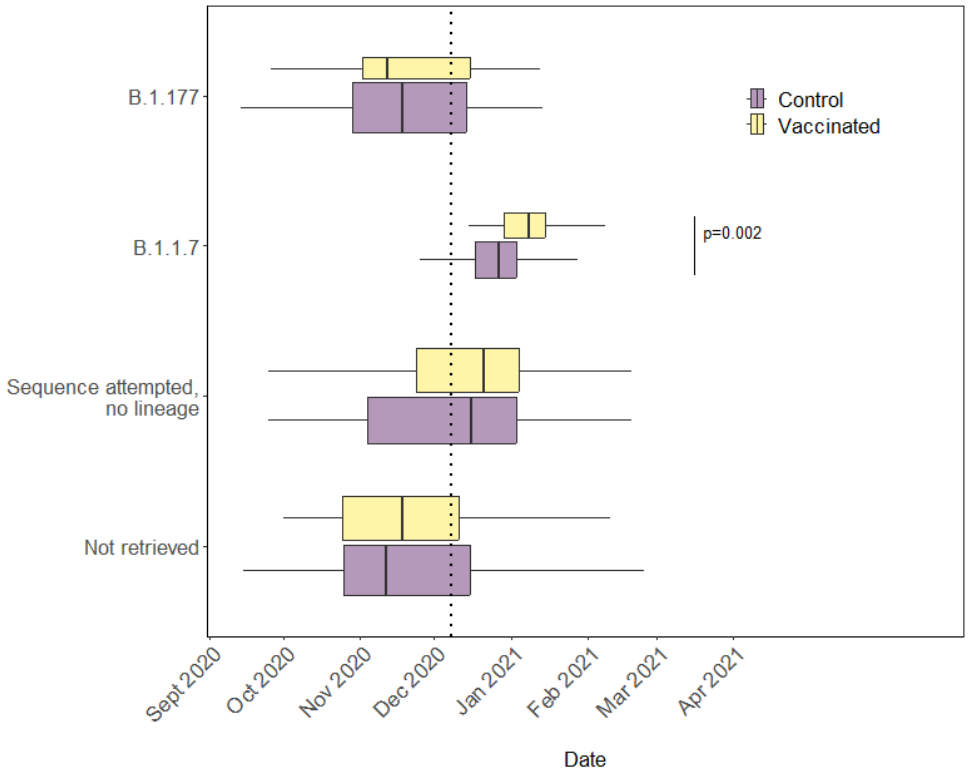


Figure 5.7: Time distribution of different sequencing result groups. Other lineages not shown. Boxplots are sized proportional to group number. The midlines of the box plots show median date of

first positive swab for each vaccination and lineage group, and the lower and upper bounds of the boxes correspond to the 25th and 75th percentiles. The whiskers show the most extreme value no further than 1.5 x IQR above or below the 75th or 25th percentile. All statistical comparisons were performed and adjusted using Bonferroni correction, p values only shown when adjusted p values <0.05. Dotted black line represents week when UK national vaccination programme began (8th December 2020).

5.3.3 Lineage, vaccine, and symptom status

For the following analyses cases of the most common alternative lineage, B.1.177 were compared with B.1.1.7 cases. Most cases for which a lineage was identified were symptomatic and the proportion of symptomatic cases were similar between the two lineages in the control and vaccinated participants (Table 5.4). There was no difference in the proportion of cases with primary symptomatic infection (experience of fever $\geq$37.8°C, cough, shortness of breath, anosmia or ageusia during episode) between lineages (B.1.177 controls: 78/109 (72%) vs B.1.1.7 controls: 43/58 (73%), p = 0.7), B.1.177 vaccinated: 16/21 (76%) vs B.1.1.7 vaccinated: 19/27 (70%), p=0.5 by Pearson's chi-squared test). Amongst symptomatic participants there was no difference in WHO severity score between B.1.177 and B.1.1.7.

Table 5.4: Comparison of symptom status by lineage and vaccine status Primary symptoms include fever >37.8°C, cough, shortness of breath, anosmia and ageusia. Comparison of symptom status by Fisher's exact test. Comparison of primary symptoms by Pearson's chi-square test.

	Controls			**Vaccinated**		
	B.1.177 N=109 (%)	B.1.1.7 N=58 (%)	p value	B.1.177 N=20 (%)	B.1.1.7 N=27 (%)	p value
Symptomatic	86 (79%)	48 (83%)	>0.9	17 (81%)	21 (78%)	0.064
Asymptomatic	7 (6.4%)	3 (5.2%)		2 (10%)	0	
Unknown	16 (15%)	7 (12%)		1 (5%)	6 (22%)	
Primary symptom experienced	78 (72%)	43 (74%)	0.7	16 (80%)	19 (70%)	0.5

5.3.4 Viral load

The infection minimum Ct, as a surrogate for peak viral load, for symptomatic participants for whom a lineage was obtained was compared with participants for whom a lineage result was not available (Figure 5.8). The infection minimum Ct value in participants for whom a lineage was obtained was significantly lower (median 17.57, IQR: 14.96-23.67) and therefore the viral load was higher compared with those for whom sequencing was attempted but no lineage result was obtained (median 23.57, IQR: 16.92-30.76, Wilcoxon rank sum adj. $p<0.001$). For those participants for whom a sequence was not attempted (no positive swab from their infection was retrieved) the infection minimum Ct was not significantly different to those for whom no lineage result was achieved.

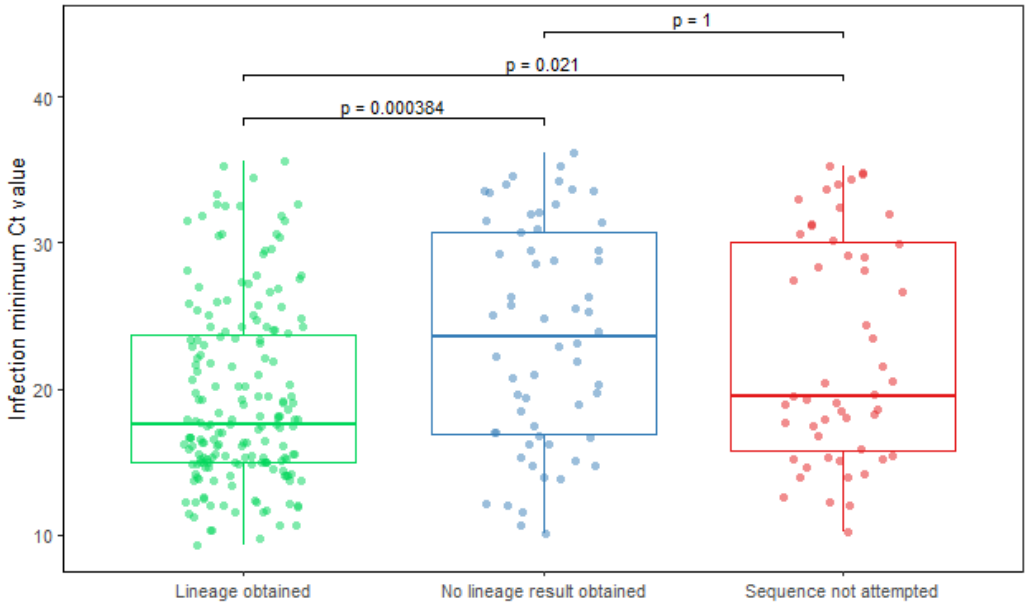


Figure 5.8: Infection minimum Ct value by lineage result. Symptomatic participants who had Ct values available, n=296. Infection minimum Ct values were not available for 60 participants and therefore are not included in this figure. Sequence not attempted = no sample retrieved. Each point represents a single participant. The midlines of the boxes show medians, and the lower and upper bounds of the boxes correspond to the 25th and 75th percentiles. The whiskers show the most extreme value no further than 1.5 x IQR above or below the 75th or 25th percentile. Groups were compared using Wilcoxon rank sum testing with Bonferroni correction for multiple tests.

The infection minimum Ct for each participant (the lowest Ct value for any gene target from all swabs obtained during the participant's infection episode) was then compared between lineage groups and by vaccination status for participants who had symptomatic infection (Figure 5.9).

Infection minimum Ct values for B.1.1.7 cases were no different for B.1.177 cases for controls (B.1.177 controls median: 17.59, IQR 15.09-24.12 vs B.1.1.7 controls median: 16.24, IQR 13.41-21.67) or for vaccinees (B.1.177 vaccinated median: 21.25, IQR 15.8-24.0 vs B.1.1.7 vaccinated median: 16.5, IQR 15.34-21.83) by Wilcoxon rank sum test (adj. p value = 1 for both comparisons). When all participants regardless of vaccine status were

considered there remained no significant difference in the infection minimum Ct value for B.1.177 and B.1.1.7 (B.1.177 median: 17.83, IQR 15.09-24.12 vs B.1.1.7 median: 16.31, IQR 14.38-21.78, adj. p value =1).

Within the vaccinee group, infection minimum Ct values from swabs from participants with sequence attempted but no lineage result (31.36, IQR 22.39-33.53) were significantly higher than for participants with confirmed B.1.1.7 infections (B.1.1.7: 16.5, IQR: 15.3-21.83, Wilcoxon rank sum adj. p=0.0057). This was not seen in the control group where there were no significant differences between sequencing result groups.

When directly comparing controls and vaccinees by sequencing result group, there was no significant difference in infection minimum Ct value between controls and vaccinees except for those cases when sequence was attempted but no lineage result was obtained where controls had significantly lower infection minimum Ct values (therefore higher viral load) than vaccinees (controls: 20.78, IQR 16.18-26.28, vs vaccinees: 31.36, IQR 22.39-33.53, Wilcoxon rank sum adj. p=0.012).

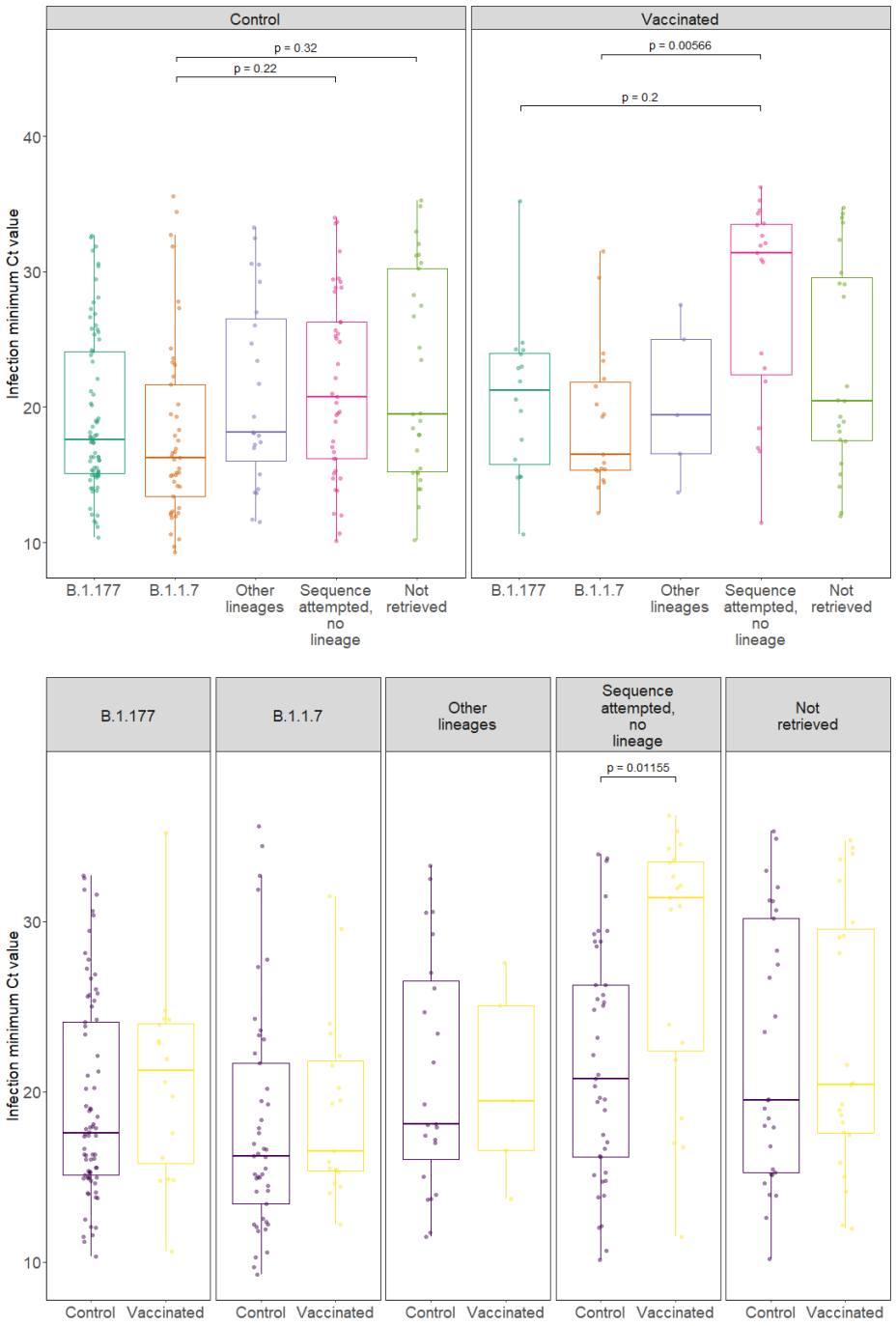


Figure 5.9: Infection minimum Ct value by sequencing result and vaccination status. Symptomatic participants who had Ct values available, n=296. Infection minimum Ct values were not available for 60 participants and therefore are not included in these figures. 'Other lineages' includes all identified lineages other than B.1.177 and B.1.1.7. Not retrieved = no swab from infective episode was retrieved for sequencing. Each point represents a single participant. The midlines of the boxes show medians, and the lower and upper bounds of the boxes correspond to the 25th and 75th percentiles. The whiskers show the most extreme value no further than 1.5 x IQR above or below the 75th or 25th percentile. Groups were compared using Wilcoxon rank sum testing with Bonferroni correction for multiple tests. All comparisons were performed but for clarity comparisons with an adjusted p value of 1 are omitted from the plot.

5.3.5 Duration

The duration of NAAT positivity was compared between sequencing result groups and by vaccination status for participants who had symptomatic infection (Figure 5.10).

There were no significant differences in the duration of NAAT positivity between sequencing result groups for the vaccinated group (top right panel, Figure 5.10). In contrast, for the control group all sequencing groups had a longer duration of positive swabs compared with those where no sample was retrieved (top left panel, Figure 5.10). When directly comparing symptomatic controls and vaccinees within lineage group, the duration of positive swabs was longer for controls compared with vaccinees for B.1.177 (controls median duration: 2 weeks, IQR 1-3 weeks, vs vaccinated median duration 1 week, IQR 1-2 weeks, Wilcoxon rank sum adj. p=0.019), but not for B.1.1.7 (controls median duration:2.5 weeks, IQR 1.75-4 weeks, vs vaccinated median duration 1 week, IQR 1-3 weeks, Wilcoxon rank sum adj. p=0.106).

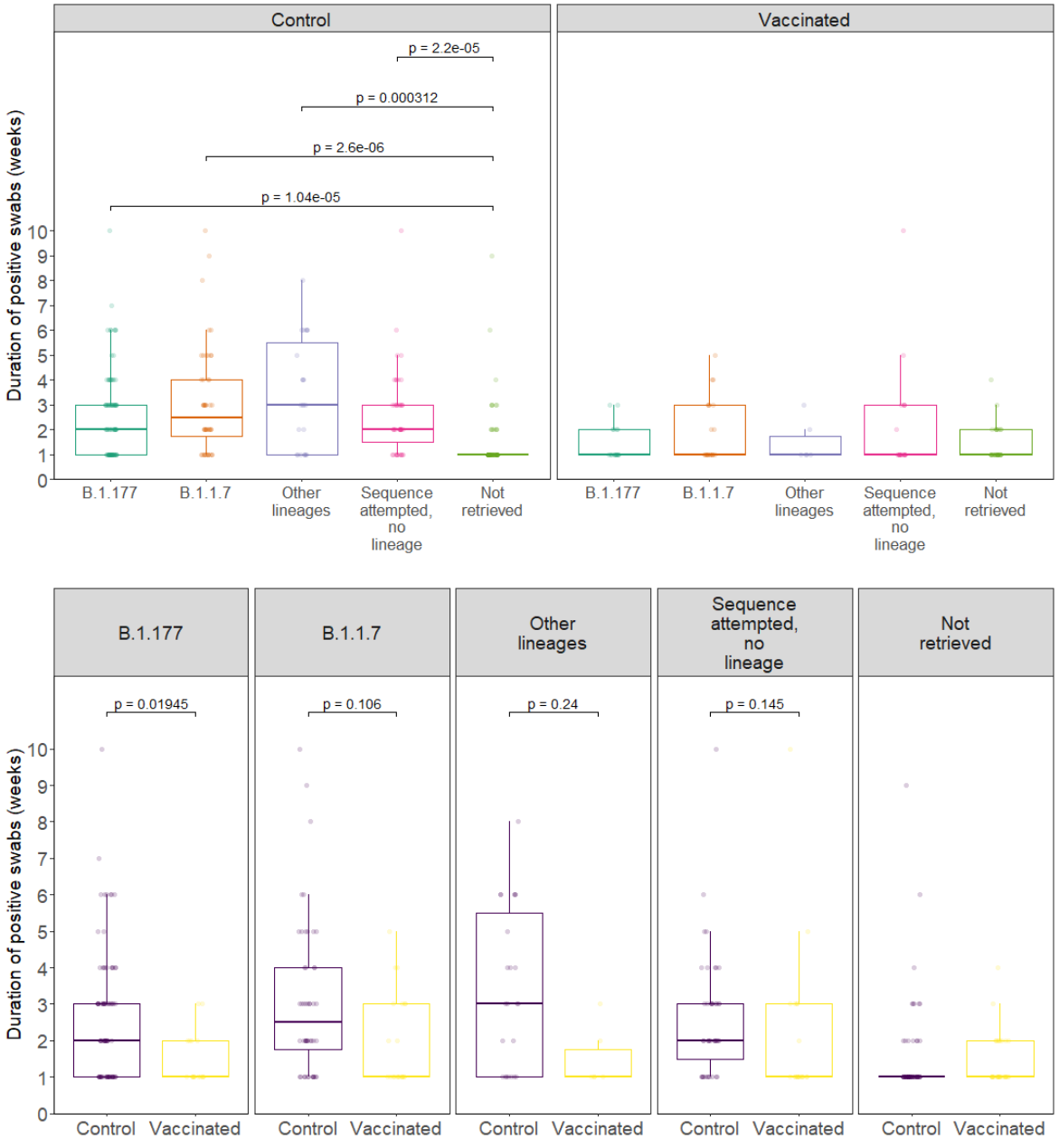


Figure 5.10: Duration of positive swabs for symptomatic participants by sequencing result and vaccination status (n=356). 'Other lineages' includes all identified lineages other than B.1.177 and B.1.1.7. Not retrieved = no swab from infective episode was retrieved for sequencing. Each point represents a single participant. The midlines of the boxes show medians, and the lower and upper bounds of the boxes correspond to the 25th and 75th percentiles. The whiskers show the most extreme value no further than 1.5 x IQR above or below the 75th or 25th percentile. Groups were compared using Wilcoxon rank sum testing with Bonferroni correction for multiple tests. All comparisons were performed but for clarity comparisons with an adjusted p value of 1 are omitted from the plot.

5.3.6 Efficacy

The consort diagram for participants included in the efficacy analyses is shown in Figure 5.11 and for the lineages identified by participant in Figure 5.12. Vaccine efficacy for ChAdOx1 nCoV-19 was calculated using sequencing result and gene target data for different symptom groups (Table 5.5, Table 5.6).

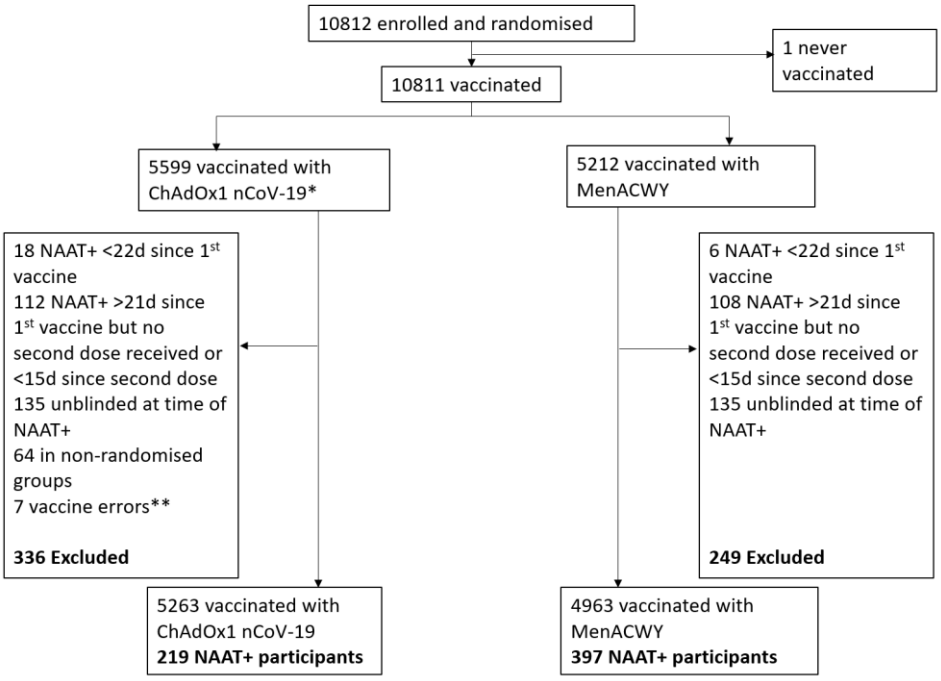


Figure 5.11: Consort diagram for participants in the COV002 trial included in efficacy analyses. NAAT+: positive nucleic acid amplification test (NAAT) *includes all participants who received at least one dose of ChAdOx1 nCoV-19. **Vaccine errors were those participants who received a different vaccine for their first and second dose, 4 participants received a ChAdOx1 nCoV-19 vaccine for their first dose and MenACWY for their second dose, and three received MenACWY for their first dose and ChAdOx1 nCoV-19 for their second dose.

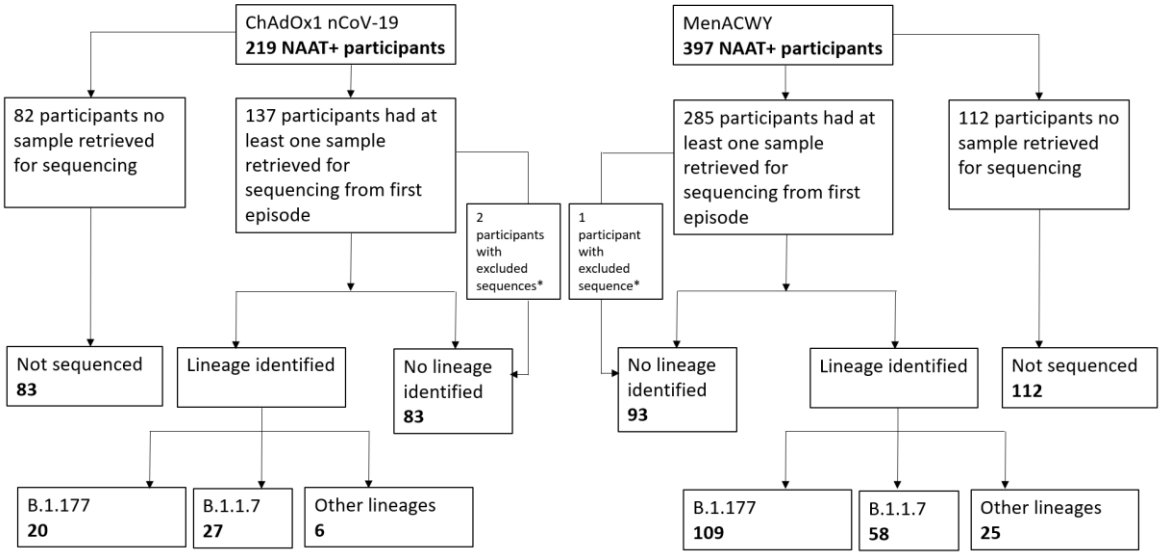


Figure 5.12: Lineage outcomes for NAAT positive participants. Of those that were excluded 3 participants had only one sample for sequencing and the results were excluded due to mixed lineages identified in sequencing data (1 participant in the vaccinated group) and lack of SGTF on PCR with sequencing identifying B.1.1.7 (1 participant in each arm). Other lineages as outlined in the legend of Table 5.2.

Vaccine efficacy against primary symptomatic COVID-19 infection caused by sequence confirmed B.1.1.7 was 58.3% (95% CI 29.1-75.5) and 80.7% (95% CI 67.1-88.6) for cases caused by B.1.177 (Table 5.5). Vaccine efficacy was higher for B.1.177 than for B.1.1.7 in all symptom groups and for any NAAT+ infection. Vaccine efficacy against primary symptomatic infection using S gene target failure as a proxy for B.1.1.7 (during a specific timeframe, see Table 5.6) was comparable to the figure obtained using sequencing results (56.2% (95% CI 30.7-72.3) but the efficacy for any NAAT positive infection was lower for SGTF samples (41.3% (95% CI 18.4-57.8) than for confirmed B.1.1.7 cases

As most symptomatic individuals experienced at least one primary symptom there was little difference between the vaccine efficacy for primary symptomatic infections versus any symptomatic infections by sequencing result. Vaccine efficacy was lower in all symptom

groups for those cases in whom sequencing was attempted but no sequence result was obtained or where samples were not retrieved for sequencing.

No sequence result was available for the large majority (83%) of asymptomatic or unknown infections. Vaccine efficacy for asymptomatic infections caused by confirmed B.1.177 (83.6% (95% CI 54.7-94.1) was similar to efficacy for primary symptomatic infections caused by B.1.177 (80.7% (95% CI 67.1-88.6). However, in the small number of asymptomatic infections confirmed as being caused by B.1.1.7 (n=16) efficacy estimates had wide confidence intervals which spanned zero (43.4% (95% CI -49.6-78.6)) in contrast to confirmed asymptomatic infections caused by B.1.177 for which there was evidence of efficacy (83.6% (95% CI 54.7-94.1). There was no evidence of vaccine efficacy for asymptomatic cases where no sample could be retrieved or where sequencing was attempted and failed.

Table 5.5: Vaccine efficacy against B.1.177 and B.1.1.7 using sequencing data. For sequencing results: 'Other lineages' group includes all identified lineages other than B.1.177 and B.1.1.7, no sequence result is for those cases where a sequence was attempted for at least one swab but a sequence sufficient to assign a lineage was not obtained, not known is where no swab from the infection was retrieved for sequencing.

	Cases	**Vaccinees**	**Controls**	**Vaccine efficacy**	**Lower CI**	**Upper CI**
		5263	4963			
Primary symptomatic (sequencing results)						
B.1.177	94	16	78	80.7	67.1	88.6
B.1.1.7	62	19	43	58.3	29.1	75.5
Other lineages	26	4	22	82.9	52.5	93.8
No sequence result	49	13	36	65.9	36.5	81.7
Not sequenced	74	22	52	60.1	34.8	75.6
Total	305	74	231	69.8	60.9	76.7
Any symptomatic infection (sequencing results)						
B.1.177	103	17	86	81.4	68.9	88.8
B.1.1.7	69	21	48	58.7	31.6	75.1
Other lineages	29	6	23	75.4	41.3	89.7
No sequence result	63	20	43	56.1	26	74
Not sequenced	92	31	61	52.1	26.5	68.7
Total	356	95	261	65.7	56.8	72.8
Asymptomatic or unknown infection (sequencing results)						
B.1.177	26	3	23	87.7	61.6	96.1
B.1.1.7	16	6	10	43.4	-49.6	78.6
Other lineages	2	0	2			
No sequence result	113	63	50	-18.8	-71.5	17.7
Not sequenced	103	52	51	3.9	-40.9	34.4
Total	260	124	136	14.0	-9.3	32.3
Any NAAT positive infection (sequencing results)						
B.1.177	129	20	109	82.7	72.3	89.2
B.1.1.7	85	27	58	56.1	31.1	72
Other lineages	31	6	25	77.4	46.4	90.5
No sequence result	176	83	93	15.8	-12.7	37.2
Not sequenced	195	83	112	30.1	7.5	47.2
Total	616	219	397	48.0	39	55.7

Table 5.6: Vaccine efficacy using SGTF in combination with sequencing data as a proxy for B.1.1.7. SGTF = S-gene target failure. SGTF analyses were restricted to cases that occurred after 1st November 2020 after which SGTF became a good proxy for B.1.1.7.

	Cases	Vaccinees	Controls	Vaccine efficacy	Lower CI	Upper CI
		5263	4963			
Primary symptomatic infection						
Non SGTF	112	21	91	78.2	65.2	86.4
SGTF	82	26	56	56.2	30.7	72.3
Not known	38	9	29	70.7	39.3	85.9
Total	232	56	176	70.0	59.6	77.7
Any NAAT positive infection						
Non SGTF	197	54	143	64.4	51.5	73.9
SGTF	146	56	90	41.3	18.4	57.8
Not known	131	61	70	17.8	-15.4	41.5
Total	474	171	303	46.8	36.1	55.7

5.3.7 Consensus genome analysis

5.3.7.1 Phylogenetic analysis

To explore whether sequences from vaccinees may cluster on the phylogenetic tree the consensus phylogeny (Figure 5.13) was analysed for phylogenetic signal related to vaccination status using the *D* statistic ($D = 0.66$, $p(D>0) < 0.05$, $p(D<1) < 0.05$). The *D* statistic showed a significant phylogenetic signal for vaccination status, indicating that vaccination status was significantly different from random distribution on the phylogeny. One explanation for the association between vaccination status and phylogenetic relatedness demonstrated by the *D* statistic could be the lower efficacy of vaccination against B.1.1.7. Since B.1.1.7 arose late in 2020 and has a greater phylogenetic divergence from the ancestral genome relative to older lineages, an over-representation of vaccinated individuals within this clade would drive the phylogenetic association.

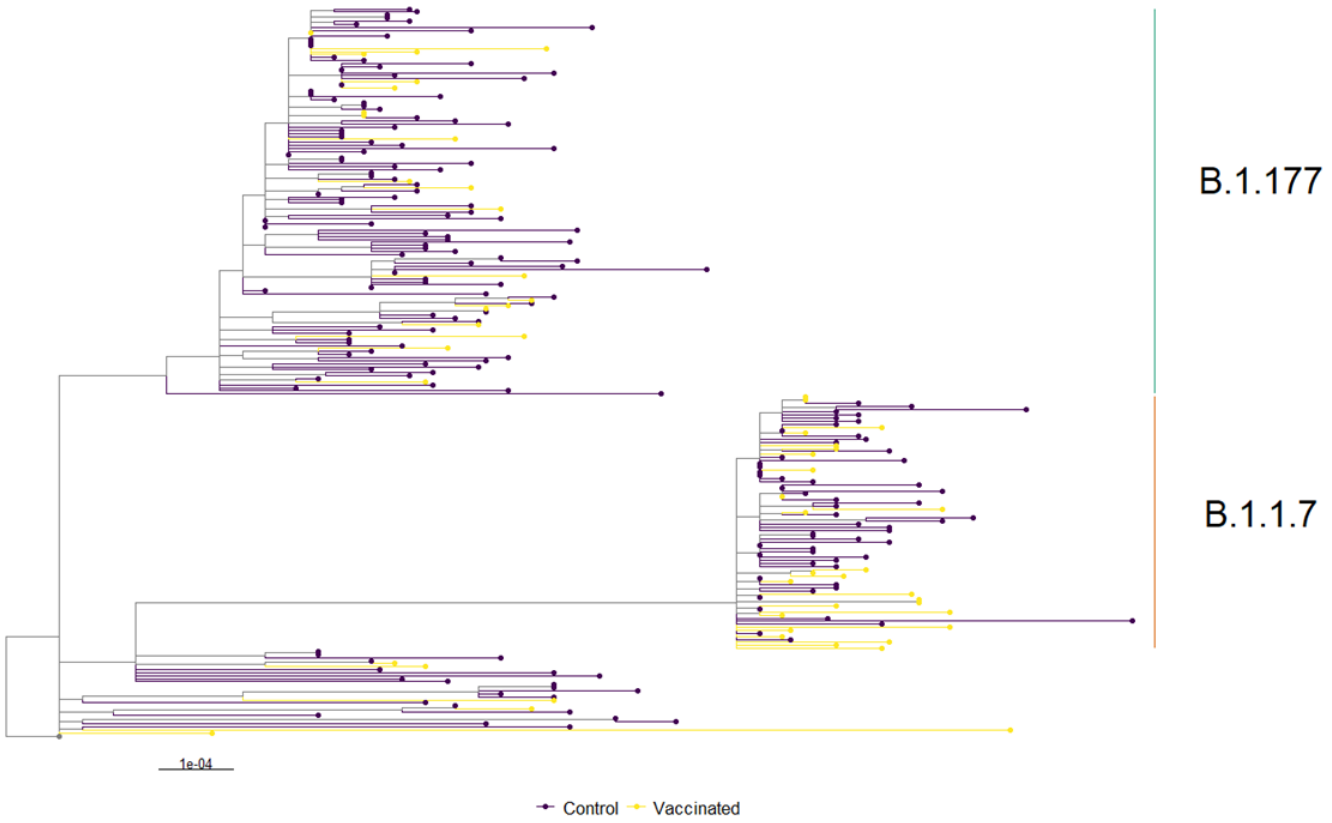


Figure 5.13: Maximum likelihood consensus phylogenetic tree by vaccination status from blinded COV002 participants (n=240). Rooted by Wuhan-1 NC_045512.2_ Only genomes with at least 40% coverages are included, three sequences were of low quality and were removed due to >50% ambiguity.

5.3.7.2 Nucleotide and amino acid substitutions by lineage

To assess whether controls and vaccinees were infected by viruses with different numbers of mutations within a given lineage, nucleotide and amino acid substitutions were analysed for regions of the genome. Overall, there were a very small number of mutations of nucleotides, in keeping with the slow rate of intra-host viral evolution regardless of vaccination status. As anticipated, sequences from the B.1.1.7 lineage demonstrated a greater number of amino acid substitutions in keeping with the known lineage-defining mutations. There were no differences in the absolute number of observed nucleotide or amino acid substitutions within lineage when controls and vaccinees were compared within lineage. This was true for total number of substitutions, the spike protein and RBD, as well as for other regions of the genome (Table 5.7, Table 5.8). When controlled for by gene length for B.1.177 the spike (especially within the NTD) and nucleocapsid genes had the greatest number of amino acid substitutions, ORF10 the greatest number of nucleotide substitutions. For B.1.1.7 ORF8 had the greatest number of both amino acid and nucleotide substitutions.

Table 5.7: Median amino acid substitutions by vaccination status (IQR). Substitutions in Envelope, Membrane, ORF3a, ORF6, ORF7a, ORF 7b, ORF9b, ORF10 not included in the table as median (IQR) = 0 (0,0) across all groups. 1 sequence (B.1.1.7 control) was insufficient for substitution analysis.

		B.1.177				**B.1.1.7**			
	Gene Length	**Controls**	**Vaccinated**	**Controls**	**Vaccinated**	**Controls**	**Vaccinated**	**Controls**	**Vaccinated**
		n=109	**n=20**	**Normalised by gene length**	**Normalised by gene length**	**n=57**	**n=27**	**Normalised by gene length**	**Normalised by gene length**
Total amino acid substitutions	29903	7 (6,9)	8 (6,9)	0.00023	0.00027	21 (19,22)	21 (18,22)	0.00070	0.00070
ORF1a	13218	1 (0,2)	1 (0,3)	0.00008	0.00008	4 (3,5)	4 (3,5)	0.00030	0.00030
ORF1b	8088	1 (1,2)	1 (1,2)	0.00012	0.00012	1 (1,2)	1 (1,2)	0.00012	0.00012
Spike	3822	3 (2,3)	3 (3,3)	0.00078	0.00078	7 (7,7)	7 (7,8)	0.00183	0.00183
RBD	223	0 (0,0)	0 (0,0)	-	-	1 (1,1)	1 (1,1)	0.00448	0.00448
NTD	292	2 (1,2)	2 (1,2)	0.00685	0.00685	0 (0,0)	0 (0,0)	-	-
ORF8	366	0 (0,0)	0 (0,0)	-	-	3 (3,4)	3 (3,3)	0.00820	0.00820
Nucleocapsid	1260	1 (1,1)	1 (1,2)	0.00079	0.00079	4 (4,4)	4 (4,4)	0.00317	0.00317

Table 5.8: Median nucleotide substitutions by vaccination status (IQR). Substitutions in Envelope, ORF3a, ORF6, ORF7a, ORF 7b, not included in the table as median (IQR) = 0 (0,0) across all groups. 1 sequence (B.1.1.7 control) was insufficient for substitution analysis.

	Gene Length	B.1.177				B.1.1.7			
		Controls	Vaccinated	Controls	Vaccinated	Controls	Vaccinated	Controls	Vaccinated
		n=109	n=20	Normalised by gene length	Normalised by gene length	n=57	n=27	Normalised by gene length	Normalised by gene length
Total nucleotide substitutions	29903	18 (17,20)	20 (17,21)	0.00060	0.00067	32 (30, 33)	32 (30,34)	0.00107	0.00107
ORF1a	13218	5 (4,7)	5 (4,9)	0.00038	0.00038	8 (7,8)	8 (7,8)	0.00061	0.00061
ORF1b	8088	2 (2,3)	2 (2,3)	0.00025	0.00025	5 (4,5)	5 (4,5)	0.00062	0.00062
Spike	3822	3 (3,4)	4 (3,4)	0.00078	0.00105	7 (7,7)	7 (7,8)	0.00183	0.00183
RBD	223	0 (0,0)	0 (0,0)	-		1 (1,1)	1 (1,1)	0.00448	0.00448
NTD	292	2 (1,2)	2 (1,2)	0.00685	0.00685	0 (0,0)	0 (0,0)	-	-
Membrane	669	1 (1,1)	1 (1,1)	0.00149	0.00149	0 (0,0)	0 (0,0)	-	-
ORF8	366	1 (1,1)	1 (0,1)	0.00273	0.00273	3 (3,4)	3 (3,3)	0.00820	0.00820
Nucleocapsid	1260	1 (1,2)	1 (1,2)	0.00079	0.00079	7 (7,7)	7 (7,7)	0.00556	0.00556
ORF10	117	1 (1,1)	1 (1,1)	0.00855	0.00855	0 (0,0)	0 (0,0)	-	-

Substitution detection however is confounded by both the coverage of the genome attained at the time of sequencing and by viral load which may be influenced by vaccination. Median genome coverage was similar between controls and vaccinated participants for B.1.177 (B.1.177 controls: 29799 (IQR: 28644, 29874), B.1.177 vaccinated: 29765 (IQR: 28663, 29796) and for B.1.1.7 (B.1.1.7 controls: 29884 (IQR: 29662, 29903), B.1.1.7 vaccinated: 29823 (IQR: 28164, 29900) with no significant differences in genome coverage by vaccination group within lineage when tested using Wilcoxon rank sum testing. When just sequences from controls were considered, B.1.1.7 control sequences had significantly more coverage compared with B.1.177 controls (W=2308.5, adj. p=0.036). To try to explore if any differences in coverage may have influenced results for different groups, a calculated nucleotide substitution rate was calculated from the root to tip distance and then divided by genome coverage for each sequence. Using calculated substitution values, there remained no significant difference by vaccination status within lineage by Wilcoxon rank sum testing (Figure 5.14).

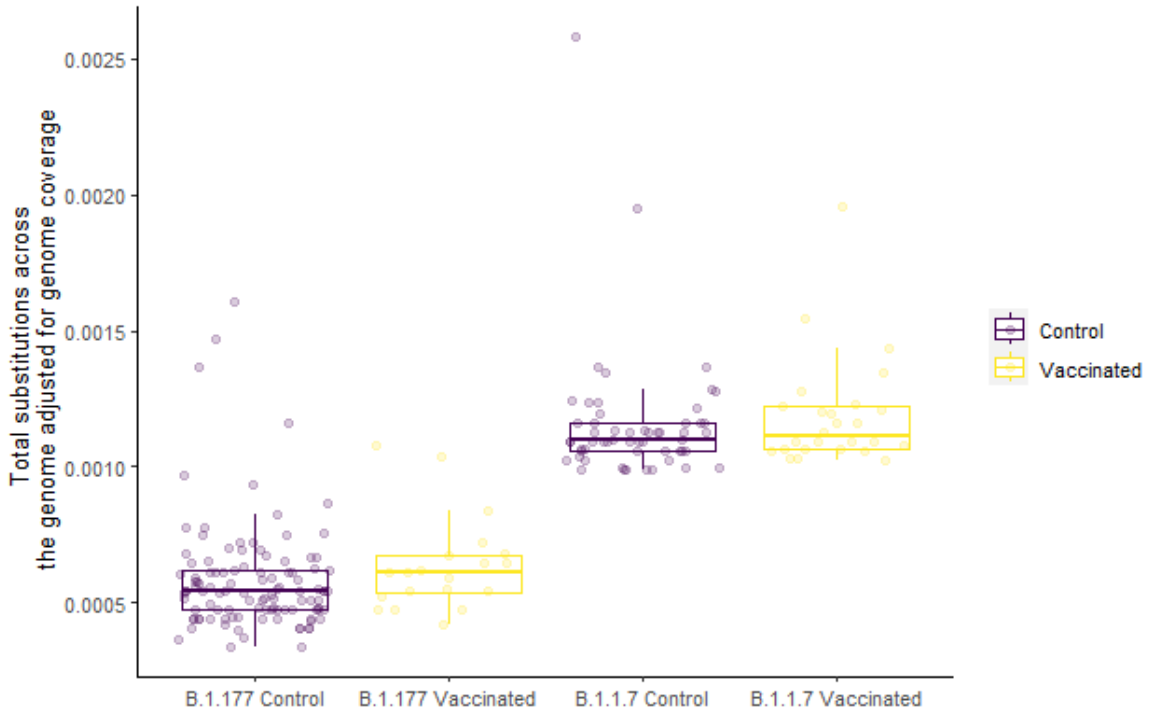


Figure 5.14: Total calculated nucleotide substitutions by vaccination status and lineage, adjusted to genome coverage (n=210).

At lower viral loads, a greater number of substitutions may be detected due to preferential amplification of small quantities of virus, and more sequencing errors are made in low viral load samples. To assess whether there was evidence of increased substitutions at low viral loads, the calculated total substitution rate (adjusted for genome coverage) was plotted for each Ct group (Figure 5.15). Across Ct groups for B.1.177 values were similar therefore it is unlikely that this had a significant influence of sequences from this lineage. At higher Ct groups for B.1.1.7 the calculated substitution rate was higher but small numbers in each group precluded meaningful statistical comparison and therefore it is unknown if this was a significant issue for B.1.1.7 samples of low viral load meaning that substitutions may have been overestimated for these few samples with low viral load.

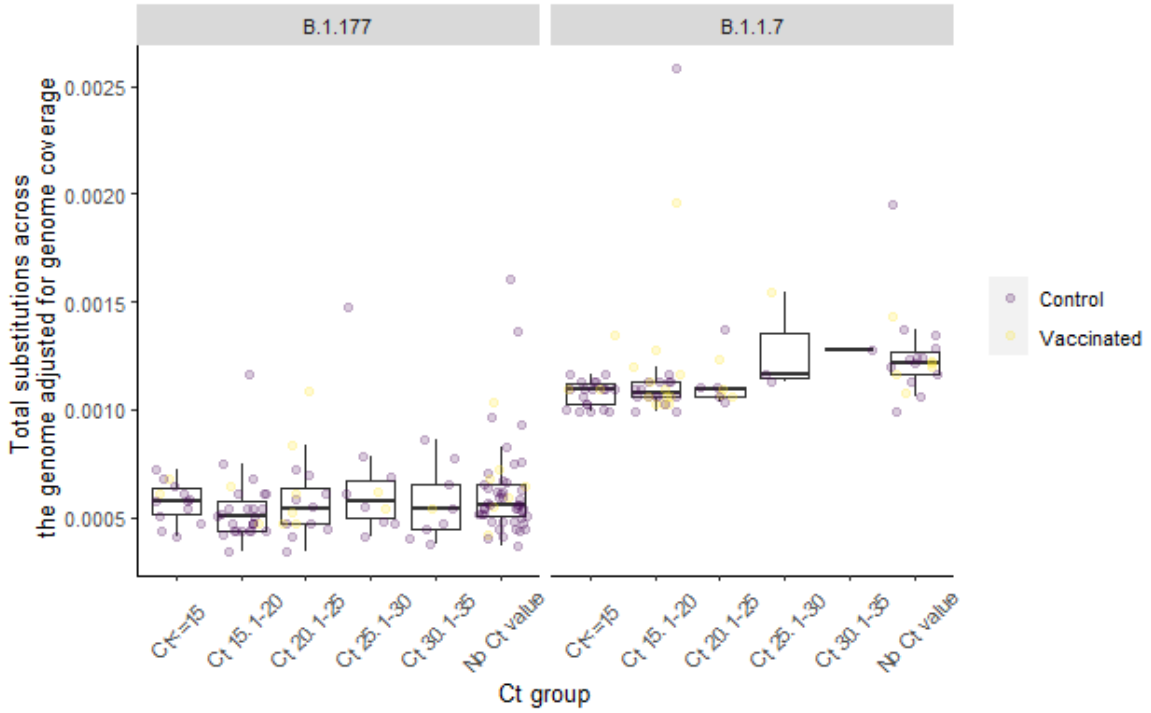


Figure 5.15: Total substitution rate adjusted for genome coverage by Ct group. The midlines of the boxes show medians, and the lower and upper bounds of the boxes correspond to the 25th and 75th percentiles. The whiskers show the most extreme value no further than 1.5 x IQR above or below the 75th or 25th percentile. Individuals plotted as dots coloured by vaccination status. Vaccination groups too small to plot by Ct group.

Amino acid substitutions occurring by vaccination group shown in Figure 5.16 and Figure 5.17. When individual mutations were examined, as expected, D614G (and accompanying P314L mutation which is in linkage disequilibrium with D614G) was detected in almost all sequences from both lineages. No B.1.1.7 lineage defining mutations were found in any of the B.1.177 sequences. Neither E484K nor K417T were detected in any sequences.

Most mutations in B.1.177 were sporadic, with no difference between controls and vaccinated individuals apart from three substitutions. For non-lineage defining substitutions

(excluding D614G/P314L) that occurred in 10% or more of either vaccine group proportions occurring in each vaccination group were compared using Fisher's exact tests (Table 5.9).

Table 5.9: Comparison of high frequency amino acid substitutions occurring in control and vaccinated participants. Adjusted p includes Bonferroni correction.

Substitution	**Gene**	**Lineage**	**Controls N (% of group)**	**Vaccinated N (% of group)**	**p**	**Adj. p**	**Associations**
L18F	S	B.1.177	53 (49)	10 (50)	1	1	Antibody escape[401] Found in P.1 (Gamma) and B.1.351 (Beta)
A3623S	ORF1a	B.1.177	6 (5.5)	4 (20)	0.048	0.386	
S5L	ORF7b	B.1.177	13 (12)	4 (20)	0.301	1	
V3718F	ORF1a	B.1.177	5 (4.6)	3 (15)	0.107	0.107	
T325I	N	B.1.177	2 (1.8)	2 (10)	0.113	0.113	T cell epitope[165]
E910D	ORF1b	B.1.177	1 (0.9)	2 (10)	0.063	0.500	
P1936S	ORF1b	B.1.177	1 (0.9)	2 (10)	0.063	0.500	
K3839R	ORF1a	B.1.177	10 (9)	1 (5)	1	1	
K1383R	ORF1b	B.1.1.7	21 (37)	7 (26)	0.458	1	
L730F	ORF1a	B.1.1.7	12 (21)	1 (3.7)	0.053	0.263	
K68	ORF8	B.1.1.7	12 (21)	3 (11)	0.366	1	
L18F	S	B.1.1.7	0	3 (11)	0.031	0.153	Antibody escape[401]
Q3966R	ORF1a	B.1.1.7	6 (11)	2 (7.4)	1	1	

For B.1.177, A3623S (in ORF1a) was found in 5.5% of B.1.177 controls and 20% of B.1.177 vaccinees however this was not significant after adjusting for multiple testing (Fisher's exact test, adj. p = 0.386).

For B.1.1.7, few sequences showed additional, non-lineage defining mutations and no differences were found between controls and vaccinees. L18F was found more often in B.1.1.7 vaccinated participants compared with B.1.1.7 controls however this was not significant after adjusting for multiple testing (11% compared with 0%, Fisher's exact test, adj. p=0.153).

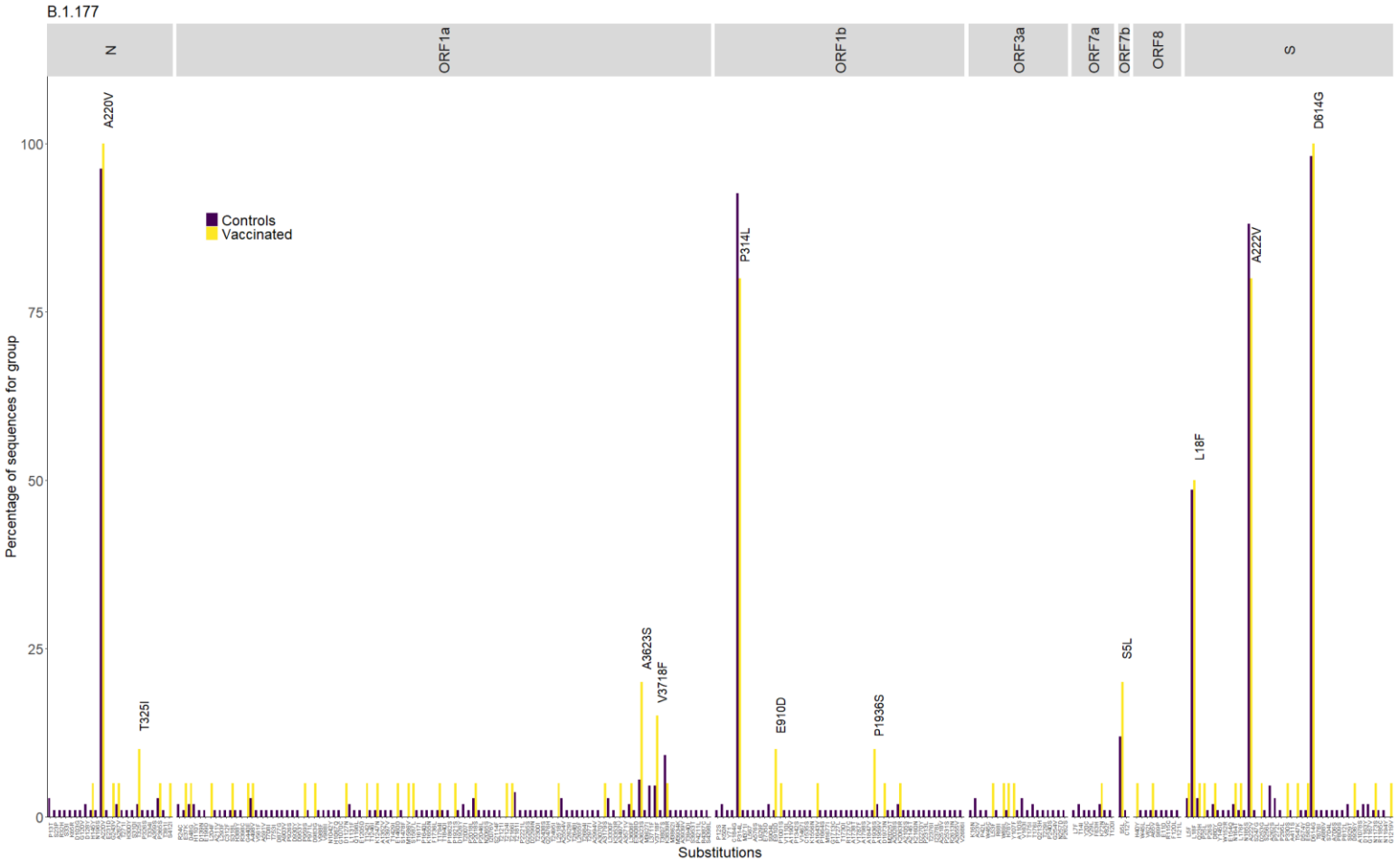


Figure 5.16: Amino acid substitutions by vaccination status in B.1.177 sequences. Mutations with greater than 10% frequency within vaccination group are labelled. Lineage defining mutations may not be detected by Nextclade due to incomplete coverage, but probable clade was estimated by identifying the clade of the nearest neighbour on a reference tree. [394]

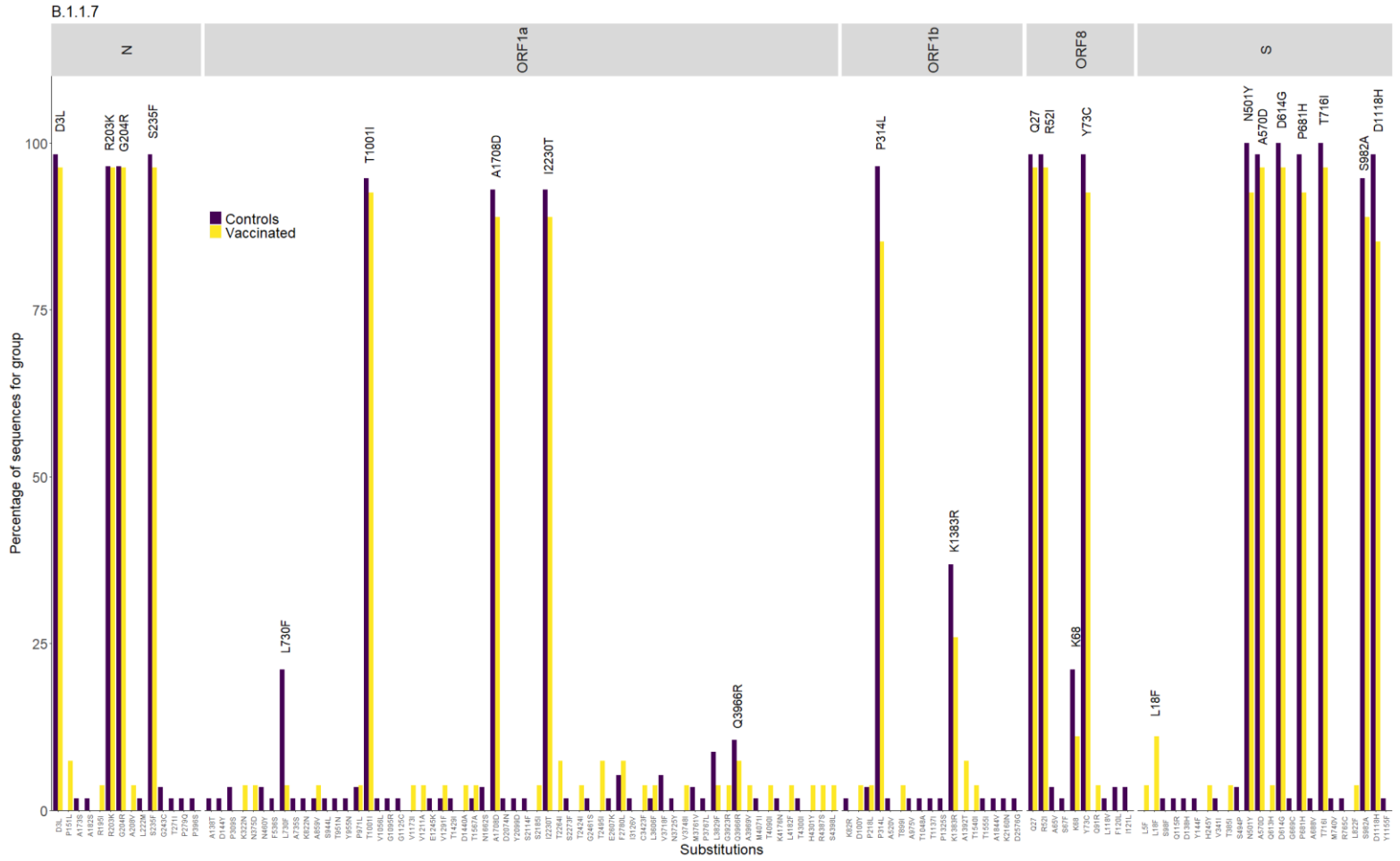


Figure 5.17: Amino acid substitutions by vaccination status in B.1.1.7 sequences. Mutations with greater than 10% frequency within vaccination group are labelled. Lineage defining mutations may not be detected by Nextclade due to incomplete coverage, but probable clade was estimated by identifying the clade of the nearest neighbour on a reference tree. [394]

5.3.8 Intrahost viral diversity

There were 150 sequences from 150 participants for B.1.177 (n=79) or B.1.1.7 (n=71) that were sequenced using Illumina in the blinded analysis dataset. A subset of samples was sequenced using the veSEQ protocol (n=108/150, 72%).

5.3.8.1 Viral load and minor allele detection

For samples using this method the number of unique reads in the samples can be used to estimate viral load. [40,402] Ct values were available for 138/150 (92%) samples. There were 8 samples that did not have either a Ct value or quantitative read data. As viral load can influence the detection of alternative allele frequencies these were explored. [40,378,383] The relationship between Ct values and unique mapped reads was also demonstrable in this dataset (Figure 5.18).

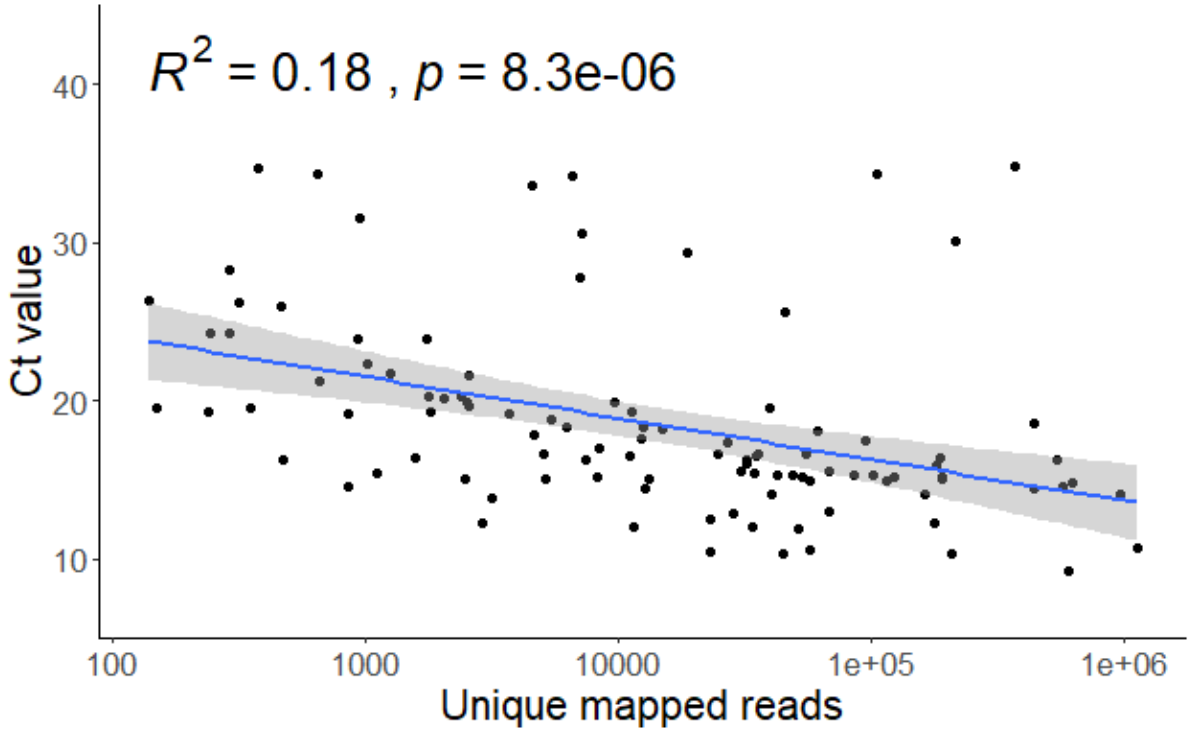


Figure 5.18: Relationship between Ct value and unique mapped reads. n=104 for whom both values were available.

