Social Contact Patterns During the COVID-19 Pandemic: Implications for Public Health and Hospital Infection Control

Jessica Rose Elizabeth Bridgen

A thesis submitted for the degree of

Doctor of Philosophy

Faculty of Health and Medicine

Lancaster Medical School

Lancaster University

July 2023

Social Contact Patterns During the COVID-19 Pandemic: Implications for Public Health and Hospital Infection Control

Jessica Rose Elizabeth Bridgen

Faculty of Health and Medicine

A thesis submitted for the degree of *Doctor of Philosophy*

Social contact patterns are an important driver of respiratory epidemics. Human behaviour is likely to change during an outbreak; this may be due to imposed control measures or as a personal-risk judgement. Following the emergence of the COVID-19 pandemic in 2020, numerous interventions were implemented in the UK to curb community transmission of SARS-CoV-2 by reducing social contact. It was therefore important to quantify contact patterns during the pandemic to understand how social networks had changed and identify key routes of transmission. Chapters 2 and 3 of this thesis outline two cross-sectional population-based surveys which quantified and characterised social contact patterns in different populations during the pandemic. Social contact patterns are quantified at the national scale in Chapter 2, following the relaxation of pandemic restrictions in July 2020. We investigated the association of demographic characteristics and behaviour, such as shielding and self-isolating, with non-household mixing. Chapter 3 describes an occupational study conducted in December 2020. In this study, we quantified social contact patterns of home delivery drivers at delivery depots and with customers, and identified the protective measures that they adopted during the pandemic. In Chapter 4 we outline a statistical framework to identify the role of hospital structure and staff interactions in nosocomial transmission of SARS-CoV-2 during the first wave of the pandemic. We present an efficient method to infer epidemiological event times and quantify relative routes of transmission, while accounting for the intricacies of the staff-patient contact network. This thesis demonstrates how social contact data can provide

insight into adherence to non-pharmaceutical interventions, identify subgroups of the population which may be at a greater risk of infection, and quantify relative routes of transmission in high-risk settings. We identify the wider implications of social contact patterns during the COVID-19 pandemic for public health and hospital infection control.

Contents

List of Tables

List of Figures

List of Abbreviations

CDC	Centers for Disease Control and Prevention
CI	Confidence Interval
CrI	Credible Interval
COVID-19	Coronavirus Disease 2019
HCAI	Healthcare Associated Infections
ILM	Individual-level Model
IPC	Infection Prevention and Control
LFS	Labour Force Survey
LFT	Lateral Flow Test
MCMC	Markov chain Monte Carlo
MERS	Middle East Respiratory Syndrome
NPI	Non-Pharmaceutical Intervention
ONS	Office for National Statistics
PCR	Polymerase Chain Reaction
PHEIC	Public Health Emergency of International Concern
PPE	Personal Protective Equipment
RAM	Random Access Memory
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronarivus 2
SPI-M-0	Scientific Pandemic Influenza Group on Modelling, Operational sub-group
UKHSA	UK Health Security Agency
WHO	World Health Organisation

Acknowledgements

Firstly, I would like to thank my supervisors Prof. Chris Jewell and Dr. Jon Read for their continued support and guidance these past few years. Their enthusiasm and confidence has enabled me to develop my research in new directions. I would like to thank Jon for providing me with countless opportunities to participate in COVID-19 research efforts, and for always being equipped with an inspired acronym. I would also like to thank Chris for such an enthusiastic and patient introduction to Bayesian statistics. I look forward to collaborating with you both in future.

During my PhD studies I have had the opportunity to collaborate with academics and scientists in other institutions. I would like to thank my co-authors at the University of Manchester for motivating the second study of this thesis and providing valuable insights into occupational health. I would also like to thank Dr. Joe Lewis, Prof. Miriam Taegtmeyer and Dr. Stacy Todd at Liverpool University Hospitals NHS Foundation Trust for sharing their clinical expertise, without which the final study of this thesis would not have been possible. A secondment to UKHSA provided me with a fresh perspective to the pandemic response. Many thanks to everyone in the Outbreak Surveillance Team, it was a privilege to be part of such a supportive and capable team.

I have thoroughly enjoyed my time in the CHICAS research group at Lancaster University. I would like to thank all of the Lancaster Medical School PhD students for making these past few years so enjoyable, in particular: Charlotte, Cían, Annie, Alex, Jessie, Rachael, Fran and Yawen. Our dubiously decorated hallway-come-office became home for coffee rounds, whiteboard scribbles and statistics chats.

I would like to thank Henry for his unwavering support and encouragement. Finally, I would like to thank my family and friends, who provided well-timed distractions from PhD work and lent an enthusiastic ear when needed.

Declaration

This thesis is the result of my own work and has not been submitted in support of an application for another degree at this or any other university.
The main body of this thesis contains approximately 27,568 words.

The following chapters have been published in peer-reviewed journals:

Chapter 2 (Paper 1)
Social mixing patterns in the UK following the relaxation of COVID-19 pandemic restrictions, July–August 2020: a cross-sectional online survey.
Bridgen, J. R. E., Jewell, C., Read, J. M.
Published in BMJ Open (2022)
10.1136/bmjopen-2021-059231.

Chapter 3 (Paper 2)
Contact patterns of UK home delivery drivers and their use of protective measures during the COVID-19 pandemic: a cross-sectional study.
Bridgen, J. R. E., Wei, H., Whitfield, C., Han, Y., Hall, I., Jewell, C. P.,
van Tongeren, M. J., and Read, J. M.
Published in BMJ Occupational and Environmental Medicine (2023)
10.1136/oemed-2022-108646.

List of Papers

Paper 1. Social mixing patterns in the UK following the relaxation of COVID-19 pandemic restrictions, July–August 2020: a cross-sectional online survey.

Bridgen, J. R. E., Jewell, C., Read, J. M.
Published in BMJ Open (2022) 10.1136/bmjopen-2021-059231.
Contribution: Lead author, jointly designed and conducted study, carried out the statistical analyses and writing of the paper.

Paper 2. Contact patterns of UK home delivery drivers and their use of protective measures during the COVID-19 pandemic: a cross-sectional study.

Bridgen, J. R. E., Wei, H., Whitfield, C., Han, Y., Hall, I., Jewell, C. P., van Tongeren, M. J., and Read, J. M.
Published in BMJ Occupational and Environmental Medicine (2023) 10.1136/oemed-2022-108646.
Contribution: Lead author, jointly designed and conducted study, carried out the statistical analyses and writing of the paper.

Paper 3. A Bayesian approach to identifying the role of hospital structure and staff interactions in nosocomial transmission of SARS-CoV-2.

Bridgen, J. R. E., Lewis, J, M., Todd, S., Taegtmeyer, M., Read, J. M and Jewell, C. P.
Contribution: Lead author, jointly designed the study, conducted the modelling and inference, and writing of the paper.

COVID-19 Impact Statement

The direction of this thesis changed greatly with the emergence of the COVID-19 pandemic. I began this PhD in October 2019 to research the role of contact networks in influenza transmission. Much of the research presented in this thesis was conducted in response to needs for additional data or analyses during the pandemic, to support the department's research and wider efforts. An element of this research was designing and conducting a UK study of social contact to quantify and characterise social mixing patterns at different stages of the pandemic. The first round of the survey was conducted in July 2020, the findings from which are presented in Chapter 2 (Paper 1) of this thesis. A second round of the survey was released in November 2020. Shortly following the release of each round of the survey, I conducted preliminary analyses of the responses and presented the findings as reports to SPI-M-O. The second round of the survey is being analysed as part of a masters project that I am co-supervising, the students of which are in the process of preparing a research publication. In addition, I contributed to the following research publication as part of the department's COVID-19 research efforts:

- Read, J.M., **Bridgen, J.R.E**., Cummings, D.A.T., Ho, A., and Jewell, C.P. Novel coronavirus 2019-nCoV COVID-19: early estimation of epidemiological parameters and epidemic size estimates. Philos. Trans. R. Soc. Lond. B Biol. Sci. 376.1829 (2021) [1].

In the course of completing this research, I was also seconded part-time to the Outbreak Surveillance Team at UKHSA to assist with the pandemic response. As part of this secondment I contributed to various research publications. Listed are research publications that are not being submitted as part of this degree but were carried out during the time of this PhD:

- Blomquist, P.B., **Bridgen, J**., Bray, N., O'Connell, A., West, D., Groves, N., Gallagher, E. et al.
  Enhancing epidemiological surveillance of the emergence of the SARS-CoV-2 Omicron variant using spike gene target failure data, England, 15 November to 31 December 2021. Eurosurveillance, 27(11) (2022) [2].

- Webster, H.H., Nyberg, T., Sinnathamby, M.A., Abdul Aziz, N., Ferguson, N., Seghezzo, G., Blomquist, P.B., **Bridgen, J**. et al.
  Hospitalisation and mortality risk of SARS-COV-2 variant omicron sub-lineage BA.2 compared to BA.1 in England. Nat Commun 13, 6053 (2022) [3].

- Harman, K., Nash, S.G., Webster, H.H., Groves, N., Hardstaff, J., **Bridgen, J**., Chand, M. et al.
  Comparison of the risk of hospitalisation among BA.1 and BA.2 COVID-19 cases treated with sotrovimab in the community in England. Influenza Other Respi Viruses. 2023; 17(5):e13150 (2023) [4].

Chapter 1

Introduction

Social interactions have important implications for the transmission of respiratory infectious diseases at the population level [5–7]. These interactions can be represented as a social contact network that may change over time. The dynamic nature of human behaviour indicates that contact patterns will inevitably change during an outbreak. Contact patterns may change due to imposed control measures or due to individuals choosing to reduce their personal risk of infection [8].

With the emergence of the COVID-19 pandemic in 2020, a range of non-pharmaceutical interventions were implemented in the UK in an attempt to curb community transmission of SARS-CoV-2 by reducing social contact [9]. It was therefore important to quantify social contact patterns during the pandemic to understand how social networks had changed and to parametrise mathematical models of SARS-CoV-2. Epidemic models of SARS-CoV-2 often incorporated social contact data alongside disease surveillance data to improve understanding of the transmission process and to better forecast the state of the epidemic [10, 11]. Identifying the key routes of transmission enables more effective interventions to be implemented, which is of particular importance in high-risk settings.

In Chapter 2 (Paper 1) of this thesis, we describe a cross-sectional UK social

contact study which quantifies and characterises non-household contact during the relaxation of pandemic restrictions in July 2020. Chapter 3 (Paper 2) subsequently describes an occupational contact study of UK home delivery drivers, which quantifies contact patterns and identifies the protective measures adopted by these key workers during the pandemic. Finally, in Chapter 4 (Paper 3) of this thesis we present a dynamical modelling study of SARS-CoV-2 transmission within a large UK hospital (nosocomial transmission) during the first wave of the pandemic. Routinely collected data was used to construct a dynamic patient-staff contact network. We demonstrate a statistical framework, incorporating the time-varying contact network, to conduct inference on hospital outbreak dynamics and quantify relative transmission routes of nosocomial infection. This introduction will provide a brief overview of the UK's response to COVID-19 as well as the context and methodology for each research study. Further detail is included in each chapter.

1.1 COVID-19 Pandemic

The COVID-19 pandemic quickly became one of the biggest global health challenges in modern history [12]. Control policies intertwined politics and science at a global scale. Misinformation, vaccine inequity and global supply chain disruptions were just a few of the challenges that needed to be navigated. Public health advice on how to best curb transmission was often considered alongside economic implications and the wider wellbeing of the population. In 2020, country-wide stay-at-home policies were commonplace, however, the frequency and stringency of non-pharmaceutical interventions (NPIs) varied greatly from nation to nation [13].

The global toll of COVID-19 is unknown. More than 6.9 million deaths have been recorded as of July 2023 and further ramifications of the pandemic will be seen for years to come [14]. Global case and death counts are thought to be greatly underestimated [15, 16] and the indirect health impacts, with disruptions to health

care systems, education and society more generally, are difficult to quantify.

1.1.1 Epidemiological Overview

COVID-19 is an infectious disease caused by the the novel coronavirus *severe acute respiratory syndrome coronavirus 2* (SARS-CoV-2) first detected in the Hubei province of China in December 2019. The virus is primarily transmitted through airborne routes, where close proximity to an infectious person increases the risk of transmission [17]. Indirect transmission of SARS-CoV-2 can also occur through contaminated surfaces [18]. Most often, infected individuals are symptomatic and experience mild to moderate respiratory illness, however COVID-19 can cause severe illness. Older individuals and those with underlying medical conditions were identified as being most at risk of severe disease [17]. Asymptomatic cases have been estimated to account for up to 40.5% of COVID-19 cases [19–21].

New variants of SARS-CoV-2 were introduced due to virus mutations, as similarly seen with influenza and other viruses. SARS-CoV-2 variants were classified as variants of concern if they met any of the following criteria: increase in transmissibility or detrimental change in COVID-19 epidemiology; increase in virulence or change in clinical presentation; decrease in effectiveness of public health and social measures or of available diagnostics, vaccines, and therapeutics [22]. The mean incubation period of SARS-CoV-2 decreased from 5-6 days for wild-type and the Alpha variant, to 4 days for the Delta variant, and 3 days for the Omicron variant. [17]. Infected individuals tended to be most infectious for the 2 days prior to symptom onset, and for the 5 days following [17].

1.1.2 The UK Response

On 31 January 2020, the first cases of COVID-19 were reported in the UK [23]. This coincided with the World Health Organisation (WHO) declaring the outbreak as a Public Health Emergency of International Concern (PHEIC) and subsequently a pandemic in March 2020 [24]. On 23 March 2020, the UK government introduced a range of NPIs to reduce contact between households and to fundamentally reduce community transmission of SARS-CoV-2. NPIs remained in place to varying degrees from March 2020 through to 2022 where, in England, the introduction of the 'Living with COVID-19' policy on 1 April saw the end to all remaining restrictions. By 1 April 2022, 169,218 deaths (defined as the number of people who died within 28 days of being identified as a COVID-19 case by a positive test) and 21,311,657 cases of COVID-19 had been recorded in the UK [25].

1.1.2.1 Non-Pharmaceutical Interventions

NPIs were initially the only available method to reduce community transmission of SARS-CoV-2 and were subsequently used in conjunction with COVID-19 vaccines. A myriad of interventions were introduced throughout the UK: stay-at-home orders (*lockdowns*); national and international travel restrictions; home-working; school closures; limits on social contact; face coverings; social distancing (physically remaining 2 metres apart from non-household members); self-isolation and shielding. In March 2020, individuals that were classed as clinically extremely vulnerable to severe illness from COVID-19 were advised to shield at home [26]. Shielding guidance was relaxed and strengthened according to fluctuations in national COVID-19 case numbers. The implementation and severity of interventions varied between the four nations of the UK. Three national lockdowns were imposed in England between March 2020 and January 2021, with an additional period of tiered restrictions in December 2020. The tiered system allowed for differing levels of restrictions by geographic area, accounting for case counts in the local and surrounding areas.

1.1.2.2 Community testing

Regular community testing can aid in the identification of cases and the mitigation of wide-spread transmission during an infectious disease outbreak. In the early stages of the pandemic, testing in the UK was limited to hospital patients, and contact tracing only focused on high-risk settings. Cases were otherwise identified through symptom identification. There were two primary testing streams used by NHS Test and Trace, the UK contact tracing service. Firstly, Pillar 1 was used for swab testing in UKHSA laboratories and NHS hospitals of health and care workers and those in clinical need. Secondly, Pillar 2 for community swab testing processed in laboratories or through the use of lateral flow tests (LFTs) [27]. Symptomatic individuals were advised to test for SARS-CoV-2 using a polymerase chain reaction (PCR) test which would be processed in a laboratory, whereas asymptomatic individuals could test if they were infectious using a LFT which would be followed by a confirmatory PCR test on the return of a positive result. LFTs were available free of charge from the UK government in April 2021 coinciding with the campaign to encourage the general public to test twice weekly [28]. Free universal symptomatic and asymptomatic community testing ceased on 1 April 2022 with the introduction of the 'Living with COVID-19' policy.

1.1.2.3 Vaccination

The UK's COVID-19 vaccination campaign was deployed in December 2020, administering the first COVID-19 vaccine globally [29]. Priority groups were identified for vaccination to prevent further mortality and protect health and social care staff and systems [30]. Further consideration was given to which type of vaccine and the number of doses that would be offered to different age groups and the clinically vulnerable. Vaccination against COVID-19 was not mandatory in the UK but was highly encouraged. Conversely, international travel in 2021 and 2022 was often contingent on vaccination status, requiring a traveller to be fully vaccinated or to quarantine upon arrival at their destination. In 2019, WHO declared vaccine

hesitancy as a leading threat to global health [31]. Vaccine uptake for COVID-19 was generally high in the UK, however, vaccine hesitancy levels were found to be associated with ethnicity and socioeconomic status [32–34].

1.2 Social Contact Patterns

1.2.1 Quantitative studies of Social Contact Patterns

It has been widely acknowledged that social contact patterns are an important feature of respiratory infectious disease dynamics [5, 6, 35, 36]. The heterogeneity in the number and type of contacts made by individuals are important considerations when trying to understand how respiratory diseases are transmitted within a population. Mathematical models of disease transmission are often used as a tool to forecast the state of an epidemic and to measure the potential impact of control measures. While models may incorporate a population structured by demographic factors to account for mixing patterns between population groups, they require the inclusion of finer scale behavioural data to better capture the nuances of social interactions and hence infectious disease transmission. Many studies have used population-based surveys to quantify social interactions in different settings to try to understand the contribution of social mixing patterns in the transmission of infectious diseases [37]. Cross-sectional surveys are most often used, providing a snapshot of contact patterns in a population at a particular time. A mixture of paper and online surveys are typically used for data collection, which require participants to complete a questionnaire, also referred to as a contact diary, detailing the number and type of interactions they had during a set period of time [37]. Alternatively, studies have used location monitoring to accurately account for individuals' movements without relying on recall, although these may be combined with a paper or online survey [7, 38, 39]. A study will define a contact in context of their study aims and the pathogen of interest. When capturing behaviour during

an outbreak, the transmission routes of the pathogen and the control measures that are in place during the study period will likely be accounted for when defining a contact.

Questionnaire surveys tend to have common limitations of recall and sampling bias. Participants are typically asked to recall the number of contacts they met over a set time period. Studies may inadvertently limit the number of contacts an individual can report leading to censored estimates. Participant burden can be a barrier for response. Surveys may reduce this by allowing the reporting of contacts as a group. Moreover, the recruitment process of contact studies range from open-to-all online surveys to those with strict recruitment protocols and inclusion criteria. Incentives, monetary or otherwise, may be used to aid in recruitment of hard to reach participants.

Mossong et al. conducted an early large-scale social contact survey, POLYMOD, which aimed to quantify contact patterns for the application of mathematical modelling of infectious diseases [40]. The 2005 POLYMOD study was conducted in eight European countries, including Great Britain. Participants recorded their contacts for a single day in a paper diary. A key finding from the POLYMOD study was that contact patterns are highly assortative with age, that is to say individuals mix preferentially with others of a similar age. The POLYMOD data are customarily used to quantify social contact patterns when modelling an age-structured population. Contact patterns between age groups can be visualised and quantified as a contact matrix which defines the mean rate of contact between each participant age group and each contact age group. Studies have examined contact rates further by locations of interest (e.g. school, work, inside or outside), type of contact (e.g. physical and non-physical) and contact duration (e.g. short or sustained duration) [40, 41]. The POLYMOD study intentionally oversampled children due to their important role in infectious disease transmission. The quantification of children's mixing patterns in the POLYMOD study has become

a particularly valuable resource, as future large-scale contact studies have typically not recruited children due to ethical considerations or have struggled to recruit young participants [38, 41–43]. A limitation of the POLYMOD study is that the sample size per country was relatively small. For example, 1,012 responses were recorded for Great Britain. A further limitation is that the contact diary design, recruitment, and follow-up methodologies differed by country, with different commercial companies hired to conduct the survey in each country. It is therefore difficult to make any inter-country comparisons of social mixing behaviour.

Following the POLYMOD study, Danon et al. conducted a national cross-sectional survey in 2009 to quantify social encounters in Great Britain [42]. The survey was available as a postal paper survey or as an online web-survey. Participants were asked to record their contacts on a given day, either as individuals or as large groups to facilitate reporting large numbers of contacts. Across the postal and online survey, 5,027 responses were collected; children were included in the study but were under-represented in the sample. This study found that children, public-sector and healthcare workers had the highest number of contact hours and were therefore at a greater risk of catching and transmitting infectious diseases. Danon et al. concluded that improved understanding of the links between occupations and contact networks may help to target control measures during an infectious disease outbreak. Occupation-specific questionnaires may enable us to understand how transmission risk could be reduced within these high-contact occupations in an outbreak.

In 2017, Klepac et al. conducted the BBC Pandemic study to capture detailed population contact patterns in the UK [38]. This study took a different approach than previous studies, designing an app which would approximate the location of a participant at hourly intervals over a 24 hour period. The participant would then be asked to provide details about the contacts they made during that time. The BBC Pandemic study was restricted to individuals aged 13 or over. Over a

15 month period, 40,177 responses with both geographical and contact information were collected. Children and individuals aged over 65 were underrepresented in the study sample. Fine-scale contact matrices were generated by settings (e.g. home, work, and school) and type of contact (physical or non-physical). The findings of this study were released in March 2020, following the emergence of COVID-19 in the UK, to provide further insights into social mixing and to contribute to the evidence base for decisions on potential NPIs.

Human behaviour is dynamic and should be expected to change during an infectious disease outbreak [8]. Behaviour may change due to control measures restricting or advising against certain activities, or an individual-level risk judgement. Although contact patterns in the UK have been quantified by numerous studies [38, 40, 42, 44], we would expect these to provide poor estimates of contact rates during an epidemic, especially if stringent control measures have been implemented to reduce mixing. The range of NPIs introduced in the UK during the COVID-19 pandemic in 2020 aimed to reduce transmission by changing individuals' behaviour, including social mixing. It therefore became important to quantify social mixing during this period, both to see how interventions had affected contact patterns and to provide quantification of the current contact patterns for contemporary modelling purposes.

A repeated cross-sectional study of adult social mixing patterns in the UK, CoMix, was conducted during the first lockdown, participants repeatedly answered a survey every two weeks for sixteen weeks [45]. In May 2020, the study was extended to collect data on contact patterns of children [46]. Data collection ended in March 2022. Weekly reports were released by the CoMix UK study team, with an aim to track the reproduction number, and mean contacts over time and geographical regions [47]. Ipsos, a market research company, was commissioned to conduct the online survey, promoting the survey to their existing members. Participants received compensation for each survey they completed. To ensure a representative sample, quotas were set on demographic characteristics such as age, gender, geographical

location and socioeconomic status. A limitation of this approach is that the sample is restricted to users with access to a computer, tablet or smart phone and internet access. Furthermore, the research panels of participants are fundamentally formed of internet users who have agreed to take part in online market research surveys [48], which are not necessarily representative of the population outside of the specified demographic characteristics. The initial findings of the study were a 74% reduction in the average daily number of contacts between 24 and 27 March 2020 when compared to the pre-pandemic contact survey POLYMOD [40, 45]. Contact matrices were produced for different settings (e.g. school, work, home), however contacts made with household members were not distinguished from contacts made within the home and thus did not provide a measure of non-household mixing. A limitation of this study was that each cohort of participants had a relatively small sample size, with 1,356 responses included in the initial analyses [45]. The CoMix survey was also released in Belgium and the Netherlands in March 2020. As of 2023, the CoMix survey had been implemented in 20 countries in the EU/EEA [49].

In Chapter 2 (Paper 1) of this thesis, we describe a cross-sectional social contact survey conducted to quantify and characterise mixing patterns in the UK when pandemic restrictions were relaxed in the Summer of 2020. Chapter 2 provides insight into adherence to social distancing guidelines and non-household mixing at a time of epidemic growth. At the time of writing, we were able to compare our findings to pre-pandemic contact studies [40, 42], and several studies on social mixing patterns conducted during the pandemic: the CoMix study findings for the same time period [50], and a study of contact patterns following a period of lockdown in Luxembourg [43]. The initial findings of this study were presented to SPI-M-O, providing insight into non-household contact rates by demographic characteristics and the impact of home-working on mean non-household contact rates [51]. Further details of the NPIs implemented and the SARS-CoV-2 infection landscape during the study period are included in the introduction to Chapter 2. A second round of this survey was released in November 2020 which does not form

part of this thesis. Findings of the second round are being prepared for publication as a Masters project. In the second round we collected further information on workplace contacts. Questions on non-household mixing and the distance travelled over the previous year's Christmas period were also included. These were added in response to requests from SPI-M-O for insight into Christmas contact patterns. We conducted preliminary analysis of the second round of the survey and findings were submitted as a report to SPI-M-O in December 2020.

Contact patterns can also drive occupational variation in risk of exposure during a respiratory infectious disease outbreak. High contact occupations have been associated with an increased likelihood of outbreaks [52, 53]. While contact studies often collect data on occupation, they rarely have the sample size needed to look at heterogeneity of mixing patterns by occupation. Occupational contact studies provide an opportunity to quantify contact patterns in a sector and to collect information on relevant occupation-specific behaviours which would be irrelevant in a wider study. Contact studies of this nature have mostly focused on healthcare workers [54–56]. Other high-contact occupations would benefit from a detailed study of interactions and behaviour in the workplace with respect to exposure risk.

In Chapter 3 (Paper 2) of this thesis we present an occupational contact study we conducted in December 2020, to quantify the contact patterns and the protective measures used during the pandemic by home delivery drivers. This research provides insight into the number and types of interactions delivery drivers, who were considered key workers during the pandemic, had at delivery depots and with customers. We compare the mean contact rates of home delivery drivers with the mean contact rate of members of the general population who attended their workplace [57]. A motivation of conducting this study was to parametrise a mathematical model of SARS-CoV-2 transmission in the home delivery sector, which has since been developed and published [58]. Further details on the role of home delivery drivers, as key workers, during the pandemic and the NPIs in place

during the study period are included in the introduction to Chapter 3.

Quantitative population-based surveys are only one example of social mixing data collected during the pandemic. Digital contact tracing became an important tool for rapid tracing of a large number of contacts during the pandemic. In the UK, the NHS COVID-19 app alerted individuals instantly if they had been exposed to an infected individual, using device proximity as a proxy for close contact [59]. The geographic location of users at time of exposure could help to identify high-risk settings. A limitation of this dataset and other app-based mobility datasets is that children and elderly individuals are often underrepresented. Furthermore, demographic information tends not to be captured. Another example is the ONS COVID-19 Infection Survey conducted in April 2020 [60]. The primary aim of the study was to estimate the prevalence of SARS-CoV-2 in the community, identifying symptomatic and asymptomatic infections. Socio-demographic data was collected from participants to assess which characteristics were associated with testing positive for SARS-CoV-2.

1.3 Transmission Routes

To understand which control measures are most appropriate and effective in an outbreak, it is firstly important to understand how the pathogen is transmitted. Direct transmission is the most common route of transmission of infections from person to person. Direct routes include physical contact, droplet transmission and contact with an agent in the environment. Indirect transmission may occur through biological, mechanical or airborne routes [61].

Identifying high-risk settings of transmission during an outbreak can enable more effective control strategies to be implemented. Certain settings may be at higher risk for infectious disease transmission and have different primary routes of transmission. Variation in risk may be due to the proximity and duration of contact, as found in household settings, the clinical vulnerability of individuals, or by occupation type. For example, risk factors for outbreaks in correctional facilities are close proximity, high-risk sexual behaviour and injection drug use [62]. The first reported outbreak of HIV within a correctional facility was reported in a Scottish prison in 1993, where shared needles and syringes were found to be the primary route of transmission [63]. Long-term healthcare facilities are also considered to be vulnerable settings for infectious disease transmission. Utsumi et al. found that 37 pathogens were associated with 206 outbreaks in long-term care facilities and that infection prevention programmes in these settings were generally inadequate [64].