As previously described [40] there was a non-linear relationship between unique mapped reads or Ct value and the number of iSNVs detected in a sequence (including only high confidence iSNVs with MAF of 5% or greater, Figure 5.19). The highest counts were seen at intermediate viral loads.

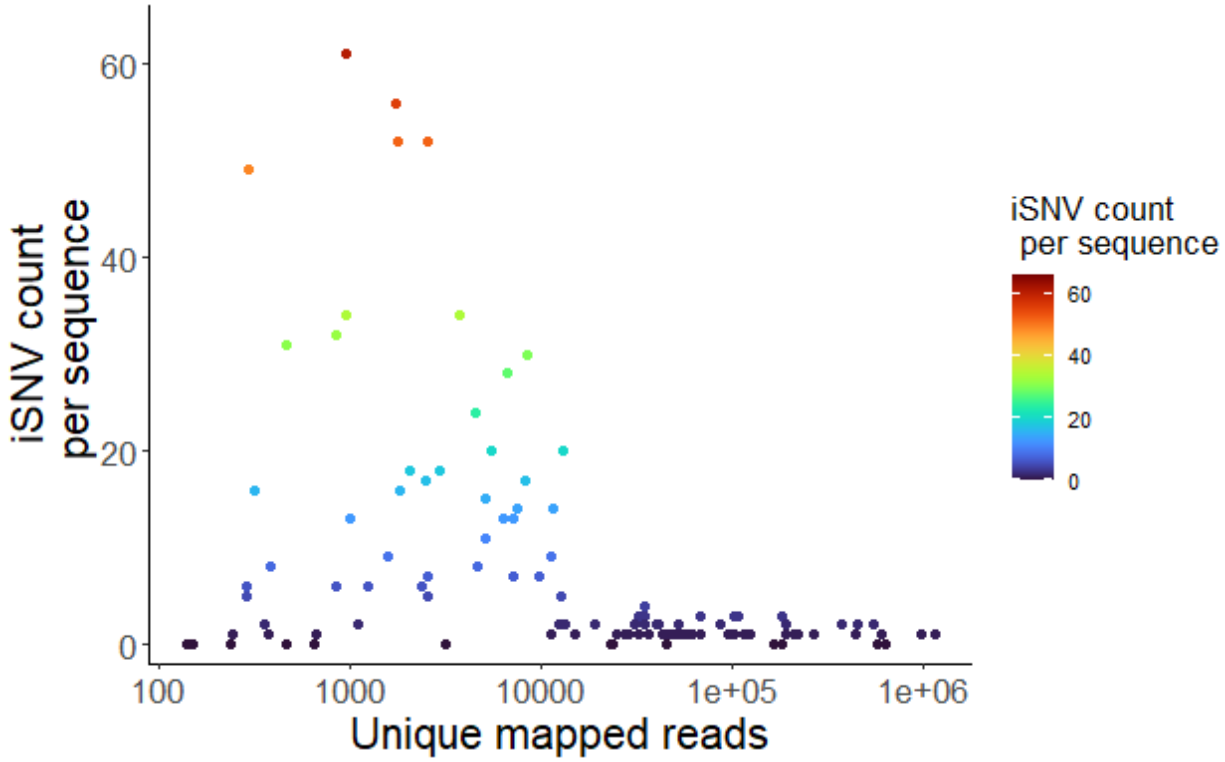


Figure 5.19: Non-linear relationship between viral load (unique mapped reads) and iSNV counts. Colour bar indicates the number of iSNVs detected in the sample to emphasise the relationship. N=108 (unique mapped read data not available for those processed using ARTIC protocol).

There was a clear relationship between increased variance of MAF and viral load (Figure 5.20). This is readily explained by the smaller pool of starting template molecules in low viral load samples: for example, with 10 starting molecules, the minimum possible MAF is 10% (1/10), and a single round of amplification – either *in vivo* or during sequencing – increases the total number of molecules (depth) to 10^2=100. As it is difficult to identify truly unique templates bioinformatically, the result is that even stringent cut-offs on MAF cannot reliably identify the true number of iSNVs in samples with low viral load. [40] For later analyses sequences were classified as high viral load (unique mapped reads >=50,000 or a Ct value <=23; n=113), low viral load (n=29), or no data where neither value was available (n=8).

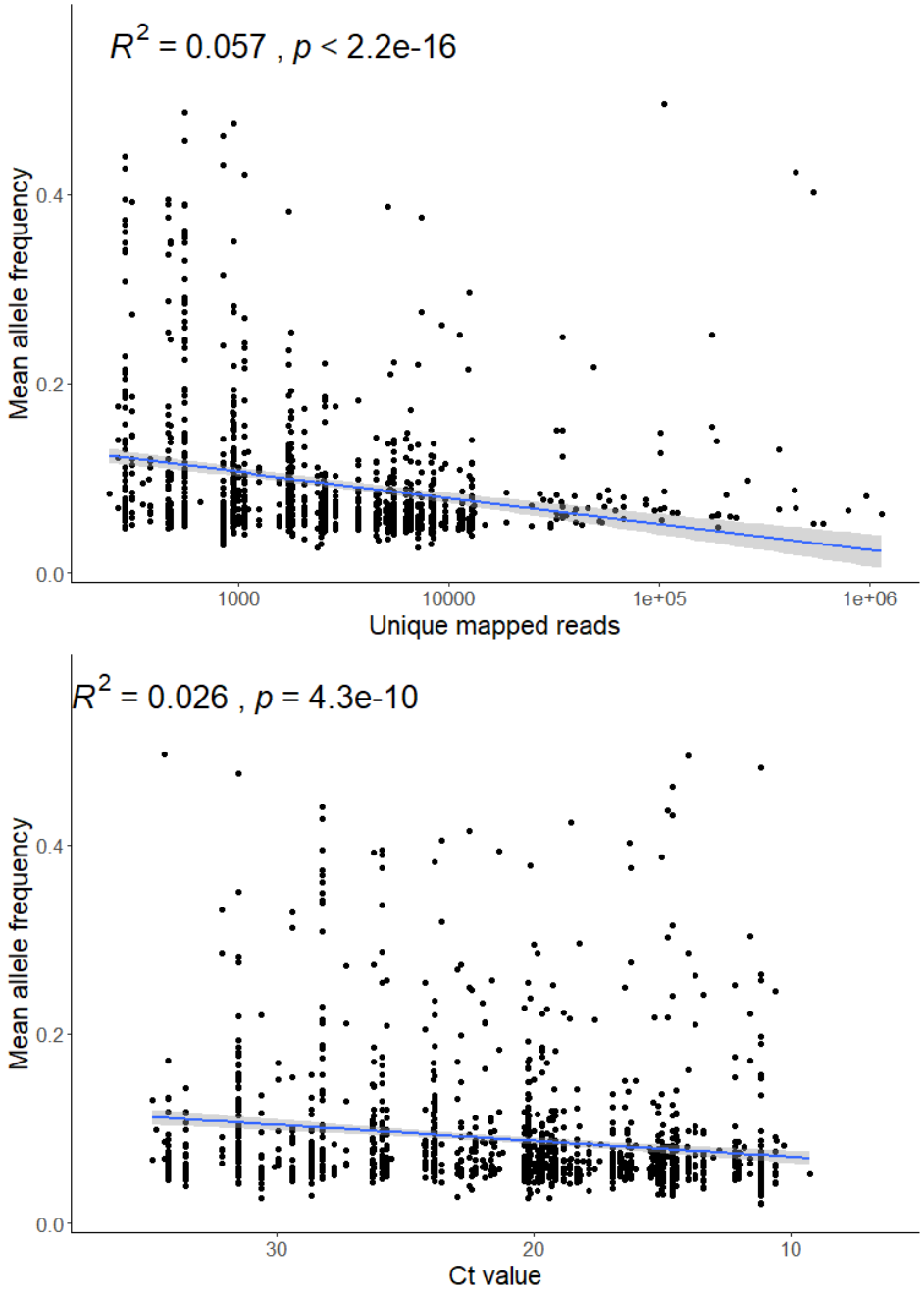


Figure 5.20: Relationship between mean allele frequency and unique mapped reads (top panel, n=1192) and Ct value (lower panel, n=1478). MAFs of all frequencies included, each point represents a single MAF.

5.3.8.2 Minor allele frequency

MAFs at >=5% were identified in 135 sequences (90%), 113/150 (75%) in high viral load samples. Using a 5% MAF cut off there were 1034 different iSNVs identified at 1008

genome positions from all samples and 625 at 613 positions in the high viral load samples. Of these 1034 iSNVs, 966 (93%) occurred in only one participant, 39 (3.8 %) that occurred in two participants and 10 (0.97%) that occurred in three participants. Correspondingly, in the high viral load samples 598/625 iSNVs (96%) occurred in one participant, 15 (2.4%) occurred in two participants and 3 (0.48%) occurred in three participants. Most participants had few iSNVs (Figure 5.21).

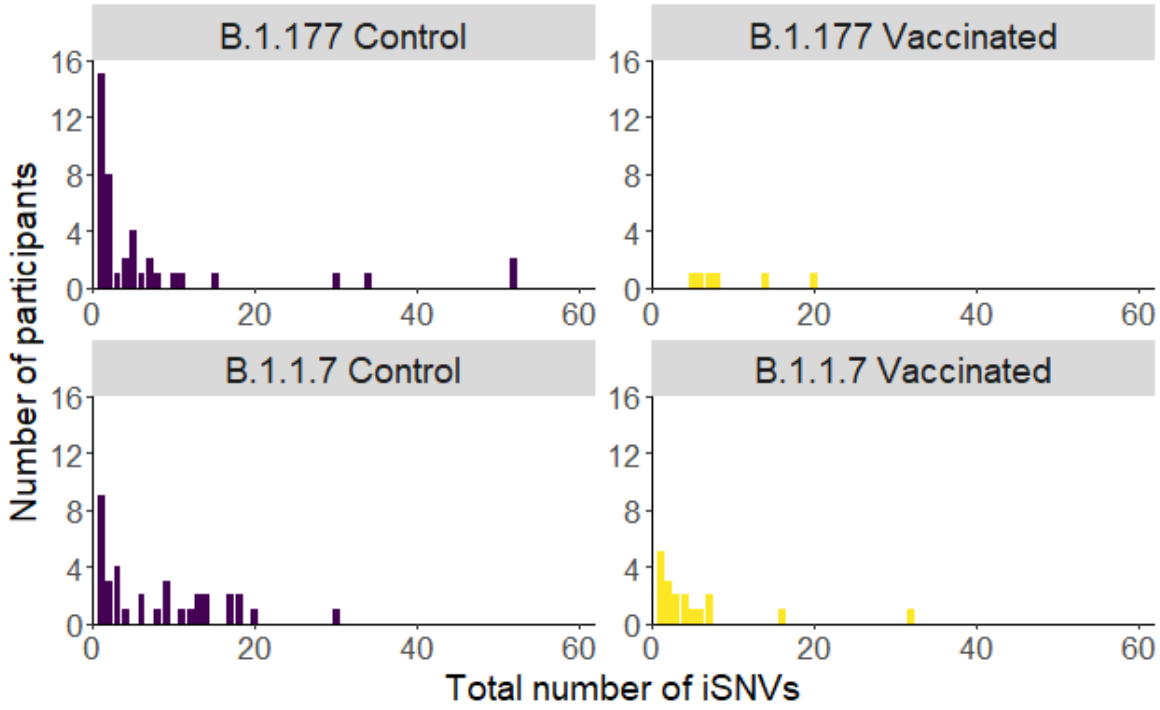


Figure 5.21: Number of iSNVs per individual in group for high viral load samples. B.1.177 controls n=45, B.1.177 vaccinated n=7, B.1.1.7 controls n=41, B.1.1.7 vaccinated n=20.

The first chronological sample for a participant that was sequenced successfully was selected therefore most samples were taken early after episode onset (either onset of symptoms or first positive swab if asymptomatic at time the first swab was taken). For B.1.177 participants the median time after onset for sampling was 1 day (IQR: 0, 4) for

controls and 0 days (IQR: 0, 2 days) for vaccinees. For B.1.1.7 participants the median time after onset for sampling was 1 day (IQR: 0, 2.75 days) for controls and 2 days (IQR: 1, 3 days) for vaccinees. There was a wide range of Ct values when the sample was taken at the start of the episode and a non-linear response with increasing interval (Figure 5.22). Ct values tended to decrease (i.e. viral load increased) if the sample was taken within a few days after onset and at longer intervals the Ct values tended to increase (i.e. viral load was lower). Similarly, to Zhang et al. [380] a non-linear relationship existed between the number of iSNVs per sequence and the interval between episode onset and sampling (Figure 5.22).

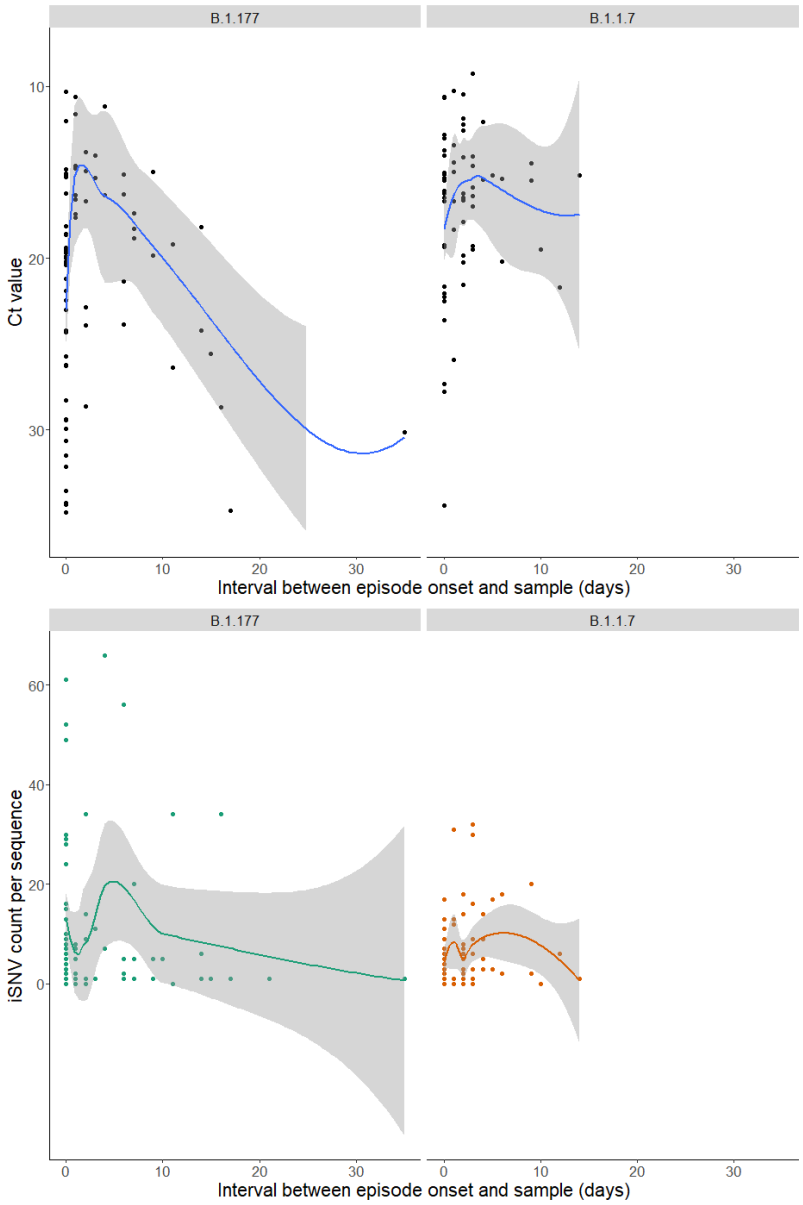


Figure 5.22: Ct values and iSNV counts and episode onset to sampling interval. Top panel: n=138, B.1.177 n=74, B.1.1.7 n=64. Lower panel: B.1.177 n=79, B.1.1.7 n=71.

Some groups have noted an increase in iSNV count over calendar time in early 2020 from cross-sectional samples. [378] No such increase in the iSNV count over time was observed in this dataset when all participants, or the high viral load subset were considered by lineage. (Figure 5.23).

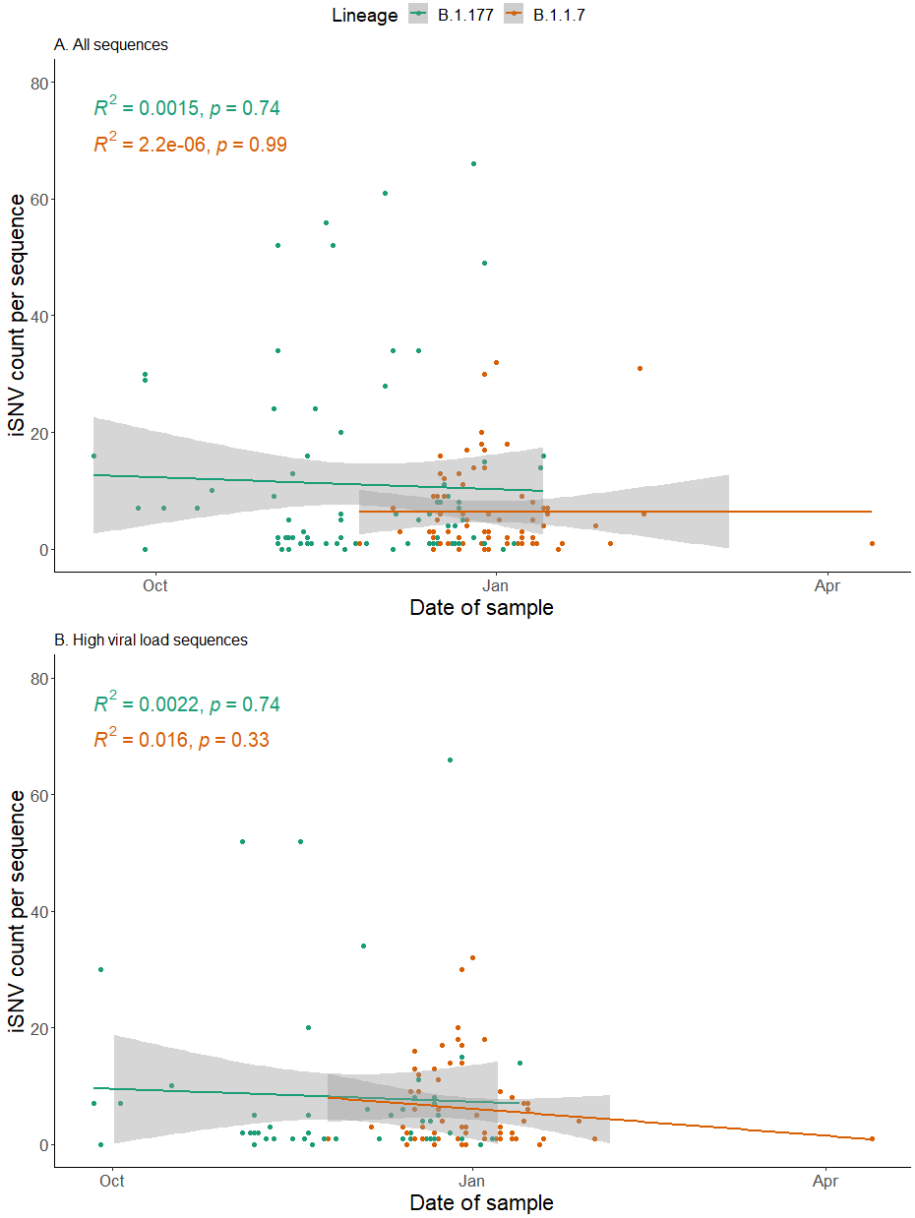


Figure 5.23: Relationship between calendar date and iSNV count per sequence. Top panel all samples (n=150), lower panel high viral load (n=113).

The median iSNV count for high viral load B.1.177 control sequences was 2 (IQR 1-6) and 7 (IQR 5.5-11) for vaccinees, and for B.1.1.7, 3 (IQR 1-12) for controls and 2.5 (IQR 1-5.25) for vaccinees. (

Table 5.10).

When just the control participants were considered, there was no significant difference in iSNV count per sequence for B.1.177 and B.1.1.7 sequences (all sequences: W=1660.5, p=0.78, high viral load sequences: W=992, p=0.55).

There was no significant difference in iSNV count per sequence between controls and vaccinated participants for B.1.177 when compared by Wilcoxon rank sum testing for all sequences and when restricted to high viral load samples (all sequences: W=341.5, p=0.25, high viral load sequences: W98.5, p=0.11). This was also true for B.1.1.7 sequences (all sequences: W=575.5, p=0.53, high viral load sequences: W=454, p=0.5).

Table 5.10: Number of unique iSNVs and median iSNV count by vaccination status and stratified by viral load status.

All samples						
Group	**n**	**Number of unique iSNVs**	**Median MAF**	**Median iSNV count per participant (IQR)**	**Range iSNV count per participant**	
					Min	**Max**
B.1.177 Controls	66	575	0.070	3.5 (1-12.5)	0	66
B.1.177 Vaccinated	13	122	0.074	7 (5-14)	0	34
B.1.1.7 Controls	50	277	0.067	4.5 (1-9)	0	30
B.1.1.7 Vaccinated	21	110	0.072	3 (1-6)	0	32
High viral load samples						
Group	**n**	**Number of unique iSNVs**	**Median MAF**	**Median iSNV count per participant (IQR)**	**Range iSNV count per participant**	
					Min	**Max**
B.1.177 Controls	45	281	0.067	2 (1-6)	0	66
B.1.177 Vaccinated	7	47	0.075	7 (5.5-11)	0	20
B.1.1.7 Controls	41	246	0.067	3 (1-12)	0	30
B.1.1.7 Vaccinated	20	82	0.070	2.5 (1-5.25)	0	32

Minor allele frequencies (using only those with $\geq$5% frequency) were grouped to give each sequence a mean MAF and cMAF value then grouped by vaccination status and lineage. Whilst the median values were higher in both lineages for MAF and cMAF values for the vaccinated group (particularly for B.1.1.7) there was no statistical difference by Wilcoxon

rank sum testing in either the median MAF or median cMAF values between vaccination status groups by a lineage. To ensure this was not secondary to viral load this was repeated for the high viral load samples and whilst the same trend emerged it was not statistically significant (Figure 5.24).

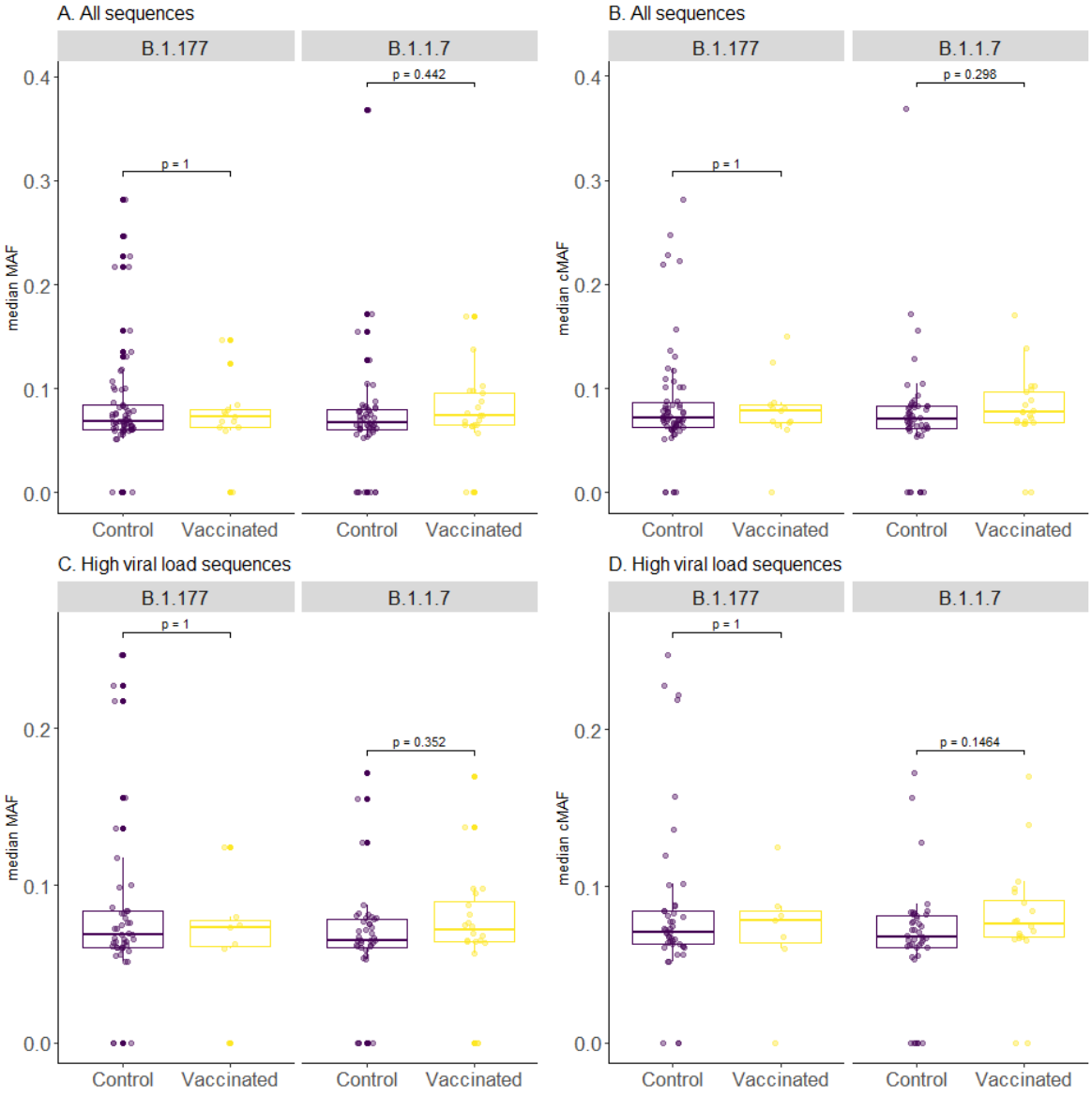


Figure 5.24: Median MAF and cMAF by lineage and vaccination status from all sequences (plots A and B) and high viral load samples (plots C and D). The midlines of the boxes show medians, and the lower and upper bounds of the boxes correspond to the 25th and 75th percentiles. The whiskers show the most extreme value no further than 1.5 x IQR above or below the 75th or 25th percentile. Groups were compared using Wilcoxon rank sum testing.

The distribution of iSNVs across the genome were examined to see how the frequency of iSNVs differed across different genes (Figure 5.25). There was wide variation across the genome and between groups. The highest numbers of iSNVs were observed in ORF1a, ORF1b, S and N in both lineages. However, when normalised by gene length the highest frequency of iSNVs in the B.1.177 lineage, were observed in ORF7a, ORF7b, N, and ORF10. Amongst B.1.1.7 length the highest frequency of iSNVs was seen in ORF7b, N and ORF10.

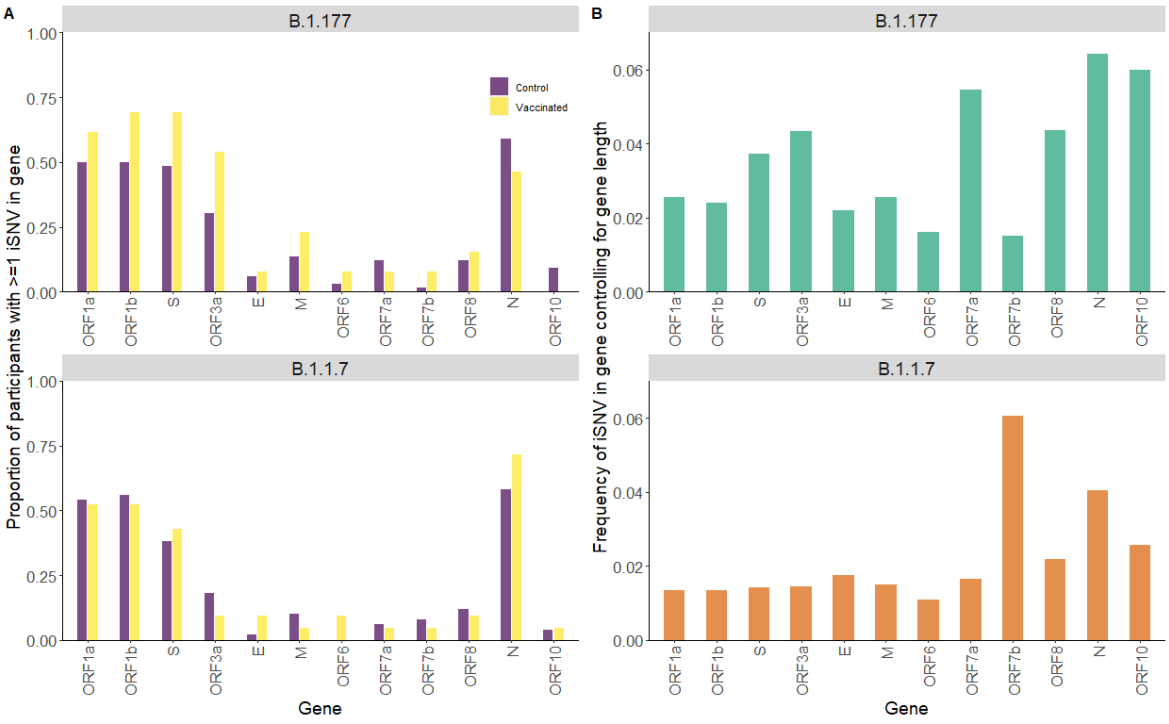


Figure 5.25: iSNVs by gene. A) Proportion of participants in each group with 1 or more iSNVs in the gene separated by lineage. B) Frequency of SNV in gene per site controlled for by gene length separated by lineage (all participants).

The most frequently observed iSNVs found by lineage are shown in Table 5.11. The highest frequency iSNVs were notably associated with sequencing protocol (e.g. T28469G and veSEQ), associated with diversity in the synthetic viral controls or had been identified

as high frequency sites by other groups. Some iSNVs were observed to occur in combination with another in more than one participant, for example G16197G and C25075T both occurred in two participants from the same study site and the sequences are adjacent on the tree; it is possible that this represents a transmission cluster.

Table 5.11: Most frequent iSNVs observed for B.1.177 and B.1.1.7. Numbers in brackets denote high VL samples only). N=number of sequences (which also is equivalent to participants). Shaded cells show iSNVs identified by other groups as highly shared, associated with synthetic control diversity or high frequency with reference in the final column.

A. B.1.177

iSNV	**Gene**	**n**	**Methodological Associations**		**Controls**	**Vaccinated**	**Ref**
			veSEQ	ARTIC	66 (45)	13 (7)	
T28469G	N	37 (27)	37 (27)	0	32 (25)	5 (2)	[40]
T25202G	S	22 (16)	0	22 (16)	18 (12)	4 (4)	[40,385]
A26137C	ORF3a	21 (14)	0	21 (14)	16 (10)	5 (4)	[379,385]
C13824A	ORF1b	20 (16)	0	20 (16)	16 (12)	4 (4)	
T514G	ORF1a	15 (9)	0	15 (9)	12 (7)	3 (2)	[379]
G25218T	S	7 (6)	0	7 (6)	7 (6)	0	[385]
A2712C	ORF1a	6 (2)	0	6 (2)	6 (2)	0	[40,385]
A2729C	ORF1a	6 (5)	0	6 (5)	5 (4)	1 (1)	
A25216T	S	5 (3)	0	5 (3)	5 (3)	0	
C25214A	S	5 (2)	0	5 (2)	4 (1)	1 (1)	[385]
T26761G	M	5 (3)	0	5 (3)	4 (2)	1 (1)	
C25231A	S	4 (1)	0	4 (1)	4 (1)	0	[385]
G29645G	ORF10	4 (3)	3 (2)	1 (1)	4 (3)	0	
A25201C	S	4 (2)	0	4 (2)	3 (1)	1 (1)	
A2730C	ORF1a	4 (4)	0	4 (4)	3 (3)	1 (1)	

B. B.1.1.7

iSNV	Gene	n	Methodological Associations		Controls	Vaccinated	Ref
			veSEQ	ARTIC	50 (41)	21 (20)	
T28469G	N	37 (27)	37 (27)	0	23 (22)	14 (13)	40
A26137C	ORF3a	6 (4)	0	6 (4)	5 (3)	1 (1)	379 385
C13824A	ORF1b	8 (6)	0	8 (6)	5 (3)	3 (3)	
T25202G	S	8 (6)	0	8 (6)	5 (3)	3 (3)	40,385
T514G	ORF1a	4 (3)	0	4 (3)	3 (2)	1 (1)	379
A29320G	N	3 (1)	1 (1)	2 (0)	2 (0)	1 (1)	40
T2949G	ORF1a	3 (2)	3 (2)	0	1 (1)	2 (1)	40,385
A28095T	ORF8	2 (1)	0	2 (1)	2 (1)	0	
G15168A	ORF1b	2 (0)	0	2 (0)	2 (0)	0	
G23642T	S	2 (0)	0	2 (0)	2 (0)	0	
T23617G	S	2 (0)	0	2 (0)	2 (0)	0	
T23620G	S	2 (0)	0	2 (0)	2 (0)	0	
T23622G	S	2 (0)	0	2 (0)	2 (0)	0	
T27264C	ORF6	2 (1)	2 (1)	0	0	2 (1)	379
T8987C	ORF1a	2 (2)	2 (2)	0	2 (2)	0	

5.4 DISCUSSION

In this chapter I have described the lineages obtained from SARS-CoV-2 infections in the COV002 trial until June 2021 and described the efficacy of the ChAdOx1 nCoV-19 vaccine against infections caused by the B.1.1.7 lineage and the most common lineage identified in our dataset circulating in mid-late 2020, B.1.177. In this exploratory analysis I have shown that ChAdOx1 nCoV-19 vaccine efficacy estimate against primary symptomatic COVID-19 caused by the B.1.1.7 (Alpha) variant was 58.3% (95% CI 29.1-75.5) which was lower than for B.1.177 (80.7%, 95% CI 67.1-88.6). This may account for the observed phylogenetic clustering by vaccination status, due to lower efficacy there are more vaccinees contributing to the B.1.1.7 clade of the tree compared with other areas of the tree. By comparing B.1.1.7 infections with the most common alternative lineage (B.1.177) I have also demonstrated that

there was no difference in symptom status and no significant difference in infection minimum Ct value or duration of positive swabs for symptomatic infections between the two lineages in our dataset.

The overall vaccine efficacy of ChAdOx1 nCoV-19 against primary symptomatic infections for this dataset concurred well with the interim analysis of ChAdOx1 nCoV-19 trials conducted on three continents. [275] Efficacy against primary symptomatic infections caused by B.1.177 lineage was higher than the overall efficacy, and higher than the recorded efficacy against B.1.1.7. The vaccine efficacy for B.1.1.7 was 22% lower for primary symptomatic disease compared with B.1.177. When using SGTF during the time period for which this served as a good proxy for B.1.1.7 (though this approach only led to a modest increase in putative B.1.1.7 cases), the vaccine efficacy estimate and confidence intervals from the larger number of cases was similar to that obtained using sequencing results suggesting this is a robust finding. These vaccine efficacy figures agree with *in vitro* work suggesting reduced, but not absent neutralisation activity of ChAdOx1 nCoV-19 elicited antibodies against B.1.1.7 virus compared with the Victoria lineage. [57] Consistent with this finding that ChAdOx1 nCoV-19 vaccination remains efficacious against B.1.1.7, a UK test negative study has since estimated that two doses of ChAdOx1 nCoV-19 had an effectiveness of 74.5% (95% CI, 68.4 to 79.4) against symptomatic COVID-19 caused by B.1.1.7. [403] Regarding other vaccination platforms, mRNA vaccine effectiveness is preserved against B.1.1.7. [404,405] though in the phase 3 for the recombinant nanoparticle protein vaccine a 10% reduction in efficacy was seen for B.1.1.7 infections compared with non-B.1.1.7 infections in a similar manner to this data. [298]

The duration of positive swabs was significantly longer in symptomatic controls compared with vaccinees for B.1.177 but not for B.1.1.7, however there was no significant difference in infection minimum Ct value from symptomatic participants by vaccination status for confirmed B.1.177 or B.1.177. The duration results may suggest that ChAdOx1 nCoV-19 vaccination may have a lower or no impact on duration of positive swabs for infections caused by the B.1.1.7 variant compared with earlier lineages in symptomatic infections. The B.1.177 finding would not have been predicted from the original macaque data which suggested that vaccinees might show similar shedding to controls. [170] However, the African green monkey model, which may resemble human infection more closely, demonstrated increased upper airway viral replication of B.1.1.7 compared with the D614G progenitor suggesting B.1.1.7 may be better adapted to the upper airway [365] and therefore vaccination may have been less efficacious in preventing shedding even if the infective episode was not accompanied by illness. The duration of positive swabs in symptomatic infections was not affected by variant is in keeping with the findings of some, but not other groups, [406,407] albeit in markedly different cohorts.

At the consensus level there was no observed difference between vaccination groups in nucleotide or amino acid substitutions for either lineage. Intra-host diversity was found in most high viral load samples (75%) but the diversity within any individual sequence was low in keeping with the findings of other groups. [40,378,381,383,384] The absolute numbers of iSNVs per participant, MAF and cumulative MAF as measures of how often the second most dominant allele occurs and the number of sites that have evidence of sub-consensus diversity were the same for controls and vaccinees infected with either B.1.177 or B.1.1.7. B.1.1.7 vaccinees were infected later than their control counterparts however the lack of effect of calendar time on iSNV count across the cohort supports the finding that vaccinees did not

have a higher incidence of iSNVs. These findings do not provide evidence for the hypothesis that vaccinated individuals may be more likely to have a higher rate of consensus and intrahost diversity secondary to vaccine mediated selective pressure.

One other non-randomised study conducted with samples from Hong Kong has considered the effect of vaccination on intrahost viral diversity. [384] Their finding of higher iSNV incidence in VOCs compared with non-VOC sequences was, in contrast not found in our study as the iSNV count in the control participants was not significantly different between B.1.177 and B.1.1.7. When comparing vaccinees with controls, I did not find a difference in the number of iSNVs per participant, median MAF or cMAF between vaccination status groups for either of the lineages in this study however Gu et al (2023) did find higher rates of viral diversity in 2 dose Comirnaty vaccinees compared with unvaccinated individuals infected by Delta. This group also found a reduction in viral diversity in 3 dose Comirnaty vaccinees compared 2 dose (Comirnaty or CoronaVac) vaccinees infected by Omicron. This demonstrates that intrahost diversity may be influenced by variant, but their findings are confounded by time since vaccination for the two variants and the possibility of significant demographic differences between the unvaccinated and vaccinated groups. B.1.1.7 infections came later in our study than B.1.177 infections so the interval between vaccination and infection was greater for B.1.1.7 infections, however one of the strengths of this unique study is that individuals were randomised to vaccine allocation so the host characteristics between groups, aside from vaccination, are well matched.

There was evidence of a range of iSNV density across the genome suggesting that different sites along the genome may be subject to different selective pressures. Other studies have found varying iSNV densities across genes but the specific findings between studies

often differ. [40,378,381,384] This, along with the high proportion (96% in high viral load samples) of iSNVs only occurring in a single participant suggest that many mutations are sporadic rather than mutational hotspots or under specific selective pressures.

The lack of difference between both consensus and sub-consensus viral diversity from control and vaccinated participants is in keeping with the knowledge that initially the viral diversity of SARS-CoV-2 in the population was very low and the molecular clock slow. This was true for the period during which these samples were taken and the only period for which RCT populations existed prior to the rapid roll out of vaccinations in high income countries, and the successive advent of Delta, BA.1 and subsequent Omicron variants resulting in multiple waves of infection producing populations with heterogeneous exposure and immune profiles. The data presented here do not provide any evidence that individuals vaccinated with ChAdOx1 nCoV-19 were infected with viruses that were any different to controls. It is notable however that the B.1.1.7 vaccinees who became infected did so later than controls, providing early evidence to suggest that vaccine mediated immunity is relative. Vaccination provided protection from infection, but the degree of protection was dependent on circulating variant, as shown by the difference between vaccine efficacy calculated for B.1.177, which exceeded overall trial vaccine efficacy findings and the lower efficacy value obtained for B.1.1.7. However, this vaccine mediated immunity appears to be incomplete and can be overcome by increasing exposure, as suggested by the later temporal distribution of vaccinated cases vs controls during the B.1.1.7 wave. As the force of infection in the population increased during this acute wave, [363] the cumulative risk of exposure would have increased for both vaccinees and controls, and the vaccinees eventually also became infected. The evidence from this randomised controlled trial is the most direct evidence for purely a host (and therefore vaccine) mediated effect given that participants were

randomised to allocation and therefore were exposed to the same risk of infection as their counterparts in the other allocation arm.

Regarding individual mutations, no mutations occurring in at least 10% of an intervention arm were found to be significantly different between arms after corrections for multiple testing. L18F was commonly seen in both controls and vaccinees in B.1.177 and more commonly in vaccinated cases of B.1.1.7 compared with controls. L18F is found in the P.1 (Gamma) and B.1.351 (Beta) VOCs which also emerged in 2020. L18F is notable as it is found in the antigenic region of the NTD and may be associated with escape from NTD-binding antibodies. [401] T325I (in nucleocapsid) which is associated with partial impairment of T cell killing ability [408] was seen more frequently in B.1.177 vaccinees compared with controls but this was not statistically significant. A3623S (in ORF1a) was also found more commonly in vaccinees but has no known biological associations. These three mutations observed as more frequent in vaccinees could be due to positive selective pressure but should not to be over-interpreted given the small sample size especially in the vaccinated group.

There are several alternative explanations for having found no difference in within host viral diversity between vaccine groups. Vaccination status may not affect intrahost diversity early in infection when most of these samples were taken. Studies using strict methodology show that the transmission bottleneck is narrow, whereby only minimal viral diversity is transmitted from donor to recipient. [40,385] Therefore, it is possible that vaccine-mediated immunity, once breached and resulting in infection in the recipient, plausibly does not affect viral evolution within-host in the timeframe before sample collection in this study for B.1.177 as 75% of samples were collected within 4 days. Sequences were not all collected at the same point after episode onset which may have acted as a confounding factor with

respect to viral load but there may also be other confounding factors. These are reflected in the conflicting literature: some groups have found iSNV count within individuals did not change with time after symptom onset, [383] but in contrast to another study which found an increase over time using a linear model to account for the impact of viral load [378] and another where iSNV count peaked after approximately 5-6 days and then declined. [380] In this study non-linear relationships between Ct values or iSNV count and time of sampling post onset were observed. No change in iSNV count was seen over the course of the epidemic period study for either lineage.

There are also reasons as to why, if a difference exists, it was not detected in this dataset. First, the sample sizes within each group are small, limiting power to detect small differences between groups. Given that SARS-CoV-2 viral diversity is known to be low, detection of a small difference between controls and vaccinees would require large numbers unlikely to be feasible in a trial of an efficacious vaccine. Second, the study of within-host viral diversity in SARS-CoV-2 is challenging in terms of methodology, notwithstanding a trial within a pandemic context. Many studies of SARS-CoV-2 select high quality, high viral load sequences for these types of analyses however the trial cohort in this study had a restricted number of possible samples at the outset. Different sequencing protocols, investigator selected definition of MAF limits and minimum read depths can vastly affect viral diversity results and can lead to flawed interpretation as a result. [385,409] As demonstrated, protocol can affect the detection of specific iSNVs, for example T28469G is associated with veSEQ and T25202G with ARTIC. These iSNVs are commonly detected in samples but represent methodological artefact secondary to diversity in commercially available viral synthetic controls rather than true diversity. [40] Many studies define a different list of high frequency sites that occur in a high number of sequences in their cohort and

exclude them from analyses, and though whilst there are recurrent iSNVs that feature, they are not consistent. [40,381,385] The sites identified at higher frequency in this dataset could represent hypermutable sites or positive convergent selection, [383] additionally it is possible that these could relate to contamination or transmission of minor variants for which there is insufficient epidemiological evidence to support or refute. In this analysis strict definitions requiring >100 reads per site with 20 reads to support an iSNV were used to minimise false positives but true low frequency iSNVs may have been excluded using this approach. An additional benefit of a case-control design is that exclusion of specific sites is not necessary in the context of this study, because artefactual and method-specific iSNVs would, by definition, occur in both vaccinees and controls. This approach confirms that the previously identified method-specific iSNVs, such as T28469G are indeed observed in both groups in expected frequencies.

Specific to this trial there are additional issues that necessitate cautious interpretation. The samples used for intrahost diversity analyses were a subset of all sequences available, which in themselves were a subset of participants. The included sequences were chosen because they were sequenced using Illumina platforms, however two types of sequencing protocol (veSEQ and ARTIC) were used, introducing methodological heterogeneity. Samples from study sites or from Lighthouse laboratories could have been sequenced at the same time especially when samples were consolidated for batch sequencing, so contamination is impossible to exclude entirely. On the other hand, mixed and high-diversity infection is also plausibly associated with high exposure, and many of the included participants were frontline healthcare workers who could be part of transmission clusters.

There are limitations to these findings. Whilst every effort was made to retrieve each sample from trial participants, some were not retrievable, particularly from large government and clinical laboratories that were processing huge volumes of tests at the height of epidemiological waves. The results presented here emphasise the difficulty of comparing sequences from individuals where there are multiple confounding factors: as shown in Chapter 4 the vaccine is likely to have reduced the viral load and duration of shedding in symptomatic individuals. Controls may therefore have had more positive swabs which means there was a greater chance a sample being successfully retrieved and a higher chance of a sample closest to peak viral load being taken. A control participant therefore may have an increased chance of one of their samples progressing to being sequenced, particularly when there were constraints on the sequencing pipelines high viral load sequences were preferentially selected as a higher viral load increases the chance of being sequenced successfully. Therefore, there is an inherent bias in place which may overestimate vaccine efficacy. This is demonstrated by only 24% of vaccinees having at least one swab successfully sequenced (compared with 48% of controls) and 37% of vaccinees having had at least one swab sent for attempted sequencing which was unsuccessful (compared with 23% of controls) and no other swab from the same infection which returned a sequence. The reduced duration of positive swabs seen in vaccinated individuals would also contribute bias by reducing the number of swabs available for retrieval by an infected vaccinated individual. Accordingly, 38% of vaccinees had no swab retrieved for sequencing compared with 28% of controls. This could be partially addressed by using PCR data in terms of using SGTF as a proxy for B.1.1.7 infections for efficacy calculations however samples with a high Ct value cannot reliably be identified as SGTF because at low viral loads, gene drop out is common.

Multiple sequencing sites led to an unavoidable heterogeneity in terms of sequencing methodologies and protocols, which limited the number of sequences suitable for sub-consensus analysis. For intrahost viral diversity analyses control participant may therefore have a higher viral load, therefore more unique reads and therefore a lower MAF as a result, making any interpretation between vaccination groups limited. As demonstrated viral load affects the detection of intrahost diversity, with highest levels of iSNVs detected when viral load is intermediate. To control for this, analyses were performed for all samples and for the subset with high viral loads however there is inherent bias in doing so given vaccination can impact viral load.

The efficacies calculated in this chapter are informative but cannot be formally compared as these sample sizes are confounded by factors mentioned above, and individuals in an RCT cannot be randomised to infection with a particular variant. Moreover, our ChAdOx1 nCoV-19 participants mostly received their vaccinations over the summer of 2020 hence the timing of the B.1.177 infections occurred a significantly shorter time after their second vaccine dose, compared to B.1.1.7 infections. Natural infection with an early variant appeared to confer a median of 7 months of protection against SARS-CoV-2 reinfection, [179] however it is now recognised that the protective effects of vaccination to mild disease wanes within weeks of the second dose.

In addition to the limitations mentioned, a further caveat is that natural selection analysis (scaled ratio of nonsynonymous to synonymous variants, dN/dS) to further understand the type of selection present was not performed on these sequences. The ratio of nonsynonymous to synonymous substitutions has been found to be higher for SARS-CoV-2 than for other human coronaviruses, with evidence of an excess of C->U mutations hypothesised to be due

to RNA editing enzymes, [410] however since there were no identified differences between controls and vaccinee sequences it would be plausible that the selective pressure was similar for both vaccine allocation groups in these cross-sectional samples taken early in infection.

There are several relevant questions that this study could not explore. This study was not able to investigate the impact of other VOCs and the effects of hybrid immunity on intrahost diversity. I was unable to look for associations with illness severity and iSNVs due to the mild spectrum of disease within our cohort as there have been different findings with respect to severity and viral diversity, with some not finding an association [383] and others suggesting there may be detectable differences between severity groups. [381] These samples were all from the upper respiratory tract so precluding comparison between anatomical sites which would have been interesting to compare iSNV profiles and consistency during infection in different compartments, especially with sequences from stool if these had been available for these participants. [377] Future work could include analysis of demographic characteristics and intrahost viral diversity and whether there are differences between vaccine groups and changes in viral diversity from longitudinal samples, for those participants who had more than one sequences obtained during their infective episode.

This work shows how genomic analysis within a vaccine trial can be used to predict real-world vaccine effectiveness and detect potential difference in vaccine efficacy between viral genomic variants. It also demonstrates the challenges of implementing and delivering genomic sequencing within a pandemic vaccine trial. Studies such as this can underpin policy decisions prior to effectiveness data becoming available. This study also demonstrates the need to incorporate sampling, planning, and budgeting for genome sequencing into

protocols, particularly for novel pathogens where the molecular clock and selective pressures are initially unknown.

6 ChAdOx1 nCoV-19 vaccination and the detection of virus and immunoglobulin in the stool of SARS-CoV-2 infected individuals

6.1 INTRODUCTION

Early in the SARS-CoV-2 pandemic, evidence emerged that SARS-CoV-2 infection may involve organs beyond the respiratory tract. Patients with COVID-19 were found to have gastrointestinal symptoms in addition to, or instead of, classical respiratory symptoms or fever. [411-413] Viral RNA was identified in the stool of infected patients. In a minority of stool samples, live virus was isolated, [81,87,411,414-417] raising the possibility of transmission via infected stool. [418] However, the role of the gastrointestinal tract in SARS-CoV-2 infection remains poorly understood.

The relationship between gastrointestinal symptoms and presence of detectable RNA in stool is not clear; symptoms such as nausea, vomiting and diarrhoea do not neatly relate to stool RNA detection [176,411,413,417] and are complicated by other factors such as severe illness, hospitalisation, and administration of antibiotics in COVID-19 patients. Many studies that have included investigation of stool SARS-CoV-2 viral dynamics have occurred in hospitalised patients as these samples are logistically difficult to collect from community settings (especially with public health restrictions in place) where most mild to moderate disease occurs. More recent data suggests that in hospitalised patients at least,

gastrointestinal symptoms are associated with a reduction in both severity and mortality, though this is in contrast with early reports. [114,419]

The SARS-CoV-2 virus (and other beta coronaviruses) enter cells by binding to ACE-2 protein and being taken up by ACE-2 receptors expressed by cells. Genes encoding the ACE-2 protein, plus proteases *TMPRSS2* and *TMPRSS4* that play a role in viral entry are expressed by enterocytes in the small and large bowel. [14] ACE-2 expression has been demonstrated in small bowel biopsy specimens thereby providing a mechanism for the cells of the GI tract to be infected. [419,420] Models of human small intestinal organoids have demonstrated the role of *TMPRSS2* and *TMPRSS4* in SARS-CoV-2 entry into cells, [421] and an increase in both RNA and infectious virus, followed by apoptosis of infected enterocytes after SARS-CoV-2 infection. [422]

It is now well recognised that SARS-CoV-2 RNA is detectable in stool samples with a meta-analysis of studies reporting a mean shedding duration of 12.5 to 17.2 days in stool. [80,417] There have been repeated instances of SARS-CoV-2 RNA detected in stool after pharyngeal RNA is no longer detectable. [411,413] This supports the possibility of active replication in gastrointestinal tissues and not just the passage of RNA swallowed from the respiratory tract. [87,411,414,417]

Less consistent has been the detection of live virus in stool samples. This is in contrast to other sample types such as saliva and upper respiratory tract swabs where live virus has been cultured more readily. [87] *In vitro* work using simulated large intestinal fluid suggests that inactivation of the virus may occur within the colonic lumen explaining why infectious virus may be less readily detected than RNA. [421]

Gastrointestinal immunological responses to SARS-CoV-2 have been observed in stool and tissue specimens. [176,419] Britton and colleagues detected SARS-CoV-2 anti-RBD protein IgA substantially above background level in the stool of 11% of their hospitalised cohort, all of whom had severe COVID-19 disease. However, this group noted that stool IgA titres were lower than serum IgA titres, and that there were no significant differences in infected patients compared with controls. [176] They note that control specimens collected prior to the pandemic exhibited high levels of background signal. [176] This study occurred in a cohort in the US in early 2020, as vaccine trials were being initiated. The ability of parenterally administered vaccines for SARS-CoV-2 to induce a mucosal immune response in the gastrointestinal tract is unknown, but evidence from saliva and upper respiratory tract samples and other pathogens such as typhoid, rotavirus and polio suggest this is plausible. [423-427] As current parenteral vaccines do not provide absolute protection from infection or block viral shedding, efforts to further understanding of mucosal immunology may contribute to alternative or complimentary vaccine strategies.

A number of SARS-CoV-2 vaccines are currently licensed for use in the UK and US (including most recently bivalent formulations), and all use the spike protein as their immunogen except one inactivated whole virus vaccine (Valneva VLA2001). Serum neutralising IgG antibodies, particularly anti-spike antibodies have been shown to correlate with protection from symptomatic SARS-CoV-2 disease. [348] Other antibody isotypes are also likely to play a role in SARS-CoV-2, particularly immunoglobulin A which is detectable in the plasma soon after infection and has neutralising capabilities. [428] Despite mucosal surfaces being the sites of entry and infectious shedding of the SARS-CoV-2 virus, the role of secretory immunoglobulin A (sIgA) is less well established.

During the late Spring of 2020, animal studies of the ChAdOx1 nCoV-19 vaccine were being conducted in parallel with recruitment of participants for human trials, with key safety data required from the animal studies to permit first-in-human dosing. In a vaccine-challenge study of ChAdOx1 nCoV-19 in rhesus macaques, the active vaccine was shown to be immunogenic. [170] Three groups of animals were included: a prime group, a prime-boost group and a ChAdOx1 GFP group as controls. Vaccinated animals showed reduced viral load in lower respiratory tract tissues and no vaccinated animals had pneumonia compared with pneumonia observed in half of the control animals. Nasal shedding of the virus was similar between vaccinated and control animals. When gastro-intestinal tissues were analysed, while no histological differences observed, viral genomic and sub-genomic RNA was higher in prime-boost animals compared with both control and prime-only vaccinated animals (see Figure 6.1). No live virus was detectable from gastrointestinal tissues.

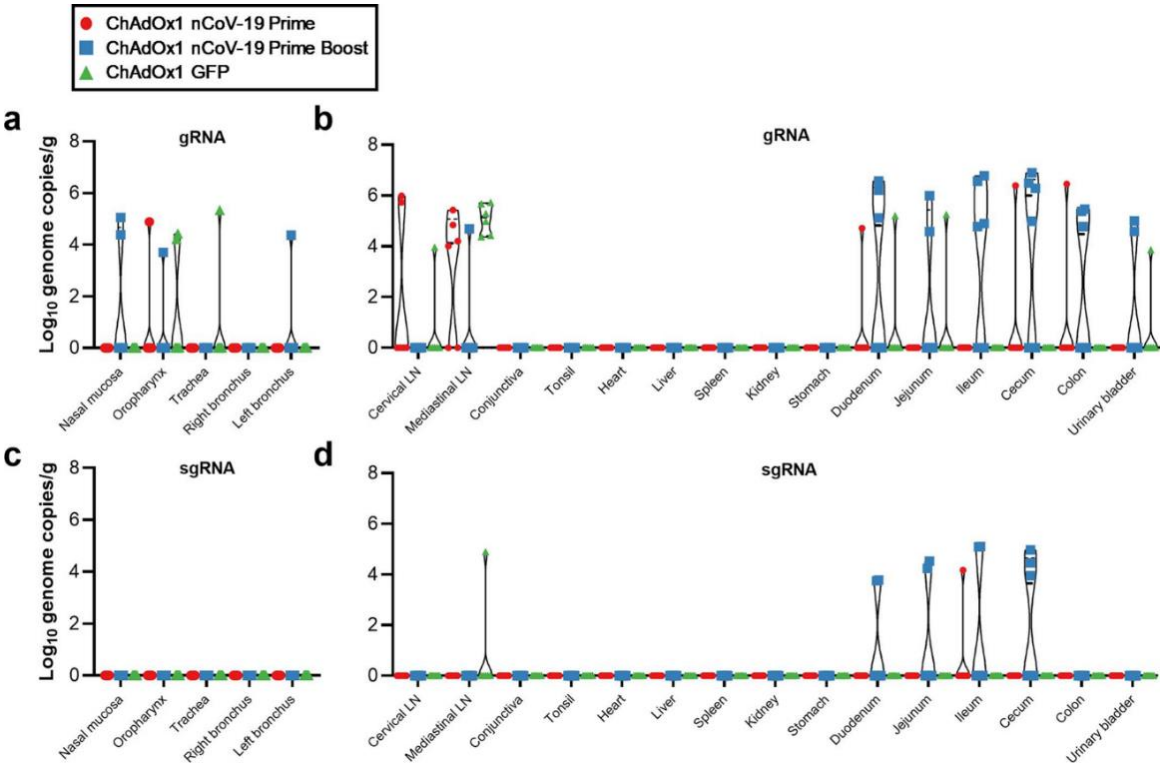


Figure 6.1: Viral load in tissues of rhesus macaques challenged with SARS-CoV-2 at 7 DPI (Days Post Infection). A. Viral gRNA in respiratory tissues excluding lung tissue. B. Viral gRNA in non-respiratory tissues. C. Viral sgRNA in respiratory tissues excluding lung tissue. D. Viral sgRNA in non-respiratory tissues. Figure and legend from extended data in van Doremalen et al. [170] Reproduction licensed under CC BY 4.0. [1]

These findings raised the concern that a two-dose ChAdOx1 nCoV-19 vaccine schedule may promote viral shedding in infected individuals, at a time when the relative contribution of faeco-oral transmission was unknown.

This study was established in response to these animal data and sought primarily to evaluate viral shedding in stool.

The objectives of this study were:

- To measure viral shedding in the stool in SARS-CoV-2 infected trial participants.
- To compare rates of viral stool shedding in control and ChAdOx1 nCoV-19 vaccinated groups.
- To compare stool SARS-CoV-2 viral load in control and ChAdOx1 nCoV-19 vaccinated groups.
- To measure if anti-spike IgG and IgA were detectable in stool of infected participants.
- To evaluate the relationship between immunoglobulin levels, vaccination status and stool PCR result.

6.2 METHODS

Stool samples were collected from participants in the COV001 and COV002 trials (see methods chapter 2) who tested positive for SARS-CoV-2. Comparison of viral shedding between vaccine and comparator arms was an exploratory endpoint for COV001 and a tertiary endpoint of COV002. Optional informed consent for the provision of stool on the receipt of a positive SARS-CoV-2 PCR test was added in Substantial Amendment 9, protocol version 9.0 dated 30 July 2020 for COV001 and in Substantial Amendment 5, protocol version 7.0, dated 18 June 2020 for COV002.

My role was the set up and management of the stool collection pathway, immunoglobulin sample processing and management and analysis of the data.

6.2.1 Participants

During the timeframe of stool sampling participant enrolment and vaccination was ongoing. Participant vaccination status were categorised as:

- Vaccinated if participants had received two doses of the ChAdOx1 nCoV-19 vaccine at least 14 days before the start of the episode of SARS-CoV-2 infection (as judged by the independent Endpoint Committee (see methods chapter 2).
- Partially vaccinated if a participant had received a single dose of ChAdOx1 nCoV-19 vaccine at least 21 days before the start of the episode of SARS-CoV-2 infection.
- Controls if the participant had received at least one dose of control vaccine more than 21 days before the start of the episode of SARS-CoV-2 infection.