Infection surveillance data does not by itself provide sufficient detail of an outbreak to disentangle the key drivers of transmission. To quantify the relative routes of transmission further data needs to be utilised. For example, social contact data can be used to improve understanding of direct contact transmission, and measures of airflow are important for airborne transmission.

1.3.1 Transmission within healthcare settings

Healthcare associated infections (HCAIs), or nosocomial infections, are infections that occur either as a direct result of healthcare interventions or from being in contact with a healthcare setting [65]. HCAIs are prevalent globally and are a serious burden in terms of morbidity, mortality, and financial costs [66]. A WHO global report on infection prevention and control (IPC) in 2022 found that effective IPC programmes can reduce healthcare infections by up to 70% [67]. Although many European countries have established IPC programmes, HCAIs still represent a significant burden of communicable diseases [68]. Prevention of nosocomial infections is critical to patient safety, with hospitalised patients often at heightened risk of serious outcomes from infections contracted in hospital [66]. Hospital outbreaks can additionally pose a significant risk of onward community transmission [69].

Hospitals in England are required to have an infection prevention protocol in place and participate in HCAI surveillance programmes [70]. Prominent HCAIs in the UK include *Methicillin-resistant Staphylococcus aureus* (MRSA) and *Clostridioides difficile* (C. difficile) [65]. The epic3 guidelines provide comprehensive recommendations for preventing HCAI in English NHS hospitals, with five standard interventions outlined: hospital environmental hygiene, hand hygiene, use of PPE, safe use and disposal of sharps, and principles of asepsis (the absence of potentially pathogenic microorganisms) [71]. Additional interventions may be implemented during a hospital-contained outbreak of an infectious disease or a large-scale epidemic. The severity, transmissibility, and transmission routes of the pathogen are key considerations when devising a control strategy. Interventions may include more stringent hygiene and PPE practices, quarantining of suspected and confirmed cases, pharmaceutical interventions, increased air filtering, and routine testing of patients and staff [72–74].

Middle East respiratory syndrome coronavirus (MERS) is an example of a pathogen which can be easily transmitted in healthcare settings. As of May 2023, 2604 cases of MERS were reported globally with the majority of cases occurring in Saudi Arabia [75]. Nosocomial infections accounted for approximately a third of MERS cases globally [76]. In 2015, all cases of an outbreak of MERS in South Korea were linked to healthcare facilities [77]. Both intra-hospital transmission and hospital-to-hospital transmission, caused by the transferring of infected patients, were reported [78]. Poor hospital IPC standards, unrecognised super-spreading events, and inadequate monitoring of contacts were reported as the cause of the outbreak [79]. Transmission of MERS was curtailed several weeks in to the outbreak through the introduction of rapid isolation measures, contact tracing, quarantine protocols, and school closures to reduce nosocomial transmission and prevent onward community transmission [72, 73].

HCAIs of SARS-CoV-2 have been reported globally. During the early stages of the pandemic, nosocomial infections of SARS-CoV-2 were estimated for up to 15.8% of patients with a confirmed infection in the UK and 44.0% in China [80, 81]. Hospital patients were found to be particularly vulnerable to adverse outcomes of SARS-CoV-2 [82, 83]. Moreover, healthcare workers had an increased risk of infection [84, 85]. Initially, routine testing of hospital patients and staff was unavailable in the UK; a triage system was recommended instead to cohort suspected symptomatic COVID-19 patients to be subsequently tested [86]. Asymptomatic patients would therefore not be screened and were likely a source of nosocomial infection [83]. Similarly, asymptomatic infections amongst healthcare workers was a risk for nosocomial transmission; Treibel et al. found that, in a London hospital, 27% of infected healthcare workers were asymptomatic [87]. Inadequate PPE for healthcare workers was another potential cause of nosocomial infection [88, 89]. Rickman et al. found that nosocomial infection rates in a London teaching hospital were reduced following the implementation of comprehensive IPC measures [83].

Identifying routes of infectious disease transmission within the complex structure of healthcare settings is challenging. One of the initial difficulties is correctly classifying infections as hospital-acquired or community-acquired. The Centers for Disease Control and Prevention's (CDC) National Healthcare Safety Network, a surveillance system for HCAIs, defines a HCAI as an infection that occurs more than 48 hours after admission [90]. While the time a patient tests positive for an infection is known, this is not generally a good indicator of when a patient was infected. A relatively simple method of classifying infections that have a short incubation period is to measure the time interval between admission and symptom onset. However, this relies on the timely detection of symptoms, which may not be pronounced. Moreover, this method is less effective if a patient can present asymptomatically. Another challenge is defining and quantifying the network of transmission opportunities within a hospital. To identify nosocomial transmission routes of a respiratory infection, contact between patients, hospital staff, and visitors would need to be considered. This may include direct contact between individuals but also indirect contact.

Mathematical models have been used to identify transmission routes of nosocomial infections for numerous pathogens and to evaluate the effectiveness of IPC interventions [91–93]. A constraint of many of these studies is that they are either limited by population size, only considering patients and staff on a particular ward following an outbreak, or the intricacies of the staff-patient contact network are not modelled. Common transmission routes considered are: healthcare workers to patients, patients to patients, and patients to healthcare workers. A lack of consensus remains regarding the primary routes of nosocomial transmission of SARS-CoV-2 [94–96]. A limitation of real-world outbreak data is that an epidemic process is only partially observed. Hospital surveillance data may provide the time a patient tests positive for an infection but the time a patient is initially infected and the time a patient ceases to be infectious is usually unknown. Inferring these unobserved event times while considering a time-varying contact network is difficult both statistically

and computationally.

In Chapter 4 (Paper 3) of this thesis we introduce a method to infer unobserved epidemiological event times and quantify relative routes of nosocomial transmission. This research identifies routes of transmission of SARS-CoV-2 in a large UK hospital, providing insight into the effectiveness of infection control measures and outbreak dynamics at the ward level. Further details on modelling and fitting models of nosocomial transmission and the exact restrictions which were in place during the study period are included in the introduction to Chapter 4.

1.4 Infectious Disease Modelling

Modelling has become an increasingly important tool when studying infectious disease dynamics [97]. Epidemiological modelling dates back to the 18th century, with Bernoulli's seminal paper on inoculation against smallpox [98, 99]. Dynamical systems approaches were applied to epidemiology in the 1900s, and since then we have seen many conceptual and technical advancements in theoretical and applied epidemiology [100]. Infectious disease modelling is fundamentally used to improve understanding of disease dynamics and forecast the state of an outbreak [100].

Emerging outbreaks present additional challenges for modelling, such as data availability and unknown epidemiological parameters of novel pathogens. While an outbreak is ongoing, real-time modelling can be used for forecasting disease dynamics; estimating key epidemiological parameters; identifying transmission routes; evaluating interventions. Most recently, we have seen how real-time modelling of SARS-CoV-2 was used to improve understanding of a novel pathogen and inform public health policy globally [101–103]. The COVID-19 pandemic brought infectious disease modelling into the limelight. In the UK, an operational subgroup of *Scientific Advisory Group for Emergencies* (SAGE) was formed, *Scientific Pandemic Influenza Group on Modelling, Operational* (SPI-M-O), to provide expert advice to the government on SARS-CoV-2 based on infectious disease modelling and epidemiology. Specific scenarios were modelled and communicated to the public regularly to explain the effect of easing or tightening restrictions [103]. Another recent example of real-time modelling was during the Ebola outbreak in 2014. Infectious disease models were used to estimate the reproduction number of the pathogen, and as the outbreak progressed, more complex mechanistic models were used to model different transmission routes and settings [97].

When modelling infectious disease dynamics, consideration should be given to model selection; a balance needs to be struck between accuracy and complexity

[100]. Infectious disease models can be broadly classified as either mechanistic or statistical. Statistical models have been widely used for prospective detection of infectious disease outbreaks and geospatial analyses [104, 105]. A mechanistic approach is usually adopted when modelling the transmission process of an infectious disease, and can be thought of as a simplified representation of a real-world process. Customarily, a compartmental model is used, where a population is separated into states (or compartments) according to their disease and infection status [100]. Elements of network modelling may be incorporated into compartmental models to describe behaviour within a population [5]. Similarly, statistical methods are often used in combination with mechanistic approaches. For example, to estimate an epidemiological parameter(s) of a pathogen from surveillance data, a mechanistic transmission model may be used to describe the dynamic epidemic process. A statistical framework would then be used to fit the model to the available data and infer the parameter(s) of interest [106]. In this thesis, we will focus on mechanistic approaches to infectious disease modelling and statistical inference methods.

1.4.1 Deterministic Models

Deterministic models can be used to describe the population-level dynamics of an infectious disease. A population is divided into states based on its disease status and assumptions are made surrounding the nature and rate of transfer between states [107]. The works of Ross (1916) [108], and Ross and Hudson (1917) [109, 110] were the first to describe the transmission process of an infectious disease by a system of ordinary differential equations. This work was later extended by Kermack and McKendrick in 1927 which led to the development of the SIR model [111]. The SIR model is the simplest compartmental model, describing a population in terms of three disease states: *susceptible to infection (S)*, *infected and infectious (I)*, and *removed from the population (R)*. The rate at which an individual transitions from the S state to the I state is generally referred to as the force of infection (λ), and

is often one of the parameters of interest when modelling disease dynamics. The force of infection is determined by the prevalence of infected individuals, the contact structure of the population, and the probability of becoming infected given contact. The system of ordinary differential equations for the SIR model is as follows:

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -\frac{\beta SI}{N}, \tag{1.1}$$

$$\frac{\mathrm{d}I}{\mathrm{d}t} = \frac{\beta SI}{N} - \gamma I, \tag{1.2}$$

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \gamma I. \tag{1.3}$$

where S and I represent the number of individuals in the susceptible state and the number of individuals in the infected state at time t respectively. β denotes the effective contact rate of the disease and N is the total population size; an infected individual makes contact with βN individuals per unit time. γ is the mean recovery rate for an individual, which describes the rate at which an individual transitions from the I state to the R state. The force of infection is defined here as $\lambda = \frac{\beta I}{N}$.

The principle of the deterministic SIR model forms the basis of many epidemic models, on account of its simplicity and flexibility [112]. There are several important assumptions that are made by the traditional SIR model. Firstly, that the population mixes homogeneously. Secondly, that the population is closed, that is to say there are no births, deaths or migration. There have been many extensions to the traditional model which avoid the assumptions described. The inherent heterogeneity in social mixing of human populations is an important aspect of respiratory disease transmission, as discussed in section 1.2.1. A common approach to including heterogeneous mixing in a compartmental model is to structure the population by age. Age-specific contact rates, often quantified from population-based surveys, can be defined and incorporated into the force of infection [113]. Age

structure is a particularly important consideration for childhood infectious diseases and vaccination strategies [97, 114]. Schenzle demonstrates the benefit of an age-structured model to investigate measles transmission within schools [115].

In a structured population, the force of infection is defined for each group of individuals, to account for the underlying contact structure. For example, a group may be defined by age, household or workplace membership. Given a population with two groups i and j, the force of infection for group i can be defined as:

$$\lambda_i(t) = \tau \sum_j c_{ji} \frac{I_j(t)}{N_j} \tag{1.4}$$

where τ denotes the transmission rate given contact, and c_{ji} represents the average contact rate between individuals in group j and individuals in group i (i.e. the number of individuals in group j contacted by an individual of group i per unit time).

When studying disease dynamics in a large population over a short time period or those of a fast spreading pathogen, births and deaths may not be an important consideration. However, including a birth rate and death rate is important when modelling pathogens which are endemic in a population, such as malaria and HIV. Anderson et al. emphasise the importance of including birth and death rates in HIV models, particularly when investigating the impact of HIV on population growth and demographic structure [116]. States may also be removed or added to the traditional SIR model to include features such as incubation period, hospitalisation, vaccination, and asymptomatic infection. For pathogens which have a notable incubation period, such as Mpox and HIV, an SEIR model may be most appropriate. The additional compartment further categorises the population into an *infected but not yet infectious (E)* state, enabling the incubation period to be explicitly modelled; a graphical representation can be found in Figure 1.1. Similarly, if a pathogen is prone to reinfect individuals, an SIRS model configuration may be most appropriate,

defining a transition rate from the R compartment back into the S compartment.

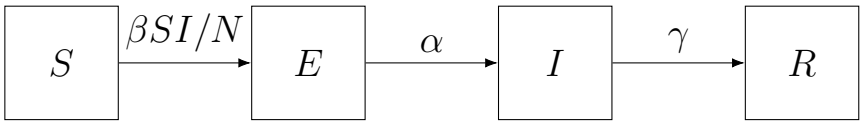

Figure 1.1: Flow diagram of a SEIR model.

1.4.2 Stochastic Approaches

The models discussed so far have been deterministic in nature; given the same starting conditions, the modelled trajectory of the epidemic would be identical in repeated simulations. However, epidemic processes in the real world are stochastic in nature [117]. Stochastic approaches incorporate an element of randomness into a model. This is particularly important when modelling small populations or a small number of infections in a large population. One method of incorporating stochasticity is by adding noise to the deterministic equations governing the epidemic process. An alternative method is to add demographic stochasticity by taking an event-driven approach to modelling an epidemic process [100]. In this section, we will focus on how Gillespie's Direct Algorithm can be used as a data generating method for stochastic compartmental models.

Gillespie's Direct Algorithm was one of two stochastic simulation algorithms introduced by Gillespie in 1977 [118]. The premise of the Direct Algorithm is to determine the time until the next event occurs. In an epidemic compartmental model, an event can be thought of as a transition between disease states. The time until the next event is assumed to be exponentially distributed and is computed from the total event rate (the sum of the independent event rates). An event is chosen at random with respect to the transition rate probabilities. The Direct Algorithm can be used for models of population-level dynamics or individual-level dynamics [100]. Individual-level dynamics tend to be modelled using stochastic approaches, as they are generally concerned with population heterogeneity. Event-driven approaches can

be particularly slow to compute, with a runtime that linearly scales with population size.

Stochastic approaches to epidemic modelling are more realistic than deterministic approaches [119]. However, more realistic modelling approaches are more computationally complex, which can be a constraint for real-world applications.

1.4.3 Individual-level Models

Compartmental models can be extended beyond population-level dynamics. Two common extensions to the compartmental approach are metapopulation models and individual-level models (ILMs). Metapopulation models describe interactions and movements between subpopulations whereas ILMs extend compartmental models to an individual-level framework.

An ILM tracks the disease status of each individual. Moreover, the force of infection is defined at the individual level, enabling further complexity to be incorporated into the model [100]. For example, heterogeneous mixing can be modelled by including social network data and spatial information. Neal and Roberts's ILM of the 1861 Hagelloch measles epidemic demonstrates the advantages of incorporating heterogeneity by including individuals' classroom and household location data [120]. Spatial location, households and classrooms were all found to be important considerations for the transmission of measles in a population. Similarly, Kretzschmar and Morris demonstrate the use of detailed contact networks for modelling the spread of sexually transmitted diseases [121]. ILMs have also been used to model disease dynamics in animal populations. Notably, an ILM was used by Keeling et al. to model the UK Foot and Mouth Disease (FMD) outbreak in 2001 at the farm-level [122]. Detailed spatiotemporal data was available, consisting of the location of all UK farms, the livestock composition of each farm, and infection surveillance data. With the inclusion of this fine-grained data in the model, spatial

and individual heterogeneities could be considered over the course of the epidemic while exploring the effectiveness of different control strategies.

Parameter inference for ILMs can be difficult [123]. To generate detailed epidemic simulations in the absence of fine-grained population data, large-scale synthetic populations can be created from sociodemographic information (e.g. population demography, household size, occupational data) [124–128]. To model transmission dynamics in different settings, each individual of the synthetic population could be assigned to a household structure, and a workplace or school location [7]. This approach is often used for contingency planning, evaluating control strategies ahead of a potential future outbreak. The main drawback of this approach is that a multitude of assumptions need to be made concerning transmission risk in different social settings, and the frequency and type of contact in different locations (such as workplaces, schools and households) [7]. A further challenge of an individual-level approach to simulating epidemics is the computational cost, particularly when trying to fit the models to surveillance data.

1.5 Inference for Infectious Disease Models

One of the core motivations behind modelling infectious disease dynamics is to better understand the transmission process. This may be in regards to quantifying traits of a specific pathogen, identifying high-risk settings, or exploring which control measures are most effective. In order to understand these aspects of disease transmission it is vital to be able to fit models to real-world data, these may be surveillance data from an outbreak and/or social contact data, to infer model parameters using statistical methods [129]. There are two apparent complications when conducting inference on real-world data that require consideration; social networks are dynamic and an epidemic process is only partially observed. This section will introduce a Bayesian approach to inference of stochastic epidemic

models.

1.5.1 Bayesian Inference

A key distinction between frequentist and Bayesian approaches is the way in which unknown model parameters are handled [130]. In frequentist statistics, model parameters are treated as fixed quantities. In contrast, a Bayesian approach considers unknown model parameters to be random variables, enabling the uncertainty of parameter estimates to be quantified. The Bayesian framework is particularly suited to epidemic data, as unobserved epidemiological event times can be treated as unknown parameters to be estimated [131].

A fundamental concept in Bayesian statistics is Bayes' theorem [132]. For a parameter θ given data y it follows that the posterior distribution can be defined as:

$$\pi(\theta|y) = \frac{f(y|\theta)\ \pi(\theta)}{\int f(y|\theta)\ \pi(\theta) d\theta} \tag{1.5}$$

where $\pi(\theta)$ denotes the prior distribution and $f(y|\theta)$ is the likelihood. In Bayesian statistics, parameters of interest are initially each assigned a prior distribution. A prior distribution represents an a priori belief about the parameter before taking the data into account. Depending on the strength of the belief, the informativeness of a prior distribution can vary, from complete uncertainty to relative certainty [133]. Meanwhile the likelihood function characterises the information from the data y, and can be defined as the probability of observing data y given the distribution of θ [134]. The posterior distribution contains all of our updated knowledge of θ given the observed data, crucially providing both a point estimate of θ but also the uncertainty surrounding that estimate. A difficulty of calculating the posterior is that it is often not possible to evaluate the integral, known as the normalising factor, particularly in high dimensions.

Markov chain Monte Carlo (MCMC) algorithms avoid the need to evaluate the integral by sampling from:

$$\pi(\theta|y) \propto f(y|\theta)\,\pi(\theta) \tag{1.6}$$

MCMC methods are often used for fitting stochastic epidemic models to partially observed data [131]. However, this is a very computationally intensive method, particularly when considering large populations or populations at an individual level. McKinley et al. provide an overview of alternative inference methods for when calculating a likelihood is computationally infeasible, concluding that *where possible* data augmentation and MCMC are still the preferred option [135]. As the volume and granularity of available data increases, more sophisticated and efficient computing approaches are required to conduct inference and perform large-scale analyses. As such, high performance computing and optimisation techniques are being increasingly employed when conducting large-scale simulations and inference on epidemic models [136–141].

1.5.2 Markov-chain Monte Carlo

MCMC methodology has had a tremendous impact on Bayesian statistics [142]. In essence, MCMC algorithms simulate random samples from a target distribution $\pi(\cdot)$, using Markov chains. Critically for epidemic models, this target distribution may be a high-dimensional posterior distribution which would otherwise by analytically intractable [143].

From the perspective of infectious disease modelling, MCMC methods provide a means to conduct inference on complex stochastic models of partially (or fully) observed epidemics [131]. Data augmentation can be implemented using MCMC methods, and is an important aspect of inference for epidemic models. Typical

surveillance data of an epidemic will contain a timestamp of when individuals tested positive for an infection, however the time that an individual initially became infected and the time they ceased to be infectious is often unobserved. These epidemiological event times are key parameters in epidemic models, particularly if the incubation period and infectiousness period is unknown for the pathogen of interest. In an MCMC algorithm, these missing event times can be included as additional parameters to be estimated [131, 144–146]. There are a wide variety of MCMC algorithms available which vary in complexity. In the following section we will introduce one of the most well known MCMC methods, the Metropolis-Hastings algorithm [147, 148].

1.5.2.1 Metropolis-Hastings algorithm

A Metropolis-Hastings algorithm explores a state-space by sampling from a proposal distribution [142]. The proposal distribution should be relatively easy to sample from and ideally as similar to the target distribution as possible. A sampled value is drawn, and is either accepted as the next state of the chain or rejected with respect to an acceptance probability. If run for a sufficient number of iterations, the Markov chain should converge to the target distribution. The rate of convergence and mixing time is highly dependent on the proposal distribution [142]. A burn-in period is generally defined, where a specified number of initial values of the chain are discarded, with the thought that the starting value and values shortly following may not represent the target distribution well. The Metropolis-Hastings algorithm is described in Algorithm 1.

Algorithm 1: Metropolis-Hastings algorithm

Input: $q(\cdot)$ proposal distribution, $\pi(\cdot)$ a target distribution
Initialise: $\theta^{(0)}$, $i = 0$, n number of iterations

while $i \leq n - 1$ **do**
  Propose $\theta^* \sim q(\cdot|\theta^{(i)})$
  Compute $\alpha(\theta^{(i)}, \theta^*) = 1 \ \wedge \frac{\pi(\theta^*)q(\theta^{(i)}|\theta^*)}{\pi(\theta^{(i)})q(\theta^*|\theta^{(i)})}$
  if $Uniform(0, 1) < \alpha(\theta^{(i)}, \theta^*)$ **then**
    $\theta^{(i+1)} = \theta^*$
  else
    $\theta^{(i+1)} = \theta^{(i)}$
  end
  $i = i + 1$
end

The Metropolis-Hastings algorithm can be extended to define a joint conditional posterior distribution for each parameter of interest and unobserved data [131]. In the case of a higher dimensional posterior distribution, the outlined Metropolis-Hastings algorithm (Algorithm 1), would implement a single block update, whereby all proposed parameter values would either be accepted or rejected as the next state of the chain. An alternative method is to use a Metropolis-within-Gibbs algorithm [144]. A Metropolis-within-Gibbs algorithm updates each parameter value or a specified block of parameters sequentially with separate Metropolis-Hastings samplers. Chapter 4 (Paper 3) outlines a novel statistical framework utilising the Metropolis-within-Gibbs algorithm, to efficiently estimate transmission rate parameters and unobserved event times for an individual-level stochastic transmission model with fine-grained contact data incorporated into the model.

1.6 Thesis Aims and Structure

The main aim of this thesis was to identify the implications of social contact patterns for public health and hospital infection control during the COVID-19 pandemic. This thesis takes two different approaches. The first two studies detail findings from primary data collection studies which aimed to quantify and characterise social contact patterns in different populations. The final study takes a Bayesian approach to quantifying relative transmission routes of nosocomial infection in a large UK hospital during the first wave of pandemic, utilising routinely collected data to account for the patient-staff contact network. The primary aims for each study are listed below:

1. To quantify and characterise non-household contact rates in the UK following the relaxation of pandemic restrictions in July 2020 (Chapter 2).
2. To quantify contact patterns of home delivery drivers and identify the protective measures they adopted during the pandemic in the UK (Chapter 3).
3. To quantify relative transmission routes of nosocomial infection in a large UK hospital during the first wave of COVID-19 (Chapter 4).

Thesis Structure

This thesis is arranged into three research papers (Chapters 2-4). The first two papers presented have been published and the final paper has been circulated to co-authors ahead of submission.

In Chapter 2 (Paper 1) of this thesis, we present the findings from the first round of a UK social contact study conducted in July 2020. We quantify and characterise non-household contact and identify the effects of shielding and isolating on social contact patterns following the relaxation of COVID-19 pandemic restrictions.

Chapter 3 (Paper 2) of this thesis presents the findings of an occupational contact study conducted in December 2020. This cross-sectional study quantifies contact patterns of UK home delivery drivers and identifies protective measures adopted during the pandemic.

In Chapter 4 (Paper 3) we present a dynamical modelling study of nosocomial transmission of SARS-CoV-2 during the first wave of the pandemic. This study provides a Bayesian framework for conducting inference on hospital outbreak dynamics to estimate unobserved epidemiological event times and quantify the relative contribution of transmission routes for nosocomial infections.

Chapter 2

Paper 1. Social mixing patterns in the UK following the relaxation of COVID-19 pandemic restrictions, July–August 2020: a cross-sectional online survey

BMJ Open

Jessica R.E. Bridgen[1], Chris P. Jewell[2], Jonathan M. Read[1]

[1] Centre for Health Informatics, Computing, and Statistics, Lancaster Medical School, Lancaster University, Lancaster, UK

[2] Department of Mathematics and Statistics, Lancaster, UK Lancaster University

2.1 Abstract

Objectives: To quantify and characterise non-household contact and to identify the effect of shielding and isolating on contact patterns.

Design: Cross-sectional study.

Setting and participants: Anyone living in the UK was eligible to take part in the study. We recorded 5143 responses to the online questionnaire between 28 July 2020 and 14 August 2020.

Outcome measures: Our primary outcome was the daily non-household contact rate of participants. Secondary outcomes were propensity to leave home over a 7 day period, whether contacts had occurred indoors or outdoors at locations visited, the furthest distance travelled from home, ability to socially distance and membership of support bubble.

Results: The mean rate of non-household contacts per person was 2.9 d^{-1}. Participants attending a workplace (adjusted incidence rate ratio (aIRR) 3.33, 95%CI 3.02 to 3.66), self-employed (aIRR 1.63, 95%CI 1.43 to 1.87) or working in healthcare (aIRR 5.10, 95%CI 4.29 to 6.10) reported significantly higher non-household contact rates than those working from home. Participants self-isolating as a precaution or following Test and Trace instructions had a lower non-household contact rate than those not self-isolating (aIRR 0.58, 95%CI 0.43 to 0.79). We found limited evidence that those shielding had reduced non-household contacts compared with non-shielders.

Conclusion: The daily rate of non-household interactions remained lower than prepandemic levels measured by other studies, suggesting continued adherence to social distancing guidelines. Individuals attending a workplace in-person or employed as healthcare professionals were less likely to maintain social distance and had a higher non-household contact rate, possibly increasing their infection risk.

Shielding and self-isolating individuals required greater support to enable them to follow the government guidelines and reduce non-household contact and therefore their risk of infection.

2.2 Introduction

On 31 January 2020, the first two cases of COVID-19 were recorded in the UK, followed by a rapid rise in identified cases and hospitalised patients. On 23 March 2020, a range of social distancing measures were implemented across the UK (lockdown), aiming to reduce interpersonal contact between households and reduce transmission of SARS-CoV-2. Schools were closed to pupils, with the exception of children of key workers. People were only allowed to leave their homes to shop for basic necessities, to exercise once a day, for medical reasons and to travel to work if working from home was not possible [1]. By July 2020, many businesses, including shops, restaurants and pubs, had reopened. Support bubbles had been introduced, allowing for a single-adult household to interact with another household of any size [2]. International travel was permitted, following the introduction of travel corridors on 10 July 2020, which enabled passengers to travel to England from certain countries without self-isolating [3]. The UK government's 'Eat Out to Help Out' scheme, which ran from 3 August 2020 to 31 August 2020, encouraged people to dine out [4]. Some social distancing restrictions remained in place, including maintaining a 2m distance between individuals (excluding household members or members of a support bubble), the wearing of face coverings on public transport and in shops, and limits on how many people could meet indoors and outdoors [5–8]. While some people in the UK began to return to work, schools remained closed. A marked decrease in case incidence was seen during April 2020, and cases remained low until the onset of the second wave in August 2020.

Epidemics are largely driven by social mixing patterns and their quantification

is useful for transmission modelling purposes, as well as assessing adherence to regulations and identifying sociodemographic factors associated with heterogeneities in contact rate [9–11]. The apparent association between social distancing restrictions and reduced case incidence indicates that a nuanced understanding of how individuals' contact patterns vary could inform behavioural interventions for the remainder of the outbreak. Previous contact studies have provided estimates for age-specific contact rates in Great Britain and the UK [11–13]. A cross-sectional survey of UK adults during the lockdown beginning in March found a substantial reduction in daily contact between people [14].