Participants who had received a single dose of ChAdOx1 nCoV-19 or control vaccine less than 21 days before the start of the episode of SARS-CoV-2 infection were excluded from analyses. Unless otherwise defined ChAdOx1 nCov-19 dosage groups are considered together. Control participants who had received a single dose of control vaccine were for these analyses considered with control participants who had received two doses of MenACWY. Due to the timing of sampling all participants who gave a stool sample were blinded to their vaccine allocation.

Symptomatic infections were defined as outlined in 4.2.1 Trial methods and data collection.

6.2.2 Sample collection

Sample collection was initiated by blinded site personnel on receipt of positive upper airway SARS-CoV-2 NAAT result for a given participant. Participants that had consented to

providing a stool sample in the event of a positive SARS-CoV-2 NAAT test (regardless of symptoms), were sent instructions on how to request a stool collection kit and arrange courier collection. Participants were asked to provide a stool sample approximately 7 days post onset of symptoms, if present, or positive NAAT test if asymptomatic. Originally plans included the possibility of collecting second samples from participants. Due to the blinded nature of the sampling and how quickly cases of SARS-CoV-2 infection were occurring it was decided that prioritising the samples from as many different participants as possible was optimal to increase the likelihood of having samples from both control and vaccinated individuals. This sub-project occurred during August-November 2020 prior to subject and observer unblinding and efficacy analyses, so there was no prior knowledge of the distribution of vaccinated and unvaccinated individuals within the sample group.

Stool samples were collected from participants' residential addresses by courier due to isolation guidance for all infected individuals in the UK at that time. Samples were collected in sterile containers (Covidien Commode Specimen Collector) and kept cool (2-8°C) in transit using a NanoCool™ cooling system. Samples were shipped to International Health Management Associates (IHMA) in Illinois, USA where they were homogenised and frozen at -80°C without additive. Frozen aliquots were shipped to processing laboratories in the US and UK for SARS-CoV-2 PCR testing, immunoglobulin assays and microbiome analyses (these are ongoing and not included in this thesis).

Symptom information was recorded in electronic e-diaries and study visits during symptomatic episodes for participants that met swabbing criteria (see methods chapter 2).

6.2.3 Sample processing

6.2.3.1 RNA extraction and PCR

SARS-CoV-2 extraction and RT-qPCR analysis was performed by PPD (Wilmington, USA). Briefly, SARS-CoV-2 was detected and quantified in stool samples using an in-house assay designed for multiple human sample types including stool. This assay detected the N gene of SARS-CoV-2 and Rnase P, a human reference gene to verify RNA purification. [429]

Genome copy equivalents (GCE) per gram were calculated as per the formula below:

$$\frac{GCE}{g} = \frac{GCE}{RXN} \times \frac{RXN}{5\mu L\ Eluate} \times \frac{100\mu L\ eluate}{200\mu L\ Stool\ Suspension} \times \frac{1000\mu L\ Stool\ Suspension}{1mL} \times \frac{1mL}{0.1g\ Stool}$$

The assay limits of quantification were a lower limit of quantification of 1.00 x 10^4 GCE/g (average Ct result of 45.00) and an upper limit of quantification of 1.00 x 10^9 GCE/g.

6.2.3.2 Supernatant preparation

Supernatants from stool samples were isolated for ELISA assays by Emily Thornton and Dana Costigan using a method developed by the Thornton group.

Briefly, samples were thawed and approximately 100 mg of sample was transferred to faecal homogenisation tubes (ZR BashingBead™ Lysis Tubes (2 mm)) containing 1 ml antibody isolation buffer. Antibody isolation buffer made from Tris 12.114g, NaCl 8.766g, Urea 60.06g, CaCl2 1.1098g, Citric acid monohydrate 21.014g added to 950mL distilled water, adjusted to pH 8 and then autoclaved. Then, 5g BSA and 0.01g Thimerosal were added to 50mL distilled water and added to the autoclaved 950mL.

Stool samples were then vortexed to homogenise the stool with the buffer. Homogenised samples were then centrifuged. The supernatant was then divided into aliquots which were frozen at -80 °C.

6.2.3.3 ELISA assays

This assay was developed, adapted and performed by Emily Thornton and Dana Costigan from methods described by Britton et al. [176]

Nunc 96-well MaxiSorp™ Plates (Invitrogen, Thermo Fisher Scientific, Waltham, USA) were coated with 50 µl of 2 µg/ml SARS-CoV-2 Spike protein (Recombinant SARS-CoV-2 Spike His Protein, R&D Systems®, Abingdon, UK) suspended in PBS, sealed, and incubated overnight at 4 °C. Plates were then washed with PBS plus 0.05% Tween™-20 (PBS-T) and plates blocked with 3% non-fat milk in PBS-T and sealed for 1 hour at room temperature.

Supernatants were thawed and centrifuged. Samples were added to a 96-well round-bottom plate. Samples were mixed with equal volume PBS-T with 1% non-fat milk and then serially diluted threefold for eight iterations. 100 µl of the diluted samples were added to the washed plate and incubated at room temperature for 2 hours.

Plates were then washed with PBS-T before the relevant detection antibody was added to each well:

100 µl goat anti-human IgG-horseradish peroxidase and conjugated secondary antibody (1: 3000, Thermo Fisher Scientific, Waltham, USA)

100 µl anti-IgA-horseradish peroxidase (1:8000, SouthernBiotech™, Birmingham, USA)

Plates were kept at room temperature for 2 hours. Following a final wash step, plates were developed using 100 µl of TMB-A and TMB-B substrates in equal ratios (BioLegend®, San Diego, US). After 20 minutes, 50 µl of Stop solution (BioLegend®, San Diego, US) was added to terminate the reaction. Plates were read at 450 nm within 30 minutes. On each plate a known highly positive serum sample was included as a positive control.

6.2.4 Sample size

Null hypothesis: Sample size calculation for number of participants required to show that (prime plus boost) vaccinees and controls did not differ in their rate of stool shedding of SARS-CoV-2 RNA during acute SARS-CoV-2 infection.

Alternative hypotheses:

Vaccinees will show increased stool shedding of SARS-CoV-2 RNA (based on animal data showing RNA found in intestinal tissues amongst prime-boost animals)

Vaccinees will show reduced stool shedding of SARS-CoV-2 RNA (based on human data showing reduced upper airway shedding)

Anticipated proportion of RNA shedding in control group:

Data from a heterogeneous array of studies from data collected in 2020 found SARS-CoV-2 RNA to be detected in the stool of 15.3 to 48.1% of infected individuals (Table 6.1).

Table 6.1: Studies investigating stool shedding of SARS-CoV-2 RNA from infected participants.

Reference	% of participants with SARS-CoV-2 RNA detected in stool	n
[413]	15.3%	9/59
[83]	18% (95% CI, 4–43%)	3/17
[430]	36%	93/258
[431]	43% (95% CI, 34-52%)	511/1402 pooled prevalence from meta-analysis of 30 studies (random effects model)
[413]	48.1% (95% CI, 38.3–57.9	68/138 pooled prevalence from meta-analysis of 12 studies (random effects model)

Based on the above data, the range of sample sizes required to have power >80% to detect a difference between the control and vaccine groups of 25-40% when proportion of RNA stool shedding in the control group was 0.2-0.4 are shown in Table 6.2 and displayed in Figure 6.2.

Table 6.2: Sample sizes required for the detection of a range of differences between groups if the proportion of stool RNA shedders in the control group is 0.2-0.4. Power rounded to 2 decimal places.

Power	Number in vaccine group (N1)	Number in control group (N2)	Total sample size (N)	Proportion of shedders in vaccine group (P1)	Proportion of shedders in control group (P2)	Difference between groups
1.00	20	200	220	0	0.4	-0.4
1.00	20	100	120	0	0.4	-0.4
1.00	20	80	100	0	0.4	-0.4
1.00	20	60	80	0	0.4	-0.4
1.00	40	100	140	0	0.4	-0.4
1.00	40	80	120	0	0.4	-0.4
1.00	40	60	100	0	0.4	-0.4
1.00	100	80	180	0	0.4	-0.4
1.00	100	100	200	0	0.4	-0.4
1.00	100	60	160	0	0.4	-0.4
1.00	20	200	220	0	0.35	-0.35
1.00	20	100	120	0	0.35	-0.35
0.99	20	80	100	0	0.35	-0.35
0.98	20	60	80	0	0.35	-0.35
1.00	40	100	140	0	0.35	-0.35
1.00	40	80	120	0	0.35	-0.35
1.00	40	60	100	0	0.35	-0.35
1.00	100	80	180	0	0.35	-0.35
1.00	100	100	200	0	0.35	-0.35
1.00	100	60	160	0	0.35	-0.35
1.00	20	200	220	0	0.3	-0.3
0.98	20	100	120	0	0.3	-0.3
0.96	20	80	100	0	0.3	-0.3
0.93	20	60	80	0	0.3	-0.3
1.00	40	100	140	0	0.3	-0.3
1.00	40	80	120	0	0.3	-0.3
0.99	40	60	100	0	0.3	-0.3
1.00	100	100	200	0	0.3	-0.3
1.00	100	80	180	0	0.3	-0.3
1.00	100	60	160	0	0.3	-0.3
0.97	20	200	220	0	0.25	-0.25
0.90	20	100	120	0	0.25	-0.25
0.87	20	80	100	0	0.25	-0.25
0.83	20	60	80	0	0.25	-0.25
0.99	40	100	140	0	0.25	-0.25
0.99	40	80	120	0	0.25	-0.25
0.97	40	60	100	0	0.25	-0.25
1.00	100	100	200	0	0.25	-0.25
1.00	100	80	180	0	0.25	-0.25
1.00	100	60	160	0	0.25	-0.25

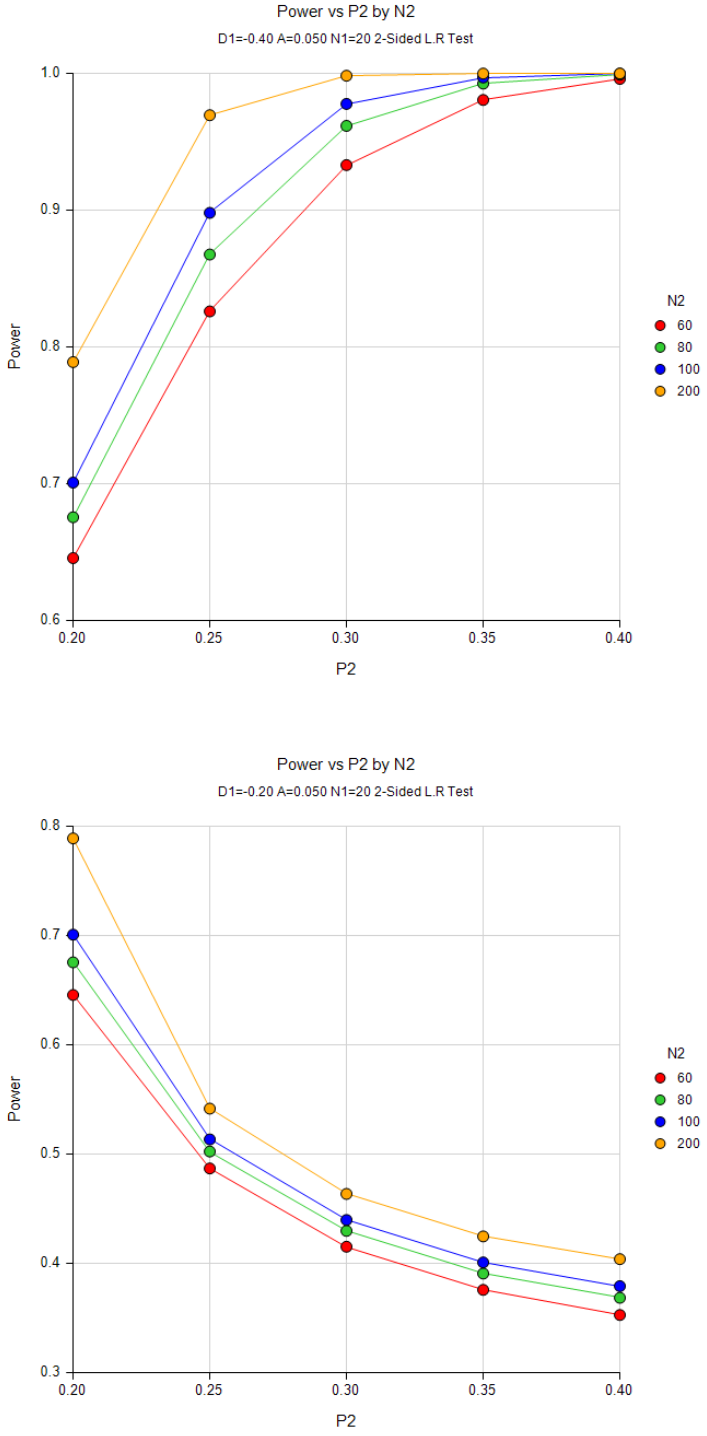


Figure 6.2: Demonstration of change in sample size required to detect differences between groups. Sample size required when vaccinee sample size (N1) is 20, to detect a 40% (shown in top panel) or a 20% (lower panel) difference between groups at a range of P2 proportions. Power is the probability of rejecting a false null hypothesis. N1 (vaccine) and N2 (control) are the number of participants sampled from each population. N is the total sample size, N1 + N2. P1 is the proportion for N1 (vaccinated group) P2 is the proportion for N2 (control group). D1 is the difference P1 – P2 assumed for power and sample size calculations.

If the RNA shedding rate, in the control population is assumed to be approximately 30-40% and the aim of the study is to detect a difference between vaccine and control groups of 30-40% then the sample size of 20 in the vaccinated group and 60 in the control group would provide 93-99% power to detect this difference.

The upper limit of the sample size for the trial was set at 100 samples due to financial and practical constraints and to account for those who may be partially vaccinated during the study period.

Test statistic is the two-sided Likelihood Ratio test, alpha = 0.05, calculations performed using PASS version 16.0.12 (NCSS, Kaysville, Utah, USA).

6.2.5 Statistical analyses

Demographic differences and dose group comparisons were performed using Wilcoxon rank sum testing for continuous variables, Pearson's chi-squared for categorical variables and Fisher's exact tests where group numbers were $\leq$5. Symptom status and stool PCR result were compared using Fisher's exact tests with Benjamini-Hochberg FDR method for multiple hypothesis testing. Factors associated with detectable SARS-CoV-2 RNA in stool were analysed using generalised linear models for binary data. Predictors of stool SARS-CoV-2 viral load were analysed using a multiple linear regression model. Model assumptions were tested and only models that met the assumptions were retained.

6.3 RESULTS

6.3.1 Participant characteristics

Stool samples were collected from 99 participants in the COV001 and COV002 trials between 19th August and 10th November 2020. All participants who had a confirmed SARS-CoV-2 positive NAAT result from an upper respiratory tract sample and who had given consent for stool sampling were asked to provide a specimen.

There were no significant demographic differences between those who gave a stool sample and those who did not during the stool sampling timeframe (Table 6.3) suggesting those who were sampled were broadly representative of the whole cohort.

Table 6.3: Characteristics of participants who were SARS-CoV-2 PCR positive during the stool sampling timeframe. CV disease= cardiovascular disease. Statistical comparison tests: Wilcoxon rank sum test: age, BMI; Pearson's Chi-squared test: sex, healthcare worker, cardiovascular and respiratory disease, episode symptom status; Fisher's exact test: diabetes, vaccination status

Characteristic	**Overall n = 349**	**Stool sample not received n = 250**	**Stool sample received n = 99**	**p value**
Age, Median (IQR)	39 (29, 51)	38 (29, 51)	42 (29, 52)	0.3
Female, n (%)	198 (57%)	148 (59%)	50 (51%)	0.14
Healthcare worker, n (%)	239 (68%)	173 (69%)	66 (67%)	0.6
BMI, Median (IQR)	25.7 (23.2, 29.6)	25.6 (23.2, 30.1)	25.8 (23.1, 29.0)	>0.9
Comorbidities. N (%)				
CV disease	41 (12%)	29 (12%)	12 (12%)	0.9
Respiratory disease	46 (13%)	31 (12%)	15 (15%)	0.5
Diabetes	5 (1.4%)	5 (2.0%)	0 (0%)	0.3
Episode symptom status, n (%)				0.3
Symptomatic	221 (63%)	157 (63%)	64 (65%)	
Asymptomatic	61 (17%)	48 (19%)	13 (13%)	
Unknown	67 (19%)	45 (18%)	22 (22%)	
Primary symptoms	194 (56%)	136 (54%)	58 (59%)	0.5
Vaccine status at PCR+ episode start, n (%)				0.6
Control	213 (61%)	156 (62%)	57 (58%)	
Partially vaccinated	58 (17%)	38 (15%)	20 (20%)	
Vaccinated	66 (19%)	48 (19%)	18 (18%)	
Excluded	12 (3.4%)	8 (3.2%)	4 (4.0%)	

6.3.2 SARS-CoV-2 RNA was detectable in stool from participants of all vaccine status groups

Demographics of trial participants who provided stool samples are shown in Table 6.4. Of all first stool samples, 35 (35%) had detectable SARS-CoV-2 RNA and 22 of these were from unvaccinated participants (63%). Participants who had a stool that was PCR positive for SARS-CoV-2 were younger (36 years old, IQR: 28-47) than those whose stool tested negative (44 years old, IQR: 32-54). Otherwise, there were no significant demographic difference in univariate analyses between those who provided a stool that was PCR positive for SARS-CoV-2 compared with those that were negative.

Table 6.4: Demographics of participants who provided a stool sample. CV disease= cardiovascular disease. Statistical comparison tests: Wilcoxon rank sum test: age, BMI, interval from episode start to stool sampling; Pearson's Chi-squared test: sex, healthcare worker, cardiovascular and respiratory disease, episode symptom status; Fisher's exact test: ethnicity, diabetes, vaccination status. Values in bold indicate a p value<0.05.

	Stool PCR result			
Characteristic	**Overall N = 99**	**Not detected N = 64**	**Detected N = 35**	**p value**
Age, Median (IQR)	42 (29, 52)	44 (32, 54)	36 (28, 47)	**0.036**
Female, n (%)	50 (51%)	35 (55%)	15 (43%)	0.3
Healthcare worker, n (%)	66 (67%)	40 (62%)	26 (74%)	0.2
BMI, Median (IQR)	25.8 (23.1, 29.0)	25.8 (23.1, 29.2)	25.8 (23.8, 28.8)	>0.9
Comorbidities, n (%)				
CV disease	12 (12%)	9 (14%)	3 (8.6%)	0.5
Respiratory disease	15 (15%)	7 (11%)	8 (23%)	0.11
Diabetes	0 (0%)	0 (0%)	0 (0%)	>0.9
Vaccine status at PCR+ episode start, n (%)				0.3
Control	57 (58%)	35 (55%)	22 (63%)	
Partially vaccinated	20 (20%)	11 (17%)	9 (26%)	
Vaccinated	18 (18%)	15 (23%)	3 (8.6%)	
Excluded	4 (4.0%)	3 (4.7%)	1 (2.9%)	
Interval between episode start date and stool sample/days, Median (IQR)	8.0 (6.0, 11.0)	9.0 (6.0, 12.2)	8.0 (5.5, 10.5)	0.092
Primary symptom experienced, n (%)	58 (59%)	29 (45%)	29 (83%)	**<0.001**

Most participants were negative on their baseline serology for SARS-CoV-2 (92/99 were negative (93%), missing data for 4 participants). Participants were mainly healthcare workers and had few comorbidities. The median day of sampling after episode start was similar between positive and negative samples.

78% (77/99) of participants who provided at least one stool sample were either unvaccinated or had only received a single dose of vaccine at the time of their stool sample. Four participants had only received a single dose of ChAdOx1 nCoV-19 or control vaccine less than 21 days prior to becoming PCR positive and were excluded from analyses. There was a non-significant trend towards negative stool in vaccinees.

Participants in the partially vaccinated and vaccinated groups received either low dose or standard dose ChAdOx1 nCoV-19 (Table 6.5, for details see methods chapter 2). In the vaccinated group there was no difference in the dose interval between those who had PCR positive or negative stool.

Table 6.5: Dose groups in vaccinated participants. LD= low dose, SD= standard dose. Fisher's exact test: vaccine dose group, Wilcoxon rank sum test: dose interval

	Stool PCR result		
Vaccinated	**Not detected N = 15**	**Detected N = 3**	**p value**
Vaccine dose group, n (%)			>0.9
LD/LD	1 (6.7%)	0 (0%)	
LD/SD	4 (27%)	1 (33%)	
SD/SD	10 (67%)	2 (67%)	
Dose interval (days), Median (IQR)	68 (53, 80)	68 (64, 76)	>0.9
Partially Vaccinated	**Not detected N = 11**	**Detected N = 9**	**p value**
Vaccine dose group, n (%)			0.7
LD	2 (18%)	0 (0%)	
LD/SD	2 (18%)	1 (11%)	
SD	3 (27%)	3 (33%)	
SD/SD	4 (36%)	5 (56%)	

Four participants provided a second stool sample. For one participant their first sample was delayed in transit and had to be discarded thus their second stool sample was considered as their first, leaving one participant in each of control, partially vaccinated and vaccinated groups. The median time for the second sample was 18 days (IQR: 14-27.5). One partially vaccinated participant had a PCR positive first stool on day 12 after episode onset and then returned a PCR negative stool from day 18. For the other two individuals both their first and second stool samples were PCR negative.

6.3.3 Symptomatic infection was associated with stool sample SARS-CoV-2 PCR positivity

Participants with symptomatic infection were more likely to have a PCR positive stool compared with asymptomatic participants (Fisher's exact test, p=0.0152) or those with unknown symptom status (Fisher's exact test, p=0.0138). There was no difference between those who were asymptomatic and unknown symptom status (Table 6.6).

Table 6.6: Symptom status and stool PCR result. Figures shown in bold are significant <0.05. Participants who had been given their first vaccination (ChAdOx1 nCoV-19 or MenACWY) <22 days prior to their SARS-CoV-2 infection were excluded, n=4, 3 of whom were symptomatic, one was asymptomatic. P values from Fisher's exact test with Benjamini-Hochberg FDR method for multiple hypothesis testing

	Stool PCR result			
Characteristic	**Not detected N = 61 (%)**	**Detected N = 34 (%)**	**Pairwise comparison**	**Adjusted p-value**
Symptom status				
Symptomatic	31 (51%)	30 (88%)	Symptomatic-Asymptomatic	**0.0152**
Asymptomatic	11 (18%)	1 (2.9%)	Asymptomatic-Unknown	1
Unknown	19 (31%)	3 (8.8%)	Symptomatic-Unknown	**0.0138**

There were no participants who were asymptomatic who had a positive stool PCR in the control or fully vaccinated groups. One asymptomatic participant in the partially vaccinated group was stool PCR positive (Figure 6.3).

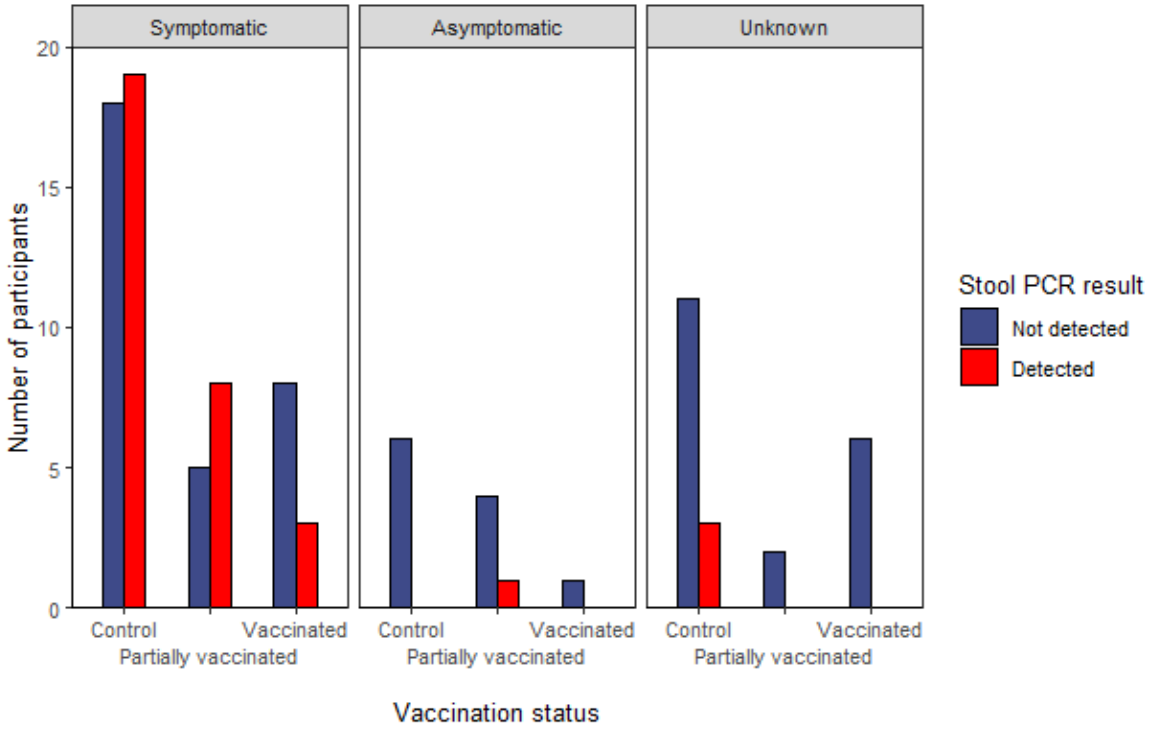


Figure 6.3: Stool SARS-CoV-2 PCR result, symptom and vaccination status. Participants who had been given their first vaccination (ChAdOx1 nCoV-19 or MenACWY) <22 days prior to their SARS-CoV-2 infection are not shown, n=4, 3 of whom were symptomatic, one was asymptomatic.

In a generalised linear model for binary data, participants who had detectable SARS-CoV-2 RNA in their stool were more likely to have a symptomatic infection (adjusted OR: 19.05, 95% CI: 2.02-179.7) and had a shorter interval between PCR positive episode start and stool sampling (adjusted OR: 0.85, 95% CI: 0.75-0.97). Age, sex, and vaccine status were not significant (Table 6.7).

Table 6.7: Generalised linear model for binary data assessing predictors for having detectable SARS-CoV-2 in stool. P values <0.05 are shown in bold.

Variables		**Estimate (adjusted OR)**	**95% CI**	**p value**
Sex	Female	0.41	0.14-1.20	0.103
Age (years)	Per 10 years	0.96	0.92-1.00	0.058
Symptom status				
Asymptomatic		Reference		
Symptomatic		19.05	2.02-179.7	**0.010**
Unknown symptoms		2.01	0.16-24.68	0.582
Vaccination status				
Controls		Reference		
Partially vaccinated		1.21	0.32-4.61	0.782
Vaccinated		0.33	0.07-1.66	0.178
Interval between episode start and stool sample	Days	0.85	0.75-0.97	**0.006**

6.3.4 Gastrointestinal symptoms were common but not related to stool PCR result

Detailed symptom data was available for a subset of participants: those whose global symptoms met swabbing criteria and thus had clinical symptomatic visits and were asked to complete an electronic diary. Nausea, diarrhoea, and abdominal pain were common in participants who donated a stool sample, whilst vomiting was infrequent (Table 6.8). Symptoms were typically mild.

Table 6.8: Gastrointestinal symptoms in SARS-CoV-2 infected participants

Symptom	Highest nausea grade N = 99	Highest vomiting grade N = 99	Highest diarrhoea grade N = 99	Highest abdominal pain grade N = 99
None	25 / 57 (44%)	54 / 57 (95%)	33 / 57 (58%)	38 / 57 (67%)
Mild	20 / 57 (35%)	1 / 57 (1.8%)	17 / 57 (30%)	15 / 57 (26%)
Moderate	11 / 57 (19%)	2 / 57 (3.5%)	6 / 57 (11%)	3 / 57 (5.3%)
Severe	1 / 57 (1.8%)	0	1 / 57 (1.8%)	1 / 57 (1.8%)
Unknown	42	42	42	42

There was no association between symptoms and stool PCR result (Table 6.9). There was no association between sex and gastrointestinal symptoms. The median age was older for participants with diarrhoea ($p<0.05$) and abdominal pain ($p=0.05$) by Wilcoxon rank sum test but not for nausea or vomiting.

Table 6.9: Gastrointestinal symptoms and stool PCR results. P values calculated using Pearson's Chi-squared tests (nausea, diarrhoea, abdominal pain) and Fisher's exact test (vomiting)

	Stool PCR result			
Symptom	Overall, N = 99 (%)	Not detected, N = 64 (%)	Detected, N = 35 (%)	p value
Nausea	32 / 57 (56%)	16 / 28 (57%)	16 / 29 (55%)	0.9
Unknown	42	36	6	
Vomiting	3 / 57 (5.3%)	3 / 28 (11%)	0 / 29 (0%)	0.11
Unknown	42	36	6	
Diarrhoea	24 / 57 (42%)	11 / 28 (39%)	13 / 29 (45%)	0.7
Unknown	42	36	6	
Abdominal pain	19 / 57 (33%)	8 / 28 (29%)	11 / 29 (38%)	0.5
Unknown	42	36	6	

There was no association between vaccination status and gastrointestinal symptoms.

6.3.5 Vaccination status was not associated with stool PCR result

There was no significant difference between vaccination groups and stool PCR result (Fisher's exact test, Table 6.10).

Table 6.10: Stool PCR result and vaccination status. P values calculated using Fisher's exact test with Benjamini-Hochberg FDR method for multiple hypothesis testing

	Stool PCR result			
Characteristic	**Not detected N = 61 (%)**	**Detected N = 34 (%)**	**Pairwise comparison**	**Adjusted p-value**
Vaccine status				
Control	35 (57%)	22 (65%)	Control-Partially vaccinated	0.792
Partially vaccinated	11 (18%)	9 (26%)	Partially vaccinated-Vaccinated	0.225
Vaccinated	15 (25%)	3 (8.8%)	Vaccinated-Control	0.225

There remained no significant difference in stool PCR result with vaccination status if just symptomatic participants were considered (Fisher's exact test, data not shown).

6.3.6 Stool SARS-CoV-2 viral load increased with time from episode start to sampling

The median SARS-CoV-2 viral load detectable in stool increased with the number of ChAdOx1 nCoV-19 doses received (Figure 6.4) however this was not statistically significant (Kruskal-Wallis rank sum test, p=0.2).

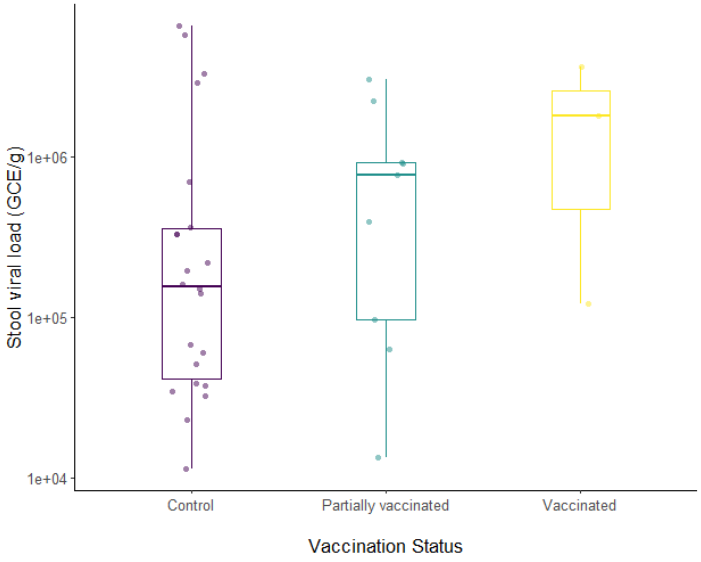


Figure 6.4: Stool SARS-CoV-2 viral load and vaccination status. GCE = genome copy equivalents

For participants with a detectable viral load there was a significant correlation between viral load and time at which the stool sample was taken after the infective episode start (Figure 6.5; Spearman's rank rho=0.44, $p<0.05$).

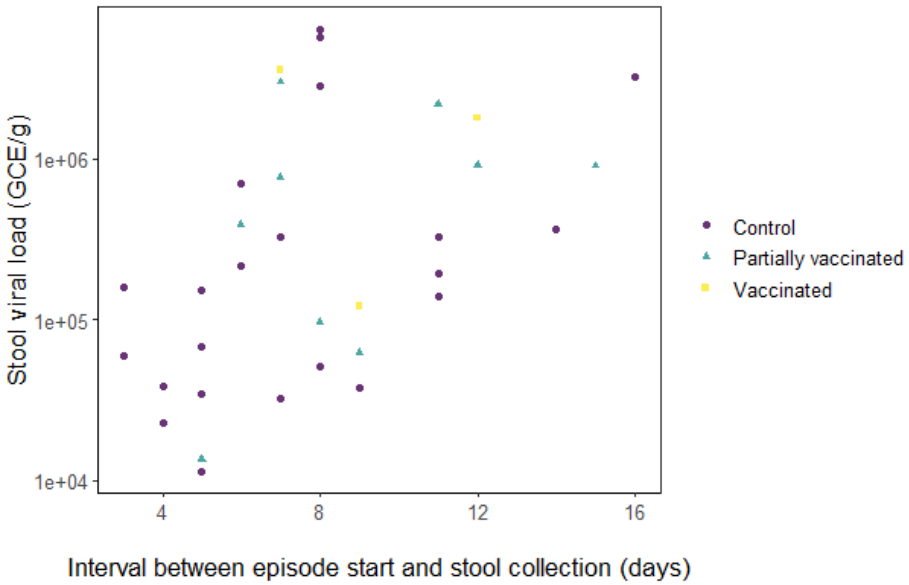


Figure 6.5: Stool SARS-CoV-2 viral load and interval between episode start and stool collection.. GCE = genome copy equivalents

Multiple linear regression was performed to see which variables predicted SARS-CoV-2 viral load in stool. The overall model was significant (Adj. R^2=0.25, F(4, 29)=3.68, p<0.05). Only increasing interval between PCR positive episode start date and stool sampling date was a significant predictor of stool viral load (adj. coefficient: 0.11, 95% 0.03-0.18, p=0.0078), vaccination status and stool IgG were not significant (Table 6.11). Age, sex, healthcare worker were removed from prior models which were not significant.

Table 6.11: Multiple linear regression model of predictors of stool viral load.

Variables		**Adj. coefficient**	**95% CI**	**p value**
Stool anti-spike IgG		1.13	-0.08, 2.33	0.067
Vaccine status				
Controls		Reference		
Partially vaccinated		0.05	-0.53, 0.63	0.863
Vaccinated		0.56	-0.32, 1.45	0.203
Interval between episode start and stool sample	Days	0.11	0.03,0.18	**0.008**

6.3.7 Anti-spike Immunoglobulin G and A was detectable in stool

All vaccination groups had detectable anti-spike IgG and A in their stool samples (Figure 6.6: Stool anti-spike IgG and IgAFigure 6.6). Pairwise Wilcoxon signed rank tests were carried out to compare immunoglobulin levels with vaccination status and stool PCR result. There was no significant difference for either IgG or IgA levels and vaccination status or stool PCR result.

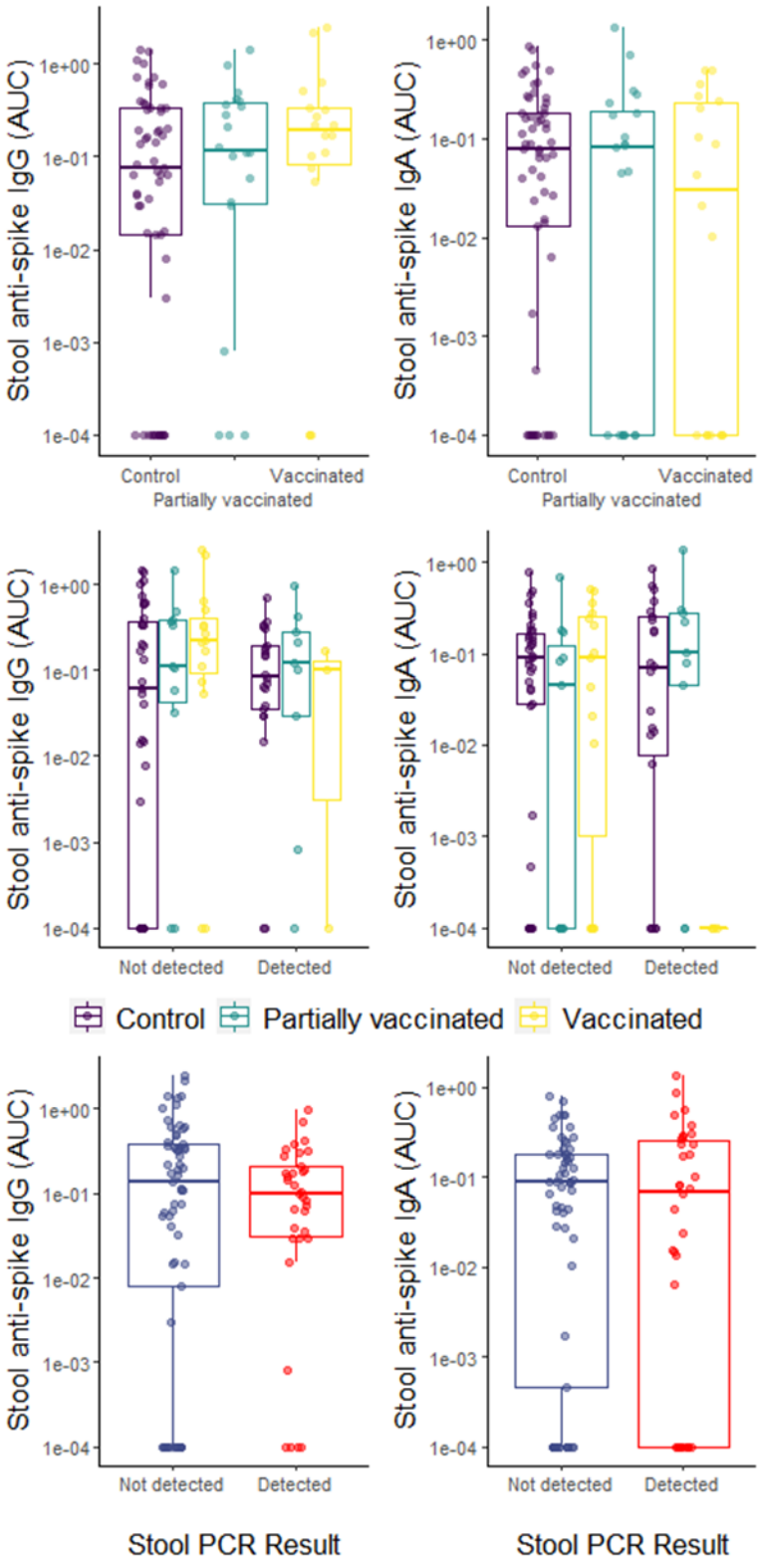


Figure 6.6: Stool anti-spike IgG and IgA by vaccination status (top panel) and vaccination status and stool PCR result (middle panel) and stool PCR result (lower panel). AUC = Area under the curve, y axes are log10 transformed.

Stool anti-spike IgG and IgA levels in some samples appeared to positively correlate but overall correlation was confounded by samples that were negative for either immunoglobulin (Figure 6.7).

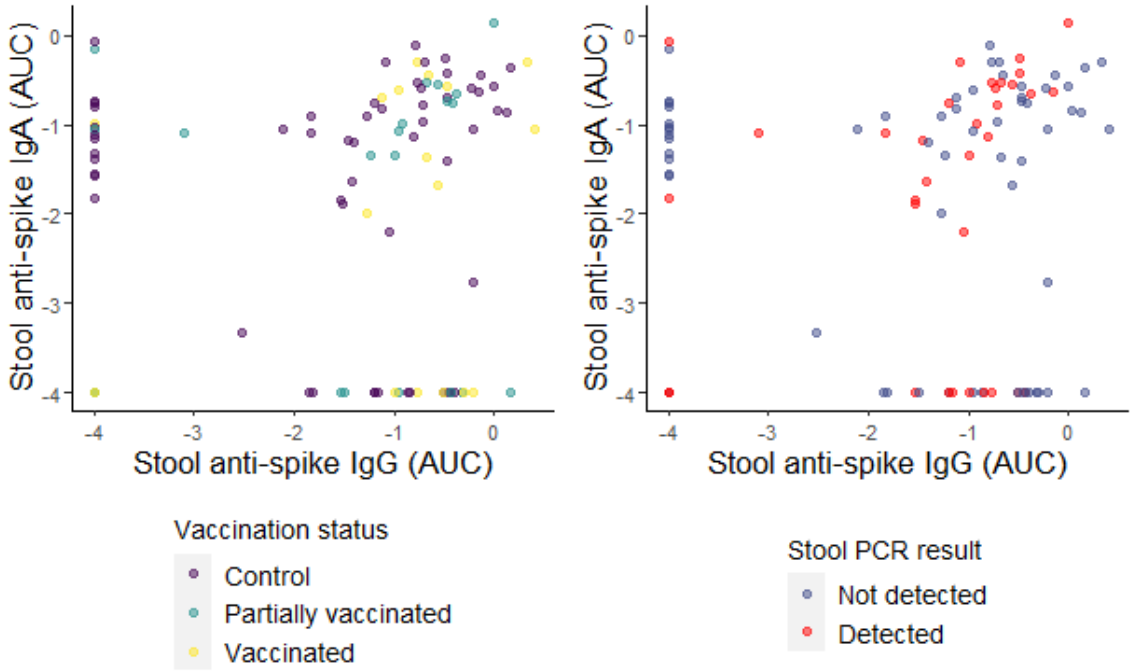


Figure 6.7: Stool anti-spike IgG and IgA by vaccination status and stool PCR result, AUC=area under the curve, both axes are log10 transformed

To see if the ratio of anti-spike IgG and IgA in the stool related to vaccine status or stool PCR positivity, the ratios for each group were compared. There was no significant difference in the anti-spike IgG and IgA ratio between different vaccination groups or stool PCR result (Figure 6.8).

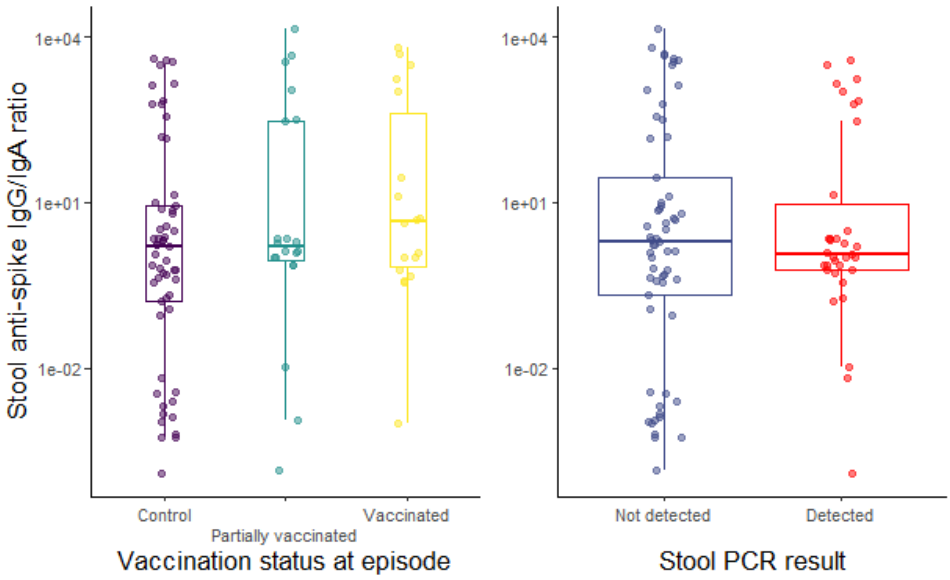


Figure 6.8: Stool anti-spike IgG/IgA ratios and vaccination status (left panel) and stool PCR result (right panel). AUC=area under the curve, y axis is log10 transformed

There was no relationship with time interval between episode start and stool sampling with stool anti-spike IgG or IgA (Figure 6.9).

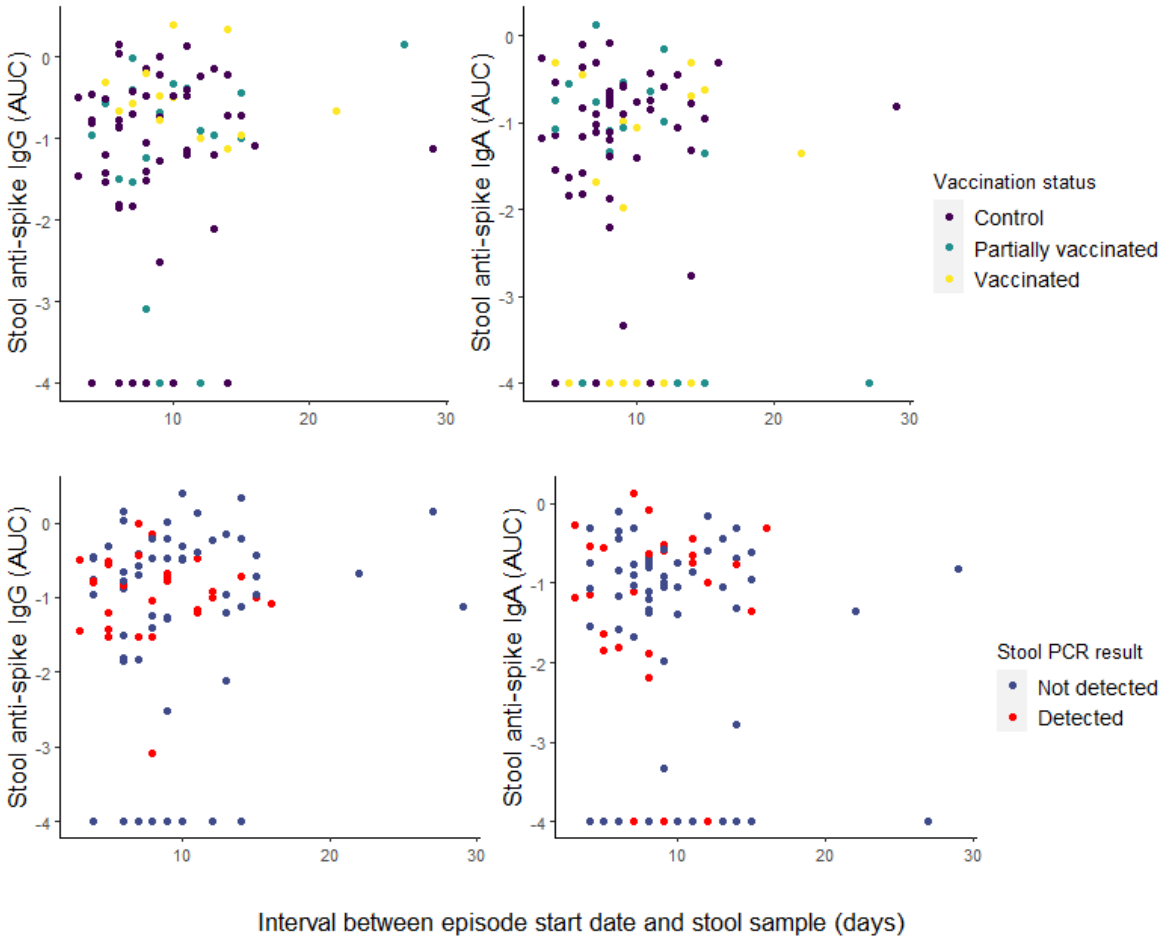


Figure 6.9: Interval between episode start date and IgG and IgA results by vaccination status and stool PCR result.

There was no relationship between stool viral load and stool IgG and stool IgA values (Figure 6.10).

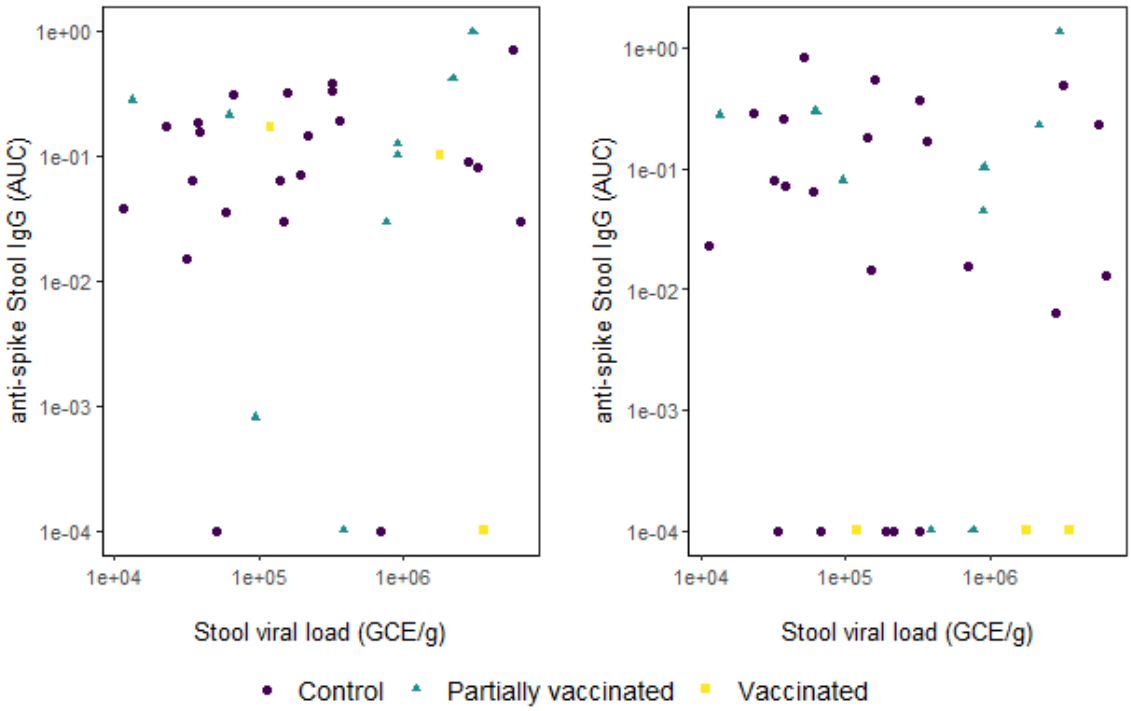


Figure 6.10: Stool viral load and stool IgG and IgA results. AUC=area under the curve, GCE = genome copy equivalents, both axes are log transformed

Three participants who returned second samples which were tested for immunoglobulin levels. For one control participant both of their samples were below the limit of detection for stool anti-spike IgG, whilst they showed an increase in stool IgA between two stool samples taken 33 days apart. Both the partially vaccinated and vaccinated participant had detectable anti-spike IgG with similar values in their first and second samples which were collected close in time (6 days and 1 day respectively). Anti-spike IgA was also detected in the first stool of these two participants however their second samples were both below the limit of detection (data not shown).

Post vaccination serum anti-spike IgG results were available for a subset of vaccinated participants that gave stool samples (Figure 6.11). There was no difference in the serum anti-spike IgG titre post boost vaccination in participants with detectable RNA in stool samples

compared with those that did not (data not shown). There was no correlation between serum anti-spike IgG after either first or second ChAdOx1 nCoV-19 vaccinations and stool anti-spike IgG at the time of infection (Spearman's correlation, p-values = 0.5689 and 0.9714 respectively). There was also no relationship between serum anti-spike IgG post vaccination and stool anti-spike IgA (data not shown).

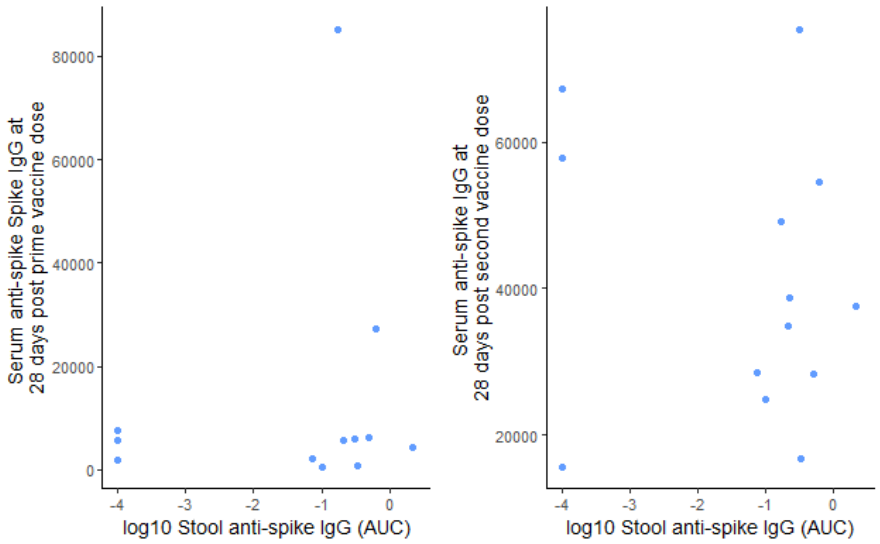


Figure 6.11: Post vaccination serum anti-spike IgG and anti-spike IgG in stool at time of infection. AUC=area under the curve. N=12 for post prime vaccination dose, n=13 for post second vaccine dose, total n of vaccinated in stool donation group=18.

6.4 Discussion

This study demonstrated that 39% (22/57) of unvaccinated, infected individuals had detectable SARS-CoV-2 RNA in their stool, in keeping with the findings of some groups but a smaller percentage compared with other studies. [176,411,413,414,417,432] There was no relationship between vaccination status and detectable SARS-CoV-2 RNA in stool.

Our unvaccinated cohort differ from those in most published and preprint studies from 2020: our participants were typically younger, participants who provided a stool sample had a WHO Covid progression score of no greater than 3 (symptomatic infection, some assistance needed) and therefore were not hospitalised. Of note the proportion of PCR positive stools in our community cohort is greater than the only other published community cohort conducted during a similar period within the UK of 18% (3/17). [83] Previous studies have not included vaccinated individuals as they were typically performed during 2020 prior to SARS-CoV-2 vaccinations becoming available. Of those in the vaccinated cohort, a smaller percentage were found to have detectable SARS-CoV-2 RNA (3/18, 16.7%) in their stool samples compared with unvaccinated individuals, however this was not statistically significant. This could be because vaccination did not reduce shedding in the stool or alternatively because the small numbers in each group did not allow detection of a small difference between the proportions of stool shedders between groups.

Gastrointestinal symptoms were common in our cohort despite mild overall disease. The finding that diarrhoea and abdominal pain were associated with increasing age is in keeping with some reports, [433] but conflicts with others. [176] Overall symptomatic infection was associated with stool RNA detection within a multivariate model, independent of vaccination status. Whilst a range of findings exist in the literature, some studies have found that symptomatic individuals have higher peak upper respiratory tract viral loads compared with asymptomatic individuals and several studies have found that asymptomatic individuals more rapidly clear their viral load compared with those who have symptomatic disease. [80] This swift clearance in asymptomatic cases may explain why there was only a single RNA positive stool in the asymptomatic group.

Whilst many studies have demonstrated prolonged viral shedding in stool, in this multivariate analysis a shorter interval between episode onset and stool sampling was associated with higher likelihood of detection. However, for stools that were PCR positive, viral load was greater with increased interval between episode onset and sampling. These data however come from single cross-sectional samples so individual trajectories in stool viral load cannot be discerned. It may be that early in infection when viral load burden is likely to be high in the upper respiratory tract, a greater proportion of individuals may have detectable RNA in stool secondary to swallowing upper respiratory tract secretions containing viral RNA. Later in the infectious process if active infection is established in the gastrointestinal tract, local replication, independent of the upper respiratory tract, may explain high levels of detectable RNA in stool. Alternatively, these participants may have had much higher viral loads in their stool at an earlier timepoint and the levels detected represent a reduction in viral load. Greater granularity would have been obtained by collecting samples at multiple timepoints from each individual, and additionally from collecting symptom data from all participants.

There was no significant difference in quantitative stool viral load between vaccine status groups. This may be due to the small number of individuals that had detectable RNA in their stool that were partially or fully vaccinated as the trend in the data from these small groups suggest that stool viral load may, counterintuitively be increased with vaccination. Were these three individuals who were fully vaccinated but had PCR positive stool unusual in some way? Post-vaccination serology was available for two, both of whom mounted a good serological response suggesting that their infection was not a result of a failure to respond to the vaccine, and from one a relatively low Ct value from their upper airway sample suggests a high viral burden within the individual. It may be that these three may

simply had a high burden of infection which may have been necessary to infect individuals who had some pre-exposure immunity. The results from stool IgG also show a similar pattern, an increasing level with each dose of vaccine.

It is possible that this is in keeping with the ChAdOx1 nCoV-19 vaccinated non-human primate data. With so few participants in each group this conclusion cannot be drawn; however, it is reasonable to suggest an alternative explanation for these findings as well as a possible mechanism by which enhanced gastrointestinal infection in vaccinated individuals may occur.

Infection in those who have received SARS-CoV-2 vaccination has, at a population level, been associated with lower viral loads from respiratory samples for variants seen early in the pandemic. [434,435] A logical hypothesis from this may be that vaccination would reduce the viral load observed in stool. As the pandemic has progressed it has become clear that immunity conferred by SARS-CoV-2 vaccination is partial rather than complete and interpretation of SARS-CoV-2 viral kinetics is exceptionally challenging; it is confounded by multiple individual and population factors including immunity, pandemic wave timing and variant dominance. [320,358,359] Whilst our data could point to vaccination being associated with excess gastrointestinal shedding, it is also compatible with these three fully vaccinated participants having particularly high systemic viral loads with consequent spillover into the gastrointestinal tract, rather than vaccination itself producing enhanced infection in the gastrointestinal tract.

Explanation of the anti-spike immunoglobulin results is more challenging as these results could result from both vaccination and current infection in vaccinated individuals, and this may vary over time since the start of infection. In unvaccinated participants, a wide

range of stool IgG levels were recorded but there was an observable median stool IgG AUC increase with each dose of ChAdOx1 nCoV-19 vaccine which may be due to the immune stimulation of the vaccine in addition to the current infection. Conversely there was no pattern observed between stool IgA expression and vaccination; this may be because vaccine doses have produced a bias towards IgG response or that stool is a suboptimal specimen from which to discern the activity of IgA at the mucosal surface. There was no relationship between levels of stool IgG or A and stool viral loads which may suggest that once infected, immunoglobulin levels did not affect viral shedding or alternatively, the sensitivity for immunoglobulins in stool, particularly IgA may be low, as has been observed for saliva. [174] Other groups have found high background levels in pre-pandemic control samples. [176] It may also be possible that the predominant T_H1-type response induced by ChAdOx1 nCoV-19 vaccination influences an individual's relative expression of IgG and IgA in the event of infection. For example, vaccine directed isotype switching to IgG1 or IgG3 isotypes which may mean that the measurable IgA may be different in unvaccinated individuals in whom their first presentation of the spike antigen was via natural mucosal infection compared with those who receive a parenteral vaccine. The lack of relationship between post-vaccination serum timepoints and stool anti-spike IgG at the time of infection in vaccinated participants is challenging to interpret because these samples are not contemporaneous, these data are further complicated by infection as a further immune stimulus in addition to vaccination at the point of stool collection. Alternatively, serum IgG post vaccination may not be a good predictor of IgG in stool at the time of infection. Analysis of serum taken at infection in these individuals would aid in the interpretation of these data, these samples are available but have not yet been analysed.