We conducted a cross-sectional online survey between 28 July 2020 and 14 August 2020 to measure the mobility of people living in the UK, which locations people were frequenting and the number of non-household contacts people were making. We aimed to quantify non-household contact behaviour and adherence to self-isolation and shielding guidance. Non-household contacts were the focus of this study, as these were the interactions affected by the social distancing measures implemented during the study period. The study period coincided with the start of the second wave of SARS-CoV-2 infection in the UK, when hospital admissions for COVID-19 were at their lowest rate since April [15].

2.3 Methods

2.3.1 Survey Methodology

Data collection was conducted through an anonymous online questionnaire; the study was branded the CoCoNet (COVID-19 Contact Network) survey. The survey was open to anyone living in the UK at the time of the survey. There was no lower age limit for participation, with children under 13 required to complete the survey with a parent or guardian. The inclusion criteria for participants were that they

completed the question on residency location and that they were resident in the UK at the time of the survey.

The survey was promoted through a university press release, engagement with the media, and posts on social media directing potential participants to the study website: `https://www.lancaster.ac.uk/health-and-medicine/research/coconet-study/`

Demographic information from participants, including age, sex, ethnicity, home location (first part of postcode) and their employment or school situation, was collected. Participants were asked about their household size, as well as the formation and size of support bubbles they may belong to. Participants were asked about their activities on the previous day (the contact reporting day), including whether they left their household and the number and characteristics of non-household contacts encountered. The questionnaire is presented in Appendix A.4 and the dataset is publicly available [16].

To reduce participant burden, a triage question on how many people participants had met the previous day determined the level of information collected on contacts. Participants reporting fewer than 15 contacts were asked to estimate the age of each contact they made, whether they met the contact indoors or outdoors, and if anyone from their household had also met that contact the same day. Participants who reported 15 or more contacts were asked to estimate the number of contacts made with different age groups, and whether they had met most of their contacts indoors or outdoors.

Responses recorded between 00:00BST 28 July and 18:00BST 14 August 2020 were included in the analysis. Partial responses to the survey were analysed if the first compulsory question asking which part of the UK a participant resided in was answered. If a participant exited the online survey early, we used their responses up to and including the last question they saw.

2.3.2 Primary and Secondary Outcome Measurements

Our primary outcome was non-household contact rate. A non-household contact was defined as someone with whom the participant had a face-to-face conversation excluding members of their own household. A participant who remained at home could still make non-household contacts by having visitors to their home.

Secondary outcomes were whether contacts occurred indoors or outdoors, propensity to leave home over a 7-day period, ability to socially distance, locations visited, furthest distance travelled from home and membership of support bubble.

2.3.3 Descriptive Analysis

Representativeness was assessed by visual comparison of participant demographics with respective Office for National Statistics 2019 mid-year estimates [17, 18]. The mean number of non-household contacts was calculated and stratified by age, sex and household size, and was compared with reported values from other social contact surveys. Adherence to social distancing guidance was assessed by calculating the proportion of participants who left home in the past 7 days, the distribution of furthest distance travelled in the past 7 days, and the proportion of participants who felt able to maintain a recommended physical distance during contact with others. Non-responses were excluded from analyses.

2.3.4 Age-specific mixing rates

To calculate age-specific mixing patterns we first defined matrix C_{ij}, where C_{ij} was the number of non-household contacts reported between participant age group i and

contact age group j. The mean contact rate per age group (M_{ij}) was given by:

$$M_{ij} = \frac{C_{ij}}{N_i} \tag{2.1}$$

where N_i was the number of participants in age group i. As a measure of uncertainty, we calculated confidence intervals by taking 1,000 bootstrapped samples of participants.

Similarly, age-specific non-household contact rates were derived from the POLYMOD data [11]. We calculated the percentage decrease of age-specific non-household contact rates between the POLYMOD data and the CoCoNet data.

2.3.5 Predictors of Contact Frequency

To identify characteristics of the participant associated with their rate of daily non-household contact, we fitted a negative binomial model to the daily number of non-household contacts reported by participants. Explanatory variables included in the model a priori were: age; sex; ethnicity; nation of residence (England, Northern Ireland, Scotland or Wales); household size; dwelling type; whether the contact reporting day was a weekend or week day; whether the participant had left their home on the contact reporting day; participant's working situation; participant's COVID-19 circumstance. To support our hypothesis-driven choice of model parameters, we also conducted a forward stepwise model selection process, with our previously selected explanatory variables used as candidate variables; see supplemental materials. Statistical analyses were conducted using R v.4.0.2 [19].

2.3.6 Patient and Public Involvement Statement

Patients or the public were not involved in the design, or conduct, or reporting, or dissemination plans of our research. However, as the online survey was promoted via social media, members of the public were free to further promote it via social media links.

2.4 Results

2.4.1 Participant Demographics

We received 5383 survey responses recorded between 28 July 2020 and 14 August 2020; 5143 responses met our inclusion criteria [16]. Most participants were aged 40-59 (55.3%, 2813/5090) (Table 2.1, Figure 2.1A). We recorded fewer responses from participants in the youngest age groups, 0-9 years (0.1%, 5/5090) and 10-19 years (0.7%, 38/5090) and in the oldest age group, aged 80+ (0.4%, 21/5090). Males, non-white ethnicities and residents of Northern Ireland and Wales were under-represented in our sample.

2.4.2 Mobility

We found 33.7% (95%CI 32.4% to 35.0%) of participants left their home every day over a 7-day period; Table 2.2. Over the same time period, most participants travelled less than 10 miles from home, but some longer-range travel (50+ miles) occurred.

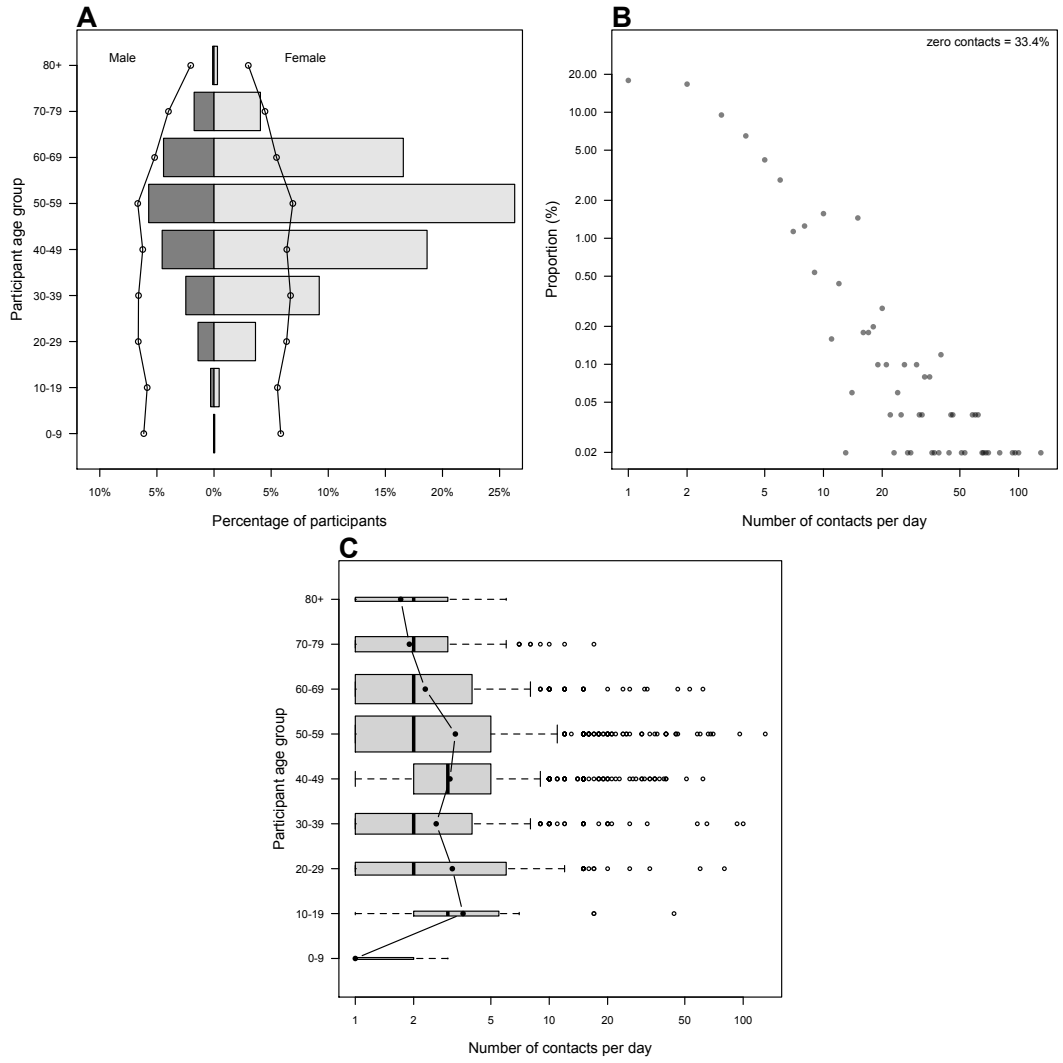

Figure 2.1: (A) Age and sex distribution of participants (bars) with ONS 2019 mid-year estimates (lines). (B) Degree distribution of non-zero contacts. (C) Distribution of reported non-zero contact rate by age group (dashed dotted line represents mean non-household contact rate). Note, log scale of x-axis in (B, C) and y-axis in (B).

Table 2.1: Participant demography and UK ONS 2019 mid-year estimates.

	No. of participants (%)	**UK ONS mid-year estimates (2019)* (%)**
Age group (N=5090)†		
0–9	5 (0.1)	12.0
10–19	38 (0.7)	11.4
20–29	256 (5.0)	13.0
30–39	598 (11.7)	13.3
40–49	1183 (23.2)	12.6
50–59	1630 (32.0)	13.6
60–69	1065 (20.9)	10.7
70–79	294 (5.8)	8.4
80+	21 (0.4)	5.0
Sex (N=5090)†		
Female	4017 (78.9)	50.6
Male	1051 (20.6)	49.4
Prefer not to say	22 (0.4)	–
Ethnicity (N=5090)		
White	4880 (95.9)	86.0
Mixed/Multiple ethnic groups	49 (1.0)	2.2
Asian/Asian British	50 (1.0)	7.5
Black/African/Caribbean/Black British	11 (0.2)	3.3
Other ethnic groups	7 (0.1)	1.0
Prefer not to say	16 (0.3)	–
No response	77 (1.5)	–
Nation (N=5143)†		
England	4714 (91.7)	84.3
Northern Ireland	33 (0.6)	2.8
Scotland	254 (4.9)	4.7
Wales	142 (2.8)	8.2
Household size (N=5073)†		
1	878 (17.3)	29.5
2	1911 (37.7)	34.5
3	987 (19.5)	15.4
4	907 (17.9)	13.9
5	287 (5.7)	4.5
6+	103 (2.0)	2.1

* Ethnicity estimates from 2011 census data.

† Question required a response from participants to progress through the online survey. N is the number of participants who provided a response to the question.

2.4.3 Non-household Contacts

A total of 14,388 non-household contacts were recorded by 5037 participants. The mean rate of non-household contacts was 2.9 d^{-1} (95%CI 2.7 to 3.0). This is a notably lower rate of non-household contact than recorded from prepandemic surveys; Appendix Table A.1. We found 33.4% (95%CI 32.1% to 34.7%) of participants made no non-household contacts. The degree distribution of non-household contacts has a long right-hand tail (95th percentile: 10 contacts d^{-1}, maximum 130 contacts d^{-1}); Figure 2.1B. We also quantified the non-household contact rate of household members of participants; see supplemental materials.

Mean non-household contact rate varied by age and was highest among 10-19 years (mean 3.6, 95%CI 1.6 to 6.5); Figure 2.1C. We found moderate assortative mixing by age, in line with both current and prepandemic contact studies (q = 0.38, 95%CI 0.17 to 0.58); Figure 2.2A. We assess assortativity using the q-statistic, where q is equal to zero for random mixing and one for assortative mixing [20]. We found that the mean daily non-household contact rate by participant age group was substantially lower when compared with prepandemic POLYMOD study; Figure 2.2B. A notable decrease in contact rate was found between people aged under 60 mixing with others aged under 60, with the largest reduction in contact rate seen across all age groups when mixing with 0-19 years; Figure 2.2B. Of all ages under 80 years old, 30-39 year olds had the highest non-household contact rate with those aged 80 or over; 4.8% (95%CIs 3.77 to 6.04) of non-household contacts reported by 30-39 years olds were with someone aged 80+.

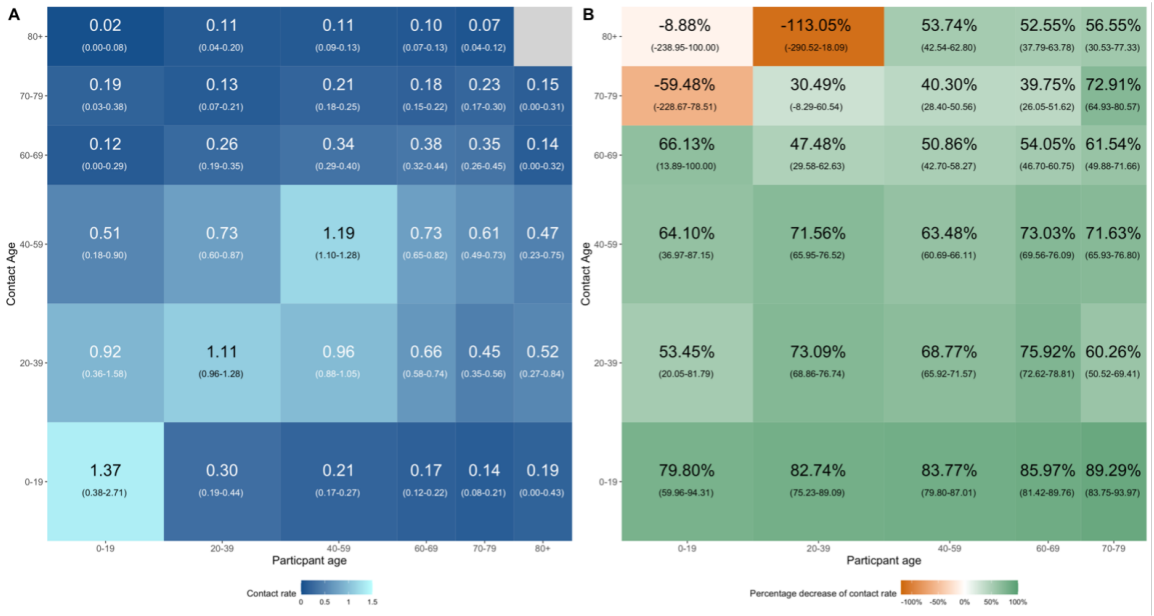

Figure 2.2: (A) Mean non-household contact rate (number of contacts per day) with different age groups reported by participant age group; bootstrapped 95% confidence intervals shown in parentheses. (B) Percentage decrease of non-household contact rate between the POLYMOD data (2005-2006) and the CoCoNet data; bootstrapped 95% confidence intervals shown in parentheses.

2.4.4 Participant Characteristics and Non-household Contact Rate

We identified the association of participant characteristics with the rate of non-household contact using a multiple regression model; Figure 2.3, Appendix Table A.2. The candidate variable dwelling type was not selected by the model selection process; Appendix Table A.3. We found no association of non-household contact rate with sex or day of the week. Contact rate varied by participant age: participants aged 30-39 (adjusted incidence rate ratio (aIRR) 0.86, 95%CI 0.76 to 0.97), aged 40-49 (aIRR 0.90, 95%CI 0.82 to 1.00) and those aged 60-69 (aIRR 0.89, 95%CI 0.79 to 1.00) reported a lower rate of contact than participants aged 50-59. We found that Asian and Asian British participants had a lower rate of contact than White participants (aIRR 0.54, 95%CI 0.36 to 0.82). Participants residing in Scotland had

a lower contact rate than those living in England (aIRR 0.80, 95%CI 0.68 to 0.95), whereas participants in Wales had a higher contact rate (aIRR 1.22, 95%CI 0.99 to 1.50).

Leaving home was associated with a higher non-household contact rate than staying at home (aIRR 5.58, 95%CI to 4.92 to 6.33). Attending a workplace (aIRR 3.33, 95%CI 3.02 to 3.66), being self-employed (aIRR 1.63, 95%CI 1.43 to 1.87) or working in healthcare (aIRR 5.10, 95%CI 4.29 to 6.10) was associated with a significantly higher rate of non-household contact than working at home.

Table 2.2: Ability of participants to social distance, membership and size of support bubbles, locations visited and mobility of participants.

	No of participants (%)
Maintaining social distance yesterday (N=3249)*	
All of the time	1910 (58.8)
More than half of the time	934 (28.7)
Less than half of the time	296 (9.1)
None of the time	89 (2.7)
Not sure	20 (0.6)
Part of a support bubble (N=5066)*	
Yes	2029 (40.1)
No	3037 (59.9)
Support bubble size (N=2011)	
1	866 (43.1)
2	560 (27.8)
3	229 (11.4)
4	201 (10.0)
5+	155 (7.7)
No response	18
Frequency of leaving home in past 7 days (N=4896)	
0 days	82 (1.7)
1 day	281 (5.7)
2 days	518 (10.6)
3 days	605 (12.4)
4 days	568 (11.6)
5 days	650 (13.3)
6 days	537 (11.0)
7 days	1650 (33.7)
Not sure	5 (0.1)
No response	30
Locations visited yesterday (N=4034)	
Someone's home	615 (15.2)
School or workplace	612 (15.2)
Doctor's surgery or healthcare facility	182 (4.5)
Supermarket or convenience store	1473 (36.5)
Other shops or retail spaces	596 (14.8)
Restaurant, café or pub	553 (13.7)
For a walk or exercise	2178 (54.0)

Table 2.2 continued from previous page

	No of participants (%)
Other	808 (20.0)
No response	0
Furthest distance travelled in past 7 days (N=4913)	
Under two miles	886 (18.0)
2-9 miles	1682 (34.2)
10-19 miles	848 (17.3)
20-49 miles	669 (13.6)
50+ miles	828 (16.9)
No response	13

* Question required a response from participants to progress through the online survey. N is the number of participants who provided a response to the question.

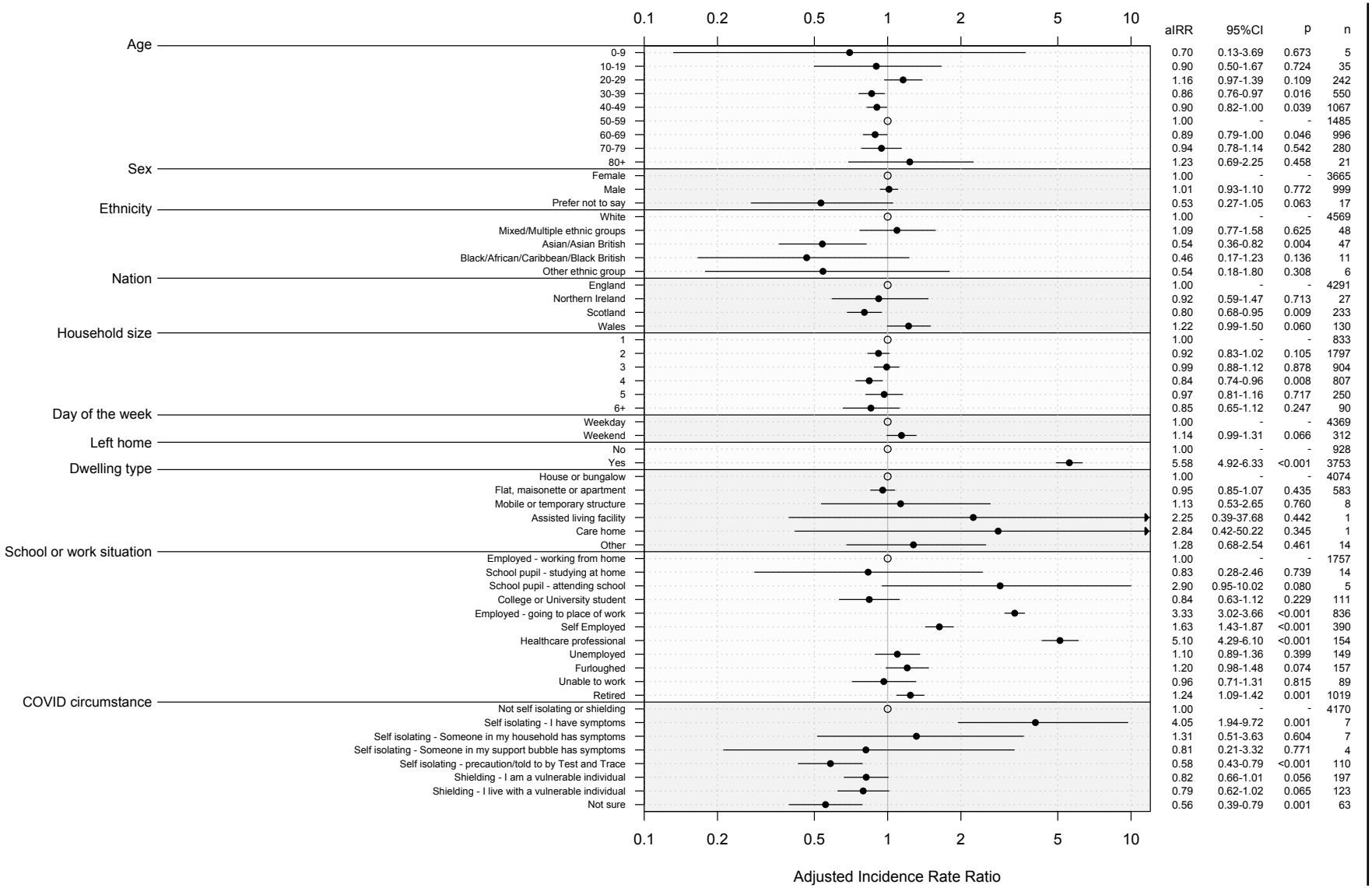

Figure 2.3: Adjusted incidence rate ratios for number of non-household contacts reported for selected variables. Hollow dots indicate reference groups.

2.4.5 Social Distancing Characteristics of Shielding and Self-isolating Individuals

There were 353 (6.9%, 353/5073) participants who reported their COVID-19 circumstance to be shielding, either due to being a vulnerable individual or living with a vulnerable individual. In addition, 136 (2.7%, 136/5073) participants reported their COVID-19 circumstance as self-isolating. Shielding individuals tended to be older than non-shielding individuals; Appendix Table A.4.

Shielding and self-isolating participants were less likely to leave their home compared with those reporting their situation to be 'not self-isolating or shielding': 58.6% (95%CI 53.2% to 63.8%) of shielding individuals, 52.6% (95%CI 43.8% to 61.2%) of self-isolating individuals and 82.7% (95%CI 81.6% to 83.8%) of other participants reported leaving their home during the contact day; Appendix Table A.4. The majority of shielding and self-isolating participants adhered to contemporary social distancing guidelines: 70.1% (95%CI 62.5% to 76.9%) of shielding participants and 73.6% (95%CI 59.7% to 84.7%) of self-isolating participants reported maintaining social distance at all time with contacts met the previous day; Appendix Table A.4.

Shielding and self-isolating individuals made fewer contacts per day outside of the household than non-shielding or isolating individuals. The unadjusted rate of non-household contact was 1.3 d^{-1}(95%CI 1.1 to 1.5) among shielding participants, 1.2 d^{-1} (95%CI 0.7 to 2.1) for self-isolating participants and 3.1 d^{-1} (95%CI 2.9 to 3.2) for participants who were not self-isolating or shielding. After adjusting for other variables, we found vulnerable individuals shielding had a marginally lower non-household contact rate than those not shielding or self-isolating (aIRR 0.82, 95%CI 0.66 to 1.01). Those self-isolating as a precaution or under test and trace instructions had a lower non-household contact rate than individuals not shielding or self-isolating (aIRR 0.58, 95%CI 0.43 to 0.79); Figure 2.3. Individuals who reported as self-isolating with symptoms had a higher rate of non-household contact than

those not self-isolating or shielding (aIRR 4.05, 95%CI 1.94 to 9.72). However, a single participant in this group reported a very large number of contacts on their contact day. This is not necessarily an example of non-adherence to social distancing guidance, as contact day and current day are different days. Our questionnaire design asked about contact on the day prior to completing the survey, which would be the day of their current COVID-19 situation. When we exclude this individual from our analysis, we found no significant difference in contact rate; Appendix Table A.5.

2.4.6 Ability to Maintain Social Distancing

Participants were asked how much of the time they were able to maintain social distance from everyone they had met the previous day, excluding members of their household and support bubble. We found 58.8% (95%CI 57.1% to 60.5%) of participants felt able to maintain social distancing at all times, while 2.7% (95%CI 2.2% to 3.4%) felt unable to maintain social distance at any time. We found that age and employment situation were associated with being able to 'maintain social distance more than half of the time'; Appendix Table A.6. Participants aged 30-39 felt less able to maintain social distance more than half of the time compared with 50-59 years (adjusted OR (aOR) 0.66, 95%CI 0.46 to 0.95). Healthcare professionals (aOR 0.26, 95%CI 0.17 to 0.40) and those attending their workplace in-person (aOR 0.71, 95%CI 0.53 to 0.96) were less likely to be able to maintain social distance than those working from home.

2.4.7 Location of Encounters

Transmission risk of SARS-CoV-2 is thought to be greater in enclosed, non-ventilated spaces and lower in outdoor environments [21]. To assess how interactions

may be distributed across these settings, we asked participants reporting fewer than 15 individual contacts whether each contact was made indoors or outdoors, and asked all participants if they met all or the majority of contacts indoors or outdoors. The distribution of contacts by indoor/outdoor setting was bimodal: nearly half of participants reported meeting all of their non-household contacts indoors (48.8%, 95%CI 47.0% to 50.6%), while 33.7% (95%CI 32.1% to 35.4%) of participants reported meeting all of their non-household contacts outdoors. We also explored the non-household contacts of participants that remained at home (visitors) and the characteristics associated with visiting another household; supplemental Appendix Table A.7.

2.5 Discussion

We found the daily rate of social contact was considerably lower than that measured prior to 2020 in similar but non-identical studies, despite our study period corresponding to a time when the COVID-19 pandemic social distancing restrictions were at their most relaxed during 2020 in the UK [11–13, 22]. The CoMix study of UK social contact rates reported a greatly reduced rate in March 2020 which increased during summer 2020, with the highest rate of contact recorded in August remaining markedly lower than prepandemic contact rate estimates [23]. Social contact studies outside of the UK also reported low daily contact rates in 2020 [24–26]. A similar increase in contact rate following lockdown was observed by Latsuzbaia et al. in Luxembourg [27].

Contact rates and ability to follow social distancing guidelines were associated with age and occupation. The older age groups (70-79, 80+), those at highest risk of severe COVID-19 outcomes, had the lowest non-household contact rates, and they mixed most often 20-59 years. Individuals attending a workplace, or those self-employed or working in healthcare, had a higher daily non-household contact rate

than those working from home, representing additional potential infection risk. A small proportion of participants reported making a large number (more than 50) of non-household contacts; these were exclusively participants who reported their employment situation as either attending their workplace in-person or working as a healthcare professional. Although the UK government was encouraging people to return to work at this time, we found that a high proportion of employed individuals (70.0%, excluding healthcare workers and those self-employed) continued to work from home [28]. In contrast to prepandemic contact surveys, we found no significant association between non-household contact rate and day of the week [11, 13].