There are several limitations to this study. This study showed that a smaller proportion of vaccinees shed SARS-CoV-2 RNA in their stool compared with controls, but this was non-significant. This study was underpowered for the detection of a small difference between groups. This study only had 35% power to detect a 20% difference between groups with shedding in the control group at 40%. Detection of a smaller effect size (e.g. 20%) would have required a total sample size of fully vaccinated participants and controls 2-3 times larger than this study with a higher number of vaccinees. Ideally, the sample size would have included more vaccinees: for example, if there had been 100 participants in each group, there would have been 88% power to detect 20% difference between groups if the proportion of RNA detected in the control group was 40%. However, it is unclear in this context what, if any, size of difference in shedding between groups is clinically meaningful. Due to staffing constraints at sites and the need to protect the blinded primary endpoint analyses, these stool samples were collected in a blinded manner. This is indicated in the random composition of the groups. For these reasons, single samples from more individuals were prioritised rather than multiple samples from fewer participants to maximise likely numbers in each group. Due to the timing of the specimen collection some participants were only partially vaccinated (had received one vaccine and had sufficient time since that dose to have mounted an immune response prior to SARS-CoV-2 episode start) reducing the number of participants in each group, however this did allow this group to be included in the analyses. Participants from different dose groups were included in this stool study, however the dose interval is likely to be more relevant than the dose in terms of immunogenicity and this did not differ between those in whom stool had RNA detected or not. [276] When the study was set up the viral and antibody kinetics in the stool were unknown so seven days following episode onset

or PCR positivity was chosen as a pragmatic timepoint. It is likely that the optimal timings for assessment of RNA and immunoglobulins are different.

With respect to immunoglobulin analysis at the gastrointestinal mucosal surface, stool is a complex and therefore noisy sample, however the alternative, biopsy specimens, are invasive and difficult to obtain. The analysis of the immunoglobulin findings of this project was complicated by the original objective being to determine viral shedding in infected participants. No baseline samples were obtained from participants, and control samples in general from uninfected unvaccinated and uninfected vaccinated individuals were not sought. Therefore, whilst acknowledging findings of high pre-pandemic and uninfected control background levels of IgA in stool and saliva, it is possible to cautiously interpret the control participants' results as their immune response to their current infection. [174,176] However, the results of those who had received one or more vaccine doses are more complex to attribute, as they may represent the immune consequence of the vaccine or infection, or a combination of these. Lastly, the samples were chilled and transported for several days prior to being processed and frozen at -80°C. This may have influenced PCR and immunoglobulin results, though there is evidence from PCR studies of upper respiratory tract samples that samples have reliably amplified across a range of storage temperatures and intervals to processing. [436,437]

There are, however, strengths to this project. This was a moderately sized, randomised controlled sub-study to look at stool PCR results and the effect of viral vectored COVID-19 vaccination on stool shedding. It was conducted prior to the emergence of the Alpha variant in the UK, so although the RNA isolates from stool have not been sequenced the SARS-CoV-2 viruses causing these infections were likely not to be variants of concern which would

have further complicated the analysis. Further work could include the measurement of cytokines in stool and infectivity assays to look for live virus in stool samples from vaccinated individuals. Microbiome analyses on these samples are ongoing to look at the associations between the composition of microbiota, viral shedding, and vaccination response.

Overall, the results from this project are inconclusive but suggest that ChAdOx1 nCoV-19 may influence both stool SARS-CoV-2 shedding and the expression of immunoglobulin in stool. The results from the stool components of the ISARIC and INSTINCT studies, as well as further mucosal studies including the ChAdOx1 nCoV-19 vaccine are awaited and will provide further complimentary data on the gastrointestinal involvement and mucosal immune response in SARS-CoV-2 infections in a range of disease severities and vaccine combinations.

7 Concluding remarks and future directions

The SARS-CoV-2 virus has resulted in the biggest pandemic of the last century, and approaching three years since first identified, transmission continues. Scientific knowledge on SARS-CoV-2 and COVID-19, the disease it causes, has accumulated at a remarkable rate due to the focus of the global scientific community on a single pathogen. Despite the fastest development and deployment of multiple effective vaccines in history, the SARS-CoV-2 pandemic has had wide-ranging health, social and economic consequences for the global community with an uncertain future trajectory.

The data included in this thesis was generated at an early stage in the pandemic in the UK, with participants at recruitment typically being SARS-CoV-2 and vaccine naïve. The current situation in the UK is very different. By July 2022, 73.5% of UK adults had received at least a primary two dose SARS-CoV-2 vaccine schedule and many have also received one or two additional booster doses. [438] There have been multiple waves of infection and ongoing high levels of transmission since these data were collected including infections caused by Delta, and more recently, Omicron variants. It has become clear that vaccine and infection mediated immunity is relative and wanes rapidly. The consequence of this is a heterogeneous population with dynamic immunity and exposure status making observational and randomised trials increasingly difficult. The pandemic and our knowledge of SARS-CoV-2 has evolved however many of the fundamental questions remain similar, particularly the impact (and future) of vaccination on transmission and disease, duration of immunity and the effect of new SARS-CoV-2 variants.

7.1 SUMMARY OF PRINCIPAL FINDINGS

Pre-clinical animal data suggested that SARS-CoV-2 vaccination may confer protection from lower airway disease but not upper airway infection. [170,171] In this thesis via longitudinal NAAT sampling of vaccine trial participants I sought to establish the impact of ChAdOx1 nCoV-19 vaccination on SARS-CoV-2 viral parameters by examining individuals who became infected within a randomised controlled vaccine trial.

7.1.1 Longitudinal sampling of vaccine trial participants

In Chapter 3 I describe symptomatic and asymptomatic upper airway sampling from ChAdOx1 nCoV-19 vaccine trial participants, who were typically working-age, white adults, of whom 65% were employed in health and social care. Over the study period over 10% of participants returned at least one SARS-CoV-2 NAAT positive swab. Individuals who tested positive were more likely to be younger and work in health or social care. High rates of infection in participants during the study period mirrored national data. I was able to demonstrate high adherence to intensive upper airway sampling over the course of a year with adherence being higher in older age groups and in female participants.

7.1.2 Application of longitudinal sampling to investigate symptom status, viral load and duration of PCR positivity and vaccination

By sampling asymptomatic individuals, this work confirmed that whilst ChAdOx1 nCoV-19 vaccine-mediated immunity protected against symptomatic disease [275] infection still occurred amongst vaccinated individuals. Infections in control participants were more likely to be symptomatic, be associated with a primary symptom (fever $\geq$37.8°C, cough, shortness of breath, anosmia and ageusia) and overall were more severe as measured by the WHO

progression score (Chapter 4). Correspondingly, for participants with symptomatic infection, cycle thresholds (an inverse proxy for viral load) were lower and duration of NAAT positive swabs was longer in controls compared with vaccinated participants but there was no difference in either parameter for those with infections that were asymptomatic or of unknown symptom status.

7.1.3 Viral parameters and vaccine efficacy by SARS-CoV-2 variant

Viral samples from infected trial participants were prospectively collected. Sequencing was supported by COG-UK consortium partners. During the trial period studied, the two most common lineages identified in participants were B.1.177 and B.1.1.7 (Alpha). There were few demographic differences between participants who became infected with one variant compared with the other, though a greater proportion of B.1.177 infections were in healthcare workers compared with B.1.1.7. Due to the timing of the B.1.1.7 wave, infections caused by this variant were later after second vaccinations compared with B.1.177 infections.

When the B.1.1.7 (Alpha) variant emerged during the study period, vaccine efficacy estimates for two different variants (B.1.177 and B.1.1.7) could be calculated (Chapter 5). These figures are caveated by the fact that trial participants are not randomised to infection by different variants: variants emerge in different epidemiological and temporal contexts and thereby limit direct statistical comparison between variants. The vaccine efficacy estimate was lower for primary symptomatic infections caused by sequence-confirmed B.1.1.7 compared with B.1.177 (58.3% (95% CI 29.1-75.5), and 80.7%, 95% CI 67.1-88.6, respectively), but it was not abolished. This result echoes *in vitro* evidence of reduced neutralisation from vaccine induced antibodies for B.1.1.7 compared to the Victoria lineage.[57]

7.1.4 Genomic analyses by vaccination status

There were very few nucleotide and amino acid mutations at consensus level in both B.1.177 and B.1.1.7 samples due to the low rate of viral evolution of SARS-CoV-2 and there were no differences between sequences retrieved from controls or vaccinees. Intrahost viral diversity was seen in 75% of high viral load samples but with a small number of iSNVs per sequence, with no difference between controls and vaccinees in median number of iSNVs, mean allele frequency (MAF) or cumulative MAF. iSNVs were distributed throughout the genome.

7.1.5 SARS-CoV-2 infection and the gastrointestinal tract

The role of the gastrointestinal tract in SARS-CoV-2 infection remains unclear. Sampling stool from a subset of infected trial participants allowed insights into stool shedding (Chapter 6). These data showed that mild gastrointestinal symptoms were common in our participants but unrelated to detectable SARS-CoV-2 RNA in stool. Participants who experienced symptomatic infection were more likely to have SARS-CoV-2 RNA detected in their stool samples. Vaccination status was not associated with detectable stool RNA. Stool IgG and IgA were detectable but not related to stool RNA result.

7.2 LIMITATIONS

The work presented in this thesis has limitations which must be appreciated in context. In a pandemic scenario there is the necessity to maximise knowledge gain in the most efficient and feasible way. It is equally important to recognise the hierarchy of priorities inherent in a phase 2/3 efficacy trial. Only basic understanding of the fundamentals of the pathogen existed at trial inception (for example: dynamics, disease, testing, and transmission)

therefore early versions of the protocol reflect the information available at that time. These factors, coupled with the stringent and essential requirements of a randomised controlled vaccine trial make studying additional endpoints within an evolving pandemic setting more difficult.

One of the strengths of the data is the detail with which symptoms were collected, however the assessment of asymptomatic symptom status could have been improved. Prospective, longitudinal symptom records from first positive test (as performed in observational studies conducted in contained settings [126,439]) would have allowed distinction between pre-symptomatic and infected individuals who remained asymptomatic. However, this approach may have been problematic as it may have affected the way in which participants presented to the symptomatic pathway which captured the primary endpoint for the trial. An alternative approach might have been asking all participants about symptoms fourteen days after their first positive test as a compromise between prospective questioning and retrospective recall bias, though in this scenario participants might have encountered a barrier to reporting symptoms if they had not divulged them to their study team previously. The contribution of asymptomatic and pre-symptomatic infected individuals to transmission within communities is well recognised [129] so for future vaccine studies or pandemics this issue will need to be addressed.

The regular longitudinal self-sampling was only requested from participants once a week which may have missed some infections. Equally, by only asking for weekly sampling, assessment of cycle threshold changes and duration of PCR positivity between vaccination status groups was relatively coarse. This could have been overcome either by asking participants to test twice weekly but any advantage in doing so may have been offset by

reduced adherence, alternatively for any participants testing positive they could have been asked to perform intensive sampling (for example alternate days) for 2 weeks after first testing positive to gain better resolution of viral dynamics during the acute infection. Both strategies would have had considerable financial, practical, and logistical implications but may have improved the amount of information gained.

Symptomatic visits were, by necessity, performed at sites to allow clinical review, testing, and follow-up care to be provided by local teams. The disadvantage of this approach was that clinical sites used a wide range of NAAT assays, not least to provide redundancy in case of assay or reagent supply chain issues during the pandemic. This provided analogue NAAT results to the trial from symptomatic swabs, precluding comparison of Ct values to those from weekly self-swabs. This could have been overcome by taking an additional sample at the symptomatic visits for processing via the Lighthouse laboratories, however the burden on clinical delivery teams was high and this could not be incorporated into the protocol.

Retrieving positive samples obtained from participants for sequencing was established early in the trial. The trial was very fortunate to receive the support of the COG-UK consortium to enable this. However, retrieval of sequences required a labour-intensive process involving the recovery of samples from high throughput settings often by individuals unrelated to the trial. Samples were then sequenced using different platforms and protocols which limits more in-depth genomic analysis. Dedicated samples for genomic analysis to be processed by a single facility on one platform and protocol would be the ideal for a future study of a similar type. This could be achieved by providing each participant with a sample pack for days 1, 3, 5, 7 and 10 to start taking as soon as they receive a positive result and

then returned to site. As these results are not time sensitive, they could be batched and processed thereby reducing running costs.

Stool sampling was performed in a blinded manner to look at shedding post vaccination however these samples were used to explore other questions. Unfortunately, the samples included relatively few who had been fully vaccinated and if resources had allowed, an unblinded team recruiting to this sub-study may have permitted directed sampling to allow more vaccinees to be included. Baseline sampling of stool at post vaccination follow up of such a large trial population in case of subsequent infection would have been unfeasible but would have provided information as to the presence of gut anti-spike IgG and A following vaccination alone. Serial sampling post infection including at 6 months post infection would also have elucidated the relationship between episode onset and stool viral load and long-term shedding with vaccination.

7.3 FURTHER WORK

There are other opportunities to explore with the data collected. The large dataset of longitudinal samples allows further analyses including reinfections, assessment of false positives and the contribution of asymptomatic sampling to estimates of vaccine efficacy.

Lastly, stool microbiome analysis has been performed by collaborators so it will be possible to integrate these results with clinical data to assess potential associations between microbiota, detectable SARS-CoV-2 RNA, and ChAdOx1 nCoV-19 vaccination status.

7.4 IMPLICATIONS AND FUTURE DIRECTIONS

SARS-CoV-2 infections caused by Omicron variants BA.4 and BA.5 infections were high during summer 2022 in the UK, and whilst death rates have not returned to the pre-vaccine era, there have been a number of periods of increased hospitalisations in patients testing positive for COVID-19 so far during 2022. [440] This is despite high vaccine uptake and detectable anti-SARS-CoV-2 spike IgG antibodies in almost all of the adult population. [440] The currently familiar SARS-CoV-2 vaccines are only the first generation of this technology. Bivalent vaccines incorporating the Wuhan and Omicron strains have recently been licensed and will be used in 2020 autumn UK booster campaigns. Recent data suggest that both vaccine and infection history can have a significant impact on the immune response to future infections, [441] and in all likelihood, future vaccinations. There is also much still to be learnt about infection and immune response at the mucosal surface and how this might be attenuated or abolished by vaccination.

Vaccine efforts to increase mucosal immunity and disrupt transmission from upper airway viral shedding, as well as induce systemic immunity have included intranasal delivery of viral vectored vaccines. Some candidates have shown promise in small preclinical trials, [442,443] but human phase 1 trial results are awaited. Should these prove suitably immunogenic, the assessment of efficacy of transmission blocking vaccines in humans may be indicated. This would be challenging but could be attempted in a high transmission setting by using some of the methods in this thesis such as regular asymptomatic sampling and viral load measurement to assess vaccine candidate impact.

Further SARS-CoV-2 variants are inevitable, making sequencing a key part of many future studies as determining the infecting variant will be important information with which

to interpret the immune response. The role of intrahost viral diversity and its relationship to outcomes and immunophenotypes is an area which may help understanding in how variants emerge. Some vaccine manufacturers have made variant specific vaccines, but, even with rapid genome acquisition and manufacturing processes, this is not a sustainable approach. It does not typically produce vaccines to market within a variant wave and as variants become more antigenically distinct the response elicited may not provide broad protection. Vaccine efforts have pivoted towards multi-valency, incorporating both ancestral and more recent strains by updating existing vaccines and pan-variant coronavirus vaccines employing multiple antigenic targets. [444]

The role of the gastrointestinal tract in SARS-CoV-2 infection remains unclear. SARS-CoV-2 RNA can remain detectable in stool in adults for months after initial mild to moderate infection. [445] Accumulating evidence implies that the gut may have a role in pathologies related to COVID-19. A recent study demonstrated the presence of detectable SARS-CoV-2 RNA in stool and antigenaemia from patients with the multisystem inflammatory syndrome in children (MIS-C, also known as paediatric inflammatory multisystem syndrome, temporarily associated with COVID-19, PIMS-TS). This condition develops after the acute infection when virus is often not detectable in respiratory samples. This suggests anatomically distinct viral persistence long after initial exposure to the virus (seen in other viruses including Ebola). [446] Another study performed in Hong Kong showed gut microbiome dysbiosis in patients with post-acute COVID-19 syndrome (PACS or long COVID). [447] There is much that remains to be understood about gut involvement and contribution to pathology in SARS-CoV-2, it may however in time provide an additional avenue for therapeutic or preventative approaches.

7.5 FINAL WORDS

The studies included in this thesis demonstrate the wealth of information that can be achieved about a novel pathogen and vaccination by intensively sampling vaccine trial participants. Animal models were good predictors of what occurred in the human airway following vaccination. The findings in this thesis in terms of reduced viral load and duration of PCR positive status after vaccination have been demonstrated in large population studies after vaccination roll out demonstrating the robustness of this work. The COVID-19 pandemic has provided a unique research context in terms of funding, scientific focus, and large numbers of highly motivated and engaged participants. It has also emphasised the probability that another human pandemic or 'Disease X' will emerge. Understanding a new pathogen relies on high-quality, systematically collected data. The work presented in this thesis demonstrates an approach of longitudinal sampling encompassing pathogen detection and sequencing that can inform and shape research proposals to another pandemic scenario to allow a swift, comprehensive, and effective response to Disease X.

8 References

1. Creative Commons. CC BY 4.0. Available at: https://creativecommons.org/licenses/by/4.0.

2. Wu F, Zhao S, Yu B, et al. A new coronavirus associated with human respiratory disease in China. *Nature* 2020; **579**(7798): 265-9.

3. Oude Munnink Bas B, Sikkema Reina S, Nieuwenhuijse David F, et al. Transmission of SARS-CoV-2 on mink farms between humans and mink and back to humans. *Science* 2021; **371**(6525): 172-7.

4. Hobbs EC, Reid TJ. Animals and SARS-CoV-2: Species susceptibility and viral transmission in experimental and natural conditions, and the potential implications for community transmission. *Transboundary and Emerging Diseases* 2021; **68**(4): 1850-67.

5. Holmes EC, Goldstein SA, Rasmussen AL, et al. The origins of SARS-CoV-2: A critical review. *Cell* 2021; **184**(19): 4848-56.

6. Pekar JE, Magee A, Parker E, et al. The molecular epidemiology of multiple zoonotic origins of SARS-CoV-2. *Science* 2022; **377**(6609): 960-6.

7. Gorbalenya AE, Baker SC, Baric RS, et al. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nature Microbiology* 2020; **5**(4): 536-44.

8. Rota Paul A, Oberste MS, Monroe Stephan S, et al. Characterization of a Novel Coronavirus Associated with Severe Acute Respiratory Syndrome. *Science* 2003; **300**(5624): 1394-9.

9. Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus ADME, Fouchier RAM. Isolation of a Novel Coronavirus from a Man with Pneumonia in Saudi Arabia. *New England Journal of Medicine* 2012; **367**(19): 1814-20.

10. Tang D, Comish P, Kang R. The hallmarks of COVID-19 disease. *PLOS Pathogens* 2020; **16**(5): e1008536.

11. Lu R, Zhao X, Li J, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The Lancet* 2020; **395**(10224): 565-74.

12. Rastogi M, Pandey N, Shukla A, Singh SK. SARS coronavirus 2: from genome to infectome. *Respiratory Research* 2020; **21**(1): 318.

13. Folegatti PM, Bittaye M, Flaxman A, et al. Safety and immunogenicity of a candidate Middle East respiratory syndrome coronavirus viral-vectored vaccine: a dose-escalation, open-label, non-randomised, uncontrolled, phase 1 trial. *The Lancet Infectious Diseases* 2020; **20**(7): 816-26.

14. Schoeman D, Fielding BC. Coronavirus envelope protein: current knowledge. *Virology Journal* 2019; **16**(1): 69.

15. Dai L, Gao GF. Viral targets for vaccines against COVID-19. *Nature Reviews Immunology* 2021; **21**(2): 73-82.

16. Brant AC, Tian W, Majerciak V, Yang W, Zheng Z-M. SARS-CoV-2: from its discovery to genome structure, transcription, and replication. *Cell & Bioscience* 2021; **11**(1): 136.

17. Zhang J, Xiao T, Cai Y, Chen B. Structure of SARS-CoV-2 spike protein. *Current Opinion in Virology* 2021; **50**: 173-82.

18. Cai Y, Zhang J, Xiao T, et al. Distinct conformational states of SARS-CoV-2 spike protein. *Science* 2020; **369**(6511): 1586-92.

19. Hoffmann M, Kleine-Weber H, Schroeder S, et al. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* 2020; **181**(2): 271-80.e8.

20. Cascella M RM, Aleem A, et al. Features, Evaluation, and Treatment of Coronavirus (COVID-19). Updated 2022 Jun 30. https://www.ncbi.nlm.nih.gov/books/NBK554776/ Figure contributed by Rohan Bir Singh, MD.

21. Hillen HS. Structure and function of SARS-CoV-2 polymerase. *Current Opinion in Virology* 2021; **48**: 82-90.

22. Gordon CJ, Tchesnokov EP, Woolner E, et al. Remdesivir is a direct-acting antiviral that inhibits RNA-dependent RNA polymerase from severe acute respiratory syndrome coronavirus 2 with high potency. *Journal of Biological Chemistry* 2020; **295**(20): 6785-97.

23. V'kovski P, Kratzel A, Steiner S, Stalder H, Thiel V. Coronavirus biology and replication: implications for SARS-CoV-2. *Nature Reviews Microbiology* 2021; **19**(3): 155-70.

24. Johns Hopkins University & Medicine. Coronavirus Resource Centre. 2022. https://coronavirus.jhu.edu/ (accessed 8th September 2022).

25. The Economist and Solstad (corresponding author) S. The pandemic's true death toll [online]. 2021. https://www.economist.com/graphic-detail/coronavirus-excess-deaths-estimates (accessed 3rd February 2022).

26. ONS. Coronavirus (COVID-19) latest insights: Deaths. 2022. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/articles/coronaviruscovid19latestinsights/deaths (accessed 13th January 2022).

27. Our World in Data. 2021. https://ourworldindata.org/coronavirus/country/brazil (accessed 8th June 2021).

28. Byun WS, Heo SW, Jo G, et al. Is coronavirus disease (COVID-19) seasonal? A critical analysis of empirical and epidemiological studies at global and local scales. *Environmental Research* 2021; **196**: 110972.

29. Wu Y, Jing W, Liu J, et al. Effects of temperature and humidity on the daily new cases and new deaths of COVID-19 in 166 countries. *Science of The Total Environment* 2020; **729**: 139051.

30. Sera F, Armstrong B, Abbott S, et al. A cross-sectional analysis of meteorological factors and SARS-CoV-2 transmission in 409 cities across 26 countries. *Nature Communications* 2021; **12**(1): 5968.

31. Ghafari M, du Plessis L, Raghwani J, et al. Purifying Selection Determines the Short-Term Time Dependency of Evolutionary Rates in SARS-CoV-2 and pH1N1 Influenza. *Molecular Biology and Evolution* 2022; **39**(2): msac009.

32. Duchene S, Featherstone L, Haritopoulou-Sinanidou M, Rambaut A, Lemey P, Baele G. Temporal signal and the phylodynamic threshold of SARS-CoV-2. *Virus Evolution* 2020; **6**(2): veaa061.

33. Volz E, Hill V, McCrone JT, et al. Evaluating the Effects of SARS-CoV-2 Spike Mutation D614G on Transmissibility and Pathogenicity. *Cell* 2021; **184**(1): 64-75.e11.

34. Tegally H, Wilkinson E, Giovanetti M, et al. Detection of a SARS-CoV-2 variant of concern in South Africa. *Nature* 2021; **592**(7854): 438-43.

35. Rambaut A, Loman N, Pybus O, et al. Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. 2020. https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563 (accessed 7th September 2022).

36. Hadfield J, Megill C, Bell SM, et al. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* 2018; **34**(23): 4121-3.

37. Choi B, Choudhary MC, Regan J, et al. Persistence and Evolution of SARS-CoV-2 in an Immunocompromised Host. *New England Journal of Medicine* 2020; **383**(23): 2291-3.

38. Kemp SA, Collier DA, Datir RP, et al. SARS-CoV-2 evolution during treatment of chronic infection. *Nature* 2021; **592**(7853): 277-82.

39. Braun KM, Moreno GK, Wagner C, et al. Acute SARS-CoV-2 infections harbor limited within-host diversity and transmit via tight transmission bottlenecks. *PLOS Pathogens* 2021; **17**(8): e1009849.

40. Lythgoe Katrina A, Hall M, Ferretti L, et al. SARS-CoV-2 within-host diversity and transmission. *Science* 2021; **372**(6539): eabg0821.

41. Barton MI, MacGowan SA, Kutuzov MA, Dushek O, Barton GJ, van der Merwe PA. Effects of common mutations in the SARS-CoV-2 Spike RBD and its ligand, the human ACE2 receptor on binding affinity and kinetics. *Elife* 2021; **10**: e70658.

42. Liu Y, Liu J, Plante KS, et al. The N501Y spike substitution enhances SARS-CoV-2 infection and transmission. *Nature* 2022; **602**(7896): 294-9.

43. Korber B, Fischer WM, Gnanakaran S, et al. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell* 2020; **182**(4): 812-27.e19.

44. Davies NG, Jarvis CI, van Zandvoort K, et al. Increased mortality in community-tested cases of SARS-CoV-2 lineage B.1.1.7. *Nature* 2021; **593**(7858): 270-4.

45. Zhou D, Dejnirattisai W, Supasa P, et al. Evidence of escape of SARS-CoV-2 variant B.1.351 from natural and vaccine-induced sera. *Cell* 2021; **184**(9): 2348-61.e6.

46. Jangra S, Ye C, Rathnasinghe R, et al. SARS-CoV-2 spike E484K mutation reduces antibody neutralisation. *The Lancet Microbe* 2021; **2**(7): e283-e4.

47. Wu H, Xing N, Meng K, et al. Nucleocapsid mutations R203K/G204R increase the infectivity, fitness, and virulence of SARS-CoV-2. *Cell Host & Microbe* 2021; **29**(12): 1788-801.e6.

48. Agerer B, Koblischke M, Gudipati V, et al. SARS-CoV-2 mutations in MHC-I-restricted epitopes evade CD8+ T cell responses. *Science Immunology* 2021; **6**(57): eabg6461.

49. 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.

50. Bedford T, Hodcroft, E. B. & Neher, R. A. Updated Nextstrain SARS-CoV-2 clade naming strategy. 2021. https://www.nature.com/articles/s41564-021-00932-w#ref-CR5 (accessed 5th February 2022).

51. Konings F, Perkins MD, Kuhn JH, et al. SARS-CoV-2 Variants of Interest and Concern naming scheme conducive for global discourse. *Nature Microbiology* 2021; **6**(7): 821-3.

52. World Health Organisation. World Health Organization Best Practices for the Naming of New Human Infectious Diseases, 2015.

53. du Plessis L, McCrone JT, Zarebski AE, et al. Establishment and lineage dynamics of the SARS-CoV-2 epidemic in the UK. *Science (New York, NY)* 2021; **371**(6530): 708-12.

54. Hodcroft EB, Zuber M, Nadeau S, et al. Spread of a SARS-CoV-2 variant through Europe in the summer of 2020. *Nature* 2021; **595**(7869): 707-12.

55. Bal A, Destras G, Gaymard A, et al. Two-step strategy for the identification of SARS-CoV-2 variant of concern 202012/01 and other variants with spike deletion H69-V70, France, August to December 2020. *Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin* 2021; **26**(3): 2100008.

56. Public Health England. Investigation of novel SARS-COV-2 variant. Variant of Concern 202012/01 Technical briefing 5, 2021.

57. Supasa P, Zhou D, Dejnirattisai W, et al. Reduced neutralization of SARS-CoV-2 B.1.1.7 variant by convalescent and vaccine sera. *Cell* 2021; **184**(8): 2201-11.e7.

58. Plante JA, Liu Y, Liu J, et al. Spike mutation D614G alters SARS-CoV-2 fitness. *Nature* 2021; **592**(7852): 116-21.

59. Yurkovetskiy L, Wang X, Pascal KE, et al. Structural and Functional Analysis of the D614G SARS-CoV-2 Spike Protein Variant. *Cell* 2020; **183**(3): 739-51.e8.

60. Walker AS, Vihta K-D, Gethings O, et al. Tracking the Emergence of SARS-CoV-2 Alpha Variant in the United Kingdom. *New England Journal of Medicine* 2021; **385**(27): 2582-5.

61. Volz E, Mishra S, Chand M, et al. Transmission of SARS-CoV-2 Lineage B. 1.1. 7 in England: Insights from linking epidemiological and genetic data. *MedRxiv* 2021: 2020.12. 30.20249034.

62. Davies Nicholas G, Abbott S, Barnard Rosanna C, et al. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science* 2021; **372**(6538): eabg3055.

63. Frampton D, Rampling T, Cross A, et al. Genomic characteristics and clinical effect of the emergent SARS-CoV-2 B.1.1.7 lineage in London, UK: a whole-genome sequencing and hospital-based cohort study. *The Lancet Infectious Diseases* 2021; **21**(9): 1246-56.

64. New and Emerging Respiratory Virus Threats Group. NERVTAG paper on COVID-19 variant of concern B.1.1.7, 2021.

65. Challen R, Brooks-Pollock E, Read JM, et al. Risk of mortality in patients infected with SARS-CoV-2 variant of concern 202012/1: matched cohort study. *BMJ* 2021; **372**: n579.

66. Dejnirattisai W, Huo J, Zhou D, et al. SARS-CoV-2 Omicron-B.1.1.529 leads to widespread escape from neutralizing antibody responses. *Cell* 2022; **185**(3): 467-84.e15.

67. Pouwels KB, Pritchard E, Matthews PC, et al. Effect of Delta variant on viral burden and vaccine effectiveness against new SARS-CoV-2 infections in the UK. *Nature Medicine* 2021; **27**(12): 2127-35.

68. Andrews N, Stowe J, Kirsebom F, et al. Covid-19 Vaccine Effectiveness against the Omicron (B.1.1.529) Variant. *New England Journal of Medicine* 2022.

69. Shrotri M, Krutikov M, Nacer-Laidi H, et al. Duration of vaccine effectiveness against SARS-CoV-2 infection, hospitalisation, and death in residents and staff of long-term care facilities in England (VIVALDI): a prospective cohort study. *The Lancet Healthy Longevity* 2022; **3**(7): e470-e80.

70. Andrews N, Stowe J, Kirsebom F, et al. Covid-19 Vaccine Effectiveness against the Omicron (B.1.1.529) Variant. *New England Journal of Medicine* 2022; **386**(16): 1532-46.

71. World Health Organization. Transmission of SARS-CoV-2: implications for infection prevention precautions: scientific brief, 09 July 2020.

72. Greenhalgh T, Jimenez JL, Prather KA, et al. Ten scientific reasons in support of airborne transmission of SARS-CoV-2. *Lancet (London, England)* 2021; **397**(10285): 1603-5.

73. Li Y, Qian H, Hang J, et al. Probable airborne transmission of SARS-CoV-2 in a poorly ventilated restaurant. *Building and Environment* 2021; **196**: 107788.

74. Miller SL, Nazaroff WW, Jimenez JL, et al. Transmission of SARS-CoV-2 by inhalation of respiratory aerosol in the Skagit Valley Chorale superspreading event. *Indoor Air* 2021; **31**(2): 314-23.

75. Lin J, Yan K, Zhang J, et al. A super-spreader of COVID-19 in Ningbo city in China. *Journal of Infection and Public Health* 2020; **13**(7): 935-7.

76. Chia PY, Coleman KK, Tan YK, et al. Detection of air and surface contamination by SARS-CoV-2 in hospital rooms of infected patients. *Nature Communications* 2020; **11**(1): 2800.

77. van Doremalen N, Bushmaker T, Morris DH, et al. Aerosol and Surface Stability of SARS-CoV-2 as Compared with SARS-CoV-1. *New England Journal of Medicine* 2020; **382**(16): 1564-7.

78. Marsalek J. Reframing the problem of the fomite transmission of COVID-19. *J Hosp Infect* 2021; **115**: 133-4.

79. Lewis D. COVID-19 rarely spreads through surfaces. So why are we still deep cleaning. *Nature* 2021; **590**(7844): 26-8.

80. Cevik M, Tate M, Lloyd O, et al. SARS-CoV-2, SARS-CoV, and MERS-CoV viral load dynamics, duration of viral shedding, and infectiousness: a systematic review and meta-analysis. *The Lancet Microbe* 2021; **2**(1): e13-e22.

81. Wang W, Xu Y, Gao R, et al. Detection of SARS-CoV-2 in Different Types of Clinical Specimens. *JAMA* 2020; **323**(18): 1843-4.

82. Jones DL, Baluja MQ, Graham DW, et al. Shedding of SARS-CoV-2 in feces and urine and its potential role in person-to-person transmission and the environment-based spread of COVID-19. *Science of The Total Environment* 2020; **749**: 141364.

83. Vaselli NM, Setiabudi W, Subramaniam K, et al. Investigation of SARS-CoV-2 faecal shedding in the community: a prospective household cohort study (COVID-LIV) in the UK. *BMC Infectious Diseases* 2021; **21**(1): 784.

84. Néant N, Lingas G, Le Hingrat Q, et al. Modeling SARS-CoV-2 viral kinetics and association with mortality in hospitalized patients from the French COVID cohort. *Proceedings of the National Academy of Sciences* 2021; **118**(8): e2017962118.

85. Cevik M, Marcus JL, Buckee C, Smith TC. Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Transmission Dynamics Should Inform Policy. *Clinical Infectious Diseases* 2021; **73**(Supplement_2): S170-S6.

86. Glenet M, Lebreil A-L, Heng L, et al. Asymptomatic COVID-19 Adult Outpatients identified as Significant Viable SARS-CoV-2 Shedders. *Scientific Reports* 2021; **11**(1): 20615.

87. Wölfel R, Corman VM, Guggemos W, et al. Virological assessment of hospitalized patients with COVID-2019. *Nature* 2020; **581**(7809): 465-9.

88. Killingley, B., Mann, A.J., Kalinova, M. *et al.* Safety, tolerability and viral kinetics during SARS-CoV-2 human challenge in young adults. *Nat Med* **28**, 1031–1041 (2022).

89. Johansson MA, Quandelacy TM, Kada S, et al. SARS-CoV-2 Transmission From People Without COVID-19 Symptoms. *JAMA Network Open* 2021; **4**(1): e2035057-e.

90. Plucinski MM, Wallace M, Uehara A, et al. Coronavirus Disease 2019 (COVID-19) in Americans Aboard the Diamond Princess Cruise Ship. *Clinical Infectious Diseases* 2021; **72**(10): e448-e57.

91. Arons MM, Hatfield KM, Reddy SC, et al. Presymptomatic SARS-CoV-2 Infections and Transmission in a Skilled Nursing Facility. *New England Journal of Medicine* 2020; **382**(22): 2081-90.

92. Green MA, García-Fiñana M, Barr B, et al. Evaluating social and spatial inequalities of large scale rapid lateral flow SARS-CoV-2 antigen testing in COVID-19 management: An observational study of Liverpool, UK (November 2020 to January 2021). *The Lancet Regional Health - Europe* 2021; **6**: 100107.

93. Viana R, Moyo S, Amoako DG, et al. Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa. *Nature* 2022; **603**(7902): 679-86.

94. Dahdouh E, Lázaro-Perona F, Romero-Gómez MP, et al. Ct values from SARS-CoV-2 diagnostic PCR assays should not be used as direct estimates of viral load. *Journal of Infection* 2021; **82**(3): 414-51.

95. Mustafa Hellou M, Górska A, Mazzaferri F, et al. Nucleic acid amplification tests on respiratory samples for the diagnosis of coronavirus infections: a systematic review and meta-analysis. *Clinical Microbiology and Infection* 2021; **27**(3): 341-51.

96. Vogels CBF, Brito AF, Wyllie AL, et al. Analytical sensitivity and efficiency comparisons of SARS-CoV-2 RT–qPCR primer–probe sets. *Nature Microbiology* 2020; **5**(10): 1299-305.

97. Jefferson T, Spencer EA, Brassey J, Heneghan C. Viral Cultures for Coronavirus Disease 2019 Infectivity Assessment: A Systematic Review. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2021; **73**(11): e3884-e99.

98. Jaafar R, Aherfi S, Wurtz N, et al. Correlation Between 3790 Quantitative Polymerase Chain Reaction–Positives Samples and Positive Cell Cultures, Including 1941 Severe Acute Respiratory Syndrome Coronavirus 2 Isolates. *Clinical Infectious Diseases* 2020; **72**(11): e921-e.

99. Peto T, Affron D, Afrough B, et al. COVID-19: Rapid antigen detection for SARS-CoV-2 by lateral flow assay: A national systematic evaluation of sensitivity and specificity for mass-testing. *EClinicalMedicine* 2021; **36**: 100924.

100. Brümmer LE, Katzenschlager S, Gaeddert M, et al. Accuracy of novel antigen rapid diagnostics for SARS-CoV-2: A living systematic review and meta-analysis. *PLOS Medicine* 2021; **18**(8): e1003735.

101. Pekosz A, Parvu V, Li M, et al. Antigen-Based Testing but Not Real-Time Polymerase Chain Reaction Correlates With Severe Acute Respiratory Syndrome Coronavirus 2 Viral Culture. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2021; **73**(9): e2861-e6.

102. Theel Elitza S, Slev P, Wheeler S, et al. The Role of Antibody Testing for SARS-CoV-2: Is There One? *Journal of clinical microbiology*; **58**(8): e00797-20.

103. Bobrovitz N, Arora RK, Cao C, et al. Global seroprevalence of SARS-CoV-2 antibodies: A systematic review and meta-analysis. *PLoS One* 2021; **16**(6): e0252617.

104. Lippi G. Potential drawbacks of SARS-CoV-2 seroprevalence surveys. *Journal of Hospital Infection* 2021; **110**: 206.

105. Crits-Christoph A, Kantor RS, Olm MR, et al. Genome sequencing of sewage detects regionally prevalent SARS-CoV-2 variants. *MBio* 2021; **12**(1): e02703-20.

106. Wurtzer S, Marechal V, Mouchel JM, et al. Evaluation of lockdown effect on SARS-CoV-2 dynamics through viral genome quantification in waste water, Greater Paris, France, 5 March to 23 April 2020. *Euro Surveill* 2020; **25**(50).

107. Hillary LS, Farkas K, Maher KH, et al. Monitoring SARS-CoV-2 in municipal wastewater to evaluate the success of lockdown measures for controlling COVID-19 in the UK. *Water Research* 2021; **200**: 117214.

108. Shah S, Gwee SXW, Ng JQX, et al. Wastewater surveillance to infer COVID-19 transmission: A systematic review. *Science of The Total Environment* 2022; **804**: 150060.

109. Menni C, Valdes AM, Freidin MB, et al. Real-time tracking of self-reported symptoms to predict potential COVID-19. *Nature Medicine* 2020; **26**(7): 1037-40.

110. Wang D, Hu B, Hu C, et al. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus–Infected Pneumonia in Wuhan, China. *JAMA* 2020; **323**(11): 1061-9.

111. Pormohammad A, Ghorbani S, Baradaran B, et al. Clinical characteristics, laboratory findings, radiographic signs and outcomes of 61,742 patients with confirmed COVID-19 infection: A systematic review and meta-analysis. *Microb Pathog* 2020; **147**: 104390-.

112. da Rosa Mesquita R, Francelino Silva Junior LC, Santos Santana FM, et al. Clinical manifestations of COVID-19 in the general population: systematic review. *Wiener klinische Wochenschrift* 2021; **133**(7): 377-82.

113. Sudre CH, Keshet A, Graham MS, et al. Anosmia, ageusia, and other COVID-19-like symptoms in association with a positive SARS-CoV-2 test, across six national digital surveillance platforms: an observational study. *The Lancet Digital Health* 2021; **3**(9): e577-e86.

114. Mao R, Qiu Y, He J-S, et al. Manifestations and prognosis of gastrointestinal and liver involvement in patients with COVID-19: a systematic review and meta-analysis. *The Lancet Gastroenterology & Hepatology* 2020; **5**(7): 667-78.

115. Tan SW, Tam YC, Oh CC. Skin manifestations of COVID-19: A worldwide review. *JAAD International* 2021; **2**: 119-33.

116. COVID Symptom Study. What are the new top 5 COVID symptoms? 11th January 2022 2021. https://covid.joinzoe.com/post/new-top-5-covid-symptoms (accessed 23 June 2022).

117. Iacobucci G. Covid-19: Runny nose, headache, and fatigue are commonest symptoms of omicron, early data show. *BMJ* 2021; **375**: n3103.

118. Salje H, Tran Kiem C, Lefrancq N, et al. Estimating the burden of SARS-CoV-2 in France. *Science* 2020; **369**(6500): 208-11.

119. Nopp S, Moik F, Jilma B, Pabinger I, Ay C. Risk of venous thromboembolism in patients with COVID-19: A systematic review and meta-analysis. *Res Pract Thromb Haemost* 2020; **4**(7): 1178-91.

120. Tang N, Li D, Wang X, Sun Z. Abnormal coagulation parameters are associated with poor prognosis in patients with novel coronavirus pneumonia. *Journal of thrombosis and haemostasis* 2020; **18**(4): 844-7.

121. Long B, Brady WJ, Koyfman A, Gottlieb M. Cardiovascular complications in COVID-19. *The American Journal of Emergency Medicine* 2020; **38**(7): 1504-7.

122. Aul DR, Gates DJ, Draper DA, et al. Complications after discharge with COVID-19 infection and risk factors associated with development of post-COVID pulmonary fibrosis. *Respiratory Medicine* 2021; **188**: 106602.

123. Taquet M, Geddes JR, Husain M, Luciano S, Harrison PJ. 6-month neurological and psychiatric outcomes in 236 379 survivors of COVID-19: a retrospective cohort study using electronic health records. *The Lancet Psychiatry* 2021; **8**(5): 416-27.

124. Ayoubkhani D, Khunti K, Nafilyan V, et al. Post-covid syndrome in individuals admitted to hospital with covid-19: retrospective cohort study. *BMJ* 2021; **372**: n693.

125. Lavezzo E, Franchin E, Ciavarella C, et al. Suppression of a SARS-CoV-2 outbreak in the Italian municipality of Vo'. *Nature* 2020; **584**(7821): 425-9.

126. Kasper MR, Geibe JR, Sears CL, et al. An Outbreak of Covid-19 on an Aircraft Carrier. *New England Journal of Medicine* 2020; **383**(25): 2417-26.

127. Letizia AG, Ramos I, Obla A, et al. SARS-CoV-2 Transmission among Marine Recruits during Quarantine. *New England Journal of Medicine* 2020; **383**(25): 2407-16.

128. Sah P, Fitzpatrick MC, Zimmer CF, et al. Asymptomatic SARS-CoV-2 infection: A systematic review and meta-analysis. *Proceedings of the National Academy of Sciences* 2021; **118**(34): e2109229118.

129. Oran DP, Topol EJ. The Proportion of SARS-CoV-2 Infections That Are Asymptomatic. *Annals of Internal Medicine* 2021; **174**(5): 655-62.

130. Dorjee K, Kim H, Bonomo E, Dolma R. Prevalence and predictors of death and severe disease in patients hospitalized due to COVID-19: A comprehensive systematic review and meta-analysis of 77 studies and 38,000 patients. *PLoS One* 2020; **15**(12): e0243191.

131. O'Driscoll M, Ribeiro Dos Santos G, Wang L, et al. Age-specific mortality and immunity patterns of SARS-CoV-2. *Nature* 2021; **590**(7844): 140-5.

132. Brazeau NF, Jenks S, Fu H, et al. COVID-19 Infection Fatality Ratio: Estimates from Seroprevalence.: Imperial College London, 2020.

133. Zerbo O, Lewis N, Fireman B, et al. Population-based assessment of risks for severe COVID-19 disease outcomes. *Influenza and Other Respiratory Viruses* 2022; **16**(1): 159-65.

134. Pijls BG, Jolani S, Atherley A, et al. Demographic risk factors for COVID-19 infection, severity, ICU admission and death: a meta-analysis of 59 studies. *BMJ Open* 2021; **11**(1): e044640.

135. Fernández Villalobos NV, Ott JJ, Klett-Tammen CJ, et al. Effect modification of the association between comorbidities and severe course of COVID-19 disease by age of study participants: a systematic review and meta-analysis. *Systematic Reviews* 2021; **10**(1): 194.

136. Malik P, Patel U, Patel K, et al. Obesity a predictor of outcomes of COVID-19 hospitalized patients—A systematic review and meta-analysis. *Journal of Medical Virology* 2021; **93**(2): 1188-93.

137. Elliott J, Bodinier B, Whitaker M, et al. COVID-19 mortality in the UK Biobank cohort: revisiting and evaluating risk factors. *European Journal of Epidemiology* 2021; **36**(3): 299-309.

138. World Health Organsiation. Estimating mortality from COVID-19: Scientific Brief, 4 August 2020.

139. Diamond MS, Kanneganti T-D. Innate immunity: the first line of defense against SARS-CoV-2. *Nature Immunology* 2022; **23**(2): 165-76.

140. Panda SK, Colonna M. Innate Lymphoid Cells in Mucosal Immunity. *Frontiers in Immunology* 2019; **10**. https://doi.org/10.3389/fimmu.2019.00861

141. Gomes AMC, Farias GB, Dias-Silva M, et al. SARS-CoV2 pneumonia recovery is linked to expansion of innate lymphoid cells type 2 expressing CCR10. *European Journal of Immunology* 2021; **51**(12): 3194-201.

142. Giamarellos-Bourboulis EJ, Netea MG, Rovina N, et al. Complex Immune Dysregulation in COVID-19 Patients with Severe Respiratory Failure. *Cell Host & Microbe* 2020; **27**(6): 992-1000.e3.

143. Qin C, Zhou L, Hu Z, et al. Dysregulation of Immune Response in Patients With Coronavirus 2019 (COVID-19) in Wuhan, China. *Clinical Infectious Diseases* 2020; **71**(15): 762-8.

144. Zhou Y, Fu B, Zheng X, et al. Pathogenic T-cells and inflammatory monocytes incite inflammatory storms in severe COVID-19 patients. *National Science Review* 2020; **7**(6): 998-1002.

145. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The lancet* 2020; **395**(10223): 497-506.

146. Lucas C, Wong P, Klein J, et al. Longitudinal analyses reveal immunological misfiring in severe COVID-19. *Nature* 2020; **584**(7821): 463-9.

147. McKechnie JL, Blish CA. The Innate Immune System: Fighting on the Front Lines or Fanning the Flames of COVID-19? *Cell Host & Microbe* 2020; **27**(6): 863-9.

148. Long Q-X, Liu B-Z, Deng H-J, et al. Antibody responses to SARS-CoV-2 in patients with COVID-19. *Nature Medicine* 2020; **26**(6): 845-8.

149. Seow J, Graham C, Merrick B, et al. Longitudinal observation and decline of neutralizing antibody responses in the three months following SARS-CoV-2 infection in humans. *Nature Microbiology* 2020; **5**(12): 1598-607.

150. Liu X, Wang J, Xu X, Liao G, Chen Y, Hu C-H. Patterns of IgG and IgM antibody response in COVID-19 patients. *Emerging Microbes & Infections* 2020; **9**(1): 1269-74.

151. Post N, Eddy D, Huntley C, et al. Antibody response to SARS-CoV-2 infection in humans: A systematic review. *PLoS One* 2021; **15**(12): e0244126.

152. Wei J, Matthews PC, Stoesser N, et al. Anti-spike antibody response to natural SARS-CoV-2 infection in the general population. *Nature Communications* 2021; **12**(1): 6250.

153. Petersen LR, Sami S, Vuong N, et al. Lack of Antibodies to Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) in a Large Cohort of Previously Infected Persons. *Clinical Infectious Diseases* 2021; **73**(9): e3066-e73.

154. Vanshylla K, Di Cristanziano V, Kleipass F, et al. Kinetics and correlates of the neutralizing antibody response to SARS-CoV-2 infection in humans. *Cell Host & Microbe* 2021; **29**(6): 917-29.e4.

155. Rosati M, Terpos E, Ntanasis-Stathopoulos I, et al. Sequential Analysis of Binding and Neutralizing Antibody in COVID-19 Convalescent Patients at 14 Months After SARS-CoV-2 Infection. *Frontiers in Immunology* 2021; **12**.

156. Lumley SF, Wei J, O'Donnell D, et al. The Duration, Dynamics, and Determinants of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Antibody Responses in Individual Healthcare Workers. *Clinical Infectious Diseases* 2021; **73**(3): e699-e709.

157. Dispinseri S, Secchi M, Pirillo MF, et al. Neutralizing antibody responses to SARS-CoV-2 in symptomatic COVID-19 is persistent and critical for survival. *Nature Communications* 2021; **12**(1): 2670.

158. RECOVERY Collaborative Group. Casirivimab and imdevimab in patients admitted to hospital with COVID-19 (RECOVERY): a randomised, controlled, open-label, platform trial. Lancet. 2022 Feb 12;399(10325):665-676. doi: 10.1016/S0140-6736(22)00163-5.

159. Kreuzberger N, Hirsch C, Chai KL, et al. SARS-CoV-2-neutralising monoclonal antibodies for treatment of COVID-19. Cochrane Database Syst Rev. 2021 Sep 2;**9**(9):CD013825.

160. Diao B, Wang C, Tan Y, et al. Reduction and Functional Exhaustion of T Cells in Patients With Coronavirus Disease 2019 (COVID-19). *Frontiers in Immunology* 2020; **11**(827).

161. Rha M-S, Jeong HW, Ko J-H, et al. PD-1-Expressing SARS-CoV-2-Specific CD8+ T Cells Are Not Exhausted, but Functional in Patients with COVID-19. *Immunity* 2021; **54**(1): 44-52.e3.

162. Liu Y, Du X, Chen J, et al. Neutrophil-to-lymphocyte ratio as an independent risk factor for mortality in hospitalized patients with COVID-19. *Journal of Infection* 2020; **81**(1): e6-e12.

163. 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.

164. Le Bert N, Tan AT, Kunasegaran K, et al. SARS-CoV-2-specific T cell immunity in cases of COVID-19 and SARS, and uninfected controls. *Nature* 2020; **584**(7821): 457-62.

165. 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. *Nature Immunology* 2020; **21**(11): 1336-45.

166. Ewer KJ, Barrett JR, Belij-Rammerstorfer S, et al. T cell and antibody responses induced by a single dose of ChAdOx1 nCoV-19 (AZD1222) vaccine in a phase 1/2 clinical trial. *Nat Med* 2021; **27**(2): 270-8.

167. Geers D, Shamier Marc C, Bogers S, et al. SARS-CoV-2 variants of concern partially escape humoral but not T cell responses in COVID-19 convalescent donors and vaccine recipients. *Science Immunology* 2021; **6**(59): eabj1750.

168. Zhang H, Deng S, Ren L, et al. Profiling CD8+ T cell epitopes of COVID-19 convalescents reveals reduced cellular immune responses to SARS-CoV-2 variants. *Cell Reports* 2021; **36**(11): 109708.

169. Moss P. The T cell immune response against SARS-CoV-2. *Nature Immunology* 2022; **23**(2): 186-93.

170. van Doremalen N, Lambe T, Spencer A, et al. ChAdOx1 nCoV-19 vaccine prevents SARS-CoV-2 pneumonia in rhesus macaques. *Nature* 2020; **586**(7830): 578-82.

171. Vogel AB, Kanevsky I, Che Y, et al. BNT162b vaccines protect rhesus macaques from SARS-CoV-2. *Nature* 2021; **592**(7853): 283-9.

172. Ravichandran S, Grubbs G, Tang J, et al. Systemic and mucosal immune profiling in asymptomatic and symptomatic SARS-CoV-2–infected individuals reveal unlinked immune signatures. *Science Advances*; **7**(42): eabi6533.

173. Cervia C, Nilsson J, Zurbuchen Y, et al. Systemic and mucosal antibody responses specific to SARS-CoV-2 during mild versus severe COVID-19. *Journal of Allergy and Clinical Immunology* 2021; **147**(2): 545-57.e9.

174. Isho B, Abe Kento T, 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.

175. Smith N, Goncalves P, Charbit B, et al. Distinct systemic and mucosal immune responses during acute SARS-CoV-2 infection. *Nature Immunology* 2021; **22**(11): 1428-39.

176. Britton GJ, Chen-Liaw A, Cossarini F, et al. Limited intestinal inflammation despite diarrhea, fecal viral RNA and SARS-CoV-2-specific IgA in patients with acute COVID-19. *Scientific Reports* 2021; **11**(1): 13308.

177. Plotkin SA, Gilbert PB. Nomenclature for immune correlates of protection after vaccination. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2012; **54**(11): 1615-7.

178. Chandrashekar A, Liu J, Martinot AJ, et al. SARS-CoV-2 infection protects against rechallenge in rhesus macaques. *Science* 2020; **369**(6505): 812-7.

179. Hall VJ, Foulkes S, Saei A, et al. COVID-19 vaccine coverage in health-care workers in England and effectiveness of BNT162b2 mRNA vaccine against infection (SIREN): a prospective, multicentre, cohort study. *The Lancet* 2021; **397**(10286): 1725-35.

180. Lumley SF, O'Donnell D, Stoesser NE, et al. Antibody Status and Incidence of SARS-CoV-2 Infection in Health Care Workers. *New England Journal of Medicine* 2020; **384**(6): 533-40.

181. Addetia A, Crawford Katharine HD, Dingens A, et al. Neutralizing Antibodies Correlate with Protection from SARS-CoV-2 in Humans during a Fishery Vessel Outbreak with a High Attack Rate. *Journal of clinical microbiology* 2020; **58**(11): e02107-20.

182. McMahan K, Yu J, Mercado NB, et al. Correlates of protection against SARS-CoV-2 in rhesus macaques. *Nature* 2021; **590**(7847): 630-4.

183. Feng S, Phillips DJ, White T, et al. Correlates of protection against symptomatic and asymptomatic SARS-CoV-2 infection. *Nature Medicine* 2021; **27**(11): 2032-40.

184. Gilbert PB, Montefiori DC, McDermott AB, et al. Immune correlates analysis of the mRNA-1273 COVID-19 vaccine efficacy clinical trial. *Science* 2022; **375**(6576): 43-50.

185. Khoury DS, Cromer D, Reynaldi A, et al. Neutralizing antibody levels are highly predictive of immune protection from symptomatic SARS-CoV-2 infection. *Nature Medicine* 2021; **27**(7): 1205-11.

186. Kalimuddin S, Tham CYL, Qui M, et al. Early T cell and binding antibody responses are associated with COVID-19 RNA vaccine efficacy onset. *Med* 2021; **2**(6): 682-8.e4.

187. Polack FP, Thomas SJ, Kitchin N, et al. Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. *New England Journal of Medicine* 2020; **383**(27): 2603-15.

188. Lopez Bernal J, Andrews N, Gower C, et al. Effectiveness of the Pfizer-BioNTech and Oxford-AstraZeneca vaccines on covid-19 related symptoms, hospital admissions, and mortality in older adults in England: test negative case-control study. *BMJ* 2021; **373**: n1088.

189. Walsh EE, Frenck RW, Falsey AR, et al. Safety and Immunogenicity of Two RNA-Based Covid-19 Vaccine Candidates. *New England Journal of Medicine* 2020; **383**(25): 2439-50.

190. Moore SC, Kronsteiner B, Longet S, et al. Evolution of long-term vaccine-induced and hybrid immunity in healthcare workers after different COVID-19 vaccine regimens. Med. 2023 Mar 10;**4**(3):191-215.e9. Epub 2023 Feb 16.

191. Andrews N, Tessier E, Stowe J, et al. Duration of Protection against Mild and Severe Disease by Covid-19 Vaccines. *New England Journal of Medicine* 2022; **386**(4): 340-50.

192. Skelly DT, Harding AC, Gilbert-Jaramillo J, et al. Two doses of SARS-CoV-2 vaccination induce robust immune responses to emerging SARS-CoV-2 variants of concern. *Nature Communications* 2021; **12**(1): 5061.

193. Dejnirattisai W, Zhou D, Supasa P, et al. Antibody evasion by the P.1 strain of SARS-CoV-2. *Cell* 2021; **184**(11): 2939-54.e9.

194. Chen T, Wu D, Chen H, et al. Clinical characteristics of 113 deceased patients with coronavirus disease 2019: retrospective study. *BMJ* 2020; **368**: m1091.

195. Lee JS, Park S, Jeong HW, et al. Immunophenotyping of COVID-19 and influenza highlights the role of type I interferons in development of severe COVID-19. *Science Immunology* 2020; **5**(49): eabd1554.

196. Kreutmair S, Unger S, Núñez NG, et al. Distinct immunological signatures discriminate severe COVID-19 from non-SARS-CoV-2-driven critical pneumonia. *Immunity* 2021; **54**(7): 1578-93.e5.

197. Zhou Z, Ren L, Zhang L, et al. Heightened Innate Immune Responses in the Respiratory Tract of COVID-19 Patients. *Cell Host & Microbe* 2020; **27**(6): 883-90.e2.

198. Chen G, Wu D, Guo W, et al. Clinical and immunological features of severe and moderate coronavirus disease 2019. *The Journal of Clinical Investigation* 2020; **130**(5): 2620-9.

199. Hadjadj J, Yatim N, Barnabei L, et al. Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients. *Science* 2020; **369**(6504): 718-24.

200. Bastard P, Rosen LB, Zhang Q, et al. Autoantibodies against type I IFNs in patients with life-threatening COVID-19. *Science* 2020; **370**(6515): eabd4585.

201. Kim Y-M, Shin E-C. Type I and III interferon responses in SARS-CoV-2 infection. *Experimental & Molecular Medicine* 2021; **53**(5): 750-60.

202. Marklund E, Leach S, Axelsson H, et al. Serum-IgG responses to SARS-CoV-2 after mild and severe COVID-19 infection and analysis of IgG non-responders. *PLoS One* 2020; **15**(10): e0241104.

203. Li K, Huang B, Wu M, et al. Dynamic changes in anti-SARS-CoV-2 antibodies during SARS-CoV-2 infection and recovery from COVID-19. *Nature Communications* 2020; **11**(1): 6044.

204. Lucas C, Klein J, Sundaram ME, et al. Delayed production of neutralizing antibodies correlates with fatal COVID-19. *Nature Medicine* 2021; **27**(7): 1178-86.

205. Ni L, Ye F, Cheng M-L, et al. Detection of SARS-CoV-2-Specific Humoral and Cellular Immunity in COVID-19 Convalescent Individuals. *Immunity* 2020; **52**(6): 971-7.e3.

206. Chen Y, Tong X, Li Y, et al. A comprehensive, longitudinal analysis of humoral responses specific to four recombinant antigens of SARS-CoV-2 in severe and non-severe COVID-19 patients. *PLOS Pathogens* 2020; **16**(9): e1008796.

207. Notarbartolo S, Ranzani V, Bandera A, et al. Integrated longitudinal immunophenotypic, transcriptional, and repertoire analyses delineate immune responses in patients with COVID-19. *Science Immunology* 2021; **6**(62): eabg5021.

208. Gao Y, Chen Y, Liu M, Shi S, Tian J. Impacts of immunosuppression and immunodeficiency on COVID-19: A systematic review and meta-analysis. *Journal of Infection* 2020; **81**(2): e93-e5.

209. Baek MS, Lee M-T, Kim W-Y, Choi JC, Jung S-Y. COVID-19-related outcomes in immunocompromised patients: A nationwide study in Korea. *PLoS One* 2021; **16**(10): e0257641.

210. Akalin E, Azzi Y, Bartash R, et al. Covid-19 and Kidney Transplantation. *New England Journal of Medicine* 2020; **382**(25): 2475-7.

211. Singson JRC, Kirley PD, Pham H, et al. Factors Associated with Severe Outcomes Among Immunocompromised Adults Hospitalized for COVID-19 - COVID-NET, 10 States, March 2020-February 2022. *MMWR Morb Mortal Wkly Rep* 2022; **71**(27): 878-84.

212. Ghione P, Gu JJ, Attwood K, et al. Impaired humoral responses to COVID-19 vaccination in patients with lymphoma receiving B-cell–directed therapies. *Blood* 2021; **138**(9): 811-4.

213. Agha ME, Blake M, Chilleo C, Wells A, Haidar G. Suboptimal Response to Coronavirus Disease 2019 Messenger RNA Vaccines in Patients With Hematologic Malignancies: A Need for Vigilance in the Postmasking Era. *Open Forum Infect Dis* 2021; **8**(7): ofab353.

214. Stephanie GY, Knight RJ, Graviss EA, et al. Kidney transplant recipients rarely show an early antibody response following the first COVID-19 vaccine administration. *Transplantation* 2021; **105**(7): e72-e3.