Black and Asian individuals have been shown to be at increased risk of SARS-CoV-2 infection in comparison to White individuals, possibly due to larger households, being more likely to be employed as essential workers, and less able to work from home [29, 30]. However, after accounting for home-working, we found that individuals of Asian and Asian British ethnicity had a significantly lower non-household contact rate than White participants. This suggests that workplaces may be more dominant as a source of infection for these individuals than previously thought [31].

The majority of participants reported being able to maintain social distance from others more than half of the time and very few participants reported failing to maintain social distance at all, a similar observation made in a UK behavioural cohort [32]. Healthcare professionals and employees attending their workplace in-person were less able to maintain physical distance from people they encountered than people working from home. This highlights the increased risk of infection that some workers may face; occupations which require employees to interact closely with a large number of people are associated with an increased likelihood of exposure to COVID-19 and clusters of cases developing at a workplace [33–35].

We found some evidence of non-adherence to self-isolating and shielding guidelines, with a high proportion of self-isolating and shielding participants leaving their home

the previous day. Smith et al also found low adherence to isolation instructions among the UK population during March-August 2020 [36]. We found that a large proportion of self-isolating and shielding participants (including those living with vulnerable individuals) made non-household contacts, suggesting shielding and self-isolating individuals needed greater support to further reduce their number of interactions and to minimise infection risk.

Participants who were self-isolating as a precautionary measure, or after having been contacted by Test and Trace, reported fewer contacts than those not shielding or self-isolating. However, participants self-isolating due to experiencing symptoms or when a member of their household had symptoms did not have reduced contact rate, possibly due to the small number of participants reporting these circumstances. Participants who reported 'not sure' as their COVID-19 circumstance had a significantly lower non-household contact rate than those not self-isolating or shielding. This may have been due to a pause in shielding guidance coinciding with the release of the survey, which may have left participants unsure of their current circumstance [37–39].

This survey captured the point in time where cases were starting to consistently rise for the first time since March 2020, with the reproduction number estimated to be between 0.8 and 1.1 [15, 40–42]. The level of social mixing in the UK at the time of this survey enabled epidemic growth.

This study was likely subject to recruitment bias, as the survey was online and open to anyone living in the UK with no active recruitment process. The survey was under-represented by children, teenagers, young adults and the very elderly, as well as ethnic minorities. In particular, under-representation of the very elderly (80+) limited our ability to gain insight into mixing patterns of the age group at highest risk of severe COVID-19 disease. In addition, as we asked participants to report their contact rate, the study may have suffered from recall bias. If a participant reported meeting 15 or more contacts, information was asked about their contacts

collectively rather than as individual interactions. When grouping contacts into age groups, participants could select up to '20+' contacts for each age group, which may have led to us underestimating some participant's non-household contact rates; supplemental materials. Participants were asked about their current COVID-19 circumstance and contact behaviour for consecutive days (contacts were those made the previous day), which may bias the association of contact rate with COVID-19 circumstance. Further, due to the study design, a cross-sectional survey, we were unable to estimate changes of contact rate over time unlike studies such as CoMix. Comparisons to prepandemic contact levels in the UK are based on social contact studies conducted within the UK prior to 2020. However, these are subject to differences in study population and study design in particular sample distributions and data collection methods.

Data availability statement

Data are available in a public, open access repository (`https://www.research.lancs.ac.uk/portal/en/datasets/coconet-manuscript-data(52d69555-0092-4757-808b-997939cdcfc0).html`).

Ethics approval

Faculty of Health and Medicine Ethics Committee at Lancaster University (reference FHMREC19135). Participation in the study was voluntary, with each participant (and where appropriate parent or guardian) giving their consent before proceeding. Participants gave informed consent to participate in the study before taking part.

Acknowledgments

We would like to thank the participants of the study for providing their time and information, and Prof Julia Gog OBE and Rev Richard Coles for helping to promote the survey.

2.6 Supplementary Material

2.6.1 Variables associated with variation in non-household contact rate - model selection

To identify characteristics of the participant associated with their rate of daily non-household contact we fitted a negative binomial model to the daily number of non-household contacts reported by participants. We then used our selected explanatory variables as candidate variables for a forward stepwise model selection process, adding or removing variables based on the model's corresponding AIC score.

Participant age group, sex, ethnicity, working situation and COVID circumstance were included as explanatory variables in all models. Candidate explanatory variables were: nation of residence (England, Northern Ireland, Scotland or Wales); household size; dwelling type; whether the contact reporting day was a weekend or week day; whether the participant had left their home on the contact reporting day.

The model selection process selected the following variables from candidate variables: age, sex, ethnicity, working situation, COVID circumstance, nation, household size, day of the week, whether the participant had left their home as explanatory variables. Dwelling type was not selected as an explanatory variable. Their association with the rate of non-household contact in the fully adjusted model is shown in Appendix Table A.3.

2.6.2 Contact clustering or transitivity

Participants who made fewer than 15 contacts and did not live alone were asked if anyone in their household had met each of their contacts that same day, as a measure of clustering (also called transitivity) within social networks [43]. We estimated 40.4% (95%CI 39.1 to 41.6) of non-household contacts were also met by another

household member on the same day. The proportion of contacts encountered by participant and household members was highest if the contact was under 20 years old: contact aged 0 to 4 (77.5%, 95%CI 66.8 to 86.1); aged 5 to 9 (80.0%, 95%CI 72.7 to 86.1); aged 10 to 19 (63.2%, 95%CI 56.1 to 69.9). This indicates that non-household interactions with children tend to be made with multiple individuals from those households.

2.6.3 Visiting other households

Evidence suggests that transmission of COVID-19 often occurs within households [44]. We found that 12.2% (95%CI 11.3 to 13.1) of participants reported visiting another household. Females (aOR 1.2, 95%CI 0.98 to 1.56) and members of support bubbles (aOR 1.92, 95%CI 1.61 to 2.28) were more likely to have visited another household; Appendix Table A.7.

2.6.4 Household visits

The mean rate of contacts made with non-household members (non-household contacts) by those not leaving their home was 0.4 d^{-1} (95%CI 0.4 to 0.5), with 23.8% (95%CI 21.2 to 26.6) meeting one or more non-household contacts. This contact rate was significantly lower than for those who did leave their home (incidence rate ratio (IRR) 0.12, 95%CI 0.11 to 0.14, p-value <0.001).

2.6.5 Support bubbles

A substantial proportion (40.1%, 95%CI 38.7 to 41.4) of participants reported being part of a support bubble, with 43.1% (95%CI 40.9 to 45.3) joining with a single-person household. Males were less likely to be part of a support bubble (aOR 0.68,

95%CI 0.58 to 0.78); support bubble membership was not associated with age group or ethnicity. Support bubbles had a median (non-participant side) size of 2 (25th percentile 1, 75th percentile 3) and mean size of 2.2, (95%CI 2.1 to 2.2), and were mostly encountered two or fewer days in the past week.

2.6.6 Survey methodology limitations

When calculating the mean number of daily non-household contacts, an assumption for the maximum number of contacts was made. The survey asked participants how many contacts they had made yesterday, with the option of '0 to 15 or more'. If participants selected '15 or more', they were asked to group the contacts they had made by age, by selecting an integer between 0 and 19 or '20+' for each contact age category. To calculate the mean number of contacts, where '20+' contacts was selected, this was assumed to be 20 contacts. We may, therefore, have underestimated the maximum number of contacts of some participants and non-household contact rates.

References

[1] Cabinet Office. *Staying at home and away from others (social distancing)*. Accessed: 2020. 2020. URL: https://www.gov.uk/government/publications/full-guidance-on-staying-at-home-and-away-from-others/full-guidance-on-staying-at-home-and-away-from-others.

[2] Department of Health and Social Care. *Making a support bubble with another household*. Accessed: 2020. 2020. URL: https://www.gov.uk/guidance/making-a-support-bubble-with-another-household.

[3] Department for Transport. *Coronavirus (COVID-19): travel corridors*. Accessed: 2020. 2020. URL: https://www.gov.uk/guidance/coronavirus-covid-19-travel-corridors.

[4] GOV UK. *Register your establishment for the Eat Out to Help Out Scheme*. Accessed: 2021. July 2020. URL: https://www.gov.uk/guidance/register-your-establishment-for-the-eat-out-to-help-out-scheme.

[5] Cabinet Office. *Staying alert and safe (social distancing)*. Accessed: 2020. 2020. URL: https://www.gov.uk/government/publications/staying-alert-and-safe-social-distancing.

[6] NI direct government services. *Coronavirus (COVID-19)*. Accessed: 2020. 2020. URL: https://www.nidirect.gov.uk/campaigns/coronavirus-covid-19.

[7] Scottish government. *Coronavirus (COVID-19) Phase 3: Scotland's route map update*. Accessed: 2020. 2020. URL: https://www.gov.scot/publications/coronavirus-covid-19-framework-decision-making-scotlands-route-map-through-out-crisis-phase-3-update/pages/8/.

[8] Welsh Government. *Coronavirus (COVID-19) social distancing guidance for everyone in Wales*. Accessed: 2020. 2020. URL: https://gov.wales/coronavirus-social-distancing-guidance.

[9] Wallinga, J., Teunis, P., and Kretzschmar, M. "Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents". en. In: *Am. J. Epidemiol.* 164.10 (Nov. 2006), pp. 936–944.

[10] Read, J. M., Edmunds, W. J., Riley, S., Lessler, J., and Cummings, D. A. T. "Close encounters of the infectious kind: methods to measure social mixing behaviour". en. In: *Epidemiol. Infect.* 140.12 (Dec. 2012), pp. 2117–2130.

[11] Mossong, J., Hens, N., Jit, M., et al. "Social contacts and mixing patterns relevant to the spread of infectious diseases". en. In: *PLoS Med.* 5.3 (Mar. 2008), e74.

[12] Danon, L., Read, J. M., House, T. A., Vernon, M. C., and Keeling, M. J. "Social encounter networks: characterizing Great Britain". en. In: *Proc. Biol. Sci.* 280.1765 (Aug. 2013), p. 20131037.

[13] Klepac, P., Kucharski, A. J., Conlan, A. J. K., Kissler, S., Tang, M., Fry, H., and Gog, J. R. "Contacts in context: large-scale setting-specific social mixing matrices from the BBC Pandemic project". Feb. 2020.

[14] Jarvis, C. I., Van Zandvoort, K., Gimma, A., et al. "Quantifying the impact of physical distance measures on the transmission of COVID-19 in the UK". en. In: *BMC Med.* 18.1 (May 2020), p. 124.

[15] Public Health England. *Coronavirus (COVID-19) in the UK*. Accessed: 2020. 2020. URL: `https://coronavirus.data.gov.uk`.

[16] Bridgen, J., Jewell, C., and Read, J. *CoCoNet manuscript data*. 2021. URL: `https://www.research.lancs.ac.uk/portal/en/datasets/coconet-manuscript-data(52d69555-0092-4757-808b-997939cdcfc0).html`.

[17] Office for National Statistics. *Estimates of the population for the UK, England and Wales, Scotland and Northern Ireland, 2020*. Accessed: 2020. 2020. URL: `https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationestimatesforukenglandandwalesscotlandandnorthernireland`.

[18] Office for National Statistics. *Families and households*. Accessed: 2020. 2019. URL: `https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/families/datasets/familiesandhouseholdsfamiliesandhouseholds`.

[19] *The R project in statistical computing*. en. Accessed: 2020. URL: `https://www.r-project.org/`.

[20] Gupta, S., Anderson, R. M., and May, R. M. "Networks of sexual contacts: implications for the pattern of spread of HIV". en. In: *AIDS* 3.12 (Dec. 1989), pp. 807–817.

[21] Nishiura, H., Oshitani, H., Kobayashi, T., et al. "Closed environments facilitate secondary transmission of coronavirus disease 2019 (COVID-19)". In: *MedRxiv* (2020).

[22] Eames, K. T. D., Tilston, N. L., Brooks-Pollock, E., and Edmunds, W. J. "Measured dynamic social contact patterns explain the spread of H1N1v influenza". en. In: *PLoS Comput. Biol.* 8.3 (Mar. 2012), e1002425.

[23] Jarvis, C. I., Gimma, A., Zandvoort, K. van, Wong, K. L. M., CMMID COVID-19 working group, and Edmunds, W. J. "The impact of local and national restrictions in response to COVID-19 on social contacts in England: a longitudinal natural experiment". en. In: *BMC Med.* 19.1 (Feb. 2021), p. 52.

[24] Liu, C. Y., Berlin, J., Kiti, M. C., et al. "Rapid review of social contact patterns during the COVID-19 pandemic". In: *Epidemiology* 32.6 (2021).

[25] Coletti, P., Wambua, J., Gimma, A., et al. "CoMix: comparing mixing patterns in the Belgian population during and after lockdown". en. In: *Sci. Rep.* 10.1 (Dec. 2020), p. 21885.

[26] Zhang, J., Litvinova, M., Liang, Y., et al. "Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China". en. In: *Science* 368.6498 (June 2020), pp. 1481–1486.

[27] Latsuzbaia, A., Herold, M., Bertemes, J.-P., and Mossong, J. "Evolving social contact patterns during the COVID-19 crisis in Luxembourg". en. In: *PLoS One* 15.8 (Aug. 2020), e0237128.

[28] GOV UK. *Prime Minister's statement on coronavirus (COVID-19): 17 July 2020*. Accessed: 2021. July 2020. URL: https://www.gov.uk/government/speeches/pm-statement-on-coronavirus-17-july-2020.

[29] Sze, S., Pan, D., Nevill, C. R., et al. "Ethnicity and clinical outcomes in COVID-19: A systematic review and meta-analysis". en. In: *EClinicalMedicine* 29 (Dec. 2020), p. 100630.

[30] Public Health England. *Disparities in the risk and outcomes of COVID-19*. Tech. rep. Aug. 2020.

[31] Office for National Statistics. *Updating ethnic contrasts in deaths involving the coronavirus (COVID-19), England and Wales: deaths occurring 2 March to 28 July 2020*. Accessed: 2021. 2020. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/articles/updatingethniccontrastsindeathsinvolvingthecoronaviruscovid19englandandwales/latest.