215. Whitaker HJ, Tsang RSM, Byford R, et al. Pfizer-BioNTech and Oxford AstraZeneca COVID-19 vaccine effectiveness and immune response amongst individuals in clinical risk groups. *Journal of Infection* 2022; **84**(5): 675-83.

216. Prendecki M, Clarke C, Edwards H, et al. Humoral and T-cell responses to SARS-CoV-2 vaccination in patients receiving immunosuppression. *Annals of the Rheumatic Diseases* 2021; **80**(10): 1322.

217. Mahil SK, Bechman K, Raharja A, et al. The effect of methotrexate and targeted immunosuppression on humoral and cellular immune responses to the COVID-19 vaccine BNT162b2: a cohort study. *The Lancet Rheumatology* 2021; **3**(9): e627-e37.

218. Alexander JL, Liu Z, Muñoz Sandoval D, et al. COVID-19 vaccine-induced antibody and T-cell responses in immunosuppressed patients with inflammatory bowel disease after the third vaccine dose (VIP): a multicentre, prospective, case-control study. *The Lancet Gastroenterology & Hepatology* 2022; **7**(11): 1005-15.

219. Chodick G, Tene L, Rotem RS, et al. The Effectiveness of the Two-Dose BNT162b2 Vaccine: Analysis of Real-World Data. *Clinical Infectious Diseases* 2022; **74**(3): 472-8.

220. Embi PJ, Levy ME, Naleway AL, et al. Effectiveness of 2-Dose Vaccination with mRNA COVID-19 Vaccines Against COVID-19-Associated Hospitalizations Among

Immunocompromised Adults - Nine States, January-September 2021. *MMWR Morb Mortal Wkly Rep* 2021; **70**(44): 1553-9.

221. Slezak J, Bruxvoort K, Fischer H, Broder B, Ackerson B, Tartof S. Rate and severity of suspected SARS-Cov-2 reinfection in a cohort of PCR-positive COVID-19 patients. *Clinical Microbiology and Infection* 2021; **27**(12): 1860.e7-.e10.

222. Abu-Raddad LJ, Chemaitelly H, Coyle P, et al. SARS-CoV-2 antibody-positivity protects against reinfection for at least seven months with 95% efficacy. *EClinicalMedicine* 2021; **35**: 100861.

223. UK Health Security Agency. Weekly national Influenza and COVID-19 surveillance report: 21st April 2022.

224. Liu C, Ginn HM, Dejnirattisai W, et al. Reduced neutralization of SARS-CoV-2 B.1.617 by vaccine and convalescent serum. *Cell* 2021; **184**(16): 4220-36.e13.

225. Pulliam JRC, van Schalkwyk C, Govender N, et al. Increased risk of SARS-CoV-2 reinfection associated with emergence of Omicron in South Africa. *Science*; **376**(6593): eabn4947.

226. Drosten C, Günther S, Preiser W, et al. Identification of a Novel Coronavirus in Patients with Severe Acute Respiratory Syndrome. *New England Journal of Medicine* 2003; **348**(20): 1967-76.

227. World Health Organisation. Summary of probable SARS cases with onset of illness from 1 November 2002 to 31 July 2003 (based on data as of the 31 December 2003), 2015.

228. Li F, Li W, Farzan M, Harrison SC. Structure of SARS Coronavirus Spike Receptor-Binding Domain Complexed with Receptor. *Science* 2005; **309**(5742): 1864-8.

229. Sui J, Li W, Murakami A, et al. Potent neutralization of severe acute respiratory syndrome (SARS) coronavirus by a human mAb to S1 protein that blocks receptor association. *Proceedings of the National Academy of Sciences* 2004; **101**(8): 2536-41.

230. Roberts A, Vogel L, Guarner J, et al. Severe acute respiratory syndrome coronavirus infection of golden Syrian hamsters. *J Virol* 2005; **79**(1): 503-11.

231. Subbarao K, McAuliffe J, Vogel L, et al. Prior infection and passive transfer of neutralizing antibody prevent replication of severe acute respiratory syndrome coronavirus in the respiratory tract of mice. *J Virol* 2004; **78**(7): 3572-7.

232. Chu Y-K, Ali GD, Jia F, et al. The SARS-CoV ferret model in an infection–challenge study. *Virology* 2008; **374**(1): 151-63.

233. McAuliffe J, Vogel L, Roberts A, et al. Replication of SARS coronavirus administered into the respiratory tract of African Green, rhesus and cynomolgus monkeys. *Virology* 2004; **330**(1): 8-15.

234. Yuchun N, Guangwen W, Xuanling S, et al. Neutralizing Antibodies in Patients with Severe Acute Respiratory Syndrome-Associated Coronavirus Infection. *The Journal of Infectious Diseases* 2004; **190**(6): 1119-26.

235. Zhang L, Zhang F, Yu W, et al. Antibody responses against SARS coronavirus are correlated with disease outcome of infected individuals. *Journal of Medical Virology* 2006; **78**(1): 1-8.

236. Soo YOY, Cheng Y, Wong R, et al. Retrospective comparison of convalescent plasma with continuing high-dose methylprednisolone treatment in SARS patients. *Clinical Microbiology and Infection* 2004; **10**(7): 676-8.

237. Cheng Y, Wong R, Soo YOY, et al. Use of convalescent plasma therapy in SARS patients in Hong Kong. *European Journal of Clinical Microbiology and Infectious Diseases* 2005; **24**(1): 44-6.

238. Mair-Jenkins J, Saavedra-Campos M, Baillie JK, et al. The Effectiveness of Convalescent Plasma and Hyperimmune Immunoglobulin for the Treatment of Severe Acute Respiratory Infections of Viral Etiology: A Systematic Review and Exploratory Meta-analysis. *The Journal of Infectious Diseases* 2015; **211**(1): 80-90.

239. Gao W, Tamin A, Soloff A, et al. Effects of a SARS-associated coronavirus vaccine in monkeys. *The Lancet* 2003; **362**(9399): 1895-6.

240. Buchholz UJ, Bukreyev A, Yang L, et al. Contributions of the structural proteins of severe acute respiratory syndrome coronavirus to protective immunity. *Proc Natl Acad Sci U S A* 2004; **101**(26): 9804-9.

241. Du L, Zhao G, He Y, et al. Receptor-binding domain of SARS-CoV spike protein induces long-term protective immunity in an animal model. *Vaccine* 2007; **25**(15): 2832-8.

242. Bukreyev A, Lamirande EW, Buchholz UJ, et al. Mucosal immunisation of African green monkeys (Cercopithecus aethiops) with an attenuated parainfluenza virus expressing the SARS coronavirus spike protein for the prevention of SARS. *The Lancet* 2004; **363**(9427): 2122-7.

243. Lin J, Zhang J-S, Su N, et al. Safety and immunogenicity from a phase I trial of inactivated severe acute respiratory syndrome coronavirus vaccine. *Antiviral therapy* 2007; **12**(7): 1107.

244. Martin JE, Louder MK, Holman LA, et al. A SARS DNA vaccine induces neutralizing antibody and cellular immune responses in healthy adults in a Phase I clinical trial. *Vaccine* 2008; **26**(50): 6338-43.

245. Reusken CBEM, Haagmans BL, Müller MA, et al. Middle East respiratory syndrome coronavirus neutralising serum antibodies in dromedary camels: a comparative serological study. *The Lancet Infectious Diseases* 2013; **13**(10): 859-66.

246. World Health Organization. MERS situation update, 2022.

247. Oh MD, Park WB, Park SW, et al. Middle East respiratory syndrome: what we learned from the 2015 outbreak in the Republic of Korea. *Korean J Intern Med* 2018; **33**(2): 233-46.

248. Min C-K, Cheon S, Ha N-Y, et al. Comparative and kinetic analysis of viral shedding and immunological responses in MERS patients representing a broad spectrum of disease severity. *Scientific Reports* 2016; **6**(1): 25359.

249. Corman VM, Albarrak AM, Omrani AS, et al. Viral Shedding and Antibody Response in 37 Patients With Middle East Respiratory Syndrome Coronavirus Infection. *Clinical Infectious Diseases* 2016; **62**(4): 477-83.

250. Alshukairi AN, Zhao J, Al-Mozaini MA, et al. Longevity of Middle East Respiratory Syndrome Coronavirus Antibody Responses in Humans, Saudi Arabia. *Emerg Infect Dis* 2021; **27**(5): 1472-6.

251. Tai W, Zhao G, Sun S, et al. A recombinant receptor-binding domain of MERS-CoV in trimeric form protects human dipeptidyl peptidase 4 (hDPP4) transgenic mice from MERS-CoV infection. *Virology* 2016; **499**: 375-82.

252. Haagmans BL, van den Brand JMA, Raj VS, et al. An orthopoxvirus-based vaccine reduces virus excretion after MERS-CoV infection in dromedary camels. *Science* 2016; **351**(6268): 77-81.

253. Muthumani K, Falzarano D, Reuschel EL, et al. A synthetic consensus anti–spike protein DNA vaccine induces protective immunity against Middle East respiratory syndrome coronavirus in nonhuman primates. *Sci Transl Med* 2015; **7**(301): 301ra132-301ra132.

254. Bosaeed M, Balkhy HH, Almaziad S, et al. Safety and immunogenicity of ChAdOx1 MERS vaccine candidate in healthy Middle Eastern adults (MERS002): an open-label, non-randomised, dose-escalation, phase 1b trial. *The Lancet Microbe* 2022; **3**(1): e11-e20.

255. Koch T, Dahlke C, Fathi A, et al. Safety and immunogenicity of a modified vaccinia virus Ankara vector vaccine candidate for Middle East respiratory syndrome: an open-label, phase 1 trial. *The Lancet Infectious Diseases* 2020; **20**(7): 827-38.

256. Modjarrad K, Roberts CC, Mills KT, et al. Safety and immunogenicity of an anti-Middle East respiratory syndrome coronavirus DNA vaccine: a phase 1, open-label, single-arm, dose-escalation trial. *The Lancet Infectious Diseases* 2019; **19**(9): 1013-22.

257. Jaume M, Yip Ming S, Cheung Chung Y, et al. Anti-Severe Acute Respiratory Syndrome Coronavirus Spike Antibodies Trigger Infection of Human Immune Cells via a pH- and Cysteine Protease-Independent FcγR Pathway. *Journal of Virology* 2011; **85**(20): 10582-97.

258. Vennema H, de Groot RJ, Harbour DA, et al. Early death after feline infectious peritonitis virus challenge due to recombinant vaccinia virus immunization. *Journal of Virology* 1990; **64**(3): 1407-9.

259. Weiss RC, Scott FW. Antibody-mediated enhancement of disease in feline infectious peritonitis: Comparisons with dengue hemorrhagic fever. *Comparative Immunology, Microbiology and Infectious Diseases* 1981; **4**(2): 175-89.

260. Halstead SB, Russell PK. Protective and immunological behavior of chimeric yellow fever dengue vaccine. *Vaccine* 2016; **34**(14): 1643-7.

261. Kim HW, Canchola JG, Brandt CD, et al. Respiratory syncytial virus disease in infants despite prior administration of antigenic inactivated vaccine. Am J Epidemiol. 1969 Apr;**89**(4):422-34.

262. Yasui F, Kai C, Kitabatake M, et al. Prior Immunization with Severe Acute Respiratory Syndrome (SARS)-Associated Coronavirus (SARS-CoV) Nucleocapsid Protein Causes Severe

Pneumonia in Mice Infected with SARS-CoV1. *The Journal of Immunology* 2008; **181**(9): 6337-48.

263. Liu L, Wei Q, Lin Q, et al. Anti-spike IgG causes severe acute lung injury by skewing macrophage responses during acute SARS-CoV infection. *JCI Insight* 2019; **4**(4).

264. Tseng C-T, Sbrana E, Iwata-Yoshikawa N, et al. Immunization with SARS Coronavirus Vaccines Leads to Pulmonary Immunopathology on Challenge with the SARS Virus. *PLoS One* 2012; **7**(4): e35421.

265. Bolles M, Deming D, Long K, et al. A Double-Inactivated Severe Acute Respiratory Syndrome Coronavirus Vaccine Provides Incomplete Protection in Mice and Induces Increased Eosinophilic Proinflammatory Pulmonary Response upon Challenge. *Journal of Virology* 2011; **85**(23): 12201-15.

266. Li Y-D, Chi W-Y, Su J-H, et al. Coronavirus vaccine development: from SARS and MERS to COVID-19. *Journal of Biomedical Science* 2020; **27**(1): 104.

267. Houser KV, Broadbent AJ, Gretebeck L, et al. Enhanced inflammation in New Zealand white rabbits when MERS-CoV reinfection occurs in the absence of neutralizing antibody. *PLOS Pathogens* 2017; **13**(8): e1006565.

268. Agrawal AS, Tao X, Algaissi A, et al. Immunization with inactivated Middle East Respiratory Syndrome coronavirus vaccine leads to lung immunopathology on challenge with live virus. *Human Vaccines & Immunotherapeutics* 2016; **12**(9): 2351-6.

269. Li C, Chen Y-X, Liu F-F, et al. Absence of Vaccine-enhanced Disease With Unexpected Positive Protection Against severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) by Inactivated Vaccine Given Within 3 Days of Virus Challenge in Syrian Hamster Model. *Clinical Infectious Diseases* 2021; **73**(3): e719-e34.

270. World Health Organisation. COVID-19 vaccine tracker and landscape. 28th June 2022. https://www.who.int/publications/m/item/draft-landscape-of-covid-19-candidate-vaccines (accessed 29th June 2022).

271. European Medicines Agency. Ervebo. 2019. https://www.ema.europa.eu/en/medicines/human/EPAR/ervebo#authorisation-details-section (accessed 1st February 2022).

272. Dicks MDJ, Spencer AJ, Edwards NJ, et al. A Novel Chimpanzee Adenovirus Vector with Low Human Seroprevalence: Improved Systems for Vector Derivation and Comparative Immunogenicity. *PLoS One* 2012; **7**(7): e40385.

273. Graham SP, McLean RK, Spencer AJ, et al. Evaluation of the immunogenicity of prime-boost vaccination with the replication-deficient viral vectored COVID-19 vaccine candidate ChAdOx1 nCoV-19. *npj Vaccines* 2020; **5**(1): 69.

274. Folegatti PM, Ewer KJ, Aley PK, et al. Safety and immunogenicity of the ChAdOx1 nCoV-19 vaccine against SARS-CoV-2: a preliminary report of a phase 1/2, single-blind, randomised controlled trial. *Lancet (London, England)* 2020; **396**(10249): 467-78.

275. Voysey M, Clemens SAC, Madhi SA, et al. Safety and efficacy of the ChAdOx1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: an interim analysis of four randomised controlled trials in Brazil, South Africa, and the UK. *Lancet (London, England)* 2021; **397**(10269): 99-111.

276. Voysey M, Costa Clemens SA, Madhi SA, et al. Single-dose administration and the influence of the timing of the booster dose on immunogenicity and efficacy of ChAdOx1 nCoV-19 (AZD1222) vaccine: a pooled analysis of four randomised trials. *The Lancet* 2021; **397**(10277): 881-91.

277. Paris C, Perrin S, Hamonic S, et al. Effectiveness of mRNA-BNT162b2, mRNA-1273, and ChAdOx1 nCoV-19 vaccines against COVID-19 in healthcare workers: an observational study using surveillance data. *Clinical Microbiology and Infection* 2021; **27**(11): 1699.e5-.e8.

278. Amirthalingam G, Bernal JL, Andrews NJ, et al. Serological responses and vaccine effectiveness for extended COVID-19 vaccine schedules in England. *Nature Communications* 2021; **12**(1): 7217.

279. Thiruvengadam R, Awasthi A, Medigeshi G, et al. Effectiveness of ChAdOx1 nCoV-19 vaccine against SARS-CoV-2 infection during the delta (B.1.617.2) variant surge in India: a test-negative, case-control study and a mechanistic study of post-vaccination immune responses. Lancet Infect Dis. 2022 Apr;**22**(4):473-482.

280. Sheikh A, Robertson C, Taylor B. BNT162b2 and ChAdOx1 nCoV-19 Vaccine Effectiveness against Death from the Delta Variant. *New England Journal of Medicine* 2021; **385**(23): 2195-7.

281. Greinacher A, Thiele T, Warkentin TE, Weisser K, Kyrle PA, Eichinger S. Thrombotic Thrombocytopenia after ChAdOx1 nCov-19 Vaccination. *New England Journal of Medicine* 2021; **384**(22): 2092-101.

282. Sadoff J, Gray G, Vandebosch A, et al. Safety and Efficacy of Single-Dose Ad26.COV2.S Vaccine against Covid-19. *New England Journal of Medicine* 2021; **384**(23): 2187-201.

283. Logunov DY, Dolzhikova IV, Shcheblyakov DV, et al. Safety and efficacy of an rAd26 and rAd5 vector-based heterologous prime-boost COVID-19 vaccine: an interim analysis of a randomised controlled phase 3 trial in Russia. *The Lancet* 2021; **397**(10275): 671-81.

284. Barrett JR, Belij-Rammerstorfer S, Dold C, et al. Phase 1/2 trial of SARS-CoV-2 vaccine ChAdOx1 nCoV-19 with a booster dose induces multifunctional antibody responses. *Nat Med* 2021; **27**(2): 279-88.

285. 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.

286. Flaxman A, Marchevsky NG, Jenkin D, et al. Reactogenicity and immunogenicity after a late second dose or a third dose of ChAdOx1 nCoV-19 in the UK: a substudy of two randomised controlled trials (COV001 and COV002). *The Lancet* 2021; **398**(10304): 981-90.

287. Sobieszczyk ME, Maaske J, Falsey AR, et al. Durability of protection and immunogenicity of AZD1222 (ChAdOx1 nCoV-19) COVID-19 vaccine over 6 months. *The Journal of Clinical Investigation* 2022; **132**(18).

288. Parry H, Bruton R, Tut G, et al. Immunogenicity of single vaccination with BNT162b2 or ChAdOx1 nCoV-19 at 5–6 weeks post vaccine in participants aged 80 years or older: an exploratory analysis. *The Lancet Healthy Longevity* 2021; **2**(9): e554-e60.

289. Baden LR, El Sahly HM, Essink B, et al. Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine. *New England Journal of Medicine* 2020; **384**(5): 403-16.

290. Dickerman BA, Gerlovin H, Madenci AL, et al. Comparative Effectiveness of BNT162b2 and mRNA-1273 Vaccines in U.S. Veterans. *New England Journal of Medicine* 2021; **386**(2): 105-15.

291. Gao Q, Bao L, Mao H, et al. Development of an inactivated vaccine candidate for SARS-CoV-2. *Science* 2020; **369**(6499): 77-81.

292. Wang H, Zhang Y, Huang B, et al. Development of an Inactivated Vaccine Candidate, BBIBP-CorV, with Potent Protection against SARS-CoV-2. *Cell* 2020; **182**(3): 713-21.e9.

293. Zhang Y, Zeng G, Pan H, et al. Safety, tolerability, and immunogenicity of an inactivated SARS-CoV-2 vaccine in healthy adults aged 18–59 years: a randomised, double-blind, placebo-controlled, phase 1/2 clinical trial. *The Lancet Infectious Diseases* 2021; **21**(2): 181-92.

294. Palacios R, Batista AP, Albuquerque CSN, et al. Efficacy and safety of a COVID-19 inactivated vaccine in healthcare professionals in Brazil: the PROFISCOV study. SSRN; 2021. DOI: 10.2139/ssrn.3822780.

295. Tanriover MD, Doğanay HL, Akova M, et al. Efficacy and safety of an inactivated whole-virion SARS-CoV-2 vaccine (CoronaVac): interim results of a double-blind, randomised, placebo-controlled, phase 3 trial in Turkey. *The Lancet* 2021; **398**(10296): 213-22.

296. Kwok SL, Cheng SM, Leung JN, et al. Waning antibody levels after COVID-19 vaccination with mRNA Comirnaty and inactivated CoronaVac vaccines in blood donors, Hong Kong, April 2020 to October 2021. *Eurosurveillance* 2022; **27**(2): 2101197.

297. Zeng G, Wu Q, Pan H, et al. Immunogenicity and safety of a third dose of CoronaVac, and immune persistence of a two-dose schedule, in healthy adults: interim results from two single-centre, double-blind, randomised, placebo-controlled phase 2 clinical trials. Lancet Infect Dis. 2022 Apr;**22**(4):483-495. Epub 2021 Dec 8.

298. Heath PT, Galiza EP, Baxter DN, et al. Safety and Efficacy of NVX-CoV2373 Covid-19 Vaccine. *New England Journal of Medicine* 2021; **385**(13): 1172-83.

299. Gov.uk. COVID-19 vaccination first phase priority groups. 2021. https://www.gov.uk/government/publications/covid-19-vaccination-care-home-and-healthcare-settings-posters/covid-19-vaccination-first-phase-priority-groups (accessed 3rd February 2022).

300. Watson OJ, Barnsley G, Toor J, Hogan AB, Winskill P, Ghani AC. Global impact of the first year of COVID-19 vaccination: a mathematical modelling study. *The Lancet Infectious Diseases* 2022; **22**(9): 1293-302.

301. Eyre DW, Taylor D, Purver M, et al. Effect of Covid-19 Vaccination on Transmission of Alpha and Delta Variants. *New England Journal of Medicine* 2022; **386**(8):744-756

302. World Health Organisation. Novel Coronavirus(2019-nCoV) Situation Report - 10, 30th January 2020. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200130-sitrep-10-ncov.pdf?sfvrsn=d0b2e480_2 (accessed 30th January 2022).

303. Jefferson T, Del Mar CB, Dooley L, et al. Physical interventions to interrupt or reduce the spread of respiratory viruses. Cochrane Database Syst Rev. 2011 Jul 6;2011(7):CD006207.

304. Worby CJ, Chang H-H. Face mask use in the general population and optimal resource allocation during the COVID-19 pandemic. *Nature Communications* 2020; **11**(1): 4049.

305. Li Y, Liang M, Gao L, et al. Face masks to prevent transmission of COVID-19: A systematic review and meta-analysis. *American Journal of Infection Control* 2021; **49**(7): 900-6.

306. Bundgaard H, Bundgaard JS, Raaschou-Pedersen DET, et al. Effectiveness of Adding a Mask Recommendation to Other Public Health Measures to Prevent SARS-CoV-2 Infection in Danish Mask Wearers. *Annals of Internal Medicine* 2020; **174**(3): 335-43.

307. Talic S, Shah S, Wild H, et al. Effectiveness of public health measures in reducing the incidence of covid-19, SARS-CoV-2 transmission, and covid-19 mortality: systematic review and meta-analysis. *BMJ* 2021; **375**: e068302.

308. Leffler CT, Ing E, Lykins JD, et al. Association of Country-wide Coronavirus Mortality with Demographics, Testing, Lockdowns, and Public Wearing of Masks. *The American journal of tropical medicine and hygiene* 2020; **103**(6): 2400-11.

309. Kamal R. Mahtani CH, Jeffrey K. Aronson. What is the evidence for social distancing during global pandemics? March 19 2020. 19 March 2020. https://www.cebm.net/covid-19/what-is-the-evidence-for-social-distancing-during-global-pandemics/ (accessed 31st January 2022).

310. W. H. O. Working Group on the Clinical Characterisation and Management of Covid-19 infection. A minimal common outcome measure set for COVID-19 clinical research. *The Lancet Infectious diseases* 2020; **20**(8): e192-e7.

311. Harris PA, Taylor R, Thielke R, et al. Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics* 2009; **42**(2): 377-81.

312. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: Building an international community of software platform partners. *Journal of Biomedical Informatics* 2019; **95**: 103208.

313. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria; 2013.

314. Lillie PJ, Samson A, Li A, et al. Novel coronavirus disease (Covid-19): The first two patients in the UK with person to person transmission. *The Journal of infection* 2020; **80**(5): 578-606.

315. Richmond P, Goldblatt D, Fusco PC, et al. Safety and immunogenicity of a new Neisseria meningitidis serogroup C-tetanus toxoid conjugate vaccine in healthy adults. *Vaccine* 1999; **18**(7): 641-6.

316. Lee JKH, Lam GKL, Shin T, Samson SI, Greenberg DP, Chit A. Efficacy and effectiveness of high-dose influenza vaccine in older adults by circulating strain and antigenic match: An updated systematic review and meta-analysis. *Vaccine* 2021; **39**: A24-A35.

317. Yin M, Huang L, Zhang Y, et al. Effectiveness and safety of dual influenza and pneumococcal vaccination versus separate administration or no vaccination in older adults: a meta-analysis. *Expert Review of Vaccines* 2018; **17**(7): 653-63.

318. Wilder-Smith A, Teleman MD, Heng BH, Earnest A, Ling AE, Leo YS. Asymptomatic SARS coronavirus infection among healthcare workers, Singapore. *Emerg Infect Dis* 2005; **11**(7): 1142-5.

319. UK Government. The Health Protection (Coronavirus, Restrictions) (Self-Isolation) (England) Regulations 2020.

320. Walker AS, Pritchard E, House T, et al. Ct threshold values, a proxy for viral load in community SARS-CoV-2 cases, demonstrate wide variation across populations and over time. *Elife* 2021; **10**: e64683.

321. Gov.uk. Age Groups (2011 Census data). Published 22 August 2018. 17 August 2020. https://www.ethnicity-facts-figures.service.gov.uk/uk-population-by-ethnicity/demographics/age-groups/latest#:~:text=21.3%25%20of%20the%20overall%20population,aged%2060%20years%20and%20over (accessed 29th March 2022).

322. Flores LE, Frontera WR, Andrasik MP, et al. Assessment of the Inclusion of Racial/Ethnic Minority, Female, and Older Individuals in Vaccine Clinical Trials. *JAMA Network Open* 2021; **4**(2): e2037640-e.

323. NHS Employers. Gender in the NHS infographic. 2019. https://www.nhsemployers.org/articles/gender-nhs-infographic.

324. Lumley SF, Richens N, Lees E, et al. Changes in paediatric respiratory infections at a UK teaching hospital 2016-2021; impact of the SARS-CoV-2 pandemic. *The Journal of infection* 2022; **84**(1): 40-7.

325. Sherman AC, Babiker A, Sieben AJ, et al. The Effect of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Mitigation Strategies on Seasonal Respiratory Viruses: A Tale of 2 Large Metropolitan Centers in the United States. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2021; **72**(5): e154-e7.

326. Wang K, Eftang CN, Jakobsen RB, Årøen A. Review of response rates over time in registry-based studies using patient-reported outcome measures. *BMJ Open* 2020; **10**(8): e030808.

327. Edwards P, Roberts I, Clarke M, et al. Increasing response rates to postal questionnaires: systematic review. *BMJ* 2002; **324**(7347): 1183.

328. Galasso V, Pons V, Profeta P, Becher M, Brouard S, Foucault M. Gender differences in COVID-19 attitudes and behavior: Panel evidence from eight countries. *Proceedings of the National Academy of Sciences* 2020; **117**(44): 27285-91.

329. Romero Starke K, Reissig D, Petereit-Haack G, Schmauder S, Nienhaus A, Seidler A. The isolated effect of age on the risk of COVID-19 severe outcomes: a systematic review with meta-analysis. *BMJ Global Health* 2021; **6**(12): e006434.

330. Wright L, Steptoe A, Fancourt D. Patterns of compliance with COVID-19 preventive behaviours: a latent class analysis of 20 000 UK adults. *Journal of Epidemiology and Community Health* 2022; **76**(3): 247.

331. Lindan CP, Desai M, Boothroyd D, et al. Design of a population-based longitudinal cohort study of SARS-CoV-2 incidence and prevalence among adults in the San Francisco Bay Area. *Annals of Epidemiology* 2022; **67**: 81-100.

332. 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. *The Lancet Infectious Diseases* 2020; **20**(9): 1034-42.

333. Ho FK, Celis-Morales CA, Gray SR, et al. Modifiable and non-modifiable risk factors for COVID-19, and comparison to risk factors for influenza and pneumonia: results from a UK Biobank prospective cohort study. *BMJ Open* 2020; **10**(11): e040402.

334. Painter SD, Ovsyannikova IG, Poland GA. The weight of obesity on the human immune response to vaccination. *Vaccine* 2015; **33**(36): 4422-9.

335. Butsch WS, Hajduk A, Cardel MI, et al. COVID-19 vaccines are effective in people with obesity: A position statement from The Obesity Society. *Obesity (Silver Spring)* 2021; **29**(10): 1575-9.

336. Poletti P, Tirani M, Cereda D, et al. Association of Age With Likelihood of Developing Symptoms and Critical Disease Among Close Contacts Exposed to Patients With Confirmed SARS-CoV-2 Infection in Italy. *JAMA Network Open* 2021; **4**(3): e211085-e.

337. Wang B, Andraweera P, Elliott S, et al. Asymptomatic SARS-CoV-2 Infection by Age: A Global Systematic Review and Meta-analysis. *The Pediatric Infectious Disease Journal* 2022.

338. Tu Y-P, Jennings R, Hart B, et al. Swabs Collected by Patients or Health Care Workers for SARS-CoV-2 Testing. *New England Journal of Medicine* 2020; **383**(5): 494-6.

339. Tsang NNY, So HC, Ng KY, Cowling BJ, Leung GM, Ip DKM. Diagnostic performance of different sampling approaches for SARS-CoV-2 RT-PCR testing: a systematic review and meta-analysis. *The Lancet Infectious Diseases* 2021; **21**(9): 1233-45.

340. McCulloch DJ, Kim AE, Wilcox NC, et al. Comparison of Unsupervised Home Self-collected Midnasal Swabs With Clinician-Collected Nasopharyngeal Swabs for Detection of SARS-CoV-2 Infection. *JAMA Network Open* 2020; **3**(7): e2016382-e.

341. Wehrhahn MC, Robson J, Brown S, et al. Self-collection: An appropriate alternative during the SARS-CoV-2 pandemic. *Journal of Clinical Virology* 2020; **128**: 104417.

342. Marx GE, Biggerstaff BJ, Nawrocki CC, et al. Detection of Severe Acute Respiratory Syndrome Coronavirus 2 on Self-Collected Saliva or Anterior Nasal Specimens Compared With Healthcare Personnel–Collected Nasopharyngeal Specimens. *Clinical Infectious Diseases* 2021; **73**(Supplement_1): S65-S73.

343. Organization WH. WHO target product profiles for COVID-19 vaccines (2020).

344. Hodgson SH, Mansatta K, Mallett G, Harris V, Emary KRW, Pollard AJ. What defines an efficacious COVID-19 vaccine? A review of the challenges assessing the clinical efficacy of vaccines against SARS-CoV-2. *The Lancet Infectious diseases* 2021; **21**(2): e26-e35.

345. Lambe T, Spencer AJ, Thomas KM, et al. ChAdOx1 nCoV-19 protection against SARS-CoV-2 in rhesus macaque and ferret challenge models. *Communications Biology* 2021; **4**(1): 915.

346. Fajnzylber J, Regan J, Coxen K, et al. SARS-CoV-2 viral load is associated with increased disease severity and mortality. *Nature Communications* 2020; **11**(1): 5493.

347. Zheng S, Fan J, Yu F, et al. Viral load dynamics and disease severity in patients infected with SARS-CoV-2 in Zhejiang province, China, January-March 2020: retrospective cohort study. *BMJ* 2020; **369**: m1443-m.

348. Feng S, Phillips DJ, White T, et al. Correlates of protection against symptomatic and asymptomatic SARS-CoV-2 infection. *Nature medicine* 2021; **27**(11): 2032-40.

349. Kissler SM, Fauver JR, Mack C, et al. Viral dynamics of acute SARS-CoV-2 infection and applications to diagnostic and public health strategies. *PLOS Biology* 2021; **19**(7): e3001333.

350. Han A, Czajkowski LM, Donaldson A, et al. A Dose-finding Study of a Wild-type Influenza A(H3N2) Virus in a Healthy Volunteer Human Challenge Model. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2019; **69**(12): 2082-90.

351. Carrat F, Vergu E, Ferguson NM, et al. Time Lines of Infection and Disease in Human Influenza: A Review of Volunteer Challenge Studies. *American Journal of Epidemiology* 2008; **167**(7): 775-85.

352. Schrock JM, Ryan DT, Saber R, et al. Cohabitation With a Known Coronavirus Disease 2019 Case Is Associated With Greater Antibody Concentration and Symptom Severity in a Community-Based Sample of Seropositive Adults. *Open Forum Infect Dis* 2021; **8**(7): ofab244.

353. Poulsen A, Qureshi K, Lisse I, et al. A Household Study of Chickenpox in Guinea-Bissau: Intensity of Exposure is a Determinant of Severity. *Journal of Infection* 2002; **45**(4): 237-42.

354. Byambasuren O, Cardona M, Bell K, Clark J, McLaws M-L, Glasziou P. Estimating the extent of asymptomatic COVID-19 and its potential for community transmission: Systematic review and meta-analysis. *Official Journal of the Association of Medical Microbiology and Infectious Disease Canada* 2020; **5**(4): 223-34.

355. Buitrago-Garcia D, Egli-Gany D, Counotte MJ, et al. Occurrence and transmission potential of asymptomatic and presymptomatic SARS-CoV-2 infections: A living systematic review and meta-analysis. *PLOS Medicine* 2020; **17**(9): e1003346.

356. Meyerowitz EA, Richterman A, Bogoch II, Low N, Cevik M. Towards an accurate and systematic characterisation of persistently asymptomatic infection with SARS-CoV-2. *The Lancet Infectious Diseases* 2021; **21**(6): e163-e9.

357. Zhou R, Li F, Chen F, et al. Viral dynamics in asymptomatic patients with COVID-19. *Int J Infect Dis* 2020; **96**: 288-90.

358. Pritchard E, Matthews PC, Stoesser N, et al. Impact of vaccination on new SARS-CoV-2 infections in the United Kingdom. *Nature Medicine* 2021; **27**(8): 1370-8.

359. Hay James A, Kennedy-Shaffer L, Kanjilal S, et al. Estimating epidemiologic dynamics from cross-sectional viral load distributions. *Science* 2021; **373**(6552): eabh0635.

360. Faria NR, Mellan TA, Whittaker C, et al. Genomics and epidemiology of the P. 1 SARS-CoV-2 lineage in Manaus, Brazil. *Science* 2021; **372**(6544): 815-21.

361. Starr TN, Greaney AJ, Hilton SK, et al. Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. *Cell* 2020; **182**(5): 1295-310.e20.

362. Meng B, Kemp SA, Papa G, et al. Recurrent emergence of SARS-CoV-2 spike deletion H69/V70 and its role in the Alpha variant B.1.1.7. *Cell reports* 2021; **35**(13): 109292.

363. Volz E, Mishra S, Chand M, et al. Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature* 2021; **593**(7858): 266-9.

364. Davies NG, Abbott S, Barnard RC, et al. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science* 2021; **372**(6538): eabg3055.

365. Rosenke K, Feldmann F, Okumura A, et al. UK B.1.1.7 (Alpha) variant exhibits increased respiratory replication and shedding in nonhuman primates. *Emerging Microbes & Infections* 2021; **10**(1): 2173-82.

366. Giles B, Meredith P, Robson S, Smith G, Chauhan A. The SARS-CoV-2 B.1.1.7 variant and increased clinical severity-the jury is out. *The Lancet Infectious Diseases* 2021; **21**(9): 1213-4.

367. Graham MS, Sudre CH, May A, et al. Changes in symptomatology, reinfection, and transmissibility associated with the SARS-CoV-2 variant B. 1.1. 7: an ecological study. *The Lancet Public Health* 2021; **6**(5): e335-e45.

368. Collier DA, De Marco A, Ferreira IATM, et al. Sensitivity of SARS-CoV-2 B.1.1.7 to mRNA vaccine-elicited antibodies. *Nature* 2021; **593**(7857): 136-41.

369. Muik A, Wallisch A-K, Sänger B, et al. Neutralization of SARS-CoV-2 lineage B.1.1.7 pseudovirus by BNT162b2 vaccine–elicited human sera. *Science* 2021; **371**(6534): 1152-3.

370. Müller K, Girl P, Giebl A, et al. Emerging SARS-CoV-2 variant B.1.1.7 reduces neutralisation activity of antibodies against wild-type SARS-CoV-2. *Journal of Clinical Virology* 2021; **142**: 104912.

371. Planas D, Bruel T, Grzelak L, et al. Sensitivity of infectious SARS-CoV-2 B.1.1.7 and B.1.351 variants to neutralizing antibodies. *Nature Medicine* 2021; **27**(5): 917-24.

372. Wang P, Nair MS, Liu L, et al. Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. *Nature* 2021; **593**(7857): 130-5.

373. Fumagalli SE, Padhiar NH, Meyer D, et al. Analysis of 3.5 million SARS-CoV-2 sequences reveals unique mutational trends with consistent nucleotide and codon frequencies. *Virology Journal* 2023; **20**(1): 31.

374. Raghwani J, Rose R, Sheridan I, et al. Exceptional Heterogeneity in Viral Evolutionary Dynamics Characterises Chronic Hepatitis C Virus Infection. *PLOS Pathogens* 2016; **12**(9): e1005894.

375. Kemal KS, Foley B, Burger H, et al. HIV-1 in genital tract and plasma of women: Compartmentalization of viral sequences, coreceptor usage, and glycosylation. *Proceedings of the National Academy of Sciences* 2003; **100**(22): 12972-7.

376. Rockett R, Basile K, Maddocks S, et al. Resistance Mutations in SARS-CoV-2 Delta Variant after Sotrovimab Use. *New England Journal of Medicine* 2022; **386**(15): 1477-9.

377. Wang Y, Wang D, Zhang L, et al. Intra-host variation and evolutionary dynamics of SARS-CoV-2 populations in COVID-19 patients. *Genome Medicine* 2021; **13**(1): 30.

378. Li J, Du P, Yang L, et al. Two-step fitness selection for intra-host variations in SARS-CoV-2. *Cell Reports* 2022; **38**(2): 110205.

379. Tonkin-Hill G, Martincorena I, Amato R, et al. Patterns of within-host genetic diversity in SARS-CoV-2. *Elife* 2021; **10**: e66857.

380. Zhang Y, Jiang N, Qi W, et al. SARS-CoV-2 intra-host single-nucleotide variants associated with disease severity. *Virus Evolution* 2022; **8**(2).

381. Agius JE, Johnson-Mackinnon JC, Fong W, et al. SARS-CoV-2 Within-Host and in vitro Genomic Variability and Sub-Genomic RNA Levels Indicate Differences in Viral Expression Between Clinical Cohorts and in vitro Culture. *Frontiers in Microbiology* 2022; **13**.

382. McCrone John T, Lauring Adam S. Measurements of Intrahost Viral Diversity Are Extremely Sensitive to Systematic Errors in Variant Calling. *Journal of Virology* 2016; **90**(15): 6884-95.

383. Valesano AL, Rumfelt KE, Dimcheff DE, et al. Temporal dynamics of SARS-CoV-2 mutation accumulation within and across infected hosts. *PLOS Pathogens* 2021; **17**(4): e1009499.

384. Gu H, Quadeer AA, Krishnan P, et al. Within-host genetic diversity of SARS-CoV-2 lineages in unvaccinated and vaccinated individuals. *Nature Communications* 2023; **14**(1): 1793.

385. Martin MA, Koelle K. Comment on “Genomic epidemiology of superspreading events in Austria reveals mutational dynamics and transmission properties of SARS-CoV-2”. *Sci Transl Med* 2021; **13**(617): eabh1803.

386. Bendall EE, Callear AP, Getz A, et al. Rapid transmission and tight bottlenecks constrain the evolution of highly transmissible SARS-CoV-2 variants. *Nat Commun* 2023; **14**(1): 272.

387. Hill V, Du Plessis L, Peacock TP, et al. The origins and molecular evolution of SARS-CoV-2 lineage B.1.1.7 in the UK. *Virus Evolution* 2022; **8**(2).

388. Fogel JM, Bonsall D, Cummings V, et al. Performance of a high-throughput next-generation sequencing method for analysis of HIV drug resistance and viral load. *Journal of Antimicrobial Chemotherapy* 2020; **75**(12): 3510-6.

389. Quick J, Grubaugh ND, Pullan ST, et al. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nature protocols* 2017; **12**(6): 1261-76.

390. Wymant C, Blanquart F, Golubchik T, et al. Easy and accurate reconstruction of whole HIV genomes from short-read sequence data with shiver. *Virus evolution* 2018; **4**(1): vey007.

391. Katoh K, Misawa K, Kuma Ki, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 2002; **30**(14): 3059-66.

392. Nguyen L-T, Schmidt HA, Von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular biology and evolution* 2015; **32**(1): 268-74.

393. Yu G. Data Integration, Manipulation and Visualization of Phylogenetic Trees 1st Edition; 2022.

394. Aksamentov I, Roemer C, Hodcroft EB, Neher RA. Nextclade: clade assignment, mutation calling and quality control for viral genomes. *Journal of open source software* 2021; **6**(67): 3773.

395. Sagulenko P, Puller V, Neher RA. TreeTime: Maximum-likelihood phylodynamic analysis. *Virus Evolution* 2018; **4**(1): vex042.

396. Bull RA, Adikari TN, Ferguson JM, et al. Analytical validity of nanopore sequencing for rapid SARS-CoV-2 genome analysis. *Nature Communications* 2020; **11**(1): 6272.

397. Anderson K. gsDesign: Group Sequential Design Version 3.3.0.

398. Fritz SA, Purvis A. Selectivity in Mammalian Extinction Risk and Threat Types: a New Measure of Phylogenetic Signal Strength in Binary Traits. *Conservation Biology* 2010; **24**(4): 1042-51.

399. Orme D, Freckleton R, Thomas G, et al. The caper package: comparative analysis of phylogenetics and evolution in R. *R package version* 2013; **5**(2): 1-36.

400. Public Health England. SARS-CoV-2 variants of concern and variants under investigation in England. Technical briefing 17, 25 June 2021.

401. McCallum M, De Marco A, Lempp FA, et al. N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2. *Cell* 2021; **184**(9): 2332-47.e16.

402. Bonsall D, Golubchik T, de Cesare M, et al. A Comprehensive Genomics Solution for HIV Surveillance and Clinical Monitoring in Low-Income Settings. *Journal of clinical microbiology* 2020; **58**(10).

403. Lopez Bernal J, Andrews N, Gower C, et al. Effectiveness of Covid-19 Vaccines against the B.1.617.2 (Delta) Variant. *New England Journal of Medicine* 2021; **385**(7): 585-94.

404. Abu-Raddad LJ, Chemaitelly H, Butt AA. Effectiveness of the BNT162b2 Covid-19 Vaccine against the B.1.1.7 and B.1.351 Variants. *New England Journal of Medicine* 2021; **385**(2): 187-9.

405. Chemaitelly H, Yassine HM, Benslimane FM, et al. mRNA-1273 COVID-19 vaccine effectiveness against the B.1.1.7 and B.1.351 variants and severe COVID-19 disease in Qatar. *Nature Medicine* 2021; **27**(9): 1614-21.

406. Kissler SM, Fauver JR, Mack C, et al. Viral Dynamics of SARS-CoV-2 Variants in Vaccinated and Unvaccinated Persons. N Engl J Med. 2021 Dec 23;**385**(26):2489-2491. Epub 2021 Dec 1.

407. Calistri P, Amato L, Puglia I, et al. Infection sustained by lineage B.1.1.7 of SARS-CoV-2 is characterised by longer persistence and higher viral RNA loads in nasopharyngeal swabs. *International Journal of Infectious Diseases* 2021; **105**: 753-5.

408. de Silva TI, Liu G, Lindsey BB, et al. The impact of viral mutations on recognition by SARS-CoV-2 specific T cells. *iScience* 2021; **24**(11): 103353.

409. Popa A, Genger J-W, Nicholson MD, et al. Genomic epidemiology of superspreading events in Austria reveals mutational dynamics and transmission properties of SARS-CoV-2. *Sci Transl Med* 2020; **12**(573): eabe2555.

410. Simmonds P. Rampant C→ U hypermutation in the genomes of SARS-CoV-2 and other coronaviruses: causes and consequences for their short-and long-term evolutionary trajectories. *MSphere* 2020; **5**(3): e00408-20.

411. Chen Y, Chen L, Deng Q, et al. The presence of SARS-CoV-2 RNA in the feces of COVID-19 patients. *Journal of Medical Virology* 2020; **92**(7): 833-40.

412. Pan L, Mu M, Yang P, et al. Clinical characteristics of COVID-19 patients with digestive symptoms in Hubei, China: a descriptive, cross-sectional, multicenter study. *The American journal of gastroenterology* 2020; **115**.

413. Cheung KS, Hung IFN, Chan PPY, et al. Gastrointestinal Manifestations of SARS-CoV-2 Infection and Virus Load in Fecal Samples From a Hong Kong Cohort: Systematic Review and Meta-analysis. *Gastroenterology* 2020; **159**(1): 81-95.

414. Xiao F, Tang M, Zheng X, Liu Y, Li X, Shan H. Evidence for Gastrointestinal Infection of SARS-CoV-2. *Gastroenterology* 2020; **158**(6): 1831-3.e3.

415. Zhang W, Du R-H, Li B, et al. Molecular and serological investigation of 2019-nCoV infected patients: implication of multiple shedding routes. *Emerging Microbes & Infections* 2020; **9**(1): 386-9.

416. Xiao F, Sun J, Xu Y, et al. Infectious SARS-CoV-2 in Feces of Patient with Severe COVID-19. *Emerg Infect Dis* 2020; **26**(8): 1920-2.

417. van Doorn AS, Meijer B, Frampton CMA, Barclay ML, de Boer NKH. Systematic review with meta-analysis: SARS-CoV-2 stool testing and the potential for faecal-oral transmission. *Alimentary Pharmacology & Therapeutics* 2020; **52**(8): 1276-88.

418. Kang M, Wei J, Yuan J, et al. Probable Evidence of Fecal Aerosol Transmission of SARS-CoV-2 in a High-Rise Building. *Annals of Internal Medicine* 2020; **173**(12): 974-80.

419. Livanos AE, Jha D, Cossarini F, et al. Intestinal Host Response to SARS-CoV-2 Infection and COVID-19 Outcomes in Patients With Gastrointestinal Symptoms. *Gastroenterology* 2021; **160**(7): 2435-50.e34.

420. Lee JJ, Kopetz S, Vilar E, Shen JP, Chen K, Maitra A. Relative Abundance of SARS-CoV-2 Entry Genes in the Enterocytes of the Lower Gastrointestinal Tract. *Genes* 2020; **11**(6).

421. Zang R, Castro MFG, McCune BT, et al. TMPRSS2 and TMPRSS4 promote SARS-CoV-2 infection of human small intestinal enterocytes. *Science immunology* 2020; **5**(47).

422. Lamers MM, Beumer J, van der Vaart J, et al. SARS-CoV-2 productively infects human gut enterocytes. *Science* 2020; **369**(6499): 50-4.

423. Sheikh-Mohamed S, Isho B, Chao GYC, et al. Systemic and mucosal IgA responses are variably induced in response to SARS-CoV-2 mRNA vaccination and are associated with protection against subsequent infection. *Mucosal Immunology* 2022; **15**(5): 799-808.

424. Nahass GR, Salomon-Shulman RE, Blacker G, et al. Intramuscular SARS-CoV-2 vaccines elicit varying degrees of plasma and salivary antibody responses as compared to natural infection. *medRxiv* 2021: 2021.08.22.21262168.

425. Jin C, Hill J, Gunn BM, et al. Vi-specific serological correlates of protection for typhoid fever. *Journal of Experimental Medicine* 2020; **218**(2): e20201116.

426. Cheuvart B, Neuzil KM, Steele AD, et al. Association of serum anti-rotavirus immunoglobulin A antibody seropositivity and protection against severe rotavirus gastroenteritis. *Human Vaccines & Immunotherapeutics* 2014; **10**(2): 505-11.

427. Ogra PL, Karzon DT, Righthand F, MacGillivray M. Immunoglobulin Response in Serum and Secretions after Immunization with Live and Inactivated Poliovaccine and Natural Infection. *New England Journal of Medicine* 1968; **279**(17): 893-900.

428. Sterlin D, Mathian A, Miyara M, et al. IgA dominates the early neutralizing antibody response to SARS-CoV-2. *Sci Transl Med* 2021; **13**(577): eabd2223.

429. PPD. Real Time One-Step RT-qPCR Detection and Quantification of SARS-CoV-2 SOP, 2021.

430. Zhang Y, Chen C, Song Y, et al. Excretion of SARS-CoV-2 through faecal specimens. *Emerging Microbes & Infections* 2020; **9**(1): 2501-8.

431. Zhang Y, Cen M, Hu M, et al. Prevalence and Persistent Shedding of Fecal SARS-CoV-2 RNA in Patients With COVID-19 Infection: A Systematic Review and Meta-analysis. *Clin Transl Gastroenterol* 2021; **12**(4): e00343.

432. Wu Y, Guo C, Tang L, et al. Prolonged presence of SARS-CoV-2 viral RNA in faecal samples. *The Lancet Gastroenterology & Hepatology* 2020; **5**(5): 434-5.

433. Lin L, Jiang X, Zhang Z, et al. Gastrointestinal symptoms of 95 cases with SARS-CoV-2 infection. *Gut* 2020; **69**(6): 997.

434. Pajon R, Paila YD, Girard B, et al. Initial analysis of viral dynamics and circulating viral variants during the mRNA-1273 Phase 3 COVE trial. Nat Med. 2022 Apr;**28**(4):823-830. Epub 2022 Feb 10.

435. Levine-Tiefenbrun M, Yelin I, Katz R, et al. Initial report of decreased SARS-CoV-2 viral load after inoculation with the BNT162b2 vaccine. *Nature Medicine* 2021; **27**(5): 790-2.

436. Rogers Amy A, Baumann Russell E, Borillo Gwynngelle A, et al. Evaluation of Transport Media and Specimen Transport Conditions for the Detection of SARS-CoV-2 by Use of Real-Time Reverse Transcription-PCR. *Journal of clinical microbiology* 2020; **58**(8): e00708-20.

437. Dzung A, Cheng PF, Stoffel C, et al. Prolonged Unfrozen Storage and Repeated Freeze-Thawing of SARS-CoV-2 Patient Samples Have Minor Effects on SARS-CoV-2 Detectability by RT-PCR. *The Journal of Molecular Diagnostics* 2021; **23**(6): 691-7.

438. Our World in Data. Coronavirus (COVID-19) Vaccinations. https://ourworldindata.org/covid-vaccinations (accessed 7th July 2022).

439. White EM, Santostefano CM, Feifer RA, et al. Asymptomatic and Presymptomatic Severe Acute Respiratory Syndrome Coronavirus 2 Infection Rates in a Multistate Sample of Skilled Nursing Facilities. *JAMA Internal Medicine* 2020; **180**(12): 1709-11.

440. ONS. Coronavirus (COVID-19) latest insights. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/articles/coronaviruscovid19/latestinsights#cis (accessed 12th July 2022).

441. Reynolds Catherine J, Pade C, Gibbons Joseph M, et al. Immune boosting by B.1.1.529 (Omicron) depends on previous SARS-CoV-2 exposure. *Science*; **0**(0): eabq1841.

442. van Doremalen N, Purushotham Jyothi N, Schulz Jonathan E, et al. Intranasal ChAdOx1 nCoV-19/AZD1222 vaccination reduces viral shedding after SARS-CoV-2 D614G challenge in preclinical models. *Sci Transl Med* 2021; **13**(607): eabh0755.

443. de Swart RL, de Leeuw OS, Oreshkova N, et al. Intranasal administration of a live-attenuated recombinant newcastle disease virus expressing the SARS-CoV-2 Spike protein induces high neutralizing antibody levels and protects from experimental challenge infection in hamsters. *Vaccine* Aug 5;**40**(33):4676-4681. Epub 2022 Jul 6.

444. Dolgin E. Pan-coronavirus vaccine pipeline takes form. *Nat Rev Drug Discov* 2022; **21**: 324-6.

445. Natarajan A, Zlitni S, Brooks EF, et al. Gastrointestinal symptoms and fecal shedding of SARS-CoV-2 RNA suggest prolonged gastrointestinal infection. *Med* 2022; **3**(6): 371-87.e9.

446. Yonker LM, Gilboa T, Ogata AF, et al. Multisystem inflammatory syndrome in children is driven by zonulin-dependent loss of gut mucosal barrier. *The Journal of Clinical Investigation* 2021; **131**(14).

447. Liu Q, Mak JWY, Su Q, et al. Gut microbiota dynamics in a prospective cohort of patients with post-acute COVID-19 syndrome. *Gut* 2022; **71**(3): 544.

448. Grant AJ, Restif O, McKinley TJ, Sheppard M, Maskell DJ, Mastroeni P. Modelling within-host spatiotemporal dynamics of invasive bacterial disease. *PLoS Biol* 2008; **6**(4): e74.

449. Sheppard M, Webb C, Heath F, et al. Dynamics of bacterial growth and distribution within the liver during Salmonella infection. *Cell Microbiol* 2003; **5**(9): 593-600.

450. Goldberg LJ, Watkins HMS, Dolmatz MS, Schlamm NA. Studies on the Experimental Epidemiology of Respiratory Infections: VI. The Relationship between Dose of Microorganisms and Subsequent Infection or Death of a Host. *The Journal of Infectious Diseases* 1954; **94**(1): 9-21.

451. Meynell GG, Stocker BA. Some hypotheses on the aetiology of fatal infections in partially resistant hosts and their application to mice challenged with Salmonella paratyphi-B or Salmonella typhimurium by intraperitoneal injection. *J Gen Microbiol* 1957; **16**(1): 38-58.

452. Gama JA, Abby SS, Vieira-Silva S, et al. Immune subversion and quorum-sensing shape the variation in infectious dose among bacterial pathogens. *PLoS pathogens* 2012; **8**(2): e1002503-e.

453. Abel S, Abel zur Wiesch P, Davis BM, Waldor MK. Analysis of Bottlenecks in Experimental Models of Infection. *PLoS Pathog* 2015; **11**(6): e1004823.

454. Meynell GG. The applicability of the hypothesis of independent action to fatal infections in mice given Salmonella typhimurium by mouth. *J Gen Microbiol* 1957; **16**(2): 396-404.

455. Moxon ER, Murphy PA. Haemophilus influenzae bacteremia and meningitis resulting from survival of a single organism. *Proc Natl Acad Sci U S A* 1978; **75**(3): 1534-6.

456. Rubin LG, Moxon ER. Haemophilus influenzae Type b Colonization Resulting from Survival of a Single Organism. *The Journal of Infectious Diseases* 1984; **149**(2): 278-.

457. Pluschke G, Mercer A, Kusećek B, et al. Induction of bacteremia in newborn rats by Escherichia coli K1 is correlated with only certain O (lipopolysaccharide) antigen types. *Infection and immunity* 1983; **39**(2): 599-608.

458. Ercoli G, Fernandes VE, Chung WY, et al. Intracellular replication of Streptococcus pneumoniae inside splenic macrophages serves as a reservoir for septicaemia. *Nat Microbiol* 2018; **3**(5): 600-10.

459. Abel S, Abel zur Wiesch P, Chang HH, et al. Sequence tag-based analysis of microbial population dynamics. *Nat Methods* 2015; **12**(3): 223-6.

460. Rossi O, Dybowski R, Maskell DJ, Grant AJ, Restif O, Mastroeni P. Within-host spatiotemporal dynamics of systemic Salmonella infection during and after antimicrobial treatment. *Journal of Antimicrobial Chemotherapy* 2017; **72**(12): 3390-7.

461. Kaiser P, Slack E, Grant AJ, Hardt WD, Regoes RR. Lymph node colonization dynamics after oral Salmonella Typhimurium infection in mice. *PLoS Pathog* 2013; **9**(9): e1003532.

462. Gerlini A, Colomba L, Furi L, et al. The role of host and microbial factors in the pathogenesis of pneumococcal bacteraemia arising from a single bacterial cell bottleneck. *PLoS Pathog* 2014; **10**(3): e1004026.

463. The UK Sepsis Trust. The Sepsis Manual 5th Edition 2019.

464. Kothary MH, Babu US. Infective dose of foodborne pathogens in volunteers: a review. *Journal of food safety* 2001; **21**(1): 49-73.

465. Baker S, Holt KE, Clements AC, et al. Combined high-resolution genotyping and geospatial analysis reveals modes of endemic urban typhoid fever transmission. *Open Biol* 2011; **1**(2): 110008.

466. Perera N, Geary C, Wiselka M, et al. Mixed Salmonella Infection: Case Report and Review of the Literature. *Journal of Travel Medicine* 2007; **14**(2): 134-5.

467. Sultan BA, Fasih N, Qaiser S, et al. Mixed Salmonella infection: a report of two cases from the Aga Khan University Hospital, Karachi. *JPMA The Journal of the Pakistan Medical Association* 2013; **63**(4): 524-6.

468. Klein K, Nickel G, Nankya I, et al. Higher sequence diversity in the vaginal tract than in blood at early HIV-1 infection. *PLoS pathogens* 2018; **14**(1): e1006754-e.

469. Valley-Omar Z, Iyengar P, von Mollendorf C, et al. Intra-host and intra-household diversity of influenza A viruses during household transmissions in the 2013 season in 2 peri-urban communities of South Africa. *PLoS One* 2018; **13**(5): e0198101-e.

470. Han AX, Felix Garza ZC, Welkers MR, et al. Within-host evolutionary dynamics of seasonal and pandemic human influenza A viruses in young children. *Elife* 2021; **10**: e68917.

471. Boeras DI, Hraber PT, Hurlston M, et al. Role of donor genital tract HIV-1 diversity in the transmission bottleneck. *Proc Natl Acad Sci U S A* 2011; **108**(46): E1156-63.

472. Holt KE, Parkhill J, Mazzoni CJ, et al. High-throughput sequencing provides insights into genome variation and evolution in Salmonella Typhi. *Nat Genet* 2008; **40**(8): 987-93.

473. Tapia MD, Tennant SM, Bornstein K, et al. Invasive Nontyphoidal Salmonella Infections Among Children in Mali, 2002-2014: Microbiological and Epidemiologic Features Guide Vaccine Development. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2015; **61 Suppl 4**: S332-8.

474. Aidley J, Rajopadhye S, Akinyemi NM, et al. Nonselective Bottlenecks Control the Divergence and Diversification of Phase-Variable Bacterial Populations. *MBio* 2017; **8**(2).

475. GBD 2017 Typhoid and Paratyphoid Collaborators. The global burden of typhoid and paratyphoid fevers: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet Infect Dis*. 2019 Apr;**19**(4):369-381. Epub 2019 Feb 18.

476. Darton TC, Meiring JE, Tonks S, et al. The STRATAA study protocol: a programme to assess the burden of enteric fever in Bangladesh, Malawi and Nepal using prospective population census, passive surveillance, serological studies and healthcare utilisation surveys. *BMJ Open* 2017; **7**(6): e016283.

477. Barkume C, Date K, Saha SK, et al. Phase I of the Surveillance for Enteric Fever in Asia Project (SEAP): An Overview and Lessons Learned. *J Infect Dis* 2018; **218**(suppl_4): S188-s94.

478. John J, Bavdekar A, Rongsen-Chandola T, et al. Estimating the incidence of enteric fever in children in India: a multi-site, active fever surveillance of pediatric cohorts. *BMC Public Health* 2018; **18**(1): 594.

479. Park SE, Toy T, Cruz Espinoza LM, et al. The Severe Typhoid Fever in Africa Program: Study Design and Methodology to Assess Disease Severity, Host Immunity, and Carriage Associated With Invasive Salmonellosis. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2019; **69**(Suppl 6): S422-s34.

480. Baker S, Favorov M, Dougan G. Searching for the elusive typhoid diagnostic. *BMC Infectious Diseases* 2010; **10**(1): 45.

481. Dougan G, Baker S. Salmonella enterica Serovar Typhi and the Pathogenesis of Typhoid Fever. *Annual Review of Microbiology* 2014; **68**(1): 317-36.

482. Dunlap NE, Benjamin WH, Berry AK, et al. A 'safe-site' for Salmonella typhimurium is within splenic polymorphonuclear cells. *Microb Pathog* 1992; **13**(3): 181-90.

483. Darton TC, Zhou L, Blohmke CJ, et al. Blood culture-PCR to optimise typhoid fever diagnosis after controlled human infection identifies frequent asymptomatic cases and evidence of primary bacteraemia. *Journal of Infection* 2017; **74**(4): 358-66.

484. Hornick R, Greisman S, Woodward T, et al. Typhoid fever: pathogenesis and immunologic control. *New England journal of medicine* 1970; **283**(14): 739-46.

485. Waddington CS, Darton TC, Jones C, et al. An outpatient, ambulant-design, controlled human infection model using escalating doses of Salmonella Typhi challenge delivered in sodium bicarbonate solution. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2014; **58**(9): 1230-40.

486. Blaser MJ, Newman LS. A Review of Human Salmonellosis: I. Infective Dose. *Reviews of Infectious Diseases* 1982; **4**(6): 1096-106.

487. Saha S, Islam MS, Sajib MSI, et al. Epidemiology of Typhoid and Paratyphoid: Implications for Vaccine Policy. *Clinical Infectious Diseases* 2019; **68**(Supplement_2): S117-S23.

488. Black RE, Cisneros L, Levine MM, et al. Case-control study to identify risk factors for paediatric endemic typhoid fever in Santiago, Chile. *Bull World Health Organ* 1985; **63**(5): 899-904.

489. Gasem MH, Dolmans W, Keuter M, Djokomoeljanto R. Poor food hygiene and housing as risk factors for typhoid fever in Semarang, Indonesia. *Tropical Medicine & International Health* 2001; **6**(6): 484-90.

490. Smith AM, Smouse SL, Tau NP, et al. Laboratory-acquired infections of Salmonella enterica serotype Typhi in South Africa: phenotypic and genotypic analysis of isolates. *BMC Infectious Diseases* 2017; **17**(1): 1-9.

491. Mogasale V, Ramani E, Mogasale VV, Park J. What proportion of Salmonella Typhi cases are detected by blood culture? A systematic literature review. *Annals of clinical microbiology and antimicrobials* 2016; **15**(1): 1-8.

492. Wain J, Bay PVB, Vinh H, et al. Quantitation of bacteria in bone marrow from patients with typhoid fever: relationship between counts and clinical features. *Journal of clinical microbiology* 2001; **39**(4): 1571-6.

493. Darton TC, Jones C, Blohmke CJ, et al. Using a Human Challenge Model of Infection to Measure Vaccine Efficacy: a Randomised, Controlled Trial Comparing the Typhoid Vaccines M01ZH09 with Placebo and Ty21a. *Plos neglected tropical diseases* 2016; **10**(8): e0004926.

494. Jin C, Gibani MM, Moore M, et al. Efficacy and immunogenicity of a Vi-tetanus toxoid conjugate vaccine in the prevention of typhoid fever using a controlled human infection model of Salmonella Typhi: a randomised controlled, phase 2b trial. *Lancet (London, England)* 2017; **390**(10111): 2472-80.

495. Gibani MM, Jin C, Shrestha S, et al. Homologous and heterologous re-challenge with Salmonella Typhi and Salmonella Paratyphi A in a randomised controlled human infection model. *PLoS Negl Trop Dis* 2020; **14**(10): e0008783.

496. Gibani MM, Jones E, Barton A, et al. Investigation of the role of typhoid toxin in acute typhoid fever in a human challenge model. *Nat Med* 2019; **25**(7): 1082-8.

497. Wain J, Diep TS, Ho VA, et al. Quantitation of bacteria in blood of typhoid fever patients and relationship between counts and clinical features, transmissibility, and antibiotic resistance. *Journal of clinical microbiology* 1998; **36**(6): 1683-7.

498. Dunstan SJ, Hue NT, Han B, et al. Variation at HLA-DRB1 is associated with resistance to enteric fever. *Nature genetics* 2014; **46**(12): 1333-6.

499. Tennant SM, Toema D, Qamar F, et al. Detection of Typhoidal and Paratyphoidal Salmonella in Blood by Real-time Polymerase Chain Reaction. *Clinical Infectious Diseases* 2015; **61**(suppl_4): S241-S50.

500. Welling MM, Hensbergen AW, Bunschoten A, et al. An update on radiotracer development for molecular imaging of bacterial infections. *Clinical and Translational Imaging* 2019; **7**(2): 105-24.

501. Dobinson HC, Gibani MM, Jones C, et al. Evaluation of the Clinical and Microbiological Response to Salmonella Paratyphi A Infection in the First Paratyphoid Human Challenge Model. *Clinical Infectious Diseases* 2017; **64**(8): 1066-73.

502. UK Health Security Agency. UK SMI B 37: investigation of blood cultures (for organisms other than Mycobacterium species). 2019.

503. Public Health England. UK Standards for Microbiology Investigations ID 24: identification of Salmonella species. 2015.

504. Zhou L, Pollard AJ. A fast and highly sensitive blood culture PCR method for clinical detection of Salmonella enterica serovar Typhi. *Annals of Clinical Microbiology and Antimicrobials* 2010; **9**(1): 14.

505. Public Health England. B 30: Investigation of faecal specimens for enteric pathogens. 2014.

506. Ye J, Coulouris G, Zaretskaya I, et al. Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* 2012; **13**(1): 134.

507. Untergasser A, Cutcutache I, Koressaar T, et al. Primer3--new capabilities and interfaces. *Nucleic Acids Res* 2012; **40**(15): e115-e.

508. StataCorp. Stata Statistical Software: Release 17.: College Station, TX: StataCorp LLC.; 2021.

509. Gibani MM. Investigating Enteric Fever – Application of Controlled Human Infection Models to Understand Enteric Fever Pathogenesis and Immunity: University of Oxford DPhil thesis; 2018.

510. Wain J, Diep TS, Bay PVB, et al. Specimens and culture media for the laboratory diagnosis of typhoid fever. *The Journal of Infection in Developing Countries* 2008; **2**(06): 469-74.

511. Awofisayo-Okuyelu A, Pratt A, McCarthy N, Hall I. Within-host mathematical modelling of the incubation period of Salmonella Typhi. *R Soc Open Sci* 2019; **6**(9): 182143.

512. Vlazaki M, Huber J, Restif O. Integrating mathematical models with experimental data to investigate the within-host dynamics of bacterial infections. *Pathogens and Disease* 2019; **77**(8).

513. Motta S, Pappalardo F. Mathematical modeling of biological systems. *Briefings in Bioinformatics* 2013; **14**(4): 411-22.

514. Lees EA, Forbester JL, Forrest S, et al. Using Human Induced Pluripotent Stem Cell-derived Intestinal Organoids to Study and Modify Epithelial Cell Protection Against Salmonella and Other Pathogens. *JoVE* 2019; (147): e59478.

515. Li, K., Huang, B., Wu, M. *et al.* Dynamic changes in anti-SARS-CoV-2 antibodies during SARS-CoV-2 infection and recovery from COVID-19. *Nat Commun* **11**, 6044 (2020).

9 Appendices

Appendix 1 Oxford Vaccine Trial Group, and other COVID-19 vaccine trial acknowledgements 334

Appendix 2 COG-UK consortium acknowledgements 354

Appendix 3 SARS-CoV-2 NAAT assays used at study sites for the symptomatic pathway (COV001 & COV002) 364

Appendix 4 Part 2: *Salmonella* Typhi 366

Appendix 4.1 Full Inclusion & Exclusion criteria for the 'Bottlenecks' Study 416

Appendix 4.2 Probability calculations for detection of strains 422

Appendix 1 Oxford Vaccine Trial Group, and other COVID-19 vaccine trial acknowledgements

Marites Aban	NIHR Imperial Clinical Research Facility, London, UK
Abdullah	Hull University Teaching Hospitals NHS Trust, Hull, UK
Kushala W M Abeysekera	University Hospitals Bristol & Weston NHS Foundation Trust
Jeremy Aboagye	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Matthew Adam	Clinical Infection Research Group, Regional Infectious Diseases Unit, NHS Lothian, UK
Kirsty Adams	NIHR UCLH Clinical Research Facility, London, UK
James P. Adamson	Public Health Wales NHS Trust, Cardiff, UK
Gbadebo Adewetan	London Northwest University Healthcare, Northwick Park Hospital, London, UK
Syed Adlou	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Khalil Ahmed	Hull University Teaching Hospitals NHS Trust, Hull, UK
Aabidah Ali	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Elizabeth R. Allen	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Lauren Allen	National Infection Service, Public Health England, UK
Nada Al-Muhandis	Hull University Teaching Hospitals NHS Trust, Hull, UK
Rachel Anslow	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Edward H. Arbe-Barnes	University of Oxford Medical School, Medical Sciences Division, University of Oxford, UK
Markus P. Ariaans	Department of Infection, Immunity and Cardiovascular Disease, University of Sheffield, UK
Gavin Babbage	NIHR Southampton Clinical Research Facility, Southampton, UK
Catherine Bailey	Aneurin Bevan University Health Board, Newport, Wales, UK
Kenneth F. Baker	Department of Infection and Tropical Medicine, Newcastle upon Tyne Hospitals NHS Foundation Trust and Translational and Clinical Research Institute, Immunity and Inflammation Theme, Newcastle University
Megan Baker	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Natalie Baker	National Infection Service, Public Health England, UK
Philip Baker	University of Oxford Medical School, Medical Sciences Division, University of Oxford, UK
Ioana Baleanu	Clinical BioManufacturing Facility, Jenner Institute, University of Oxford, UK
Anna Bara	NIHR Imperial Clinical Research Facility, London, UK
Andrew S. Barr	Department of Infection and Tropical Medicine, Newcastle upon Tyne Hospitals NHS Foundation Trust, UK

Jordan R. Barrett	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Jessica Barrett	London Northwest University Healthcare, Northwick Park Hospital, London, UK
Louise Bates	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Alexander Batten	Clinical BioManufacturing Facility, Jenner Institute, University of Oxford, UK
Kirsten Beadon	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Emily Beales	Vaccine Institute, Institute of Infection & Immunity, St. Georges, University of London and St Georges University Hospitals NHS Trust, London, UK
Rebecca Beckley	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Sandra Belij-Rammerstorfer	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Jonathan Bell	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Duncan Bellamy	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Sue Belton	The University of Nottingham Health Service, Cripps Health Centre, University Park, Nottingham, UK
Adam Berg	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Eleanor Berrie	Clinical BioManufacturing Facility, Jenner Institute, University of Oxford, UK
Lisa Berry	NIHR Southampton Clinical Research Facility, Southampton, UK
Amy Beveridge	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Kevin R Bewley	National Infection Service, Public Health England, UK
Inderjeet Bharaj	London Northwest University Healthcare, Northwick Park Hospital, London, UK
Else Margreet Bijker	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Sarah Birch	Academic Directorate of Communicable Diseases and Specialised Medicine, Sheffield Teaching Hospitals NHS Foundation Trust
Kathryn Birchall	Clinical Research Facility, Sheffield Teaching Hospitals NHS Foundation Trust, UK
Olivia Bird	Vaccine Institute, Institute of Infection & Immunity, St. Georges, University of London and St Georges University Hospitals NHS Trust, London, UK
Karen Bisnauthsing	NIHR BRC at Guy's and St Thomas' NHS Foundation Trust, UK
Mustapha Bittaye	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Luke Blackwell	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Rachel Blacow	Clinical Research Facility, Queen Elizabeth University Hospital, Glasgow, UK

Heather Bletchly	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Caitlin L Blundell	Department of Biochemistry, University of Oxford, UK
Susannah R Blundell	Department of Biochemistry, University of Oxford, UK
Pritesh Bodalia	Pharmacy, University College London Hospitals NHS Trust, UK
Emma Bolam	Clinical BioManufacturing Facility, Jenner Institute, University of Oxford, UK
Elena Boland	Clinical BioManufacturing Facility, Jenner Institute, University of Oxford, UK
Nicola Borthwick	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Amy Boyd	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Penny Bradley	Department of Pharmacy, Newcastle upon Tyne Hospitals NHS Foundation Trust, UK
Tanja Brenner	Clinical BioManufacturing Facility, Jenner Institute, University of Oxford, UK
Alice Bridges-Webb	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Phillip Brown	National Infection Service, Public Health England, UK
Claire Brown	NIHR/Wellcome Trust Birmingham Clinical Research Facility, Birmingham, UK
Charlie Brown-O'Sullivan	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Emily Brunt	National Infection Service, Public Health England, UK
William Budd	NIHR Imperial Clinical Research Facility, London, UK
Jamie Burbage	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Aileen Burn	Research Directorate, Newcastle upon Tyne Hospitals NHS Foundation Trust, UK
Karen R Buttigieg	National Infection Service, Public Health England, UK
Nicholas Byard	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Ingrid Cabrera Puig	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Anna Calvert	Vaccine Institute, Institute of Infection & Immunity, St. Georges, University of London and St Georges University Hospitals NHS Trust, London, UK
Susana Camara	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Federica Cappuccini	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Melanie Carr	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Miles W Carroll	National Infection Service, Public Health England, UK
Andrew Carson-Stevens	Division of Population Medicine, School of Medicine, Cardiff University, UK
Helen R Casey	North Bristol NHS Trust, Bristol, UK

Lucia Carratala Castro	Vaccine Institute, Institute of Infection & Immunity, St. Georges, University of London and St Georges University Hospitals NHS Trust, London, UK
Katrina Cathie	University Hospital Southampton NHS Foundation Trust, UK
Jim Chadwick	National Infection Service, Public Health England, UK
Krishna Chatterjee	NIHR Cambridge Clinical Research Facility, Cambridge, UK
Irina Chelysheva	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Oliver Chester	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Sunder Chita	London Northwest University Healthcare, Northwick Park Hospital, London, UK
Jee-Sun Cho	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Liliana Cifuentes	Kennedy Institute of Rheumatology, Nuffield Department of Orthopaedics, The University of Oxford, UK
Elizabeth Clark	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Matthew Clark	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Rachel Colin-Jones	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Hayley Colton	Department of Infection and Tropical Medicine, Sheffield Teaching Hospitals NHS Foundation Trust
Sean Connarty	London Northwest University Healthcare, Northwick Park Hospital, London, UK
Naomi S. Coombes	National Infection Service, Public Health England, UK
Rachel Cooper	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Tumena Corrah	London Northwest University Healthcare, Northwick Park Hospital, London, UK
Catherine A. Cosgrove	Vaccine Institute, Institute of Infection & Immunity, St. Georges, University of London and St Georges University Hospitals NHS Trust, London, UK
Wendy E. M. Crocker	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Christopher Cunningham	Infectious Diseases Department, Cambridge and Peterborough NHS Foundation Trust, UK
Christina J Cunningham	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Brad E. Damratoski	Clinical BioManufacturing Facility, Jenner Institute, University of Oxford, UK
Zsofia Danos	Vaccine Institute, Institute of Infection & Immunity, St. Georges, University of London and St Georges University Hospitals NHS Trust, London, UK
Mehreen S Datoo	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK

Chandrabali Datta	Clinical BioManufacturing Facility, Jenner Institute, University of Oxford, UK
Hannah Davies	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Sophie Davies	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Judith Davies	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
John Davis	Research Directorate, Newcastle upon Tyne Hospitals NHS Foundation Trust, UK
Tesfaye Demissie	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Amisha Desai	Pharmacy Department, University Hospitals Birmingham NHS Foundation Trust
Claudio Di Maso	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Tanya Dinesh	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Francesca R. Donnellan	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Naomi Douglas	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Charlotte Downing	University of Oxford Medical School, Medical Sciences Division, University of Oxford, UK
Jonathan Drake	University of Oxford Medical School, Medical Sciences Division, University of Oxford, UK
Rachael Drake-Brockman	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Ruth Elizabeth Drury	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Andrew D. S. Duncan	Clinical Infection Research Group, NHS Lothian, Edinburgh, UK
Kirstine Eastick	Hull University Teaching Hospitals NHS Trust, Hull, UK
Mandy Edwards	Aneurin Bevan University Health Board, Newport, Wales, UK
Nick J. Edwards	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Frances Edwards	University Hospitals Bristol & Weston NHS Foundation Trust
Omar M. El Muhanna	Clinical BioManufacturing Facility, Jenner Institute, University of Oxford, UK
Sean C. Elias	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Branwen Ellison-Handley	Clinical Research Facility, Sheffield Teaching Hospitals NHS Foundation Trust, UK
Michael J. Elmore	National Infection Service, Public Health England, UK
Marcus Rex English	University of Oxford Medical School, Medical Sciences Division, University of Oxford, UK
Celestine Eshiwe	Hull University Teaching Hospitals NHS Trust, Hull, UK
Mutjaba Ghulam Farooq	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK

Sofiya Fedosyuk	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Sally Felle	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Susie Ferguson	Clinical Infection Research Group, NHS Lothian, Edinburgh, UK
Carla Ferreira Da Silva	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Richard Fisher	Clinical BioManufacturing Facility, Jenner Institute, University of Oxford, UK
Richard Fitzgerald	Department of Clinical Sciences, Liverpool School of Tropical Medicine, UK
James Fletcher	NIHR Imperial Clinical Research Facility, London, UK
Hazel Fofie	Vaccine Institute, Institute of Infection & Immunity, St. Georges, University of London and St Georges University Hospitals NHS Trust, London, UK
Henry Fok	NIHR BRC at Guy's and St Thomas' NHS Foundation Trust and King's College London British Heart Foundation Centre, School of Cardiovascular Medicine and Sciences, UK
Karen J Ford	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Jamie Fowler	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Emma Francis	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Sabrina Fudge	University Hospitals Bristol & Weston NHS Foundation Trust
Julie Furze	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Pablo Galian-Rubio	Clinical BioManufacturing Facility, Jenner Institute, University of Oxford, UK
Harriet Garlant	National Infection Service, Public Health England, UK
Ester German	Department of Clinical Sciences, Liverpool School of Tropical Medicine, UK
Ciaran Gilbride	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Kerry Godwin	National Infection Service, Public Health England, UK
Karishma Gokani	NIHR/Wellcome Trust Birmingham Clinical Research Facility, Birmingham, UK
Jack Goodall	London Northwest University Healthcare, Northwick Park Hospital, London, UK
Jayne Goodwin	Health and Care Research Wales, Cardiff, UK
Giacomo Gorini	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Lara Gracie	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Nicola Greenwood	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Anishka Gunawardene	London Northwest University Healthcare, Northwick Park Hospital, London, UK
Gaurav Gupta	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK

Mark Hackett	University Hospitals Bristol & Weston NHS Foundation Trust
Joseph Hamlyn	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Daniel Hammersley	The University of Nottingham Health Service, Cripps Health Centre, University Park, Nottingham, UK
Aidan T. Hanrath	Department of Infection and Tropical Medicine, Newcastle upon Tyne Hospitals NHS Foundation Trust and Translational and Clinical Research Institute, Immunity and Inflammation Theme, Newcastle University
Brama Hanumunthadu	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Stephanie A. Harris	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Clair Harris	NIHR BRC at Guy's and St Thomas' NHS Foundation Trust, UK
Thomas D. Harrison	Department of Infection and Tropical Medicine, Sheffield Teaching Hospitals NHS Foundation Trust, UK
Daisy Harrison	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Thomas C. Hart	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
John Haughney	Clinical Research Facility, Queen Elizabeth University Hospital, Glasgow, UK
Sophia Hawkins	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Ian Head	University Hospitals Bristol & Weston NHS Foundation Trust
John Aaron Henry	University of Oxford Medical School, Medical Sciences Division, University of Oxford, UK
David B Hettle	Infection Sciences, North Bristol NHS Trust, Bristol, UK
Jennifer Hill	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Lisa Hitchins	Department of Clinical Sciences, Liverpool School of Tropical Medicine, UK
Gina Hodges	Clinical BioManufacturing Facility, Jenner Institute, University of Oxford, UK
Susanne H. Hodgson	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Mimi M. Hou	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Catherine F. Houlihan	Virology Department, University College London Hospitals NHS Trust, UK and Department of Infection and Immunity, University College London, UK
Ashleigh Howard	Department of Clinical Sciences, Liverpool School of Tropical Medicine, UK
Elizabeth Howe	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Nicola Howell	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Holly E. Humphries	National Infection Service, Public Health England, UK
Katrina Hurley	University Hospitals Bristol & Weston NHS Foundation Trust

Catherine Hyams	Academic Respiratory Unit, University of Bristol, Southmead Hospital, Bristol, UK
Sabina Ikram	Vaccine Institute, Institute of Infection & Immunity, St. Georges, University of London and St Georges University Hospitals NHS Trust, London, UK
Alka Ishwarbhai	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Poppy Iveson	University of Oxford Medical School, Medical Sciences Division, University of Oxford, UK
Frederic Jackson	Clinical BioManufacturing Facility, Jenner Institute, University of Oxford, UK
Susan Jackson	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Natasha Jesudason	MRC - University of Glasgow Centre for Virus Research & Department of Infectious Diseases, Queen Elizabeth University Hospital, UK
Carina C. D. Joe	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Christopher Jones	Infection Sciences, North Bristol NHS Trust, Bristol, UK
Kathryn Jones	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Elizabeth Jones	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Reshma Kailath	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Arnab Kar	London Northwest University Healthcare, Northwick Park Hospital, London, UK
Konstantinos Karampatsas	Vaccine Institute, Institute of Infection & Immunity, St. Georges, University of London and St Georges University Hospitals NHS Trust, London, UK
Mwila Kasanyinga	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Linda J Kay	Department of Infection, Immunity and Cardiovascular Disease, University of Sheffield
Jade Keen	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Johanna Kellett Wright	Infection Sciences, North Bristol NHS Trust, Bristol, UK
Elizabeth J. Kelly	AstraZeneca BioPharmaceuticals PLC
Sarah Kelly	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
David Kerr	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Liaquat Khan	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Baktash Khozoee	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Ankush Khurana	London Northwest University Healthcare, Northwick Park Hospital, London, UK
Sarah Kidd	University Hospitals Bristol & Weston NHS Foundation Trust

Annabel Killen	University of Oxford Medical School, Medical Sciences Division, University of Oxford, UK
Jasmin Kinch	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Patrick Kinch	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Lloyd D. W. King	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Thomas B King	University of Oxford Medical School, Medical Sciences Division, University of Oxford, UK
Lucy Kingham	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Francesca Knapper	University Hospitals Bristol & Weston NHS Foundation Trust
Daniel Knott	National Infection Service, Public Health England, UK
Stanislava Koleva	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Colin W Larkworthy	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Jessica P J Larwood	University of Oxford Medical School, Medical Sciences Division, University of Oxford, UK
Alison M Lawrie	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Emily A. Lees	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Alice Lelliott	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Nana-Marie Lemm	NIHR Imperial Clinical Research Facility, London, UK
Stephanie Leung	National Infection Service, Public Health England, UK
Yuanyuan Li	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Amelia M. Lias	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Konstantinos Liatsikos	Department of Clinical Sciences, Liverpool School of Tropical Medicine and Liverpool University Hospitals NHS Foundation Trust, UK
Aline Linder	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Samuel Lipworth	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Shuchang Liu	Clinical BioManufacturing Facility, Jenner Institute, University of Oxford, UK
Xinxue Liu	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Adam Lloyd	Clinical Research Facility, NHS Lothian, Edinburgh, UK
Lisa Loew	Clinical BioManufacturing Facility, Jenner Institute, University of Oxford, UK
Raquel Lopez Ramon	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK

Jonathan C. MacDonald	Department of Gastroenterology, Queen Elizabeth University Hospital, Glasgow, UK
Gordon MacGregor	Department of Respiratory Medicine, Queen Elizabeth University Hospital, Glasgow, UK
Meera Madhavan	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Rebecca Makinson	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Garry Mallett	University of Oxford Medical School, Medical Sciences Division, University of Oxford, UK
Nicola Manning	University Hospitals Bristol & Weston NHS Foundation Trust
Kushal Mansatta	University of Oxford Medical School, Medical Sciences Division, University of Oxford, UK
Spyridoula Marinou	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Emma Marlow	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Richard P. Marshall	AstraZeneca BioPharmaceuticals PLC
Julia L. Marshall	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Moncy Mathew	Pharmacy Clinical Trials (Adult), Guy's and St Thomas NHS Foundation Trust, UK
Olga Mazur	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Andrea Mazzella	NIHR BRC at Guy's and St Thomas' NHS Foundation Trust, UK
Hugh McCaughan	Laboratory For Bacterial Evolution and Pathogenisis (LBEP), The Roslin Institute, University of Edinburgh, UK
Joanne McEwan	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Rosa Maeve McGing	Hull University Teaching Hospitals NHS Trust, Hull, UK
Joanna McGlashan	National Infection Service, Public Health England, UK
Lorna McInroy	National Infection Service, Public Health England, UK
Zoe McIntyre	NIHR Cambridge Clinical Research Facility, Cambridge, UK
Tom McLellan	ILD Service, Royal Papworth NHS Foundation Trust, Cambridge, UK
Steve McSwiggan	Clinical Infection Research Group, NHS Lothian, Edinburgh, UK
Savviz Mehdipour	NIHR Imperial Clinical Research Facility, London, UK
Patricia B. Miralhes	Clinical Microbiology and Virology Department, University College London Hospitals NHS Trust, UK
Neginsadat Mirtorabi	University of Oxford Medical School, Medical Sciences Division, University of Oxford, UK
Celia Mitton	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Fiona Moghaddas	Department of Clinical Immunology, North Bristol NHS Trust, Bristol, UK
Mariya Molai	Hull University Teaching Hospitals NHS Trust, Hull, UK

Ella Morey	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Róisín Morgans	Clinical BioManufacturing Facility, Jenner Institute, University of Oxford, UK
Susan J. Morris	Clinical BioManufacturing Facility, Jenner Institute, University of Oxford, UK
Sheila Morris	Clinical Infection Research Group, Regional Infectious Diseases Unit, NHS Lothian, UK
Helen C. Morris	NIHR Cambridge Clinical Research Facility, Cambridge, UK
Hazel Morrison	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Franca Morselli	NIHR BRC at Guy's and St Thomas' NHS Foundation Trust, UK
Gertraud Morshead	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Richard Morter	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Nathifa A. Moyo	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Mushiya Mpelembue	London Northwest University Healthcare, Northwick Park Hospital, London, UK
Ekta Mukhopadhyay	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Jilly Muller	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Alasdair P.S. Munro	NIHR Southampton Clinical Research Facility, Southampton, UK
Sarah Murphy	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Philomena Mweu	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Gurudutt Naik	Aneurin Bevan University Health Board, Newport, Wales, UK
Kush Naker	NIHR/Wellcome Trust Birmingham Clinical Research Facility, Birmingham, UK
Eleni Nastouli	Clinical Microbiology and Virology Department, University College London Hospitals NHS Trust, UK
Cecilia Njenga	NIHR Imperial Clinical Research Facility, London, UK
Andrés Noé	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Fay L Nugent	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Katie O'Brien	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Daniel O'Connor	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Blanché Oguti	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Victoria Olchawski	Clinical BioManufacturing Facility, Jenner Institute, University of Oxford, UK
Neil J Oldfield	School of Life Sciences, University of Nottingham, Nottingham, UK

Catarina Oliveira	Clinical BioManufacturing Facility, Jenner Institute, University of Oxford, UK
Peter J. O'Reilly	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Piper Osborne	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
David R. J. Owen	NIHR Imperial Clinical Research Facility, London, UK
Daniel R. Owens	NIHR Southampton Clinical Research Facility, Southampton, UK
Nelly Owino	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Mihaela Pacurar	NIHR Southampton Clinical Research Facility, Southampton, UK
Susan Palmer	Aneurin Bevan University Health Board, Newport, Wales, UK
Helena M. R. T. Parracho	Clinical BioManufacturing Facility, Jenner Institute, University of Oxford, UK
Dipak Patel	Clinical Research and Innovation Office, Sheffield Teaching Hospitals NHS Foundation Trust, UK
Maia Patrick-Smith	University of Oxford Medical School, Medical Sciences Division, University of Oxford, UK
Ruth O. Payne	Department of Infection and Tropical Medicine, Sheffield Teaching Hospitals NHS Foundation Trust and the Department of Infection, Immunity and Cardiovascular Disease, University of Sheffield, UK
Elizabeth J. Penn	National Infection Service, Public Health England, UK
Anna Pennington	Aneurin Bevan University Health Board, Newport, Wales, UK
Marco Polo Peralta Alvarez	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
James Perring	University of Oxford Medical School, Medical Sciences Division, University of Oxford, UK
Angelina Peterson	Department of Clinical Sciences, Liverpool School of Tropical Medicine, UK
Jennifer Phillips	University Hospitals Bristol & Weston NHS Foundation Trust
Lorinda Pickup	NIHR Cambridge Clinical Research Facility, Cambridge, UK
Jo Piper	NIHR Cambridge Clinical Research Facility, Cambridge, UK
Dimitra Pipini	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Mary Plank	AstraZeneca BioPharmaceuticals PLC
Sinead Plant	Clinical Infection Research Group, NHS Lothian, Edinburgh, UK
Jennifer Pooley	North Bristol NHS Trust, Bristol, UK
Ian Poulton	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Claire Powers	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
David A. Price	Department of Infection and Tropical Medicine, Newcastle upon Tyne Hospitals NHS Foundation Trust, UK
Vivien Price	NIHR/Wellcome Trust Birmingham Clinical Research Facility, Birmingham, UK
Pamela C. Proud	National Infection Service, Public Health England, UK

Samuel Provstgaard-Morys	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
David Pulido	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Sheena Quaid	London Northwest University Healthcare, Northwick Park Hospital, London, UK
Kajal Radia	University of Oxford Medical School, Medical Sciences Division, University of Oxford, UK
Durga Rajapaksa	National Infection Service, Public Health England, UK
Thurkka Rajeswaran	NIHR BRC at Guy's and St Thomas' NHS Foundation Trust, UK
Alberto San Francisco Ramos	Vaccine Institute, Institute of Infection & Immunity, St. Georges, University of London and St Georges University Hospitals NHS Trust, London, UK
Fernando Ramos Lopez	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Tommy Rampling	Clinical Microbiology and Virology Department, University College London Hospitals NHS Trust, UK
Isobel Ramsay	Infectious Diseases Department, Cambridge and Peterborough NHS Foundation Trust, UK
Jade Rand	NIHR Southampton Clinical Research Facility, Southampton, UK
Helen Ratcliffe	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Pooja Ravji	Infectious Diseases Department, Cambridge and Peterborough NHS Foundation Trust, UK
Thomas Rawlinson	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
David Rea	Clinical Research Network, West of England, UK
Ashwin Reddy	Pulmonary Vascular Diseases Unit, Cambridge and Peterborough NHS Foundation Trust, UK
Mila Resuello-Dauti	NIHR UCLH Clinical Research Facility, London, UK
Emilia Reyes Pabon	Clinical BioManufacturing Facility, Jenner Institute, University of Oxford, UK
Sarah Rhead	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Tawassal Riaz	Infection Sciences, North Bristol NHS Trust, Bristol, UK
Carla M. Ribiero	NIHR Cambridge Clinical Research Facility, Cambridge, UK
Marivic Ricamara	NIHR UCLH Clinical Research Facility, London, UK
Alex Richter	NIHR/Wellcome Trust Birmingham Clinical Research Facility & Institute of Immunology and Immunotherapy, University of Birmingham, UK
Neil D. Ritchie	Department of Infectious Diseases, Queen Elizabeth University Hospital, Glasgow, UK
Adam J. Ritchie	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Alexander J. Robbins	NIHR Imperial Clinical Research Facility, London, UK

Hannah Roberts	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Ryan E Robinson	Department of Clinical Sciences, Liverpool School of Tropical Medicine and Liverpool University Hospitals NHS Foundation Trust, UK
Sophie Roche	University of Oxford Medical School, Medical Sciences Division, University of Oxford, UK
Christine S. Rollier	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Louisa Rose	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Amy L. Ross Russell	NIHR Southampton Clinical Research Facility, Southampton, UK
Simon Royal	School of Medicine, Division of Primary Care, University of Nottingham, Nottingham, UK
Indra Rudiansyah	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Kim Ryalls	Pharmacy Department, Sheffield Teaching Hospitals NHS Foundation Trust
Charlotte E. Sabine	NIHR/Wellcome Trust Birmingham Clinical Research Facility, Birmingham, UK
Stephen Saich	NIHR Southampton Clinical Research Facility, Southampton, UK
Jessica C Sale	NIHR/Wellcome Trust Birmingham Clinical Research Facility, Birmingham, UK
Ahmed M. Salman	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Stephannie Salvador	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Amada Sanchez-Gonzalez	Department of Infection and Tropical Medicine, Newcastle upon Tyne Hospitals NHS Foundation Trust, UK
Helen Sanders	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Katherine Sanders	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Iman Satti	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Jack E. Saunders	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Caroline Saunders	NIHR Cambridge Clinical Research Facility, Cambridge, UK
Ina Schim van der Loeff	Department of Infection and Tropical Medicine, Newcastle upon Tyne Hospitals NHS Foundation Trust and Translational and Clinical Research Institute, Immunity and Inflammation Theme, Newcastle University
Ella Schofield	University of Oxford Medical School, Medical Sciences Division, University of Oxford, UK
Gavin R. Screaton	Medical Sciences, University of Oxford, UK
Samiullah Seddiqi	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK

Rameswara R. Segireddy	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Sonia Serrano	NIHR BRC at Guy's and St Thomas' NHS Foundation Trust, UK
Sifut Sethi	Hull University Teaching Hospitals NHS Trust, Hull, UK
Farah Shahi	Hull University Teaching Hospitals NHS Trust, Hull, UK
Imam Shaik	National Infection Service, Public Health England, UK
Hannah R. Sharpe	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Katherine Sharrocks	Department of Medicine, University of Cambridge, UK
Robert Shaw	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Emma Sheehan	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Amy Shepherd	Clinical Infection Research Group, Regional Infectious Diseases Unit, NHS Lothian, UK
Farah Shiham	Department of Clinical Sciences, Liverpool School of Tropical Medicine and Liverpool University Hospitals NHS Foundation Trust, UK
Sarah E. Silk	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Laura Silva-Reyes	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Nisha Singh	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Jaisi Sinha	Public Health Wales NHS Trust, Cardiff, UK
Holly E. Smith	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
David J Smith	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Catherine C Smith	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Carla Solórzano	Department of Clinical Sciences, Liverpool School of Tropical Medicine, UK
Kim Sorley	NIHR Imperial Clinical Research Facility, London, UK
Luciana Sowole	NIHR BRC at Guy's and St Thomas' NHS Foundation Trust, UK
Alexandra J Spencer	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Lisa Stockdale	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Lisa V. Stockwell	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Arabella S. V. Stuart	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Ann Sturdy	London Northwest University Healthcare, Northwick Park Hospital, London, UK
Joe Suich	Hull University Teaching Hospitals NHS Trust, Hull, UK
Natalina Sutton	Vaccine Institute, Institute of Infection & Immunity, St. Georges, University of London and St Georges University Hospitals NHS Trust, London, UK

Anna Szigeti	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Abdessamad Tahiri-Alaoui	Clinical BioManufacturing Facility, Jenner Institute, University of Oxford, UK
Farah Tahmasebi	Manchester Lighthouse Laboratory, Manchester, UK
Rachel Tanner	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Alexander W Tarr	School of Life Sciences, University of Nottingham, Nottingham, UK & NIHR Nottingham Biomedical Research Centre, Nottingham University Hospitals NHS Trust, Nottingham, UK
Richard Tarrant	Clinical BioManufacturing Facility, Jenner Institute, University of Oxford, UK
Natalie Tate	Department of Clinical Sciences, Liverpool School of Tropical Medicine, UK
Keja Taylor	Clinical BioManufacturing Facility, Jenner Institute, University of Oxford, UK
Iona Jennifer Taylor	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Justin Taylor	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Rebecca te Water Naude	University of Oxford Medical School, Medical Sciences Division, University of Oxford, UK
Kate Templeton	Clinical Infection Research Group, NHS Lothian, Edinburgh, UK
Yrene Themistocleous	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Merin Thomas	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Kelly M Thomas	National Infection Service, Public Health England, UK
Tonia M Thomas	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Julia Thompson	AstraZeneca BioPharmaceuticals PLC
Amber J Thompson	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Patrick J Tighe	School of Life Sciences, University of Nottingham, Nottingham, UK
Gerlynn Ferreras Tiongson	London Northwest University Healthcare, Northwick Park Hospital, London, UK
Adriana Tomic	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
James Towner	University of Oxford Medical School, Medical Sciences Division, University of Oxford, UK
Nguyen Tran	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Julia A. Tree	National Infection Service, Public Health England, UK
Gerardo Trillana	NIHR BRC at Guy's and St Thomas' NHS Foundation Trust, UK
Charlotte Trinham	NIHR/Wellcome Trust Birmingham Clinical Research Facility, Birmingham, UK
Rose Trivett	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK

Adam Truby	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Aadil El-Turabi	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Richard Turner	AstraZeneca BioPharmaceuticals PLC
Cheryl Turner	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Nicola Turner	NIHR UCLH Clinical Research Facility, London, UK
Bhavya Tyagi	London Northwest University Healthcare, Northwick Park Hospital, London, UK
Marta Ulaszewska	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Benjamin R. Underwood	Windsor Research Unit, Cambridge and Peterborough NHS Foundation Trust, UK
Maithili Varadarajan	Hull University Teaching Hospitals NHS Trust, Hull, UK
Marije K Verheul	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Iason Vichos	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Laura Walker	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Matthew E Wand	National Infection Service, Public Health England, UK
Sarah C. Warren	NIHR Southampton Clinical Research Facility, Southampton, UK
Marion E. E. Watson	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Ekaterina Watson	London Northwest University Healthcare, Northwick Park Hospital, London, UK
Stewart Webb	Department of Infectious Diseases, Queen Elizabeth University Hospital, Glasgow, UK
Andrea Webster	Research Directorate, Newcastle upon Tyne Hospitals NHS Foundation Trust, UK
Rowena Weighell	NIHR Cambridge Clinical Research Facility, Cambridge, UK
Jeanette H. Wells	Aneurin Bevan University Health Board, Newport, Wales, UK
Beth White	Department of Infectious Diseases, Queen Elizabeth University Hospital, Glasgow, UK
Rachel White	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Caroline White	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Paul Williams	Clinical BioManufacturing Facility, Jenner Institute, University of Oxford, UK
Rachel L Williams	Research & Innovation, North Bristol NHS Trust, Bristol, UK
Rebecca L. Winslow	NIHR/Wellcome Trust Birmingham Clinical Research Facility, Birmingham, UK
Danielle Woods	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Andrew T. Worth	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK

Daniel Wright	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Marzena Wroblewska	Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK
Xin Li Yao	Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK
Yee Ting Nicole Yim	NIHR UCLH Clinical Research Facility, London, UK
Dalila Zizi	Clinical BioManufacturing Facility, Jenner Institute, University of Oxford, UK

Acknowledgements COVID-19 vaccine trials

Advent, South Africa
Michael Breese
BioIndustry Association
Annette England
Ian McCubbin
Cell & Gene Therapy Catapult
Stephen Ward
CobraBio
Clinical Trials Research Governance Office, University of Oxford
Ronja Bahadori
Elaine Chick
Heather House
Claire Riddle
Data and Safety Monitoring Board (DSMB)
George Bouliotis
Steve Black
Elizabeth Bukusi
Cornelia Dekker
Robert Heyderman
Gregory Hussey
Paul Kaye
Bernhards Ogutu
Walter Orenstein
Sonia Ramos
Manish Sadarangani
Deloitte UK
Alex Hope
Department of Health and Social Care, UK Government
Amina Elmi
Harry Mayhew
Martin Shanahan
Department of Paediatrics, University of Oxford
Joanna Bagniewska
Elizabeth Derow
Georg A. Holländer
Samantha Vanderslott
Endpoint Evaluation Committee
Jeremy Carr
Stephen Chambers
Kim Davis
Simon Drysdale
Malick Gibani
Elizabeth Hammershaimb
Michael Harrington

Celina Jin
Seilesh Kadambari
Rama Kandasamy
Toby Maher
Jamilah Meghji
Claire Munro[1]
David Pace
Rekha R. Rapaka
Robindra Basu Roy
Daniel Silman
Gemma Sinclair
Jing Wang
Halix
The Cambridge NIHR CRF COVID Vaccine Group
The GSTT NIHR CRF COVID Vaccine Group
The Imperial CRF COVID Vaccine Group
Jenner Institute, University of Oxford
Iona Tarbet
Nuffield Department of Medicine, University of Oxford
Richard Cornall
Richard Liwicki
Denis Murphy
Elizabeth Salter
Katherine Skinner
Philip Taylor
Oto Velicka
Oxford Biomedica
Oxford Research Services (Contracts)
Carly Banner
Sally Pelling-Deeves
Gary Priest
Oxford University Hospitals NHS Trust
Monique Andersson
Laura Dunn
Bruno Holthof
Pall Europe
Public Affairs Directorate and Divisional Communication Team
Alison Brindle
Alexander Buxton
James Colman
Chris McIntyre
Steve Pritchard
Sartorius
Zander Hack
VMIC

Appendix 2 COG-UK consortium acknowledgements

The COVID-19 Genomics UK (COG-UK) Consortium

Funding acquisition, Leadership and supervision, Metadata curation, Project administration, Samples and logistics, Sequencing and analysis, Software and analysis tools, and Visualisation:
Dr Samuel C Robson PhD 13

Funding acquisition, Leadership and supervision, Metadata curation, Project administration, Samples and logistics, Sequencing and analysis, and Software and analysis tools:
Prof Nicholas J Loman PhD 41 and Dr Thomas R Connor PhD 10, 69

Leadership and supervision, Metadata curation, Project administration, Samples and logistics, Sequencing and analysis, Software and analysis tools, and Visualisation:
Dr Tanya Golubchik PhD 5

Funding acquisition, Metadata curation, Samples and logistics, Sequencing and analysis, Software and analysis tools, and Visualisation:
Dr Rocio T Martinez Nunez PhD 42

Funding acquisition, Leadership and supervision, Metadata curation, Project administration, and Samples and logistics:
Dr Catherine Ludden PhD 88

Funding acquisition, Leadership and supervision, Metadata curation, Samples and logistics, and Sequencing and analysis:
Dr Sally Corden PhD 69

Funding acquisition, Leadership and supervision, Project administration, Samples and logistics, and Sequencing and analysis:
Ian Johnston 99 and Dr David Bonsall PhD 5

Funding acquisition, Leadership and supervision, Sequencing and analysis, Software and analysis tools, and Visualisation:
Prof Colin P Smith PhD 87 and Dr Ali R Awan PhD 28

Funding acquisition, Samples and logistics, Sequencing and analysis, Software and analysis tools, and Visualisation:
Dr Giselda Bucca PhD 87

Leadership and supervision, Metadata curation, Project administration, Samples and logistics, and Sequencing and analysis:
Dr M. Estee Torok FRCP 22, 101

Leadership and supervision, Metadata curation, Project administration, Samples and logistics, and Visualisation:
Dr Kordo Saeed MD/ FRCPath 81, 110 and Dr Jacqui A Prieto PhD 83, 109

Leadership and supervision, Metadata curation, Project administration, Sequencing and analysis, and Software and analysis tools:
Dr David K Jackson PhD 99

Metadata curation, Project administration, Samples and logistics, Sequencing and analysis, and Software and analysis tools:
Dr William L Hamilton PhD 22

Metadata curation, Project administration, Samples and logistics, Sequencing and analysis, and Visualisation:
Dr Luke B Snell MSc/ MBBS 11

Funding acquisition, Leadership and supervision, Metadata curation, and Samples and logistics:
Dr Catherine Moore 69

Funding acquisition, Leadership and supervision, Project administration, and Samples and logistics:
Dr Ewan M Harrison PhD 99, 88

Leadership and supervision, Metadata curation, Project administration, and Samples and logistics:
Dr Sonia Goncalves PhD 99

Leadership and supervision, Metadata curation, Samples and logistics, and Sequencing and analysis:
Prof Ian G Goodfellow PhD 24, Dr Derek J Fairley PhD 3, 72, Prof Matthew W Loose PhD 18 and Joanne Watkins MSc 69

Leadership and supervision, Metadata curation, Samples and logistics, and Software and analysis tools:
Rich Livett MSc 99

Leadership and supervision, Metadata curation, Samples and logistics, and Visualisation:
Dr Samuel Moses MD 25, 106

Leadership and supervision, Metadata curation, Sequencing and analysis, and Software and analysis tools:
Dr Roberto Amato PhD 99, Dr Sam Nicholls PhD 41 and Dr Matthew Bull PhD 69

Leadership and supervision, Project administration, Samples and logistics, and Sequencing and analysis:
Prof Darren L Smith PhD 37, 58, 105

Leadership and supervision, Sequencing and analysis, Software and analysis tools, and Visualisation:
Dr Jeff Barrett PhD 99 and Prof David M Aanensen PhD 14, 114

Metadata curation, Project administration, Samples and logistics, and Sequencing and analysis:
Dr Martin D Curran PhD 65, Dr Surendra Parmar PhD 65, Dr Dinesh Aggarwal MRCP 95, 99, 64 and Dr James G Shepherd MBChB/MRCP 48

Metadata curation, Project administration, Sequencing and analysis, and Software and analysis tools:
Dr Matthew D Parker PhD 93

Metadata curation, Samples and logistics, Sequencing and analysis, and Visualisation:
Dr Sharon Glaysher PhD 61

Metadata curation, Sequencing and analysis, Software and analysis tools, and Visualisation:
Dr Matthew Bashton PhD 37, 58, Dr Anthony P Underwood PhD 14, 114, Dr Nicole Pacchiarini PhD 69 and Dr Katie F Loveson PhD 77

Project administration, Sequencing and analysis, Software and analysis tools, and Visualisation:
Dr Alessandro M Carabelli PhD 88

Funding acquisition, Leadership and supervision, and Metadata curation:
Dr Kate E Templeton PhD 53, 90

Funding acquisition, Leadership and supervision, and Project administration:
Dr Cordelia F Langford PhD 99, John Sillitoe BEng 99, Dr Thushan I de Silva PhD 93 and Dr Dennis Wang PhD 93

Funding acquisition, Leadership and supervision, and Sequencing and analysis:
Prof Dominic Kwiatkowski 99, 107, Prof Andrew Rambaut DPhil 90, Dr Justin O'Grady PhD 70, 89 and Dr Simon Cottrell PhD 69

Leadership and supervision, Metadata curation, and Sequencing and analysis:
Prof Matthew T.G. Holden PhD 68 and Prof Emma C Thomson PhD/FRCP 48

Leadership and supervision, Project administration, and Samples and logistics:
Dr Husam Osman PhD 64, 36, Dr Monique Andersson PhD 59, Prof Anoop J Chauhan 61 and Dr Mohammed O Hassan-Ibrahim PhD/FRCPath 6

Leadership and supervision, Project administration, and Sequencing and analysis:
Dr Mara Lawniczak 99

Leadership and supervision, Samples and logistics, and Sequencing and analysis:
Prof Ravi Kumar Gupta PhD 88, 113, Dr Alex Alderton PhD 99, Dr Meera Chand 66, Dr Chrystala Constantinidou PhD 94, Dr Meera Unnikrishnan PhD 94, Prof Alistair C Darby PhD 92, Prof Julian A Hiscox PhD 92 and Prof Steve Paterson PhD 92

Leadership and supervision, Sequencing and analysis, and Software and analysis tools:
Dr Inigo Martincorena 99, Prof David L Robertson PhD 48, Dr Erik M Volz PhD 39, Dr Andrew J Page PhD 7 and Prof Oliver G Pybus DPhil 23

Leadership and supervision, Sequencing and analysis, and Visualisation:
Dr Andrew R Bassett PhD 99

Metadata curation, Project administration, and Samples and logistics:
Dr Cristina V Ariani PhD 99, Dr Michael H Spencer Chapman MBBS 99, 88, Dr Kathy K Li MBBCh/FRCPath 48, Dr Rajiv N Shah BMBS/MRCP/MSc 48, Dr Natasha G Jesudason MBChB MRCP FRCPath 48 and Dr Yusri Taha MD/PhD 50

Metadata curation, Project administration, and Sequencing and analysis:
Martin P McHugh MSc 53, Dr Rebecca Dewar PhD 53

Metadata curation, Samples and logistics, and Sequencing and analysis:
Dr Aminu S Jahun PhD 24, Dr Claire McMurray PhD 41, Ms Sarojini Pandey MSc 84, Dr James P McKenna PhD 3, Dr Andrew Nelson PhD 58, 105, Dr Gregory R Young PhD 37, 58, Dr Clare M McCannPhD 58, 105 and Mr Scott Elliott 61

Metadata curation, Samples and logistics, and Visualisation:
Ms Hannah Lowe MSc 25

Metadata curation, Sequencing and analysis, and Software and analysis tools:
Dr Ben Temperton Ph.D. 91, Dr Sunando Roy PhD 82, Dr Anna Price PhD 10, Dr Sara Rey PhD 69 and Mr Matthew Wyles 93

Metadata curation, Sequencing and analysis, and Visualisation:
Stefan Rooke MSc 90 and Dr Sharif Shaaban PhD 68

Project administration, Samples and logistics, Sequencing and analysis:
Dr Mariateresa de Cesare PhD 98

Project administration, Samples and logistics, and Software and analysis tools:
Laura Letchford BSc 99

Project administration, Samples and logistics, and Visualisation:
Miss Siona Silveira MSc 81, Dr Emanuela Pelosi FRCPath 81 and Dr Eleri Wilson-Davies MD/FRCPath 81

Samples and logistics, Sequencing and analysis, and Software and analysis tools:
Dr Myra Hosmillo PhD 24

Sequencing and analysis, Software and analysis tools, and Visualisation:
Áine O'Toole MSc 90, Dr Andrew R Hesketh PhD 87, Mr Richard Stark MSc 94, Dr Louis du Plessis PhD 23, Dr Chris Ruis PhD 88, Dr Helen Adams PhD 4 and Dr Yann Bourgeois PhD 76

Funding acquisition, and Leadership and supervision:
Dr Stephen L Michell PhD 91, Prof Dimitris Grammatopoulos PhD/FRCPath 84, 112, Dr Jonathan Edgeworth PhD/FRCPath 12, Prof Judith Breuer MD 30, 82, Prof John A Todd PhD 98 and Dr Christophe Fraser PhD 5

Funding acquisition, and Project administration:
Dr David Buck PhD 98 and Michaela John BSc 9

Leadership and supervision, and Metadata curation:
Dr Gemma L Kay PhD 70

Leadership and supervision, and Project administration:
Steve Palmer 99, Prof Sharon J Peacock 88, 64 and David Heyburn 69

Leadership and supervision, and Samples and logistics:
Danni Weldon BSc 99, Dr Esther Robinson PhD 64, 36, Prof Alan McNally PhD 41, 86, Dr Peter Muir PhD 64, Dr Ian B Vipond PhD 64, Dr John BoYes MBChB 29, Dr Venkat

Sivaprakasam PhD 46, Dr Tranprit Saluja FRCPath/MD 75, Dr Samir Dervisevic FRCPath 54 and Dr Emma J Meader FRCPath 54

Leadership and supervision, and Sequencing and analysis:
Dr Naomi R Park PhD 99, Karen Oliver BSc 99, Dr Aaron R Jeffries Ph.D. 91, Dr Sascha Ott PhD 94, Dr Ana da Silva Filipe PhD 48, Dr David A Simpson PhD 72 and Dr Chris Williams MB BS 69

Leadership and supervision, and Visualisation:
Dr Jane AH Masoli MBChB 73, 91

Metadata curation, and Samples and logistics:
Dr Bridget A Knight PhD. 73, 91, Dr Christopher R Jones Ph.D. 73, 91, Mr Cherian Koshy MSc CSci FIBMS 1, Miss Amy Ash BSc 1, Dr Anna Casey PhD 71, Dr Andrew Bosworth PhD 64, 36, Dr Liz Ratcliffe PhD 71, Dr Li Xu-McCrae PhD 36, Miss Hannah M Pymont MSc 64, Ms Stephanie Hutchings 64, Dr Lisa Berry PhD 84, Ms Katie Jones MSc 84, Dr Fenella Halstead PhD 46, Mr Thomas Davis MSc 21, Dr Christopher Holmes PhD 16, Prof Miren Iturriza-Gomara PhD 92, Dr Anita O Lucaci PhD 92, Dr Paul Anthony Randell MBBCh 38, 104, Dr Alison Cox PhD 38, 104, Pinglawathee Madona 38, 104, Dr Kathryn Ann Harris PhD 30, Dr Julianne Rose Brown PhD 30, Dr Tabitha W Mahungu FRCPath 74, Dr Dianne Irish-Tavares FRCPath 74, Dr Tanzina Haque FRCPath PhD 74, Dr Jennifer Hart MRCP 74, Mr Eric Witele MSc 74, Mrs Melisa Louise Fenton DipHE 75, Mr Steven Liggett 79, Dr Clive Graham MD 56, Ms Emma Swindells Bsc 57, Ms Jennifer Collins BSc 50, Mr Gary Eltringham BSc 50, Ms Sharon Campbell MSc 17, Dr Patrick C McClure PhD 97, Dr Gemma Clark PhD 15, Dr Tim J Sloan PhD 60, Mr Carl Jones 15 and Dr Jessica Lynch PhD MBChB 2, 111

Metadata curation, and Sequencing and analysis:
Dr Ben Warne MRCP 8, Steven Leonard PhD 99, Jillian Durham BSc 99, Dr Thomas Williams MD 90, Dr Sam T Haldenby PhD 92, Dr Nathaniel Storey PhD 30, Dr Nabil-Fareed Alikhan PhD 70, Dr Nadine Holmes PhD 18, Dr Christopher Moore PhD 18, Mr Matthew Carlile BSc 18, Malorie Perry MSc 69, Dr Noel Craine DPhil 69, Prof Ronan A Lyons MD 80, Miss Angela H Beckett MSc 13, Salman Goudarzi PhD 77, Christopher Fearn MRes 77, Kate Cook 77, Hannah Dent BSc 77 and Hannah Paul MRes 77

Metadata curation, and Software and analysis tools:
Robert Davies 99

Project administration, and Samples and logistics:
Beth Blane BSc 88, Sophia T Girgis MSc 88, Dr Mathew A Beale PhD 99, Katherine L Bellis 99, 88, Matthew J Dorman 99, Eleanor Drury 99, Leanne Kane 99, Sally Kay 99, Dr Samantha McGuigan 99, Dr Rachel Nelson PhD 99, Liam Prestwood 99, Dr Shavanthi Rajatileka PhD 99, Dr Rahul Batra MD 12, Dr Rachel J Williams PhD 82, Dr Mark Kristiansen PhD 82, Dr Angie Green PhD 98, Miss Anita Justice MSc 59, Dr Adhyana I.K Mahanama MD 81, 102 and Dr Buddhini Samaraweera MD 81, 102

Project administration, and Sequencing and analysis:
Dr Nazreen F Hadjirin PhD 88 and Dr Joshua Quick PhD 41

Project administration, and Software and analysis tools:
Mr Radoslaw Poplawski BSc 41

Samples and logistics, and Sequencing and analysis:

Leanne M Kermack MSc 88, Nicola Reynolds PhD 7, Grant Hall BS 24, Yasmin Chaudhry BSc 24, Malte L Pinckert MPhil 24, Dr Iliana Georgana PhD 24, Dr Robin J Moll PhD 99, Dr Alicia Thornton 66, Dr Richard Myers 66, Dr Joanne Stockton PhD 41, Miss Charlotte A Williams BSc 82, Dr Wen C Yew PhD 58, Alexander J Trotter MRes 70, Miss Amy Trebes MSc 98, Mr George MacIntyre-Cockett BSc 98, Alec Birchley MSc 69, Alexander Adams BSc 69, Amy Plimmer 69, Bree Gatica-Wilcox MPhil 69, Dr Caoimhe McKerr PhD 69, Ember Hilvers MA 69, Hannah Jones 69, Dr Hibo Asad PhD 69, Jason Coombes BSc 69, Johnathan M Evans MSc 69, Laia Fina 69, Lauren Gilbert A-Levels 69, Lee Graham BSc 69, Michelle Cronin 69, Sara Kumziene-SummerhaYes MSc 69, Sarah Taylor 69, Sophie Jones MSc 69, Miss Danielle C Groves BA 93, Mrs Peijun Zhang MSc 93, Miss Marta Gallis MSc 93 and Miss Stavroula F Louka MSc 93

Samples and logistics, and Software and analysis tools:
Dr Igor Starinskij Msc MRCP 48

Sequencing and analysis, and Software and analysis tools:
Dr Chris J Illingworth PhD 47, Dr Chris Jackson PhD 47, Ms Marina Gourtovaia MSc 99, Gerry Tonkin-Hill 99, Kevin Lewis 99, Dr Jaime M Tovar-Corona PhD 99, Dr Keith James PhD 99, Dr Laura Baxter PhD 94, Dr Mohammad T. Alam PhD 94, Dr Richard J Orton PhD 48, Dr Joseph Hughes PhD 48, Dr Sreenu Vattipally PhD 48, Dr Manon Ragonnet-Cronin PhD 39, Dr Fabricia F. Nascimento PhD 39, Mr David Jorgensen MSc 39, Ms Olivia Boyd MSc 39, Ms Lily Geidelberg MSc 39, Dr Alex E Zarebski PhD 23, Dr Jayna Raghwani PhD 23, Dr Moritz UG Kraemer DPhil 23, Joel Southgate MSc 10, 69, Dr Benjamin B Lindsey MRCP 93 and Mr Timothy M Freeman MPhil 93

Software and analysis tools, and Visualisation:
Jon-Paul Keatley 99, Dr Joshua B Singer PhD 48, Leonardo de Oliveira Martins PhD 70, Dr Corin A Yeats PhD 14, Dr Khalil Abudahab PhD 14, 114, Mr Ben EW Taylor MEng 14, 114 and Mirko Menegazzo 14

Leadership and supervision:
Prof John Danesh 99, Wendy Hogsden MSc 46, Dr Sahar Eldirdiri MBBS MSc FRCPath 21, Mrs Anita Kenyon MSc 21, Dr Jenifer Mason MBBS 43, Mr Trevor I Robinson MSc 43, Prof Alison Holmes MD 38, 103, Dr James Price PhD 38, 103, Prof John A Hartley PhD 82, Dr Tanya Curran PhD 3 , Dr Alison E Mather PhD 70, Dr Giri Shankar 69, Dr Rachel Jones 69, Dr Robin Howe 69 and Dr Sian Morgan FRCPath 9

Metadata curation:
Dr Elizabeth Wastenge MD 53, Dr Michael R Chapman PhD 34, 88, 99, Mr Siddharth Mookerjee MPH 38,103, Dr Rachael Stanley PhD 54, Mrs Wendy Smith 15, Prof Timothy Peto PhD 59, Dr David Eyre PhD 59, Dr Derrick Crook 59, Dr Gabrielle Vernet MBBS 33, Dr Christine Kitchen PhD 10, Huw Gulliver 10, Dr Ian Merrick PhD 10, Prof Martyn Guest PhD 10, Robert Munn BSc 10, Dr Declan T Bradley 63, 72 and Dr Tim Wyatt 63

Project administration:
Dr Charlotte Beaver 99, Luke Foulser 99, Sophie Palmer 88, Carol M Churcher 88, Ellena Brooks MA 88, Kim S Smith 88, Dr Katerina Galai PhD 88, Georgina M McManus BSc 88, Dr Frances Bolt PhD 38, 103, Dr Francesc Coll PhD 19, Lizzie Meadows MA 70, Dr Stephen W Attwood PhD 23, Dr Alisha Davies 69, Elen De Lacy MSc 69, Fatima Downing 69, Sue Edwards 69, Dr Garry P Scarlett PhD 76, Mrs Sarah Jeremiah MSc 83 and Dr Nikki Smith PhD 93

Samples and logistics:

Danielle Leek Bsc 88, Sushmita Sridhar BS 88, 99, Sally Forrest BSc 88, Claire Cormie 88, Harmeet K Gill PhD 88, Joana Dias MSc 88, Ellen E Higginson PhD 88, Mailis Maes MPhil 88, Jamie Young BSc 88, Michelle Wantoch PhD 7, Sanger Covid Team (www.sanger.ac.uk/covid-team) 99, Dorota Jamrozy 99, Stephanie Lo 99, Dr Minal Patel PhD 99, Verity Hill 90, Ms Claire M Bewshea MSc 91, Prof Sian Ellard FRCPath 73, 91, Dr Cressida Auckland FRCPath 73, Dr Ian Harrison 66, Dr Chloe Bishop 66, Dr Vicki Chalker 66, Dr Alex Richter PhD 85, Dr Andrew Beggs PhD 85, Dr Angus Best PhD 86, Dr Benita Percival PhD 86, Dr Jeremy Mirza PhD 86, Dr Oliver Megram PhD 86, Dr Megan Mayhew PhD 86, Dr Liam Crawford PhD 86, Dr Fiona Ashcroft PhD 86, Dr Emma Moles-Garcia PhD 86, Dr Nicola Cumley PhD 86, Mr Richard Hopes 64, Dr Patawee Asamaphan PhD 48, Mr Marc O Niebel MSc 48, Prof Rory N Gunson PhD FRCPath 100, Dr Amanda Bradley PhD 52, Dr Alasdair Maclean PhD 52, Dr Guy Mollett MBChB 52, Dr Rachel Blacow MBChB 52, Mr Paul Bird MSc 16, Mr Thomas Helmer 16, Miss Karlie Fallon 16, Dr Julian Tang 16, Dr Antony D Hale MBBS 49, Dr Louissa R Macfarlane-Smith PhD 49, Katherine L Harper MBiol 49, Miss Holli Carden MSc 49, Dr Nicholas W Machin MSc 45, 64, Ms Kathryn A Jackson MSc 92, Dr Shazaad S Y Ahmad MSc 45, 64, Dr Ryan P George PhD 45, Dr Lance Turtle PhD MRCP 92, Mrs Elaine O'Toole BSc 43, Mrs Joanne Watts BSc 43, Mrs Cassie Breen BSc 43, Mrs Angela Cowell MSc 43, Ms Adela Alcolea-Medina 32, 96, Ms Themoula Charalampous MSc 12, 42, Amita Patel 11, Dr Lisa J Levett PhD 35, Dr Judith Heaney PhD 35, Dr Aileen Rowan PhD 39, Prof Graham P Taylor DSc 39, Dr Divya Shah PhD 30, Miss Laura Atkinson MSc 30, Mr Jack CD Lee MSc 30, Mr Adam P Westhorpe BSc 82, Dr Riaz Jannoo PhD 82, Dr Helen L Lowe PhD 82, Miss Angeliki Karamani MSc 82, Miss Leah Ensell BSc 82, Mrs Wendy Chatterton MSc 35, Miss Monika Pusok MSc 35, Mrs Ashok Dadrah MSc 75, Miss Amanda Symmonds MSc 75, Dr Graciela Sluga MD/MSC 44, Dr Zoltan Molnar PhD 72, Mr Paul Baker MD 79, Prof Stephen Bonner 79, Ms Sarah Essex 79, Dr Edward Barton MD 56, Ms Debra Padgett BSc 56, Ms Garren Scott BSc 56, Ms Jane Greenaway MSc 57, Dr Brendan AI Payne MD 50, Dr Shirelle Burton-Fanning MD 50, Dr Sheila Waugh MD 50, Dr Veena Raviprakash MD 17, Ms Nicola Sheriff BSc 17, Ms Victoria Blakey BSc 17, ms Lesley-Anne Williams BSc 17, Dr Jonathan Moore MD 27, Ms Susanne Stonehouse BSc 27, Dr Louise Smith 55, Dr Rose K Davidson PhD 89, Dr Luke Bedford 26, Dr Lindsay Coupland PhD 54, Ms Victoria Wright BSc 18, Dr Joseph G Chappell PhD 97, Dr Theocharis Tsoleridis PhD 97, Prof Jonathan Ball PhD 97, Mrs Manjinder Khakh 15, Dr Vicki M Fleming PhD 15, Dr Michelle M Lister PhD 15, Dr Hannah C Howson-Wells PhD 15, Dr Louise Berry 15, Dr Tim Boswell 15, Dr Amelia Joseph 15, Dr Iona Willingham 15, Dr Nichola Duckworth 60, Dr Sarah Walsh 60, Dr Emma Wise PhD 2, 111, Dr Nathan Moore PhD 2, 111, Miss Matilde Mori BSc 2, 108, 111, Dr Nick Cortes MRCP FRCPath 2, 111, Dr Stephen Kidd PhD 2, 111, Dr Rebecca Williams BMBS 33, Laura Gifford MSc 69, Miss Kelly Bicknell 61, Dr Sarah Wyllie 61, Miss Allyson Lloyd 61, Mr Robert Impey MSc 61, Ms Cassandra S Malone MSc 6, Mr Benjamin J Cogger BSc 6, Nick Levene MSc 62, Lynn Monaghan 62, Dr Alexander J Keeley MRCP 93, Dr David G Partridge FRCP FRCPath 78, 93, Dr Mohammad Raza 78, 93, Dr Cariad Evans 78, 93 and Dr Kate Johnson 78, 93

Sequencing and analysis:

Emma Betteridge BSc 99, Ben W Farr BSc 99, Scott Goodwin MSc 99, Dr Michael A Quail PhD 99, Carol Scott 99, Lesley Shirley MSc 99, Scott AJ Thurston BSc 99, Diana Rajan MSc 99, Dr Iraad F Bronner PhD 99, Louise Aigrain PhD 99, Dr Nicholas M Redshaw PhD 99, Dr Stefanie V Lensing PhD 99, Shane McCarthy 99, Alex Makunin 99, Dr Carlos E Balcazar PhD 90, Dr Michael D Gallagher PhD 90, Dr Kathleen A

Williamson PhD 90, Thomas D Stanton BSc 90, Ms Michelle L Michelsen BSc 91, Ms Joanna Warwick-Dugdale BSc 91, Dr Robin Manley Ph.D. 91, Ms Audrey Farbos MSc 91, Dr James W Harrison Ph.D. 91, Dr Christine M Sambles Ph.D. 91, Dr David J Studholme PhD. 91, Dr Angie Lackenby 66, Dr Tamyo Mbisa 66, Dr Steven Platt 66, Mr Shahjahan Miah 66, Dr David Bibby 66, Dr Carmen Manso 66, Dr Jonathan Hubb 66, Dr Gavin Dabrera 66, Dr Mary Ramsay 66, Dr Daniel Bradshaw 66, Dr Ulf Schaefer 66, Dr Natalie Groves 66, Dr Eileen Gallagher 66, Dr David Lee 66, Dr David Williams 66, Dr Nicholas Ellaby 66, Hassan Hartman 66, Nikos Manesis 66, Vineet Patel 66, Juan Ledesma 67, Ms Katherine A Twohig 67, Dr Elias Allara 64, 88, Ms Clare Pearson 64, 88, Mr Jeffrey K. J. Cheng MSc 94, Dr Hannah E. Bridgewater PhD 94, Ms Lucy R. Frost BSc 94, Ms Grace Taylor-Joyce BSc 94, Dr Paul E Brown PhD 94, Dr Lily Tong PhD 48, Ms Alice Broos BSc 48, Mr Daniel Mair BSc 48, Mrs Jenna Nichols BSc 48, Dr Stephen N Carmichael PhD 48, Dr Katherine L Smollett PhD 40, Dr Kyriaki Nomikou PhD 48, Dr Elihu Aranday-Cortes PhD/DVM 48, Ms Natasha Johnson BSc 48, Dr Seema Nickbakhsh PhD 48, 68, Dr Edith E Vamos PhD 92, Dr Margaret Hughes PhD 92, Dr Lucille Rainbow PhD 92, Mr Richard Eccles MSc 92, Ms Charlotte Nelson MSc 92, Dr Mark Whitehead PhD 92, Dr Richard Gregory PhD 92, Mr Matthew Gemmell MSc 92, Ms Claudia Wierzbicki BSc 92, Ms Hermione J Webster BSc 92, Ms Chloe L Fisher MSc 28, Mr Adrian W Signell BSc 20, Dr Gilberto Betancor PhD 20, Mr Harry D Wilson BSc 20, Dr Gaia Nebbia PhD FRCPath 12, Dr Flavia Flaviani PhD 31, Mr Alberto C Cerda MSc 96, Ms Tammy V Merrill MSc 96, Rebekah E Wilson MSc 96, Mr Marius Cotic MSc 82, Miss Nadua Bayzid BSc 82, Dr Thomas Thompson PhD 72, Dr Erwan Acheson PhD 72, Prof Steven Rushton PhD 51, Prof Sarah O'Brien PhD 51, David J Baker BEng 70, Steven Rudder 70, Alp Aydin MSci 70, Dr Fei Sang PhD 18, Dr Johnny Debebe PhD 18, Dr Sarah Francois PhD 23, Dr Tetyana I Vasylyeva DPhil 23, Dr Marina Escalera Zamudio PhD 23, Mr Bernardo Gutierrez MSc 23, Dr Angela Marchbank BSc 10, Joshua Maksimovic FD 9, Karla Spellman FD 9, KathrynMcCluggage Msc 9, Dr Mari Morgan PhD 69, Robert Beer BSc 9, Safiah Afifi BSc 9, Trudy Workman HNC 10, William Fuller BSc 10, Catherine Bresner Bsc 10, Dr Adrienn Angyal PhD 93, Dr Luke R Green PhD 93, Dr Paul J Parsons PhD 93, Miss Rachel M Tucker MSc 93, Dr Rebecca Brown PhD 93 and Mr Max Whiteley PhD 93

Software and analysis tools:

James Bonfield BSc 99, Dr Christoph Puethe 99, Mr Andrew Whitwham BSc 99, Jennifier Liddle 99, Dr Will Rowe PhD 41, Dr Igor Siveroni PhD 39, Dr Thanh Le-Viet PhD 70 and Amy Gaskin MSc 69

Visualisation:

Dr Rob Johnson PhD 39

1 Barking, Havering and Redbridge University Hospitals NHS Trust, 2 Basingstoke Hospital, 3 Belfast Health & Social Care Trust, 4 Betsi Cadwaladr University Health Board, 5 Big Data Institute, Nuffield Department of Medicine, University of Oxford, 6 Brighton and Sussex University Hospitals NHS Trust, 7 Cambridge Stem Cell Institute, University of Cambridge, 8 Cambridge University Hospitals NHS Foundation Trust, 9 Cardiff and Vale University Health Board, 10 Cardiff University, 11 Centre for Clinical Infection & Diagnostics Research, St. Thomas' Hospital and Kings College London, 12 Centre for Clinical Infection and Diagnostics Research, Department of Infectious Diseases, Guy's and St Thomas' NHS Foundation Trust, 13 Centre for Enzyme Innovation, University of Portsmouth (PORT), 14 Centre for Genomic Pathogen

Surveillance, University of Oxford, 15 Clinical Microbiology Department, Queens Medical Centre, 16 Clinical Microbiology, University Hospitals of Leicester NHS Trust, 17 County Durham and Darlington NHS Foundation Trust, 18 Deep Seq, School of Life Sciences, Queens Medical Centre, University of Nottingham, 19 Department of Infection Biology, Faculty of Infectious & Tropical Diseases, London School of Hygiene & Tropical Medicine, 20 Department of Infectious Diseases, King's College London, 21 Department of Microbiology, Kettering General Hospital, 22 Departments of Infectious Diseases and Microbiology, Cambridge University Hospitals NHS Foundation Trust; Cambridge, UK, 23 Department of Zoology, University of Oxford, 24 Division of Virology, Department of Pathology, University of Cambridge, 25 East Kent Hospitals University NHS Foundation Trust, 26 East Suffolk and North Essex NHS Foundation Trust, 27 Gateshead Health NHS Foundation Trust, 28 Genomics Innovation Unit, Guy's and St. Thomas' NHS Foundation Trust, 29 Gloucestershire Hospitals NHS Foundation Trust, 30 Great Ormond Street Hospital for Children NHS Foundation Trust, 31 Guy's and St. Thomas' BRC, 32 Guy's and St. Thomas' Hospitals, 33 Hampshire Hospitals NHS Foundation Trust, 34 Health Data Research UK Cambridge, 35 Health Services Laboratories, 36 Heartlands Hospital, Birmingham, 37 Hub for Biotechnology in the Built Environment, Northumbria University, 38 Imperial College Hospitals NHS Trust, 39 Imperial College London, 40 Institute of Biodiversity, Animal Health & Comparative Medicine, 41 Institute of Microbiology and Infection, University of Birmingham, 42 King's College London, 43 Liverpool Clinical Laboratories, 44 Maidstone and Tunbridge Wells NHS Trust, 45 Manchester University NHS Foundation Trust, 46 Microbiology Department, Wye Valley NHS Trust, Hereford, 47 MRC Biostatistics Unit, University of Cambridge, 48 MRC-University of Glasgow Centre for Virus Research, 49 National Infection Service, PHE and Leeds Teaching Hospitals Trust, 50 Newcastle Hospitals NHS Foundation Trust, 51 Newcastle University, 52 NHS Greater Glasgow and Clyde, 53 NHS Lothian, 54 Norfolk and Norwich University Hospital, 55 Norfolk County Council, 56 North Cumbria Integrated Care NHS Foundation Trust, 57 North Tees and Hartlepool NHS Foundation Trust, 58 Northumbria University, 59 Oxford University Hospitals NHS Foundation Trust, 60 PathLinks, Northern Lincolnshire & Goole NHS Foundation Trust, 61 Portsmouth Hospitals University NHS Trust, 62 Princess Alexandra Hospital Microbiology Dept., 63 Public Health Agency, 64 Public Health England, 65 Public Health England, Clinical Microbiology and Public Health Laboratory, Cambridge, UK, 66 Public Health England, Colindale, 67 Public Health England, Colindale, 68 Public Health Scotland, 69 Public Health Wales NHS Trust, 70 Quadram Institute Bioscience, 71 Queen Elizabeth Hospital, 72 Queen's University Belfast, 73 Royal Devon and Exeter NHS Foundation Trust, 74 Royal Free NHS Trust, 75 Sandwell and West Birmingham NHS Trust, 76 School of Biological Sciences, University of Portsmouth (PORT), 77 School of Pharmacy and Biomedical Sciences, University of Portsmouth (PORT), 78 Sheffield Teaching Hospitals, 79 South Tees Hospitals NHS Foundation Trust, 80 Swansea University, 81 University Hospitals Southampton NHS Foundation Trust, 82 University College London, 83 University Hospital Southampton NHS Foundation Trust, 84 University Hospitals Coventry and Warwickshire, 85 University of Birmingham, 86 University of Birmingham Turnkey Laboratory, 87 University of Brighton, 88 University of Cambridge, 89 University of East Anglia, 90 University of Edinburgh,91 University of Exeter, 92 University of Liverpool, 93 University of Sheffield, 94 University of Warwick, 95 University of Cambridge, 96 Viapath, Guy's and St Thomas' NHS Foundation Trust, and King's College Hospital NHS Foundation Trust, 97 Virology, School of Life Sciences, Queens Medical Centre, University of Nottingham, 98 Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford,99 Wellcome Sanger Institute, 100 West of Scotland Specialist Virology Centre, NHS Greater Glasgow and Clyde, 101 Department of Medicine, University of

Cambridge, 102 Ministry of Health, Sri Lanka, 103 NIHR Health Protection Research Unit in HCAI and AMR, Imperial College London, 104 North West London Pathology, 105 NU-OMICS, Northumbria University, 106 University of Kent, 107 University of Oxford, 108 University of Southampton, 109 University of Southampton School of Health Sciences, 110 University of Southampton School of Medicine, 111 University of Surrey, 112 Warwick Medical School and Institute of Precision Diagnostics, Pathology, UHCW NHS Trust, 113 Wellcome Africa Health Research Institute Durban and 114 Wellcome Genome Campus.

AMPHEUS Project

Oxford Viral Sequencing Group, Wellcome Centre for Human Genetics, University of Oxford, UK
Christophe Fraser
David Buck
Angie Green
George MacIntyre-Cockett
Paolo Piazza
John A Todd
Amy Trebes

Oxford Viral Sequencing Group, Big Data Institute, University of Oxford, UK
Laura Thomson

Appendix 3 SARS-CoV-2 NAAT assays used at study sites for the symptomatic pathway (COV001 & COV002)

Site	SARS-CoV2 NAAT assay (1)	SARS-CoV2 NAAT assay (2)	SARS-CoV2 NAAT assay (3)	SARS-CoV2 NAAT assay (4)	SARS-CoV2 NAAT assay (5)	SARS-CoV2 NAAT assay (6)
Glasgow	Hologic Aptima SARS-CoV-2/Panther	Seegene Allplex 2019-nCoV	Abbott M200 RealTime SARS-C0V-2			
Sheffield	Hologic Aptima SARS-CoV-2/Panther	In-house assay based on primers/probes from Charite protocol				
SGUL	Roche COBAS 6800/8800	Cepheid Xpert Xpress SARS-CoV-2 -rtRT-PCR	Altona RealStar SARS-CoV- 2 rRT-PCR kit			
UCL	Cepheid Xpert Xpress SARS-CoV-2 -rtRT-PCR	Hologic Aptima SARS-CoV-2/Panther	BGI rtRT-PCR	Thermofisher TaqPath COVID-19 CE-IVD RT-PCR		
Wales	Cepheid Xpert Xpress SARS-CoV-2 -rtRT-PCR	GenMark ePlex SARS-CoV-2				
Southampton	Hologic Aptima SARS-CoV-2/Panther	Prolab Viasure Real Time PCR	WHO –E-gene & Internal Control assay (in-house)	CDC assay (N1 & N2 genes) plus WHO-E gene and Internal control	Cepheid Xpert Xpress SARS-CoV-2 -rtRT-PCR	Biofire FilmArray RP2.1plus
Hull	GeneFinder COVID-19 Plus RealAmp Kit	Vela ViroKey SARS-CoV-2 RT-PCR	Anatolia Geneworks Bosphore SARS-CoV-2	Cepheid Xpert Xpress SARS-CoV-2	Hologic Aptima SARS-CoV-2/Panther	Biofire FilmArray RP2.1plus
Newcastle	Roche COBAS 6800/8800	Altona RealStar SARS-CoV- 2 rRT-PCR kit	Cepheid Xpert Xpress SARS-CoV-2 -rtRT-PCR	Mobidiag Amplidiag Easy	BioFire RP2.1-plus	

Appendix 3 SARS-CoV-2 NAAT assays

Site	SARS-CoV2 NAAT assay (1)	SARS-CoV2 NAAT assay (2)	SARS-CoV2 NAAT assay (3)	SARS-CoV2 NAAT assay (4)	SARS-CoV2 NAAT assay (5)	SARS-CoV2 NAAT assay (6)
UHBW	Hologic Aptima SARS-CoV-2/Panther	Altona RealStar SARS-CoV- 2 rRT-PCR kit				
NBT	Hologic Aptima SARS-CoV-2/Panther	Altona RealStar SARS-CoV- 2 rRT-PCR kit				
Oxford	Altona RealStar SARS-CoV- 2 rRT-PCR kit	Abbott M200 RealTime SARS-C0V-2	Cepheid Xpert Xpress SARS-CoV-2 -rtRT-PCR	BioFire COVID-19 RTPCR		
Birmingham	Hologic Aptima SARS-CoV-2/Panther	Cepheid Xpert Xpress SARS-CoV-2 -rtRT-PCR	Abbott M200 RealTime SARS-C0V-2	Prolab Viasure Real Time PCR		
Cambridge	Hologic Aptima SARS-CoV-2/Panther					
Edinburgh	Seegene Allplex 2019-nCoV					
GSTT	AusDiagnostics PCR SAR-Cov-2/ Flu/ RSV	Hologic Aptima SARS-CoV-2/Panther				
Imperial	Roche COBAS 6800/8800	Cepheid Xpert Xpress SARS-CoV-2 -rtRT-PCR	AusDiagnostics PCR SAR-Cov-2/ Flu/ RSV	Abbott M200 RealTime SARS-C0V-2	Thermofisher TaqPath COVID-19 CE-IVD RT-PCR	
Liverpool	Hologic Aptima SARS-CoV-2/Panther	Prolab Viasure Real Time PCR				
Nottingham	Altona RealStar SARS-CoV- 2 rRT-PCR kit	Abbott M200 RealTime SARS-C0V-2	Cepheid Xpert Xpress SARS-CoV-2 -rtRT-PCR	Mobidiag Amplidiag Easy		

Appendix 4 Part 2: *Salmonella* Typhi

10 Bottlenecks

10.1 Introduction ("Bottlenecks")

The study of within-host population dynamics of a pathogen is an evolving field of exploration. Identification of newer genetic and visualisation techniques has facilitated the investigation of these intricate stories. Prior to this, scientists used existing approaches to study the fundamental question of how a population of individual pathogenic units causes infection, and what happens to that population within the host. Despite this, particularly in humans, the spatiotemporal population dynamics of common human bacterial infections remain poorly understood. Understanding these complexities, coupled with immunological insights may have two key practical implications: rational vaccine design for the prevention of disease and selection of the most appropriate dosing regimens for antimicrobial agents during treatment. [448,449]

Exposure of a susceptible host to a single pathogenic unit, for example one bacterium, is typically insufficient to cause an infection but following exposure of a host to a large number of these units, an infection may be readily established. [450] Meynell and Stocker put forward two potential constructs to explain this. [451] The first being that of cooperation or synergy whereby the organisms within a population act together to cause the infection. The second is the concept of 'independent action' whereby each organism has the capacity to cause an infection. Each single unit has a very small probability of doing so but increasing the size of the inoculating population increases the probability that one (or more) of that population will succeed in causing an infection. Given that any organism has a small chance of causing infection, the most extreme consequence of this is that a single organism is responsible for the initiation of an infection within a host despite the inoculum being a much larger population. In this case it is the so called 'founder' organism that establishes the

infection and all organisms contributing to infection in the host have resulted from clonal expansion of this founder. The concept of population constriction is known as the bottleneck or independent action hypothesis. Indeed, the independent action and cooperative hypotheses may not be mutually exclusive and both mechanisms may occur at different points during the infective process. It is also plausible that populations of different pathogens may operate differently in how they cause infections which may be reflected in the infective dose. [452]

These hypotheses can be investigated using experimental models and observations can be made from clinical infections. Figure 10.1 illustrates how using markers to discern subpopulations can provide information about how the total population has changed. Figure 10.1 also emphasises both that the bottleneck can vary in size and that the replication of organisms, and any genetic variation that occurs prior to sampling, can affect the composition of the sample population. It is also important to recognise that there are multiple barriers to infection which can be dynamic, and host or pathogen related.

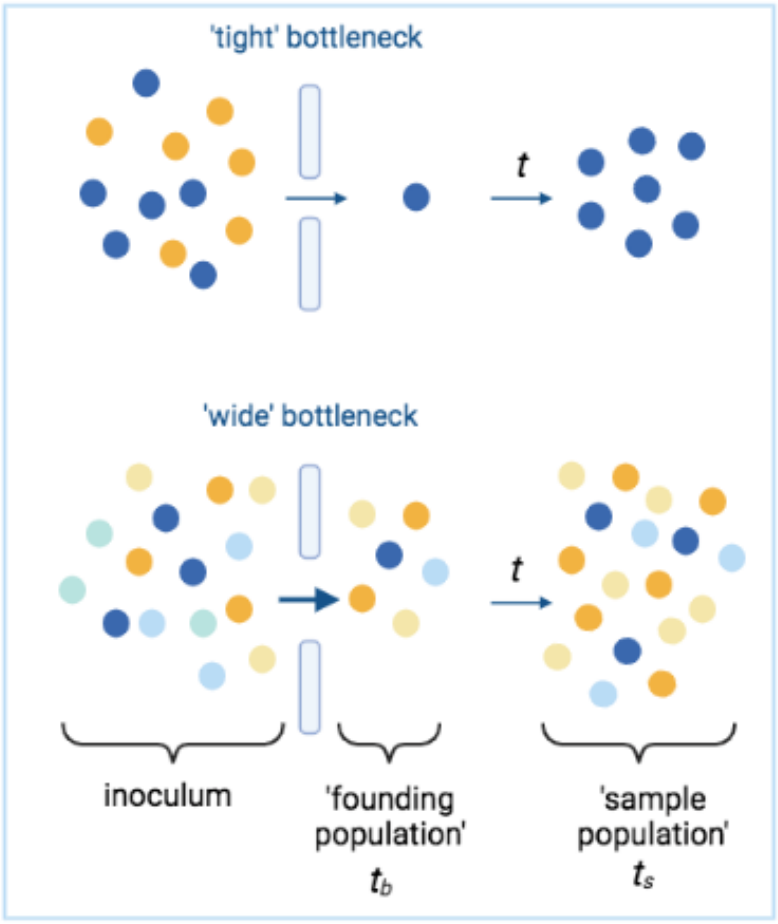


Figure 10.1: The effect of bottlenecks on genetic diversity. Bottlenecks are events or processes that produce a reduction in population size. Those individuals that survive the bottleneck typically go on to replicate and therefore to 'found' a new population. Individual pathogens are represented by coloured circles. The different colours represent a means to distinguish an individual subpopulation such as a genetic tag. Barriers to infection are represented by the solid light blue bars. A tight bottleneck can lead to a stochastic loss of subpopulations and a large departure in the ratios of different subpopulations from the inoculum. Alternatively, a wide bottleneck may produce more subtle changes such as loss of a limited subpopulations. The study of the dynamics of these subpopulations can used to study events within a host. Figure and legend adapted from Abel et al. (2015). [453] Reproduction licensed under CC BY 4.0. [1]

First, I will consider some of the important historical investigations into bacterial population dynamics before reviewing more recent work. I will then consider the use of a controlled human infection model of *Salmonella* Typhi to study population dynamics in humans.

10.1.1 Early evidence for population bottlenecks in experimental infection

Early work investigating the relationship between mortality of a variety of respiratory pathogens and infective dose in mice suggested that the minimum lethal dose was likely to be a single organism, supporting the concept of independent action. [450] In the 1950s Meynell performed experiments to explore the bottleneck hypothesis using mouse models of invasive *Salmonella* infections. [451,454] Bacterial strains with differing antibiotic resistance mechanisms, flagellar antigens or ability to ferment xylose, were used, allowing easy identification with available laboratory techniques.

Meynell inoculated groups of mice a range of doses, each dose level comprising more than one *Salmonella* strain. His hypothesis was if the principle of a single cell bottleneck were true, a single strain should be cultured from mice at death (or should form the majority of isolates) at doses either at, or less than LD_{50} (the dose at which 50% of mice die), based on the assumption that the bacteria present from the mortem blood sample were progeny of the original infecting inoculum. The data generated provided some evidence in favour of a bottleneck. Single strain blood cultures were attained from some mice when the inoculum was given orally or intraperitoneally, and abscesses contained a majority of or only a single strain. [451,454] However, there were several features which compromised interpretation of Meynell's experiments: the differential growth rate between the streptomycin sensitive and resistant strains, the use of passaged and thus attenuated strains, mixed bacteraemias in the late stages of infection and differential virulence between routes of inoculation. These limitations and the inconsistent results make it hard to draw conclusions from this early work, but it did establish that using relatively simple models of co-infection, the concepts could be explored.

A subsequent study by Moxon & Murphy tested the same hypothesis using the infant rat model of bacteraemia and meningitis caused by intranasal inoculation of *Haemophilus*

influenza type b. This study used equal proportions of two strains that differed by susceptibility to streptomycin. [455] They found convincing evidence of a bottleneck: 96.7% (58/60) of bacteraemias were caused by a single strain at lower infective doses (10^4-10^5 colony forming units, CFU). Interestingly single strains were still evident at an infective dose of 10^8 CFU but were less likely. At each dose increase, a greater proportion of mixed cultures was observed suggesting a dose-response relationship. These findings were replicated for nasopharyngeal cultures performed 48-72 hrs post challenge. [456] Intriguingly, serial blood cultures taken up to six hours post intranasal inoculation were more likely to be mixed compared with those taken days after challenge, suggesting that reduction in the number of strains did not occur at the nasopharyngeal mucosa; the bottleneck was at a subsequent stage of infection. Meningitis only resulted in those rats that became bacteraemic, supportive of haematogenous route of infection, and in 39/40 cases that had a single variant blood culture, the same strain was found in CSF. Of those where blood cultures returned both variants, 13/19 showed a single strain in their CSF, and the remainder were mixed; demonstrating that progression from bloodstream to meninges was likely to be associated with a second bottleneck at the junction between the bloodstream and the central nervous system. [455]

Further support for a bottleneck appearing beyond the mucosa was found in a new-born rat *E. coli* model of colonisation and bacteraemia. Mixed inoculum experiments produced mixed bacterial populations from mesenteric lymph nodes. This mirrored the colonising populations in the gastrointestinal tract, yet pure cultures of either variant were obtained from blood samples. [457] These results support the concept of clonal expansion of a single cell to cause the bloodstream infection but also that multiple strains are transported to regional lymph nodes.

These two experiments also demonstrate that the host is comprised of multiple anatomical compartments, such as the bloodstream and lymph nodes which may, in themselves be hosts-within-hosts. These early co-infection studies suggest that there may be distinct reservoirs of infection that contain a single strain. They also highlight that there are several stages at which a population bottleneck can take place: mucosal site of exposure, at the level of lymph node, entry into or from the bloodstream or other tissues. These studies showed that experimental models could further understanding of pathogen dynamics but were limited by available techniques.

10.1.2 Recent experimental work

Many factors influence the infectious process: infectious dose, route of administration, pathogen virulence, growth characteristics and host susceptibility. Traditionally, bacterial disease burden has been measured by looking at the host holistically. Global host bacterial burden however demonstrates little of the mechanism of how that burden has occurred in different anatomical compartments over time.

More recent methods have permitted i*n vivo* visualisation of these processes. Multicolour fluorescence microscopy and laser scanning confocal fluorescence microscopy were used in a mouse model of *Salmonella* infection to explore within-host foci of infection. [449] Following intravenous inoculation with a 1:1 ratio of two different *S.* Typhimurium strains, at both one hour and day 4 post inoculation, all 143 infected hepatic phagocytes contained only one strain or the other, but not both. Interestingly the ratio between the two strains within the population of infected phagocytes remained the same as the ratio in the inoculum. By day 4, all infected hepatic phagocytes tested had either one strain or the other present thus it is likely that the multiple bacteria present in each phagocyte resulted from clonal expansion of the original infecting cell. A similar mouse pneumococcal model also demonstrated that each infectious focus within splenic tissue was initiated by only a single

strain. [458] These results emphasise that the host contains many micro-infections, which may exist at a cellular rather than organ level.

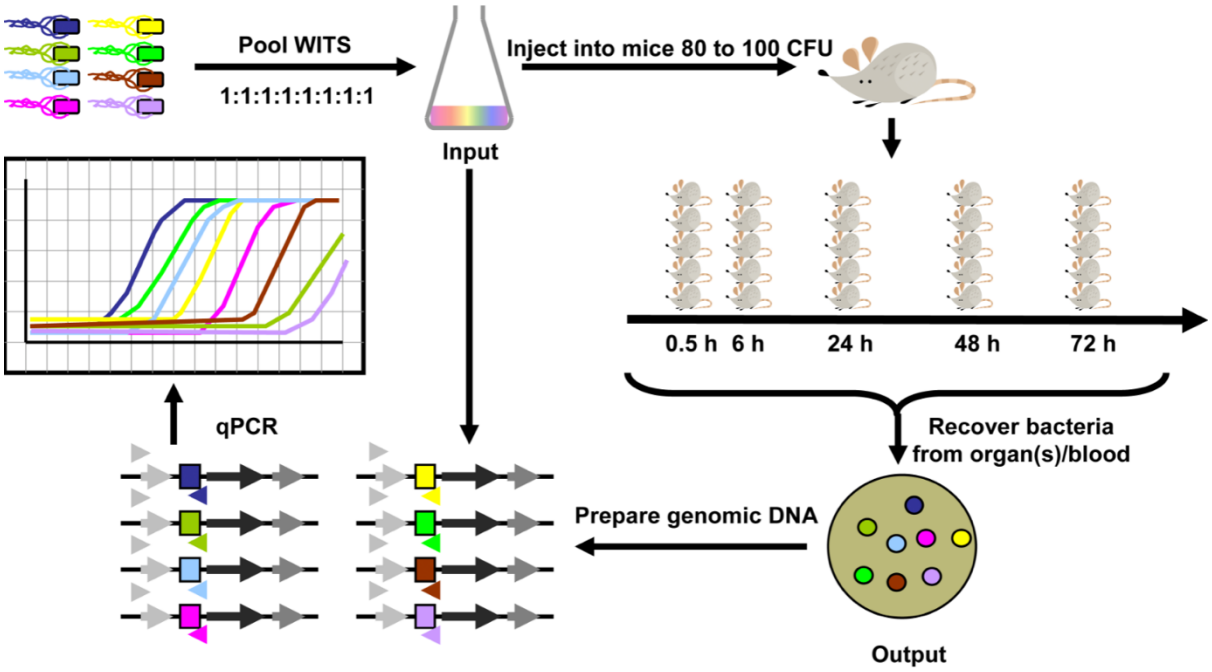


Figure 10.2: The experimental design of a WITS experiment Mice were inoculated i.v. with approximately 80 to 100 CFU of *S.* Typhimurium, being approximately equal CFU of each of eight WITS. At 0.5, 6, 24, 48, and 72 h p.i., groups of mice were killed, and the livers, spleens, "other organs" (kidneys, heart, lungs, and mesenteric lymph nodes), and blood were cultured for the presence of bacteria. All bacterial colonies were scraped from the agar plates, and chromosomal DNA was purified using standard commercial kits. The S. enterica population structure in individual organs at different times after i.v. inoculation was detected using quantitative PCR. Each biological sample was assessed by qPCR using each of the eight primer pair combinations, the number of copies of each WITS chromosomal DNA per reaction (biological sample) was calculated from standard curves for each of the WITS. Figure and legend from Grant et al. (2008). [448] Reproduction licensed under CC BY 4.0. [1]

Modern molecular tagging techniques, such as wild-type isogenic tagged strains (WITS) and sequence tag-based analysis of microbial populations (STAMP) have permitted the study of population dynamics with even greater spatiotemporal resolution. [448,459,460] Figure 10.2 shows an example of a WITS experiment within a mouse model. In these studies subpopulations of the original pathogen are defined by a unique genetic tag inserted into a non-coding region of the genome. These genetically neutral barcodes produce subpopulations that behave identically *in vivo* but can be readily differentiated, typically

using techniques such as quantitative PCR. This is preferable to using strains that have phenotypic differences such as antibiotic resistance which may have key roles with respect to virulence or growth. These methods offer the advantage over studies of co-infection as they allow the introduction of a number of strains thereby providing data from multiple subpopulations. [459] Studying multiple strains offers advantages over simple two strain co-infection experiments as they can provide information to distinguish replication and movement between compartments and can permit the demonstration of bottlenecks which may be larger than a single bacterium. [459,461]

In a mouse model of *S.* Typhimurium WITS data indicate that even during the first six hours of an infection there are changes in the prevalence of subpopulations over time within and between organs. [448] The differences in subpopulation ratios between liver and spleen were maintained until 24 hours suggestive of local foci of infection and a lack of mixing between organs at least up until 24 hours post challenge. After this timepoint haematogenous spread makes the subpopulations in liver and spleen more comparable, however when the bacterial population from an entire animal was considered some subpopulations formed a greater proportion than others. Given the virulence of each tagged strain was identical, this suggests that these populations that appeared in greater proportion had had more sites for replication. Using multiple strains gives the resolution to suggest that by chance certain subpopulations infected more hosts-within-hosts, if there were no population constriction the ratio from the inoculum should be broadly preserved in the sample population from the entire animal. Additionally using multiple strains demonstrates the stochastic nature of these infections thereby producing data for *in silico* modelling which can be used to derive information that is not directly measured. [448] Whole genome sequencing methods have also provided support for earlier observations: intergenic or silent SNPs of pneumococci cultured from mice after a 3 isogenic variant challenge revealed evidence of a monoclonal genetic populations resulting from clonal expansion from a single infecting cell. [462]

10.1.3 Clinical evidence

The population dynamics of bacterial infection in animal models may seem distant from common clinical infections. In the UK alone there are estimated to be at least 200 000 cases of sepsis causing at least 46,000 deaths per year, with a direct annual cost to the NHS of at least £1.5 billion. [463] Many patients with sepsis will have a bacteraemia. Despite the morbidity and mortality associated with bacterial blood stream infections (BSI), comparatively little is known about the pathogenesis from a pathogen population perspective. At a time of increasing antimicrobial resistance and a lack of new antimicrobial agents, understanding the basic pathogenesis of BSI is essential for efforts directed at prevention. Within-host population dynamics are one aspect of this.

Clinical infection can provide some information on population dynamics, but data are often limited as patients are already infected at the time of presentation. There are some observations that may support the idea of bottlenecks. Studies of infective doses from outbreaks, laboratory infections and controlled human infection models suggest that for some such as *Shigella dysenteriae*, *Coxiella burnetti* and *Mycobacterium tuberculosis* the dose at which 50% of individuals are infected (ID_{50}) is low [452,464] supporting the concept that a single organism, or very few organisms can be sufficient to cause infection.

It is plausible that within endemic settings individual simultaneous exposure to more than one genotype of typhoidal *Salmonella* is both possible and likely [465] yet reports of co-infections from enteric fever pathogens remain rare. [466,467] One explanation for this may be lack of detection: simple lack of access to blood cultures, or that isolates may appear monomorphic yet be genotypically different. Cases in the literature describe detection occurring when two populations are observed on susceptibility testing plates rather than at an earlier step in processing. Whilst it is unlikely that there would be equal concentrations of multiple types of bacteria each with identical virulence within a contaminated water

source, mixed infections might be predicted to occur more frequently unless there is an elimination of other strains at some stage during infection.

This introduction has focussed on bacterial infection, however there are data from many other pathogens which show evidence of bottlenecks. Pathogens especially viruses, with a high degree of genetic diversity, allow exploration of the dynamics within and between human hosts by using inherent genetic differences as markers of subpopulations. [468-470] For example, studies of heterosexual transmission of HIV have found evidence of homogeneous sequences in recipient blood samples despite the diversity of variants present in the donor genital tract, suggesting it is a single variant which founds the initial infection in the recipient. [471] Analysis of intra-host single nucleotide variants allows exploration of viruses with low genetic diversity such as SARS-CoV-2 and demonstrates a narrow transmission bottleneck within household. [40]

This approach unfortunately is restricted by the natural genetic diversity of the organism, and for many, particularly bacterial species such as *Salmonella* Typhi there is insufficient diversity to support within host studies. [472]

10.1.4 Experimental models to investigate population dynamics

There are inherent difficulties of studying infection from inoculation to infection in humans. Figure 10.3 outlines some of the key desirable features for a model to investigate population dynamics. Many of these features are only available within established animal models and whilst providing a basis for knowledge, these models may be limited in the way in which within-animal population dynamics can inform knowledge about human pathogenesis.

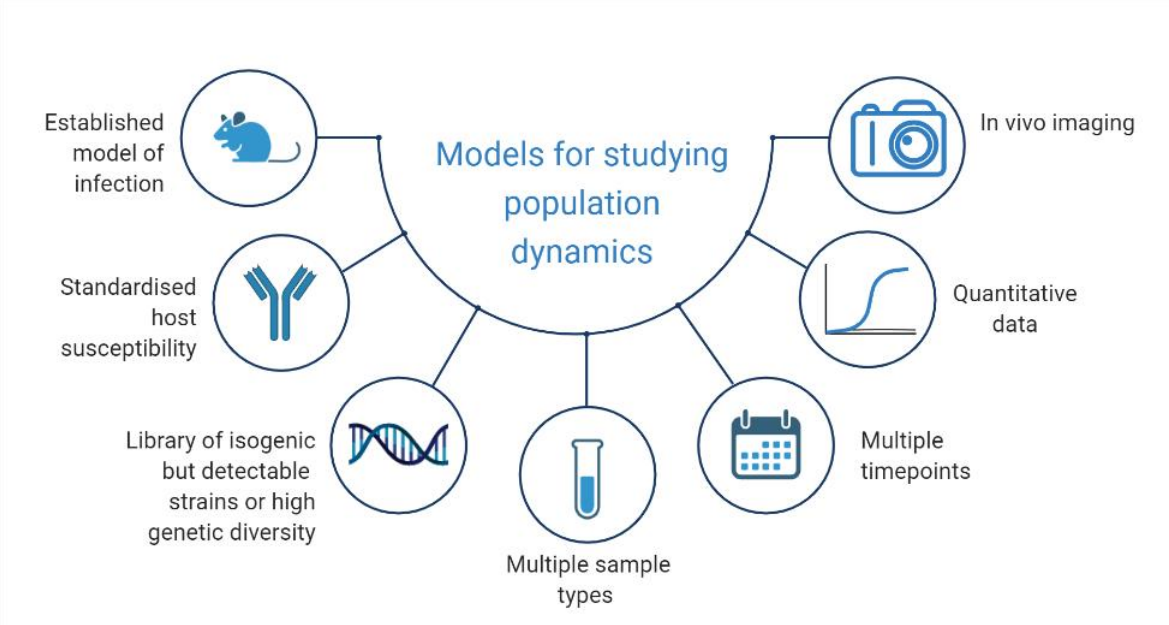


Figure 10.3: Ideal features for an experimental model to explore population dynamics.

Controlled human infection models (CHIMs) involve the exposure of healthy volunteers to specific doses of well-characterised infectious agents. These models can be used to study many aspects of the infectious process from the pathogenesis, immune response, and clinical disease, particularly in infections that are human restricted or difficult to study in the settings where they occur naturally. There are clear advantages of human infection models over observation of natural infection: defined timing, dosage and route of exposure, observation through the incubation to infection period and ability to strictly select participants. It is, however, important to remember that these too are models which have their limitations and typically have been established for purposes that may differ from studying population dynamics.

Before considering the *Salmonella* Typhi CHIM and its application to the study of population dynamics in humans I will briefly review natural *Salmonella* Typhi infection.

10.1.5 Natural *Salmonella* Typhi infection

10.1.5.1 Bacteriology

Salmonella Typhi are Gram-negative, rod-shaped bacteria which are catalase positive, oxidase negative and non-lactose fermenting. In the diagnostic laboratory they are typically identified via selective enrichment culture and the use of chromogenic or selective agar such as XLD (Xylose-Lysine-Desoxycholate) agar followed by agglutination with commercial antisera.

Salmonella Typhi is one of the species *Salmonella enterica* which is, in turn split into 6 subspecies. The subspecies *enterica* is then further subcategorised into serovars of which Typhi is one. Serovars are classified by the White-Kaufmann-Le Minor scheme, based on antibody reactions to surface LPS and flagella antigens (O, H and Vi).

Salmonella enterica subspecies *enterica* serovars can be further divided by the nature of the disease they cause. The vast majority of serovars are non-typhoidal, causing a spectrum from asymptomatic infection through to invasive disease and death especially in the very young and immunocompromised. [473] *Salmonella* Typhi and *Salmonella* Paratyphi A, B and C however can cause severe 'typhoidal' disease with mortality in up to 20% in the pre-antimicrobial era. [474]

10.1.6 Epidemiology & Burden

Globally there are estimated to be approximately 10 million cases of typhoid per year, resulting in over 130 000 deaths and 9.8 million disability-adjusted life years, with both the morbidity and mortality burden disproportionately affecting those in the first two decades of life. [475]

The systematic analysis for the Global Burden of Disease Study suggests that incidence is highest in south Asia, then southeast Asia, Sub-Saharan Africa, and Oceania. Whilst these data have many strengths the investigators emphasise the paucity of data from many regions accounting for uncertainty in these estimates. [475] Several large studies are underway to address some of these gaps, for example: STRAATA, [476] SEAP, [477] SEFI, [478] and SETA. [479]

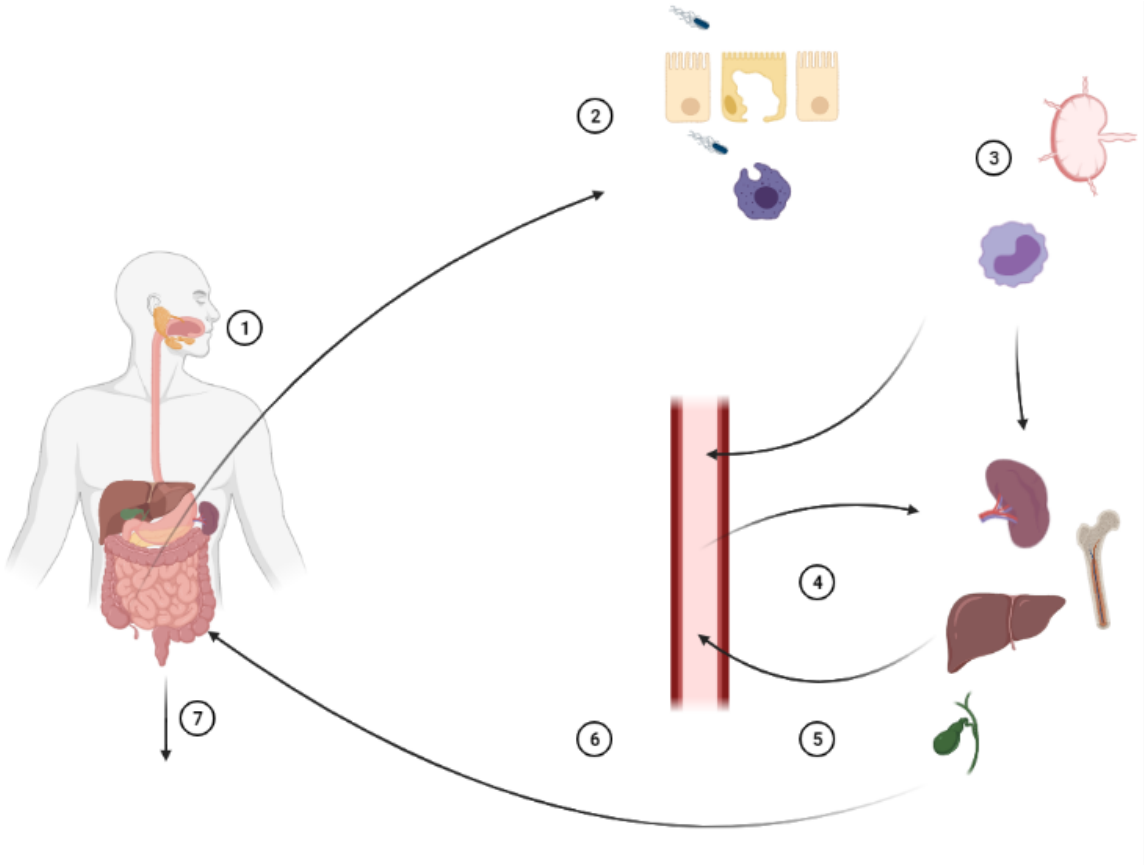


Figure 10.4: Pathogenesis of Enteric fever
1. S. Typhi bacteria are typically ingested in contaminated food or water
2. S. Typhi invade in the small intestine and cross the gut epithelium
3. S. Typhi is transferred to monocytic cells and trafficked to the reticulo-endothelial cells, a transient asymptomatic primary bacteraemia may occur
4. S. Typhi preferentially seeds to the liver spleen, gallbladder and bone marrow and can reside there in the long term, evading the immune system
5. After an incubation period S. Typhi re-emerges into the bloodstream, with a low bacterial burden, this may be associated with symptoms in the human host
6. S. Typhi may be shed via the gallbladder into the gastrointestinal tract
7. Bacteria can be intermittently shed into the stool, sometimes at high levels.
Figure and legend adapted from Baker et al. (2010). [480] Reproduction licensed under CC BY 4.0. [1]

Studies of typhoid pathogenesis are limited by human host restriction and understanding of this disease is largely based on animal models. These typically use a non-typhoidal *Salmonella* species manipulated in the model (e.g. by route of inoculation) to cause invasive infection. In natural infection bacteria are typically ingested and if they survive the gastric acid in the stomach pass into the small intestine. At the gastrointestinal epithelial surface, bacterial entry is not fully understood, but may occur via or between epithelial cells or by uptake via M cells. [481] Murine and human models of Salmonellosis suggest that within hours of inoculation there is a primary asymptomatic bacteraemia thereby disseminating inoculating bacteria around the body particularly to bone marrow and the reticulo-endothelial system. [482,483] Following the incubation period, the length of which is inversely related to dose, [484] a sustained bacteraemia accompanied by systemic symptoms occurs. *S.* Typhi has a predilection for colonising the gallbladder and intermittent shedding of bacteria into the bowel occurs (Figure 10.4).

10.1.7 Infectious Dose & Transmission

Recent controlled human infection studies have determined the ID_{50} in healthy adults in non-endemic regions to approximately of 10^3 organisms however these studies use bicarbonate solution to suppress stomach acid prior to the delivery of the challenge dose. [485] Older studies which suspended the challenge dose in milk showed an ID_{50} of 10^7 CFU with a dose dependent relationship between dose and clinical attack rate and an inverse relationship with median incubation time in days. [484] In contrast, from studies of outbreaks where the infective dose has been estimated from incubation periods, the infectious dose is lower and usually $<10^3$ CFU (reviewed in Blaser & Newman, 1982). [486] It is highly likely that in endemic settings the dose individuals receive may vary widely and, under the right circumstances a small number of organisms may be enough to cause disease.

S. Typhi is transmitted via faeco-oral spread, particularly via water contaminated with sewage, [465,487] within households, [465] and less frequently by contaminated food stuffs. [488,489] Laboratory acquired infection has also been reported. [490]

10.1.8 Disease

Typhoid fever is characterised by systemic malaise, high fevers which may be sustained for days, headache, abdominal discomfort, and gastrointestinal disturbance. Other features may include rose spots, coughing, vomiting and stupor. Untreated, clinical disease can be complicated by intestinal perforation which can result in death. In the absence of complications such as perforation, laboratory investigations may show changes in white cell counts (increased or decreased) and elevations in CRP or liver transaminases, but these are non-specific. Asymptomatic carriage of *S.* Typhi in the gallbladder, with ongoing stool shedding can occur for many years despite treatment with antimicrobials.

10.1.9 Diagnostics

Clinical diagnoses of typhoid are unreliable due to the overlap with other febrile illnesses within endemic areas. Typically, blood and stool are the most readily accessible samples taken for *S.* Typhi diagnosis, though other specimens such as bone marrow and duodenal aspirates can also be used. Microbiological diagnosis is difficult due to the low bacterial burden in samples producing relatively low sensitivity, with bone marrow culture the most sensitive of all methods, with a ratio of 13.5:1 of bacteria in the bone marrow compared with simultaneous venous blood specimen. [491,492] Currently there is a reliance on culture-based methods which are costly, both in terms of time and infrastructure. Nucleic acid amplification methods are hampered by similar issues as culture – a lack of sensitivity despite good specificity and are expensive, though faster than culture. It is worth noting the lack of sensitivity is due to the lack of DNA template in the sample due to low bacterial

count in peripheral blood; the NAAT methodology and primers are highly sensitive in themselves. DNA-based methods do not currently permit antimicrobial susceptibility testing either which culture can offer, albeit relatively slowly. Rapid point-of-care tests, which would be most suited to endemic areas, have not yet offered the required level of sensitivity and specificity to support routine clinical use.

Following this brief overview of the pathogen used in this model I will now consider the model as a means for investigating population dynamics in humans.

10.1.10 The Salmonella Typhi CHIM as a model for exploring population dynamics within a human host

10.1.10.1 The Salmonella Typhi controlled human infection model

The controlled human infection model (CHIM) of *Salmonella* Typhi was established at the University of Oxford and has been used in several studies since 2011. These studies have explored a range of questions: infective dose finding, [485] investigating the efficacy of oral[493] and parenteral[494] vaccines as well as looking at the effects of reinfection[495] and the role of the typhoid toxin. [496] Over 400 participants have undergone challenge with either *S.* Typhi or *S.* Paratyphi A and the model has been shown to be safe and reproducible.

10.1.10.2 How well suited is the CHIM to explore population dynamics?

Established, reproducible model of infection for given pathogen with ecological route of infection

To offer an insight into population dynamics the model must be well established with a clearly defined dose, reliable incubation period and means for monitoring and diagnosis. The oral route of inoculation is the same in the model as for natural infection and whilst the gastric acid suppression is artificial this appears only to reduce the CFU required to induce the same clinical attack rate, it is unknown if this would affect infection downstream in the

gastrointestinal tract. Classically the development of such models starts with the essential step of dose finding typically using dose-escalation studies with the aim of characterising a reproducible percentage of participants becoming infected. [485] Unlike animal studies where a range of doses can be used[450,451,455], the chosen dose for the *S*. Typhi CHIM is 10^4 CFU to cause a clinical attack rate of approximately 65%. This is because of the required balance of exposing the smallest number of participants possible to the pathogen, but to cause a sufficient number of infections from which to be able show differences between groups (e.g. following vaccination). Through repeated studies utilising the same core challenge procedures the attack rates in participants receiving 10^4 CFU of the same wild-type *S*. Typhi in the control groups range from 65-77% in a total of 102 participants and median time to diagnosis ranged from 6-8 days across these studies. [485,493,494,496]

The bloodstream bacterial burden of *S*. Typhi infection in the CHIM mirror observations in observational studies of uncomplicated typhoid fever in the region of one CFU per ml. [485,497] Sensitivity can be improved by increasing blood volumes. [497] Bone marrow would offer a greater sensitivity but is markedly invasive, may be less acceptable to participants, cannot be easily repeated like a blood sample and logistically difficult to implement within an ambulatory CHIM setting. Stool shedding is intermittent following challenge and burden in stool is low, correspondingly stool culture sensitivity is low. [497]

Previous challenge work showed it was possible to detect evidence of early bacteraemia using a culture-PCR based method akin to that described in the animal literature but was not predictive of clinical typhoid diagnosis within the challenge period. This suggests that early bacteraemia may be necessary but not sufficient to cause subsequent bacteraemia. [483] The culture-PCR method in this study had an overall sensitivity of 70% using a 5 ml blood sample, less than that for routine BACTEC™ cultures (87.5%) but the inclusion of the culture-PCR data increased the attack rate from 59% to 73%.

Standardised host susceptibility to given pathogen

Host vulnerability to typhoid infection is not fully understood. It is known to be related to prior exposure as suggested by a reduction of incidence with age in endemic areas, and host genetics. [498] This CHIM was carried out in a non-endemic area and participants must not have spent 6 months or more living within an endemic region to minimise the presence of host immunity. For this study participants were eligible if they had not received the licensed Vi-polysaccharide vaccine within the last 3 years. All participants had their Vi antibody performed at baseline but included regardless of results.

Large library of genetically near identical but discernible strains

The gold standard manufacture of challenge agents is Good Manufacturing Practice (GMP). Manufacturing challenge agents to GMP is a lengthy and expensive process. If additional genetic modifications are made to a wild-type organism, under schedule 2 of the Genetically Modified (Deliberate Release) Regulations 2002 an application for consent to release must be made to the Department for Environment, Food and Rural Affairs for each genetically modified organism. Confirming that the introduction of a virulence-neutral genetic barcode did not alter the disease phenotype or clinical attack rate within a CHIM is likely to be prohibitive for multiple bacterial strains. The introduction of antibiotic resistant genes is not permitted for experimental work in humans to avoid the risk of transfer of such genes from challenge agent to host flora. For these reasons, multiple isogenic but distinguishable bacterial strains for use in a CHIM do not yet exist.

In 2017, a study was conducted to investigate the role of the typhoid toxin in acute typhoid fever. [496] An isogenic typhoid toxin-deficient knockout strain was constructed from the wild-type Quailes strain that was used in the previous challenge studies. Both the wild-type and toxin-deficient strains were considered to be phenotypically comparable with no

differences in key virulence factors such as SPI-7. Participants were randomised to be challenged with identical doses of either the wild-type or toxin-deficient strain. Animal data suggested that the toxin may play a role in the manifestation of disease however this study found no difference in the clinical attack rate, time to diagnosis, clinical phenotype, or quantitative colony counts within the CHIM.

These two strains therefore presented the opportunity to use the typhoid toxin gene to act as a marker to track the presence or absence of two subpopulations within the human host.

Multiple specimen types and multiple timepoints for sampling

For participant safety, daily blood tests following challenge are performed for diagnostic purposes. This provides regular opportunities for sampling; however, blood draw volumes are necessarily restricted for safety to avoid anaemia from over-sampling, which can be exacerbated by typhoid infection. For repeated sampling blood and stool are the main practical specimen types. Additional information would be gained from bone marrow where bacterial burden is highest[492] but poses the issues already discussed. More invasive samples such as bile fluid, splenic aspirates and mesenteric lymph nodes are not attainable within a human model.

Sampling techniques producing quantitative data

Quantitation of the bacterial burden in *S.* Typhi infection has traditionally been performed using pour plates but can also be performed using a commercial collection and concentration system. [497] Whilst colonies can be obtained, as shown by previous CHIM studies, due to the paucity of bacteria in the blood, this is not a sensitive route for diagnostic testing. [485,496] Quantitative PCR has a similarly low sensitivity for the same reason. [499]

In vivo imaging

Techniques for the diagnosis of invasive bacterial infections in humans do exist and typically rely on fluorodeoxyglucose PET or radiolabelling of white cells. These are used for the diagnosis of deep-seated clinical infections, particularly in association with prosthetic material. Spatial resolution even with the addition of a further modality such as CT would be insufficient to track individual bacterial cell populations within systemic infections. Animal models typically use *ex vivo* imaging such as confocal microscopy on infected organ sections however there are multiple new modalities that offer promise in this area. [458,500]

I have described the background of experimental exploration and evaluated the use of a controlled human infection model of *S.* Typhi to study pathogen dynamics. I will now present the findings of a pilot study.

10.1.11 Aims for Part 2 of this thesis

This is the first time, to our knowledge that a controlled human infection model will have been utilised to consider whether a bottleneck occurs in *S.* Typhi infection. This builds on a body of evidence from animal work that population constrictions do occur during infection and that population dynamics can be investigated using models of infection. The overall aim of this study was to establish if there is evidence of independent action of *S.* Typhi within a human experimental model using a combination of *S.* Typhi strains. To achieve this, it was also necessary to evaluate whether the controlled human infection model (CHIM) of *S.* Typhi was suitable for exploring population dynamics in a human host.

The primary objective of this study was:

- to characterise the bacteraemia that occurs post challenge in participants who have received a combination of wild type *Salmonella* Typhi Quailes strain (WT) and a typhoid toxin-deficient isogenic mutant of *Salmonella* Typhi Quailes strain SB6000 (TN).

The null hypothesis was that both strains would be isolated from individuals inoculated with both strains.

The secondary objectives were:

- To quantitate the bacteraemia that occurs post challenge in participants who have received a combination of wild type *Salmonella* Typhi Quailes strain (WT) and a typhoid toxin-deficient isogenic mutant of *Salmonella* Typhi Quailes strain SB6000 (TN).
- To investigate stool shedding of *Salmonella* Typhi that occurs post challenge in participants who have received a combination of wild type *Salmonella* Typhi Quailes

strain (WT) and a typhoid toxin-deficient isogenic mutant of *Salmonella* Typhi Quailes strain SB6000 (TN).

10.2 METHODS ("BOTTLENECKS")

10.2.1 Study Design

The OVG2018/06 'Bottlenecks' study was a pilot observational study to assess whether inoculating human participants with a combination of *S.* Typhi strains resulted in a mixed bacteraemia. It was based on the ambulatory enteric fever control human infection ("challenge") model established in Oxford since 2011 and developed from a number of studies involving *S.* Typhi [485] and *S.* Paratyphi A [501].

10.2.1.1 Setting

The study was performed at the Centre for Clinical Vaccinology and Tropical Medicine, Churchill Hospital. Oxford, UK.

10.2.1.2 Study Population

Healthy adult participants aged 18-60 who volunteered and were eligible for inclusion were recruited to the study. Full inclusion and exclusion criteria are found in Appendix 4.1. Briefly, participants could not have received a typhoid vaccination in the preceding 3 years, nor a diagnosis of enteric fever or resided in an enteric fever endemic area for >6 months. Exclusion criteria included significant health issues, including psychiatric illness and allergies, gallbladder disease, as well as occupational and household contact with individuals who may be vulnerable to *Salmonella* infection.

10.2.1.3 Recruitment

Potential participants were approached for recruitment using a variety of methods, including mailouts, healthy volunteer databases, poster advertisements and social media. Potential

participants who expressed an interest in the study had a copy of the participant information sheet sent or emailed to them. They were initially assessed for eligibility by telephone screening by a member of the study team and if appropriate invited for an in-person screening visit.

10.2.1.4 Informed consent

At the face-to-face screening appointment participants had the opportunity to discuss the nature and rationale for the study, the implications, potential side effects and risks of taking part with a member of the study team prior to signing a written informed consent form. Continued verbal consent was also sought at each study visit.

10.2.1.5 Screening

Screening was performed to assess health, travel, occupation, and vaccine history. Participants underwent a mood assessment using the Hospital Anxiety and Depression Scale (HADS) questionnaire. Blood and urine testing was performed including full blood count, creatinine, amylase, liver function testing, HLA-B27 screen, screening for coeliac disease and immunoglobulin A, blood borne viruses and Vi titre. Participants also underwent an ECG to assess for prolonged QTc and an abdominal ultrasound to assess for gallbladder disease. Participants were also asked to consent to the study team contacting with their regular GP for provision of their medical and vaccination history to facilitate assessment of eligibility.

10.2.1.6 Sample size

The choice of sample size of up to 15 volunteers was pragmatic and based on financial and practical constraints.

10.2.1.7 Clinical Study procedures

A table of study procedures is shown in Table 10.1.

Table 10.1: Summary of study procedures. TD = typhoid diagnosis. [a]Physical examination was performed at other times in the study if clinically indicated and at the discretion of the study team. [b]For female participants only. [c]Stool samples were collected at least one week after completion of a 7-day course of antibiotics and each taken 48 hours apart, until 3 successive stool samples were negative for S. Typhi. If persistent stool shedding occurred after completion of antibiotics, participants will be issued a further course of an alternative antibiotic and the stool clearance process was then repeated. If persistently positive, they would be referred to an Infectious Diseases Consultant at the Oxford University Hospitals NHS Foundation Trust.

			Intensive challenge period (first 14 days)						Treatment		Diagnosis			Follow up	
	Screening	D-7	D0	D0+12hrs	D1 to D6	D7	D8 to D13	D14	D21	D28	TD	TD+12hrs	TD 24, 48,72, 96 hrs	D90	D180
Informed consent	x														
Consent quiz	x														
Revalidation of consent		x													
Obtain 24 hr contact details		x													
Medical history	x														
Interim medical history		x	x	x	x	x	x	x		x	x	x	x	x	x
Physical examination[a]	x										x				
Vital signs	x	x	x	x	x	x	x	x		x	x	x	x	x	x
Urine test for urinalysis	x														
Pregnancy test[b]	x	x	x					x			x				
Stool sample[c]			x	x	x	x	x	x		x	x	x	x	x	x
Blood sample	x		x	x	x	x	x	x		x	x	x	x	x	x
12-lead ECG	x														
Ultrasound scan	x														
Mood assessment	x	x				x		x			x				
Challenge			x												
Issue study pack			x												

Briefly, participants were fasted for 90 minutes before and after challenge. On the day of challenge participants were given 120mL sodium bicarbonate solution (1.75% wt/vol) to neutralise stomach acid. [485] One minute after this, participants were given 1-5 x 10^4 CFU of *S.* Typhi, comprising of equal parts *S.* Typhi Quailes strain and SB6000 suspended in sodium bicarbonate solution. Challenge doses were given within 3 hours of preparation. Participants were seen 8-12 hours after challenge administration and then daily thereafter for 14 days as per previous challenge studies. [485,494] Daily visits involved clinical observations and sampling.

Participants completed an electronic diary for 21 days after challenge involving twice daily temperature measurements and self-assessment of solicited symptoms. Participants had 24-hour phone access to a study clinician. Antibiotics (first line: ciprofloxacin 500mg twice daily for 7 days) were commenced at diagnosis or day 14 in those who remained undiagnosed.

The diagnostic criteria used were consistent with those in previous studies. [485] Briefly, blood cultures positive for *S.* Typhi or fever ≥38 °C persisting for >12 hours. Only BACTEC™ blood cultures were used to make clinical management decisions. Clinicians were not informed of blood quantitation results in real time. Severe typhoid fever was defined as recorded oral temperature of ≥40 °C, systolic blood pressure of <85 mmHg, significant lethargy or confusion, gastrointestinal bleeding or perforation, or any grade 4 laboratory abnormality.

10.2.2 Laboratory and other methods

10.2.2.1 Vi Antibody measurement

IgG responses to the Vi-antigen were measured using a commercial ELISA kit (VaccZyme, The Binding Site Ltd, Birmingham, UK) according to the manufacturer's instructions. The

lower limit of detection of this ELISA is 7.4 EU/ml. For those samples measuring below this level there were assigned a value of 3.7 EU/ml (half the lower limit of detection), any titre above 7.4 EU/ml was considered to be detected in keeping with previous studies. [496]

10.2.2.2 Challenge strains

The Quailes strain is a wild-type strain originally isolated from a chronic carrier in Baltimore MD, USA. SB6000 is a toxin-deficient strain, constructed from the Quailes strain by the Galan laboratory (Department of Microbial Pathogenesis, New Haven, Connecticut, USA) and characterised as previously described. [496]

10.2.2.3 Challenge Agent preparation

Inocula were freshly prepared on the morning of each challenge day. GMP Master stock vials of *S.* Typhi Quailes strain and *S.* Typhi SB6000 were thawed and suspended in NaHCO3 solution (1.75% wt/vol) at appropriate bacterial concentrations. The combination challenge agent was made by combining 15 ml of each diluted bacterial stock to make 30 ml challenge inoculum.

10.2.2.4 Calculation of Challenge agent dose

Confirmation of the bacterial content of master stocks and challenge dose was achieved by direct plating onto tryptone soya agar (Oxoid) for each bacterial strain. Plates were then incubated for 24 hours at 37 ˚C in 5% CO_2 and colonies were counted. The tolerance for accepting the ratio of strains was 43-57%. This figure was based on the data from the plates made from challenge agent preparation for determination of the challenge agent inoculum collected in the study 'Investigating Typhoid Fever Pathogenesis' (TYGER) (OVG2016/03). [496] The mean total colonies of the six plates were 825 with a standard deviation of 64. Assuming the two flasks containing only SB6000, or Quailes (wild-type) have equal number of bacteria, the chance of claiming deviation from a 1:1 ratio is 1% if +/- 7 % is used to

define balance of the two strains. This means that we accept the two strains are balanced in the flasks if we observed the ratio of colony counts of the two strains in the challenge confirmation plates between 43%-57%. Conversely, if the CFU counts in the SB6000 or Quailes flasks differ more greatly, defined as ≥ 60:40 or ≤ 40:60, the chance of correctly claiming deviation from a 1:1 ratio in the challenge inoculum is 91%.

10.2.2.5 Sampling

For the study endpoints at each visit after challenge for this study (until TD+12 hours or day 14, whichever was sooner) 10mls of whole blood was inoculated into:

- BACTEC™ aerobic blood culture bottle
- Lithium heparin for Ox bile culture
- WAMPOLE isolator 10 tubes for blood quantification

Other samples (such as routine haematology, biochemistry and additional BACTEC™ blood cultures) were taken at timepoints post diagnosis for safety monitoring. Additional samples (such as PBMCs) were taken for exploratory objectives which are not reported here.

Following a substantial amendment on 26th July 2019 (protocol version 4.0), blood volumes at diagnosis were doubled to 60ml and divided equally between these three specimen types to maximise sampling at predicted peak bacterial burden in the blood stream.

10.2.2.6 Automated blood cultures

Blood cultures were taken at least daily from volunteers and processed according to national standards. [502] For each culture 10 ml of fresh blood was inoculated into BACTEC™ Plus Aerobic vials (BD, New Jersey, USA) and incubated on a BACTEC™ FX System (BD, New Jersey, USA). *S.* Typhi growth and serotype were confirmed by standard microbiological methods [503]: slide agglutination according to the Kauffman-White classification and

biochemical testing (API20E, bioMerieux, Lyon, France). Isolates were transferred to agar slopes by taking two sweeps of the entire plate to capture a representative sample of colonies present on the plate.

10.2.2.7 Blood quantification

Quantitative blood cultures were performed at all visits from 12 hours post challenge until typhoid diagnosis plus 12 hours (TD +12 hrs) or day 14 in those remaining undiagnosed. Whole blood (10 ml) was inoculated directly into an ISOLATOR 10 tube (a commercial lysis-centrifugation system; Abbott Rapid Diagnostics, Waltham, USA). Samples were processed within four hours of collection. The sample was centrifuged for 30 minutes at 3000 G and the entire supernatant was then removed. The pellet is then vortexed to a homogenous suspension that was directly plated onto XLD agar (Oxoid, Waltham, USA). Plates were incubated at 37 ˚C for a minimum of 24 hrs in an aerobic incubator before colony counting. Quantity of organisms was calculated as total organism count/volume of blood (ml). Colonies were stored individually in Microbank tubes™ (see Isolate storage).

10.2.2.8 Ox bile culture

This method was adapted from the original by Zhou and Pollard.[504] Blood (10 ml) was taken from participants at all visits from 12 hours post challenge until TD +12 hrs or day 14. Culture tubes were prepared by aliquoting 20 ml TSB-3% Ox bile into 50 ml Falcon tubes under sterile conditions. On the day of blood collection 1.5 µl micrococcal nuclease (New England BioLabs, Ipswich, USA) and 5 ml blood sample were added to the culture tubes. Cultures were incubated for 24 hours (range 12-48 hours) in a shaking incubator at 37 ˚C. Following incubation, cultures were centrifuged for 20 minutes, and the supernatant was removed. The cell pellet was then re-suspended. Cell pellets were washed with 10 mM Tris-1m EDTA, this 'wash' was then added to the re-suspended pellet and centrifuged. Removal of the supernatant, the wash step and centrifugation were repeated. Pellets were stored at -80 °C until DNA extraction.

10.2.2.9 Stool culture

Stool samples were processed by OUH laboratories in accordance with standard methods.[505] Briefly, phosphate buffered saline (PBS) and selenite broth were inoculated with a sample of faecal material. XLD plates were spread with the inoculated PBS broth. Following incubation for 18-24 hours at 37 ˚C in O_2 the selenite broth was sub-cultured to *Salmonella* chromogenic agar plates which were incubated for 18-24 hours at 37 ˚C in O_2. Identification of colonies was performed as per standard methods.[503] Isolates were transferred to agar slopes by selecting individual *Salmonella* colonies to reduce the risk of collecting other organisms on the plate from a cultured stool sample.

10.2.2.10 Isolate storage

Isolates from culture-based methods were received on agar slopes and transferred for long term storage in Microbank™ tubes (Pro-Lab Diagnostics, Birkenhead, UK) for future DNA extraction. For agar slopes, a 10 µl loop was dipped into the Microbank™ tube and the loop

was the drawn across the full surface of the slope twice to collect as much material as possible.

Blood quantification colonies were stored in Microbank™ tubes directly. For single colonies, a 10 µl loop was used to collect material from a single colony taking care not to touch other colonies on the same plate.

For both types of isolate the content of the loop was then transferred to the Microbank™ tube, the tube sealed and inverted to emulsify the organism. Samples were left for 30 seconds to allow the organism to be absorbed onto the surface of the beads. The excess fluid was then removed to the level of the beads. Microbank™ tubes were stored at -80 ˚C.

10.2.2.11 DNA extraction for bacterial samples (from automated blood/stool culture and blood quantification)

All samples were cultured overnight in liquid lysogeny broth prior to extraction. All extractions were performed using 1-2 beads. For those samples where a single strain was detected the extraction process was repeated using 6-8 beads. DNA extraction was performed using the Wizard® Genomic DNA purification kit (Promega, Madison, USA) as per manufacturer's instructions. The blood culture isolates as the primary endpoint were extracted four times for each isolate.

10.2.2.12 DNA extraction from Ox bile culture

DNA was extracted from the pellet using the QIAamp® DNA blood mini kit (Qiagen, Hilden, Germany) as per the manufacturer's instructions.

10.2.2.13 Primer design, selection, and validation

The PCR primers for Salmonella Typhi Quailes strain (Q) and Salmonella Typhi SB6000 (S) were designed using PrimerBLAST[506] and Primer 3[507] according to the PltA and PltB

gene sequences. These were designed, developed, and validated by Marije Verheul, Jennifer Hill, and Jenny Clarke. Sequences are shown in Table 10.2.

Table 10.2: Primer sequences

Strain	Primer sequence	Primer name	Amplicon size	Tm (°C)	Ta (°C)
Quailes	ACCACCTTGCGTTTCCAGAT	5_for_inPltAB_Q	184	59.89	55
Quailes	TGGTGGAAGTAGTGACAGTCG	5_rev_inPltAB_Q		59.39	55
SB6000	GTGATCAACGTAACCGCCTTC	2_for_PltABcutsite_S	291	59.61	55
SB6000	ATGAGTAAGTATGTACCTGGAACT	2_rev_PltABcutsite_S		56.63	55

Primer testing was performed using Quailes and SB6000 templates to confirm the primers selectively amplify fragments of the Quailes and SB6000 genome. Confirmation of the sensitivity of the primers selected was performed at concentrations of 0.1-20 ng/µl. Confirmation that the primer sets detected both strains was subsequently performed on samples of combined Quailes and SB6000 DNA at ratios from 0.1 to 99.9% for each strain of a total concentration of 10 ng/µl.

Primers were synthesized by Thermo Fisher Scientific Inc. (Waltham, USA).

10.2.2.14 PCR procedure

Concentration of DNA in extraction samples was measured using a NanoDrop™ 2000/2000c Spectrophotometer. Ultrapure molecular grade water (Life Technologies Ltd, Carlsbad, USA) was used as blank according to manufacturer's instructions.

Samples were diluted to 10 ng/µl using Ultrapure nuclease-free water (Life Technologies Ltd, Carlsbad, USA). PCR reaction was carried out in a 50 µl volume, comprising 25 µl of OneTaq DNA polymerase (New England Biolabs, Ipswich, USA), 1 µl each of forward and reverse primers, 21 µl ultrapure nuclease free water (Life Technologies Ltd, Carlsbad, USA) and 10 µl DNA template. The following amplification steps were used: 1 cycle of 94 °C for 30 seconds, 30 cycles of 94 °C, 55 °C for 1 min, 68 °C for 1 min; and 1 cycle of 68 °C for 5 min.

The PCR amplification product (10 µl) was separated by electrophoresis at 90 V for 45 minutes on a 2% agarose gel and examined using a UV transilluminator (Syngene, Frederick, USA). A 50 bp DNA ladder (Life Technologies Ltd, Carlsbad, USA) was used as reference to determine the size of the PCR amplicons.

10.2.2.15 Algorithm for assigning results to a time point

Due to the multiple steps involved with each sample type and the risk of DNA contamination during processing rules for designating results for a given time point were applied:

- BACTEC™ culture isolates from which at least one strain should be detected: for a strain to be considered detected it must be present from samples from at least two different extractions, for a strain to be considered not detected it must be absent in samples from all four extractions.
- Quantification colonies from which only one strain should be detected: if both strains were detected on a PCR and clear evidence of both from other quantification colonies, BACTEC™ or Ox bile culture methods then it was assumed that contamination had occurred at the colony selection and bead banking stage and no further testing was performed. If both strains were detected and no evidence of both strains from other methods, PCR testing is repeated.

- Culture PCR pellet from which neither, either or both strains may be detected. These were performed twice; if results were incongruent then PCR was repeated.

For a given study time point an overall time point decision was made as to whether the composite of the results showed no strain, a single strain or both strains as detected. Timepoints were designated as mixed if there was evidence from any of the 3 assays that more than one strain was present.

10.2.3 Data capture

Clinical variables and electronic diary data were recorded using OpenClinica™ (OpenClinica, LLC) and Redcap (Vanderbilt University) research database applications.

10.2.4 Statistical analysis

Statistical analyses were performed using Stata 17 software. [508]

10.2.5 Approvals

The OVG2018/06 study was sponsored by the University of Oxford (Clinical Trials & Research Governance). Ethical approvals for the primary protocol, and any study amendments were obtained from the Hampshire A Research Ethics Committee (19/SC/0024). Approvals for deliberate release of the genetically modified strain of *S.* Typhi were obtained from the Department for Environment, Food and Rural Affairs (Reference: R18/48/01). The study was registered with clinicaltrials.gov (NCT03889067) and conducted in accordance with the principles of the Declaration of Helsinki (2013) and Good Clinical Practice Guidelines.

Both challenge strains were manufactured to a GMP standard at the Walter Reed Army Institute of Research (Silver Spring, USA) and stored as a frozen suspension in soya tryptone medium containing 10% sucrose at -80 °C before use.

10.2.6 Funding

This work was funded by Professor Richard Moxon's Emeritus fund held in the Department of Paediatrics, University of Oxford.

10.3 Results ("Bottlenecks")

10.3.1 Recruitment & Participants

Participants were enrolled between May 2019 and January 2020. Eleven participants were enrolled and ten were challenged with a combined inoculum of WT (Quailes) and toxin-deficient *S.* Typhi (SB6000). Fewer than half of the participants who underwent in person screening were eligible to take part (Figure 10.5). Demographics of challenged participants are shown in Table 10.3.

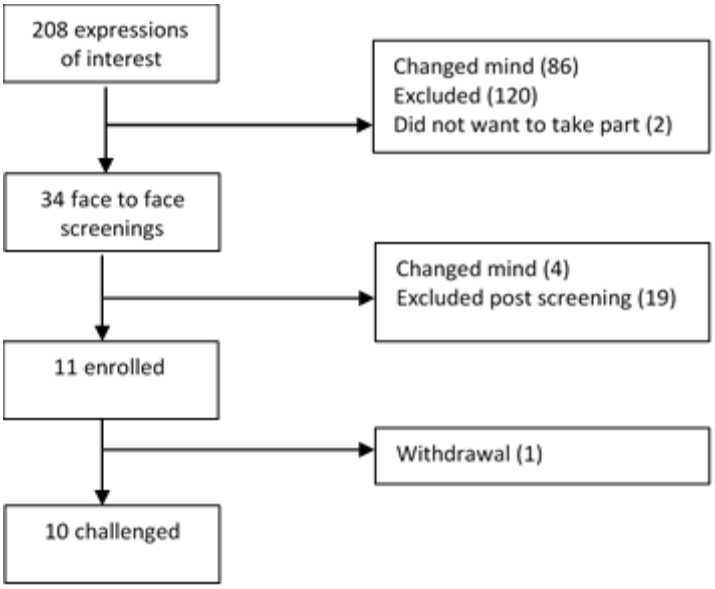


Figure 10.5: Recruitment flow chart for Bottlenecks study

Characteristic	N=10
Male, n (%)	6 (60)
Age range, y, mean±SD (range)	29.8±7.6 (19-45)
BMI, mean±SD (range)	24.8±3.2 (20.0-31.7)
Ethnicity, white British, n (%)	8 (80)
White (other), n (%)	2 (20)
Baseline detectable Vi IgG titre (>7.4 EU/mL), n (%)	5 (50)

Table 10.3: Challenge participant demographics

10.3.2 Challenge dose

All participants received challenge agents at a total dose within the target range of 1-5 x 10^4 CFU. All four batches of challenge met the predefined criteria for acceptable ratio of the two strains. There was no significant difference in the challenge dose administered in those that were diagnosed with typhoid compared with those that remains undiagnosed (Mann-Whitney U test: z= -0.557, p=0.5775).

The challenge dose and ratio of strains are shown in Table 10.4.

Table 10.4: Challenge dose and ratio of strains within challenge inoculum for each participant

Participant(s)	Challenge total dose (CFU)	Dose Quailes (CFU)	Dose SB6000 (CFU)	Percentage Q	Percentage S
9028	2.64 x 10^4	1.33 x 10^4	1.31 x 10^4	50.3	49.7
9025, 9009, 9006, 9007	2.47 x 10^4	1.15 x 10^4	1.33 x 10^4	46	54
9023	2.32 x 10^4	1.28 x 10^4	1.05 x 10^4	55	45
9013, 9002, 9014, 9022	2.01 x 10^4	1.03 x 10^4	9.8 x 10^3	51.3	48.7

10.3.3 Diagnosis of typhoid fever

Of the 10 participants who were challenged, 5 were diagnosed (50%). There was no significant difference in detectable versus undetectable Vi IgG titres between those who were diagnosed and those who remained undiagnosed (p=0.5, Fisher's exact test).

Participants were diagnosed a median of 134.64 hours (IQR: 120-168) after challenge, in keeping with previous challenge studies. Figure 10.6 shows the positive culture samples received from each participant. Two participants, 9023 and 9014, returned no positive cultures from blood or stool during the trial. All other participants returned at least two positive cultures from blood or stool. All diagnosed participants had two or more positive blood cultures and at least one positive stool culture, which preceded or coincided with their typhoid diagnosis.

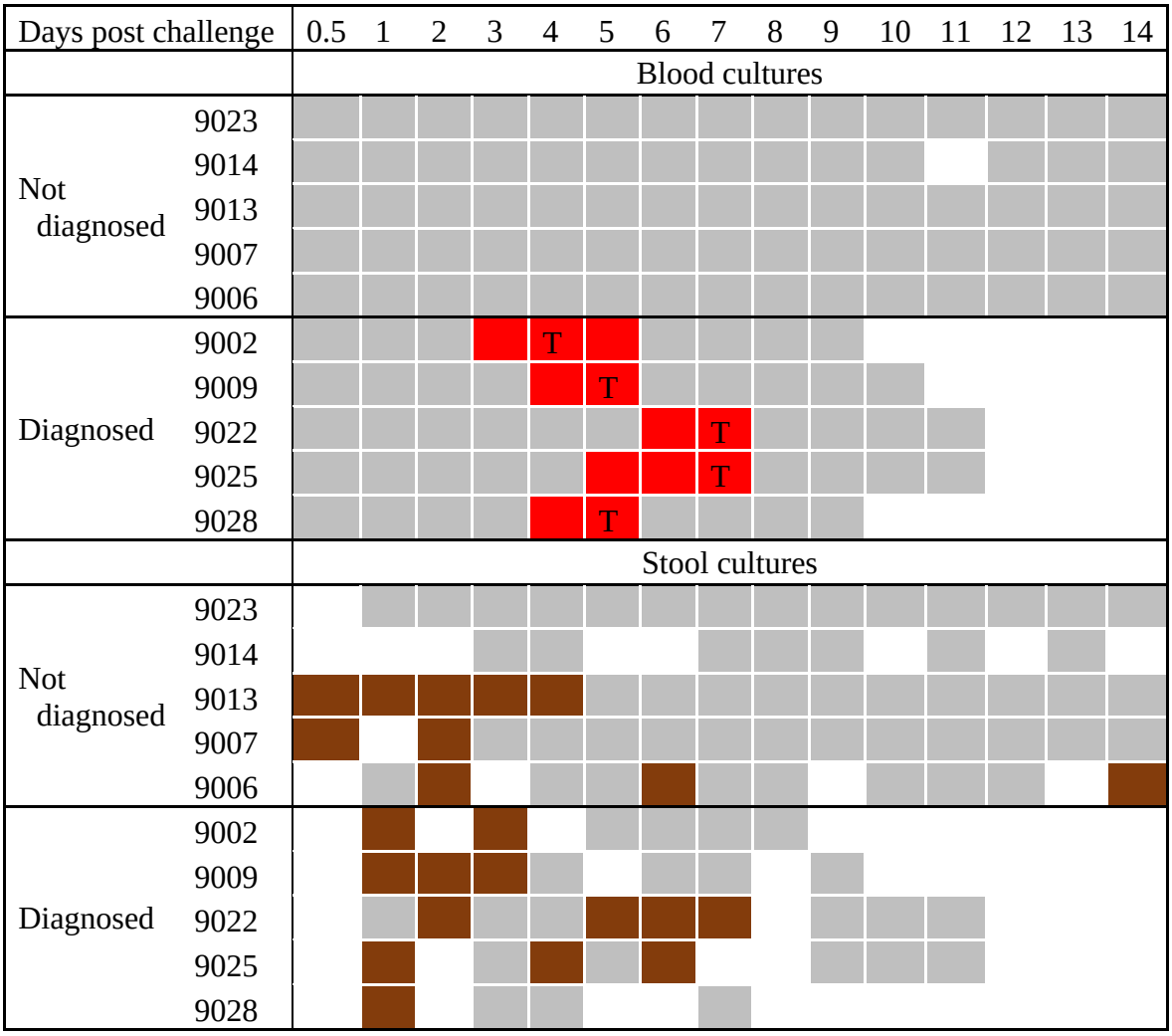


Figure 10.6: Blood and stool culture results Where multiple samples were taken on a single day any positive culture is represented even if negative cultures were also taken that day. T: treatment started. Red squares: positive blood culture, brown squares: positive stool culture, white squares: no sample, grey squares: negative culture result (blood or stool)

10.3.4 Safety

There were three adverse events of special interest in two participants. Two participants met the criteria for severe typhoid fever. Participant 9009 was diagnosed on day 5 post challenge and had temperatures >40 °C on day 6 and a grade 4 laboratory abnormality on day 7 (blood potassium result of 3.0). Participant 9002 was diagnosed on day 3 and had a grade 4 laboratory abnormality (CRP of 206.7) on day 4. Additionally participant 9002 had a positive clearance stool following the antibiotic course and required a second course of antibiotic treatment.

10.3.5 PCR results from diagnosed participants

A total of 44 BACTEC™ blood cultures from diagnosed participants were taken, 18/44 (41%) of which were positive. Of the Ox bile cultures, 11/41 (27%) samples taken were positive. Ox bile cultures were negative from participants who were not diagnosed. There were 42 samples for quantitation taken, of which 11/42 (26%) detected at least one colony of *S.* Typhi. From those 11 positive quantitative samples, there were a total of 105 colonies. There were no quantitative samples positive for *S.* Typhi colonies in non-diagnosed participants.

All participants had evidence of both SB6000 and Quailes strains in their blood samples on at least one timepoint (Figure 10.7) and for four out of five participants both strains were detected at the diagnosis timepoint when bacterial burden in the bloodstream was shown to be the highest.

All diagnosed participants had at least one positive quantitative culture, with total bacterial burden ranging from 0.5 – 2.65 CFU/ml at diagnosis (four participants, blood quantitation was negative at diagnosis for participant 9025). All diagnosed participants had at least one positive Ox bile culture however no early bacteraemias within the first 48 hours

were detected. There were eight positive Ox bile cultures across all participants; at these timepoints, the BACTEC™ blood cultures taken at the same time were also positive. In contrast, there were 14 positive BACTEC™ blood cultures, 43% of which had a negative Ox bile culture taken at the same time.

9002	D1	D2	D3	TD	TD+12	TD+24
Blood culture						
Blood culture 2						
Blood quantification						
Blood quantification 2						
Oxbile culture						
Oxbile culture 2						
Blood overall result						

9009	D3	D4	D5	TD	TD+12
Blood culture					
Blood culture 2					
Blood quantification					
Blood quantification 2					
Oxbile culture					
Oxbile culture 2					
Blood overall result					

9022	D5	D6	D7	TD	TD+12
Blood culture					
Blood culture 2					
Blood quantification					
Blood quantification 2					
Oxbile culture					
Oxbile culture 2					
Blood overall result					

9025	D4	D5	D6	TD	TD+12
Blood culture					
Blood culture 2					
Blood quantification					
Oxbile culture					
Blood overall result					

9028	D2	D3	D4	TD	TD+12
Blood culture					
Blood quantification					
Oxbile culture					
Blood overall result					

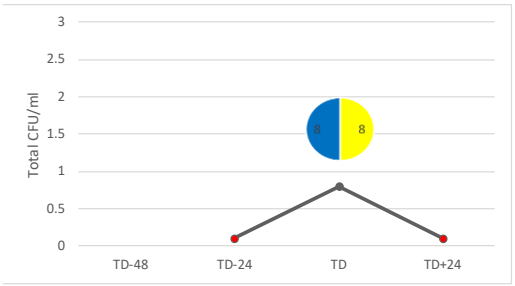
Both strains
SB6000
Quailes

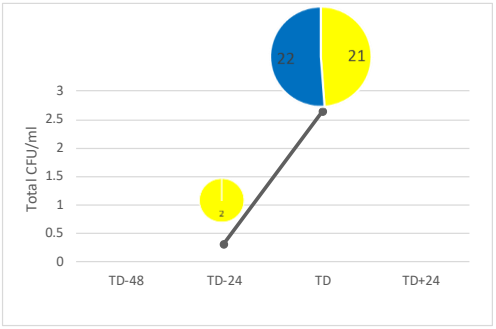
Figure 10.7: PCR results from blood samples for diagnosed participants. White squares: no sample, grey squares: negative result. TD = typhoid diagnosis visit, TD+12 = 12 hours after typhoid diagnosis visit, D3 = Day 3. The typhoid diagnosis timepoint is aligned in this figure rather than day post challenge. Participant 9028 was the first participant so only had a single sample of each type taken at diagnosis; an amendment was instituted after this participant to double blood sampling at diagnosis. Participant 9025 had only 3 of 6 samples taken at diagnosis due to difficulties with blood draw and no blood sampling at TD+12 due to anaemia.

10.3.6 Blood Quantitation results

The quantitation results from participants are shown in Figure 10.8. Participants had lower levels of bacteria in the samples taken prior to diagnosis, with a variable trajectory to peak levels which were seen in all but one participant at the point of diagnosis. Measured and estimated quantitation revealed CFU counts in keeping with previous reports of one CFU/ml in the blood. Only one participant (9002) had a positive culture result after starting treatment.

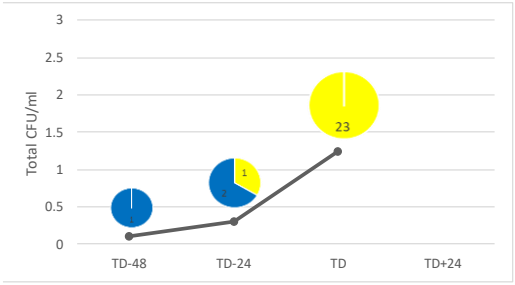
Despite all participants having samples for quantitation taken at all timepoints preceding diagnosis, only two participants had more than one positive quantification (9009, 9022). Samples from these two participants showed an increase of colonies isolated to a maximum at the point of diagnosis. For participants 9002, 9009 and 9028 approximately equal ratios of Quailes and SB6000 colonies were detected from quantitation plates at diagnosis. In participant 9022, there was a predominance of Quailes in the blood quantitation but both strains were detected in the BACTEC and Ox bile culture. Participant 9025 did not have a positive blood quantitation plate at the point of diagnosis and had only a single colony from the sample taken two days prior to diagnosis.

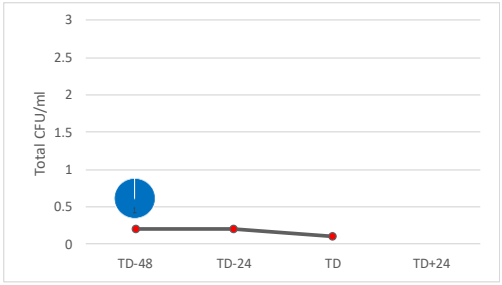




9002	TD-48	TD-24	TD	TD+24
S			8	
Q			8	
Total colonies			16	
Total CFU/ml		0.1	0.8	0.1
Overall result				

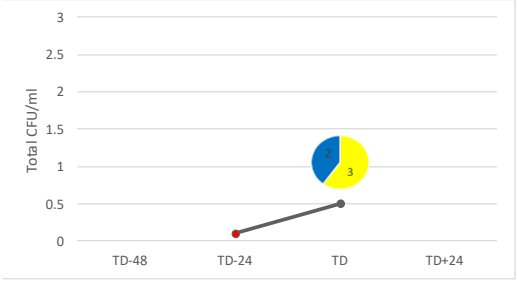
9009	TD-48	TD-24	TD	TD+24
S			22	
Q		2	21	
Total colonies		2	53	
Total CFU/ml		0.3	2.65	
Overall result				





9022	TD-48	TD-24	TD	TD+24
S	1	2		
Q		1	23	
Total colonies	1	4	23	
Total CFU/ml	0.1	0.3	1.25	
Overall result				

9025	TD-48	TD-24	TD	TD+24
S	1			
Q				
Total	1			
Total CFU/ml	0.2	0.2	0.1	
Overall result				



9028	TD-48	TD-24	TD	TD+24
S			2	
Q			3	
Total colonies			5	
Total CFU/ml		0.1	0.5	
Overall result				

Both strains	
SB6000	
Quailes	

Figure 10.8: Bacterial dynamics over time per participant. Quantification data for blood samples given in data table and plotted as line chart. Points with black marker plotted raw result from quantitative samples; those estimated by calculation are given in red. Where blood culture or Ox bile culture was positive in the absence of quantitative data a conservative estimate of 0.1 CFU/ml is used on the assumption that there was a minimum of a single organism in a 10 ml sample to yield a positive result. Pie charts over time point plots show proportion of each strain detected with raw colony numbers given. Where the number of colonies exceeds those for which individual results are available, this is because either a colony had both strains detected at PCR and thus were assumed to have been contaminated on picking, were within sites of contamination on the plate or were unavailable for PCR. Where these events occurred, these were deemed to be random and not systematic.

10.3.7 Stool PCR results

All five diagnosed participants had positive stool cultures (Figure 10.9). Three of whom showed shedding of both strains simultaneously whilst two of whom had only single strains detected in their stool cultures. Three out of the five participants who were not diagnosed during the study period had positive stool cultures as shown in Figure 10.9. From all stool cultures, 38% (8/23) showed both strains detected. Four of the eight participants who had positive stool cultures had a mixed culture in the first 48 hours after challenge. After day 4 only a single strain was detected from positive stool cultures.

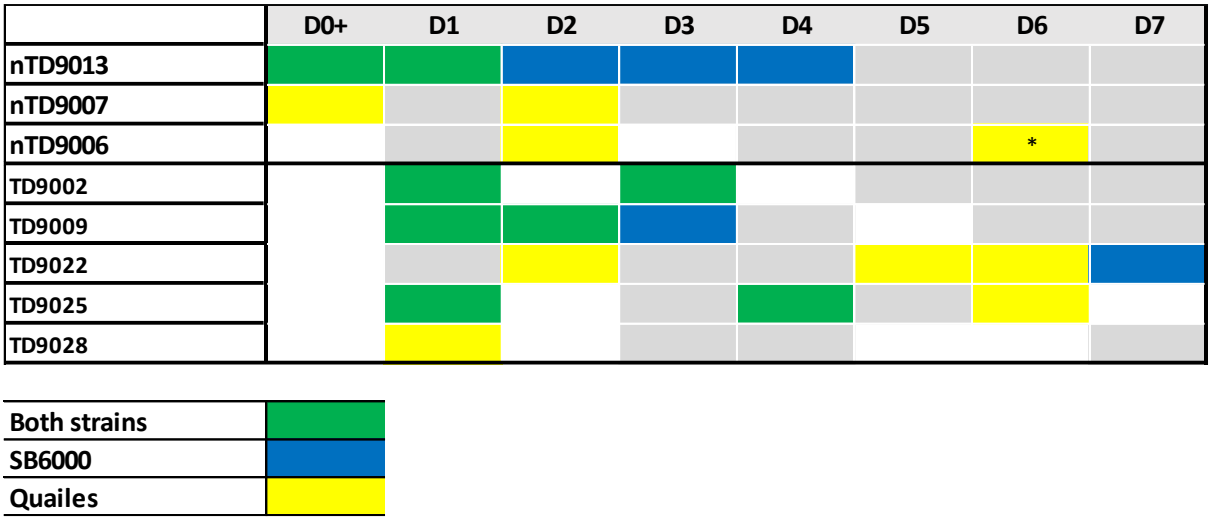


Figure 10.9: Stool culture results Participant 9006 had 3 positive stool cultures. *The day 6 sample Quailes result from culturing single beads could not be confirmed by culture of multiple beads (neither strain was detected) and day 14 sample was not conclusive from single or multiple bead culture followed by PCR and is not shown in this figure. There were no other positive stool cultures after day 7.

10.4 Discussion ("Bottlenecks")

The results of this pilot study show that when a small sample size of individuals was given an equal combination of wild-type and toxin-deficient strains of *S*. Typhi at a total dose of 1-5 x 10^4 CFU, DNA from both strains can be detected from the blood at typhoid diagnosis.

These results are in keeping with the null hypothesis that there is no evidence of a single cell bottleneck.

This pilot study highlights the many challenges of studying bacterial population dynamics within a human host. There is a considerable body of literature from both Gram-positive and Gram-negative bacterial species within animal models that suggest bottlenecks of a bacterial population within a host may occur; furthermore, within-host studies of human viral infection indicate that transmission population bottlenecks may be a feature across a range of pathogens.

Whilst all participants who became bacteraemic had both strains present at or before diagnosis, timepoints where only one strain was detected featured in all participants. There are several possibilities as to why this may have occurred. Firstly, these single strain results could be real and only one strain was present in the blood stream at the time the sample was taken. However, a more likely explanation is that at low concentrations, due to the stochastic nature of sampling it is plausible and likely that only a single bacteria may be present in a sample and therefore a single strain could only ever be detected. An example of this is the day 6 result for participant 9025 where SB6000 was detected from the BACTEC™ culture and the Quailes strain was detected from the Ox bile culture. This may appear to be a discordant and potentially incorrect result but where their blood quantitation at the same timepoint was negative their bacterial burden was clearly low and hence these seemingly conflicting results are biologically plausible. The calculations in Appendix 4.2 illustrate how changes in two of the simplest parameters, namely blood volume and bacterial burden in the bloodstream, can impact the ability to detect both strains. In the context of a low level (0.2 CFU/ml) bacteraemia for participant 9025 throughout their infection it is credible that both were present but only one was detected.

As the bacterial burden typically increases to a maximum at diagnosis it becomes more likely that, if they are present, both strains will be detected, due to the number of bacteria in each blood sample. I attempted to address this as a parameter that could be manipulated, unlike bacterial burden, amending the protocol to double the blood volume draw at the point of diagnosis, after the first participant had been challenged. This was done to increase the likelihood of having sufficient blood volume to deliver a robust result. Unlike animal models where a mouse may have all or a significant proportion of its blood volume sampled at euthanasia, constraints on blood volume sampling in humans make the study of a low-level bacteraemia difficult to study. Lastly, there were more instances where SB6000 was detected as a single strain compared with the Quailes strain. This might suggest a growth advantage by this strain or a difference in sensitivity of the primers. Primer sensitivity testing of strains in isolation and in combination does not suggest this is an issue and *in vitro* growth characteristics of both strains in LB broth have been noted to be comparable. [496,509] However, there are no data on whether there is any competitive growth advantage of either strain *within* host or when sampled simultaneously. The biology and role of the typhoid toxin, particularly within the human host is still unclear and it may be that the toxin-negative strain may behave differently within host (such as preferred location of infection) that may not influence disease phenotype observed in the TYGER study[496] but may influence an experiment such as this.

Why did this experiment fail to reveal a population bottleneck in human *S.* Typhi infection? It is of course reasonable, having acknowledged the methodological limitations of the study, to consider that there may be no population bottleneck to be found despite data from other species and organisms suggestive of one. It may also be that the nature of any population bottleneck may be wider than a single organism or localised to a particular physiological location or indeed be dynamic and influenced by a wide range of factors such dose and host susceptibility. The high-level, whole host co-infection methods used in this

pilot study did not have the resolution to address these factors further. We cannot rule out that each strain detected did not originate from singly infected cells or organs within a host.

The dose employed for this challenge study may have critically influenced whether a bottleneck was detected. Animal co-infection models showed that dose was critical in revealing a bottleneck effect, [451,454,455] and it may be that the dose in this study would only ever result in mixed infections. As discussed, the dose in this CHIM is specifically chosen with an alternative aim in mind: to infect a sufficient number of participants to keep sample sizes to a minimum, not overwhelm host defences and allow the testing of interventions such as vaccines. This dose may differ from a potentially lower dose encountered in natural infection. Indeed, much consideration was given as to whether the dose should be dropped by a log to the lower level used in the initial CHIM dose finding study. However, this was likely to result in fewer participants becoming infected and an even lower bacterial count in the blood thus further reducing confidence in any observation of a single strain bacteraemia. [485]

Measured and estimated quantitation revealed CFU counts in keeping with other attempts to quantitate *S*. Typhi in the blood. [497] The quantitative testing in this study showed that the bacterial burden increases over time to a peak at a variable trajectory. Antibiotic treatment terminated bacterial detection in the blood by any means in all but one participant. It is noted that for three participants (9002, 9009, 9028) the ratio of strains from the quantitative plates approximated the ratio in the inoculum which may suggest that the infection of reservoir sites and replication of strains within these sites during the pre-symptomatic phase occurred in equal ratios. The finding that for participant 9022 all colonies identified on the quantitation plates were Quailes whilst the culture-based methods revealed both strains to be present is the most compelling evidence that the ratio of strains may not

have been 50:50 in this individual at this timepoint and suggests that at some point in the infective process there may have been a relative bottleneck.

There are limitations to this study. The number of participants is small, and this limits the interpretation of findings. The study was stopped after 10 participants (originally planned for up to 15 participants) due to the COVID-19 pandemic. Additionally, the attack rate was lower than anticipated at 50%, producing fewer diagnosed participants and samples to study. The lower attack rate may be due to chance or variations in host susceptibility not captured by the screening process. There was however no difference in the Vi titre between those who were diagnosed and those who remained well. Blood samples were divided between three different sample types: BACTEC, quantification plates and Ox bile culture. This was a balance between sensitivity and quantitative versus qualitative results. The Ox bile culture was included specifically for detection of primary bacteraemias. However, none were detected, in contrast to the findings of a previous study by the group. [483] The 'optimal' sample for attempting to ascertain the ratio of strains in the blood without an intervening culture step that may favour a given strain were the quantitation plates. However, the number of samples that were BACTEC™ positive but failed to yield colonies on the quantitation step shows that this method is not as sensitive, presumably as bacteria perish prior to plating.

The very nature of the procedural steps for this study produces bottlenecks. At each stage of sample processing only a sample of the bacterial population is taken. To try and address this, sweeps across plates and slopes were taken to maximise the retrieval of as many organisms of the population as possible. Additionally, where any single strains were detected from PCR of the DNA extracted from microbank beads, these were repeated using 6-8 beads to ensure this was not a source of a sampling error. Furthermore, there are culture steps which may further distort results as this model does not take into account cell death and replication.

The stool results show a variety of single strain and combination strain results. The passage of the inoculum through the GI tract and direct excretion within the first few days after challenge may explain both strains detected in early samples from some participants. Attempted quantitation of *S.* Typhi in stool has shown a wide range of bacterial burden in this heterogeneous sample type. [510] Similarly to blood it is possible that when a single strain is detected this is a result of sampling, as only a small amount of stool is used for culture purposes. Isolation rates can be improved by using a larger volume, but this was beyond the scope of this study. [510]

Hypothetically a more ideal experiment to investigate population dynamics at whole host level in humans using a *S.* Typhi challenge model may be by using a collection of barcoded wild-type Quailes strains, at a range of dose levels, with a focus on taking maximal blood sample volumes for a highly sensitive, single quantitative identification method. For example, just by administering a larger number of strains it may be possible to detect if there is a quantitative loss of strains from gut to bloodstream. Single cell work looking at infection of human phagocytes from a challenge model using DNA probes or fluorescent techniques may also be able to give further resolution at a cellular level in humans. These additional experiments would not, however address the fundamental issue that bloodstream bacterial burden in typhoid is low, the key factor hindering diagnostic techniques for *S.* Typhi in the clinical setting. The CHIM model of *S.* Typhi will always be constrained by this and makes it hard to study dynamics using co-infection in this model. Low bacterial burden regardless of the number of strains used will always limit the confidence in findings that suggest a bottleneck. This could be improved by sampling greater blood volumes at the time of diagnosis or serial sampling during a prodromal phase where symptoms are present but formal diagnosis has not been made and antimicrobials have not been commenced. This would require minimising at other time points and would need to be assessed for safety implications. Alternatively, bone marrow sampling at diagnosis would be more sensitive

though this poses significant logistical difficulties even if it proved acceptable to participants.

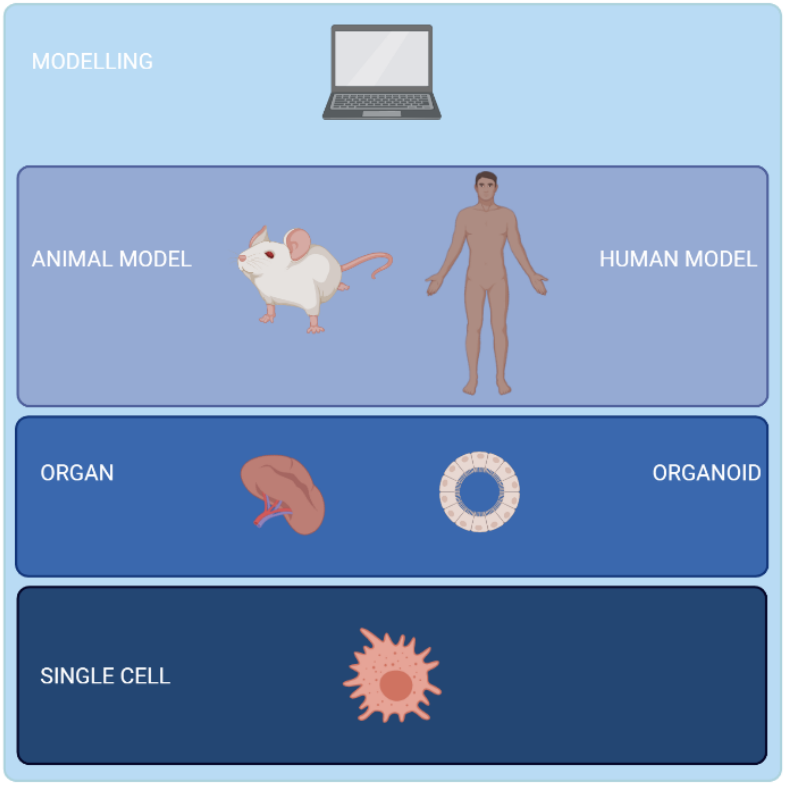


Figure 10.10: Sources of experimental and theoretical information from different within-host levels

The combination of experimental data derived from a relatively small number of human participants in combination with mathematical modelling may be the most practical way currently to further investigate some of these within host mysteries. Indeed, CHIM data and animal data have already been used to estimate additional parameters highlighting their use in hard to study processes. [448,511] The 17 parameters present in Awofisayo-Okuyelu et al's model of the incubation period of *S.* Typhi demonstrate the complexity of within-host modelling. Further work on the derivation of a mechanistic mathematical model to attempt to describe some of the parameters may help guide experimental work, this may be particularly relevant for pathogen models such as S. Typhi which are highly influenced by stochasticity. Such models may also allow the derivation of data which cannot be measured within a human host. [512,513] Figure 10.10 shows the potential different sources of information,

at different levels of resolution that could be used to provide data for models, and how modelling data could help better define experimental designs.

The potential for rich data provided by, and the caveats of human CHIM studies complement recent calls for the integration of modelling into routine observational work. [512] The bridge that remains to be forged appears to be between those able to study spatiotemporal dynamics in animals and those attempting to do the same using different techniques in humans.

FURTHER WORK

1. Single cell flow cytometry to assess whether circulating phagocytes are infected with single or multiple strains.
2. Organoid work: human-induced pluripotent stem cell (hiPSC)-derived organoids allow the modelling of *S.* Typhi epithelial invasion *in vitro*. [514] Combined with fluorescent tagging and microscopy exploration of whether both strains invade within a human organoid model may produce interesting insights into a potential bottleneck at the gastrointestinal mucosa.
3. Whilst this would take significant resources and commitment, the incorporation of genetic barcoding into challenge model strains in multiple pathogen models would provide a substantial body of information that could accrue alongside the other outputs of controlled human infection models.
4. Modelling work building on existing and future data drawing these different levels of information together to optimise experimental design, samples size and sampling strategies.

Appendix 4.1 Full Inclusion & Exclusion criteria for the 'Bottlenecks' Study

10.4.1 Inclusion Criteria

Participants must satisfy all the following criteria to be considered eligible for the study:

- Agree to give informed consent for participation in the study.
- Aged between 18 and 60 years inclusive at time of challenge.
- In good health as determined by medical history, physical examination, and clinical judgment of the study team.
- Agree (in the study team's opinion) to comply with all study requirements, including capacity to adhere to good personal hygiene and infection control precautions.
- Agree to allow his or her General Practitioner (and/or Consultant if appropriate), to be notified of participation in the study.
- Agree to allow study staff to contact his or her GP to access the participant's medical history and vaccination records.
- Agree to allow Public Health England to be informed of their participation in the study.
- Agree to give his or her close contacts written information informing them of the participant's involvement in the study and offer them voluntary screening for *S.* Typhi carriage.
- Agree to have 24-hour contact with study staff during the four weeks post challenge and are able to ensure that they are contactable by mobile phone for the duration of the challenge period until antibiotic completion.
- Agree to allow the study team to hold the name and 24-hour contact number of a close friend, relative or housemate who will be kept informed of the study

participant's whereabouts for the duration of the challenge period (from the time of challenge until completion of antibiotic course). This person will be contacted if study staff are unable to contact the participant.

- Have internet access on a daily basis to allow completion of the e-diary and real-time safety monitoring.
- Agree to avoid antipyretic/anti-inflammatory treatment from the time of challenge (Day 0) until advised by a study doctor or until 14 days after challenge.
- Agree to refrain from donating blood for the duration of the study.
- Agree to provide their National Insurance/Passport number for the purposes of TOPS registration and for payment of reimbursement expenses.
- Agree not to take part in another research study involving an IMP or challenge for the entire period of recruitment in the current study.

10.4.2 Exclusion Criteria

The participant will not be enrolled if any of the following apply:

- History of significant organ/system disease that could interfere with trial conduct or completion. Including, for example, but not restricted to:
    - Cardiovascular disease
    - Respiratory disease
    - Haematological disease[a]
    - Endocrine disorders

[a] This includes anaemia. The acceptable lower limits for [haemoglobin] are 125 g/L for female participants and 135 g/L for male participants (Guidelines for the Blood Transfusion Services in the UK, 8th edition, 2018 https://www.transfusionguidelines.org/red-book (Accessed 5th December 2018).

- Renal or bladder disease, including history of renal calculi.
- Biliary tract disease, including biliary colic, asymptomatic gallstones, or previous cholecystectomy.
- Gastro-intestinal disease including requirement for antacids, H_2-receptor antagonists, proton pump inhibitors or laxatives.
- Neurological disease
- Metabolic disease
- Autoimmune disease
- Psychiatric illness requiring hospitalisation or known or suspected current drug and/or alcohol misuse (alcohol misuse defined as an intake exceeding 42 units per week)
- Infectious disease

- Have any known or suspected impairment of immune function, alteration of immune function, or prior immune exposure that may alter immune function to typhoid resulting from, for example:
  - Congenital or acquired immunodeficiency, including IgA deficiency.
  - Human Immunodeficiency Virus infection or symptoms/signs suggestive of an HIV-associated condition
  - Receipt of immunosuppressive therapy such as anti-cancer chemotherapy or radiation therapy within the preceding 12 months or long-term systemic corticosteroid therapy
  - Receipt of immunoglobulin or any blood product transfusion within 3 months of study start.
  - History of cancer (except squamous cell or basal cell carcinoma of the skin and cervical carcinoma in situ).
- Screening positive for HLA B27 gene.

- Moderate or severe depression or anxiety as classified by the Hospital Anxiety and Depression Score (score 11 or higher on either score) at screening or before challenge that is deemed clinically significant by the study doctors[b].
- Weight less than 50kg[c].
- Presence of implants or prosthetic material.
- Anyone taking long-term medication (e.g. analgesia, anti-inflammatories, or antibiotics) that may affect symptom reporting or interpretation of the study results.
- Contraindication to ciprofloxacin or macrolide antibiotics.
- Female participants who are pregnant, lactating or who are unwilling to ensure that they or their partner use effective contraception[d] 30 days prior to enrolment/challenge and continue to do so until three negative stool samples, the first taken a minimum of 1 week after completion of antibiotic treatment and subsequent samples at least 48 hours apart each have been obtained.
- Occupational (unless willing and able to confirm that they are not working in the following environment for the period of challenge to clearance confirmation):

- Full-time, part-time, or voluntary occupations involving:

---

[b] If elevated scores are due to temporary significant life events, the questionnaire may be repeated after resolution of the event with a view to inclusion if normal.

[c] Or a Body Mass Index (BMI) that, in the opinion of the study doctors, may adversely impair the interpretation of the study results or affect the safe performance of any study procedures.

[d] As defined in sections 4.1 and 4.2 of http://www.hma.eu/fileadmin/dateien/Human_Medicines/01-About_HMA/Working_Groups/CTFG/2014_09_HMA_CTFG_Contraception.pdf (accessed 5th October 2018)

- Clinical or social work with direct contact with young children (defined as those attending pre-school groups or nursery or aged under 2 years), or
- Clinical or social work with direct contact with highly susceptible patients or persons in whom typhoid infection would have particularly serious consequences (unless willing to avoid work for at least 6 weeks until demonstrated not to be infected with *S.* Typhi in accordance with guidance from Public Health England and willing to allow study staff to inform their employer).

- Full time, part time or voluntary occupations involving:

- Commercial food handling (involving preparing or serving unwrapped foods not subjected to further heating)

- Close household contact with:
  - Young children (defined as those attending pre-school groups, nursery or those aged less than 2 years)
  - Individual(s) who is (are) immunocompromised.
- Scheduled elective surgery (including dental work) or other procedures requiring general anaesthesia during the study period.
- Participants who have participated in another research study involving an investigational product that might affect risk of typhoid infection or compromise the integrity of the study within the 30 days prior to enrolment (e.g. significant volumes of blood already taken in previous study)[e].

[e] As assessed by both participant questioning and registration with The Over Volunteering Prevention System (TOPS) database.

- Detection of any abnormal results from screening investigations (at the clinical discretion of the study team).
- Inability to comply with any of the study requirements (at the discretion of the study staff and the participant's General Practitioner).
- Any other social, psychological or health issues which, in the opinion of the study staff, may
    - Put the participant or their contacts at risk because of participation in the study,
    - Adversely affect the interpretation of the primary endpoint data,
    - Impair the participant's ability to participate in the study.
- Having previously received any typhoid vaccine within the last 3 years.
- Having been resident in an enteric fever endemic country for 6 months or more[f].
- Have previously been diagnosed with laboratory-confirmed typhoid or paratyphoid infection or been given a diagnosis compatible with enteric fever.
- Have participated in previous typhoid or paratyphoid challenge studies (with ingestion of challenge agent).
- Have a prolonged corrected QT interval (>450 milliseconds) on ECG screening.

[f] As defined by NaTHNaC Travel Health Pro (https://travelhealthpro.org.uk/): any country where typhoid vaccination is recommended before visiting.

Appendix 4.2 Probability calculations for detection of strains

To illustrate one aspect of the mathematical aspect of discerning two strains from this human challenge, the following calculations were performed. A model of human typhoid infection is naturally highly complex and relies on multiple factors and assumptions. [511] It is acknowledged that these calculations are necessarily reductionist to answer what the probability of detecting both strains in a given context is when both are present. They focus on 3 main parameters: the concentration of bacteria in the bloodstream, the volume of blood sampled and the relative proportion of two strains.

Table 10.5 illustrates the simple relationship between the volume of blood taken in a blood sample and the number of bacteria in a given sample depending on the concentration.

Table 10.5: Relationship between number of bacteria in the bloodstream and blood sample volume

	Concentration of bacteria CFU/ml					
	0.1	0.5	1	5	10	50
Blood volume/ml						
5	0.5	2.5	5	25	50	250
10	1	5	10	50	100	500
20	2	10	20	200	200	1000
50	5	25	50	500	5000	2500
100	10	50	100	1000	10000	5000

Experiments in humans are mostly performed in the ranges in the highlighted cells.

Assumptions for the following calculations:

- Concentration of bacteria per ml of whole blood is 1 CFU/ml unless otherwise stated.
- Quailes and SB6000 are in the proportion stated in the peripheral bloodstream.
- Both strains are spread equally throughout the bloodstream
- The detection method is perfectly sensitive and specific.

- Both strains are detected with equal probability
- Assuming non-replacement on selection
- Bacterial replication and death are not included in calculations.
- Samples taken at random from bloodstream.

Equal proportions of two strains: what is the probability of detecting both strains?

Assuming the probability of detecting a Quailes colony in a 10 ml blood sample is p=0.5 and the probability of detecting a SB6000 colony in a blood sample is p=0.5.

$$\frac{C_{10}^{10}}{2^{10}} = \frac{1}{1024} = 0.0009765 \approx 0.001$$

Approximately 1 in 1000 10 ml samples would show all 10 colonies of a particular single strain (when p=0.5 of each).

To calculate this for two strains therefore, this is halved to 1 in 500 to be either all Quailes or all SB6000.

Probability of having both strains in a 10ml blood sample of 10 colonies:

$$1 - 0.0009765 - 0.0009765 = 0.998047$$

Therefore, we can see that it is highly likely that if both strains are present in equal proportions, concentration of bacteria in the bloodstream is physiological and 10 ml of blood is drawn. There is however a small chance that even if these conditions are indeed true the result may falsely show that there is a single variant.

Equal proportions of two strains: impact of blood volume

Assuming the probability of detecting a Quailes colony in a 20 ml blood sample is p=0.5 and the probability of detecting a SB6000 colony in a blood sample is p=0.5.

$$\frac{C_{20}^{20}}{2^{20}} = \frac{1}{1048576} = 0.00000095 \approx 0.000001$$

Approximately 1 in 1,000,000 20 ml samples would show all 20 colonies of either single strain (when p=0.5 of each).

Probability of having both strains in a 20 ml blood sample of 20 colonies:

$$1 - 0.00000095 - 0.00000095 = 0.9999981 \approx 0.\dot{9}$$

Increasing the blood volume makes the chance of falsely finding a single strain when both are present, negligible.

Equal proportions of two strains: impact of concentration and blood volume

These calculations can then be extended to further demonstrate the relationship, as the concentration or blood volume drop the chance of detecting both if both are present, reduces (Table 10.6).

Table 10.6: Probability of detecting both strains when present in equal proportions in a range of bacterial concentrations and volume scenarios. *Cannot calculate probabilities when the total number being sampled is not an integer.

	Concentration of bacteria CFU/ml					
	0.1	0.5	1	5	10	50
Blood volume/ml						
5	*	*	0.9375	0.9999999	1	1
10	0	0.9375	0.9980469	1	1	1
20	0.5	0.9980469	0.9999981	1	1	1
50	0.9375	0.9999999	1	1	1	1
100	0.9980469	1	1	1	1	1

Unequal proportions of two strains: impact of concentration and blood volume

When the ratio of strains is no longer equal, the chance of being able to detect both strains also is affected and makes it less likely that both strains will be detected. This is less pronounced at a ratio of 60:40 at 1 CFU/ml (Table 10.7) but is much more marked at a ratio of 80:20 (Table 10.8).

Table 10.7: Probability of detecting both strains when present in a 60:40 ratio in a range of bacterial concentrations and volume scenarios. *Cannot calculate probabilities when the total number being sampled is not an integer.

	Concentration of bacteria CFU/ml					
	0.1	0.5	1	5	10	50
Blood volume/ml						
5	*	*	0.912	0.9999972	1	1
10	0	0.912	0.9938485	1	1	1
20	0.48	0.9938485	0.9999634	1	1	1
50	0.912	0.9999972	1	1	1	1
100	0.9938485	1	1	1	1	1

Table 10.8: Probability of detecting both strains when present in an 80:20 ratio in a range of bacterial concentrations and volume scenarios. *Cannot calculate probabilities when the total number being sampled is not an integer.

	Concentration of bacteria CFU/ml					
	0.1	0.5	1	5	10	50
Blood volume/ml						
5	*	*	0.672	0.9962221	0.9999857	1
10	0	0.672	0.8926257	0.9999857	1	1
20	0.32	0.8926257	0.9884708	1	1	1
50	0.672	0.9962221	0.9999857	1	1	1
100	0.8926257	0.9999857	1	1	1	1