[32] Fancourt, D., Bu, F., Mak, H. W., and Steptoe, A. "Covid-19 Social Study Results Release 25". In: *UCL* 25 (Nov. 2020).

[33] Office for National Statistics. *Which occupations have the highest potential exposure to the coronavirus (COVID-19)?* Accessed: 2020. 2020. URL: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/articles/whichoccupationshavethehighestpotentialexposuretothecoronaviruscovid19/2020-05-11.

[34] European Centre for Disease Prevention and Control. *COVID-19 clusters and outbreaks in occupational settings in the EU/EEA and the UK*. Accessed: 2021. 2020. URL: https://www.ecdc.europa.eu/en/publications-data/covid-19-clusters-and-outbreaks-occupational-settings-eueea-and-uk.

[35] Industrial Injuries Advisory Council. *COVID-19 and occupation: position paper 48*. Tech. rep. Mar. 2021.

[36] Smith, L. E., Potts, H. W. W., Amlot, R., Fear, N. T., Michie, S., and Rubin, J. "Adherence to the test, trace and isolate system: results from a time series of 21 nationally representative surveys in the UK (the COVID-19 Rapid Survey of Adherence to Interventions and Responses [CORSAIR] study)". In: *MedRxiv* (2020).

[37] Department of Health and Social Care. *Updates to the shielding programme on national and regional levels*. Accessed: 2020. 2020. URL: https://www.gov.uk/government/news/updates-to-the-shielding-programme-on-national-and-regional-levels.

[38] NI direct government services. *Coronavirus (COVID-19): pausing of shielding for extremely vulnerable people*. Accessed: 2020. 2020. URL: https://www.nidirect.gov.uk/articles/coronavirus-covid-19-pausing-shielding-extremely-vulnerable-people.

[39] Scottish government. *Shielding to be paused*. Accessed: 2020. 2020. URL: https://www.gov.scot/news/shielding-to-be-paused/.

[40] Children's Task and Finish Group. *Update on Children, Schools and Transmission*. Accessed: 2021. 2020. URL: https://www.gov.uk/government/publications/tfc-children-and-transmission-4-november-2020.

[41] Scientific Advisory Group for Emergencies. *SPI-M-O: Consensus statement on COVID-19, 12 August 2020*. Accessed: 2021. Aug. 2020. URL: https://www.gov.uk/government/publications/spi-m-o-consensus-statement-on-covid-19-12-august-2020.

[42] Scientific Advisory Group for Emergencies. *SPI-M-O Consensus Statement on COVID-19, 19 August 2020*. Accessed: 2021. Aug. 2020. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/916883/spi-m-o-consensus-statement-s0702-sage-52-200819.pdf.

[43] Wasserman, S., Faust, K., and Stanley (University of Illinois Wasserman, Urbana-Champaign). *Social Network Analysis: Methods and Applications*. en. Cambridge University Press, Nov. 1994.

[44] Leclerc, Q. J., Fuller, N. M., Knight, L. E., CMMID COVID-19 Working Group, Funk, S., and Knight, G. M. "What settings have been linked to SARS-CoV-2 transmission clusters?" en. In: *Wellcome Open Res* 5 (June 2020), p. 83.

Chapter 3

Paper 2. Contact patterns of UK home delivery drivers and their use of protective measures during the COVID-19 pandemic: a cross-sectional study

BMJ Occupational & Environmental Medicine

Jessica R.E. Bridgen[1], Hua Wei[2], Carl Whitfield[3], Yang Han[4], Ian Hall[4], Chris P. Jewell[5], Martie J A van Tongeren[6], Jonathan M. Read[1]

[1] Centre for Health Informatics, Computing, and Statistics, Lancaster Medical School, Lancaster University, Lancaster, UK

[2] School of Health Sciences, University of Manchester, Manchester, UK

[3] Immunity and Respiratory Medicine, University of Manchester, Manchester, UK

[4] Department of Mathematics, University of Manchester, Manchester, UK

[5] Department of Mathematics and Statistics, Lancaster University, Lancaster, UK

[6] Academic Health Sciences Centre, University of Manchester, Manchester, UK

3.1 Abstract

Objectives: To quantify contact patterns of UK home delivery drivers and identify protective measures adopted during the pandemic.

Methods: We conducted a cross-sectional online survey to measure the interactions of 170 UK delivery drivers during a working shift between 7 December 2020 and 31 March 2021.

Results: Delivery drivers had a mean number of 71.6 (95%CI 61.0 to 84.1) customer contacts per shift and 15.0 (95%CI 11.2 to 19.2) depot contacts per shift. Maintaining physical distancing with customers was more common than at delivery depots. Prolonged contact (more than 5 minutes) with customers was reported by 5.4% of drivers on their last shift. We found 3.0% of drivers had tested positive for SARS-CoV-2 since the start of the pandemic and 16.8% of drivers had self-isolated due to having a suspected or confirmed SARS-CoV-2 infection. In addition, 5.3% (95%CI 2.3% to 10.2%) of participants reported having worked while ill with COVID-19 symptoms, or with a member of their household having a suspected or confirmed case of COVID-19.

Conclusion: Delivery drivers had a large number of face-to-face customer and depot contacts per shift compared with other working adults during this time. However, transmission risk may be curtailed as contact with customers was of short duration. Most drivers were unable to maintain physical distance with customers and at depots at all times. Usage of protective items such as face masks and hand sanitiser was widespread.

3.2 Introduction

The delivery sector has been central in ensuring that services and supplies have remained available throughout the COVID-19 pandemic in the UK. There has been an unprecedented demand for home deliveries during the pandemic, rising sharply with the implementation of nationwide 'stay at home' orders [1]. The UK government classed delivery drivers as key workers, defined as workers critical to the COVID-19 response [2]. Shielding guidance for clinically extremely vulnerable individuals in January 2021 advised the use of food and prescription delivery services to minimise the need to leave home [3]. As non-essential businesses were forced to close for extended periods of time in 2020 and 2021, many businesses relied on online sales to generate income [4].

Transmission of SARS-CoV-2, the virus that causes COVID-19, primarily occurs through airborne routes, however, indirect transmission may occur through contaminated surfaces [5, 6]. High contact occupations are thought to be associated with an increased likelihood of employees being exposed to SARS-CoV-2 and developing clusters of cases in the workplace [7, 8]. Reducing the number of social contacts, increasing ventilation and frequent handwashing were advised methods to reduce workplace risk of exposure to SARS-CoV-2 [9]. Delivery drivers interact regularly with other employees at depots (or collection hubs) and with a large number of customers. To mitigate against infection, contact-free deliveries became widely available to minimise contact and reduce transmission risk between delivery drivers and customers [10].

The study period, 7 December 2020 to 31 March 2021, coincides with the peak and gradual decline of the second wave of COVID-19 in the UK, following the emergence of the alpha variant [11, 12]. The UK also entered a period of lockdown during this time, where 'stay at home' orders were in place, and non-essential businesses were closed to reduce transmission [13]. We aimed to quantify the contact patterns of

delivery drivers within their depot and with their customers, identifying the types of contact they were making, and to identify the protective measures adopted during the pandemic.

3.3 Methods

3.3.1 Survey methodology

We conducted a cross-sectional online survey between 7 December 2020 and 31 March 2021 to quantify behaviours thought to be associated with transmission risk of SARS-CoV-2. An anonymous online survey (the 'CoCoNet: Home Delivery Driver survey') was used for data collection. The survey design was adapted from a previous population-wide study [14]. Study participants had to meet the following inclusion criteria: a resident in the UK at the time of the survey, working as a home delivery driver and aged 18 or over. The survey was promoted through university press releases, engagement with delivery driver groups on social media (LinkedIn and Facebook) and targeted Facebook advertisements.

Survey responses received between 7 December 2020 and 31 March 2021 were included in the analysis. Partial survey responses were analysed for all questions that had been displayed to the participant.

Age, sex, ethnicity, household size and other demographic information was collected from participants. Employment information was requested, including employment type, working hours, types of items typically delivered and sick pay eligibility. The survey included questions pertaining directly to COVID-19, such as whether participants had tested positive for SARS-CoV-2, if they had to self-isolate due to suspected or confirmed SARS-CoV-2 infection and if they had worked while being ill with COVID-19 symptoms. Participants were asked to recall specific details from their last shift working as a delivery driver, including the number of customers they

met face to face, the number of individuals they had a face-to-face conversation with at their depot and their use of personal protective equipment (PPE). The survey questions can be found in Appendix B.3 and the dataset is publicly available [15].

3.3.2 Primary and secondary outcome measurements

Our primary outcome measurement was the number of contacts delivery drivers have per shift. This was stratified into contact with customers and contact with individuals (employees or customers) at a delivery depot. A contact was defined as someone whom the participant had a face-to-face conversation with. Secondary outcome measurements were: the number of deliveries per shift, the type of contact drivers were having with customers, ability to maintain physical distance, use of protective items, COVID-19-related presenteeism, the frequency of self-isolation and COVID-19 infection.

3.3.3 Data analysis

Study representativeness was assessed by comparing participant demographics with the labour force survey (LFS) estimates for delivery drivers and couriers. Quarterly LFSs for the time period November 2020 to January 2022 were aggregated for comparison, due to the relatively small sample size of delivery drivers and couriers included in each individual survey [16–21]. To assess representativeness, we compared the percentage distribution of each demographic variable from the LFS to the binomial CIs of our sample.

To identify occupational and personal characteristics associated with participants' interactions with customers, we fitted a negative binomial regression model to the number of customer contacts per shift. Explanatory variables included in the model were: age, sex, employment type, furthest distance travelled from the collection

point to a delivery, weekly working hours and the type of items delivered. The model was assessed for multicollinearity by calculating the variance inflation factor for each independent variable. Statistical analyses were conducting using R v.4.0.2.

3.4 Results

3.4.1 Participant demographics

We received 170 survey responses between 9 December 2020 and 31 March 2021, which met our inclusion criteria. Male participants accounted for 75.3% (128/170) of the sample, our survey over sampled females when compared with the aggregated LFS; Table 3.1. The majority of participants were aged 40–59 (56.5%, 96/170). Participants predominantly resided in England (81.8%, 139/170), with 1.8% (3/170) of participants residing in Northern Ireland, 10.6% (18/170) residing in Scotland and 5.9% (10/170) residing in Wales. Our sample was representative by nation when compared with the LFS. The distribution of ethnicities in our sample is broadly reflective of the sector when compared with the LFS aggregated data. Most participants in our sample were white (91.8%, 156/170) which is comparable to the LFS (92.3%, 1648/1785), however we also have representation of other ethnicities; Table 3.1.

3.4.2 Employment situation

The majority of participants reported their employment situation to be either self-employed and completely independent (54.3%, 95%CI 46.3% to 62.2%) or employed full time by one company (28.4%, 95%CI 21.6% to 36.0%); Appendix Table B.1. Over half of delivery drivers (53.1%, 95%CI 45.1% to 61.0%) reported their weekly working hours to be between 31 and 50 hours, and 22.2% (95%CI 16.1% to 29.4%)

of delivery drivers reported working more than 50 hours a week. The majority of participants reported their most recent working shift to be during the week of completing the survey (92.6%, 95%CI 87.4% to 96.1%). A small proportion of participants reported their most recent working shift to be more than a month before completing the survey (2.5%, 95%CI 0.7% to 6.2%). Most participants (68.2%, 95%CI 60.1% to 75.5%) reported that they did not receive statutory sick leave pay while working as delivery drivers.

3.4.3 Workplace interactions

The mean number of customer contacts was 71.6 (95%CI 61.0 to 84.1) per shift. We found 95.2% (95%CI 90.4% to 98.1%) of participants had brief face-to-face interactions (less than 5 minutes) with customers on their last shift, 5.4% (95%CI 2.4% to 10.4%) of participants had prolonged face-to-face interactions (more than 5 minutes) with customers and 8.2% (95%CI 4.3% to 13.8%) had entered a customer's property. We found that 61.9% (95%CI 53.5% to 69.8%) of participants were able to maintain physical distance with customers at all times during their last shift. A small proportion of participants (2.7%, 95%CI 0.7% to 6.8%) reported that they could not maintain physical distance at all on their last shift.

The mean number of contacts per shift in depot (where drivers collected items for delivery) was 27.9 (95%CI 12.2 to 55.8). This was reduced to 15.0 (95%CI 11.2 to 19.2) contacts per shift after excluding a single individual who reported making an exceptionally large number of contacts. We found that 42.4% (95%CI 34.2% to 50.9%) of participants reported that they were able to maintain physical distance from contacts at the depot at all times during their last shift. Whereas 8.3% (95%CI 4.4% to 14.1%) of participants were unable to maintain physical distance at all on their previous shift.

We found 10.5% (95%CI 6.1% to 16.4%) of participants shared a vehicle with a

Table 3.1: Participant demographics and aggregated labour force survey estimates for 'delivery drivers and couriers'.

	No of participants (percentage, 95% binomial CI)	Aggregated Quarterly labour force survey estimates for delivery drivers and couriers* November 2020–January 2022 No of participants (percentage)
Age group	N=170[†]	N=1785
18	– [‡]	5 (0.3%)
18–29	27 (15.9%, 10.74%- 22.26%)	230 (12.9%)
30–39	35 (20.6%,14.78%-27.45%)	245 (13.7%)
40–49	46 (27.1%, 20.54%-34.39%)	338 (18.9%)
50–59	50 (29.4%, 22.68%-36.87%)	530 (29.7%)
60–69	12 (7.1%, 3.70%-12.01%)	384 (21.5%)
70+	0 (0.0%, 0.00%-2.15%)	53 (3.0%)
Sex	N=170[†]	N=1785
Female	40 (23.5%, 17.37%-30.63%)	181 (10.1%)
Male	128 (75.3%, 68.11%-81.58%)	1604 (89.9%)
Prefer not to say	2 (1.2%, 0.14%-4.19%)	–
Ethnicity	N=170	N=1785
White	156 (91.8%, 86.57%-95.42%)	1648 (92.3%)
Mixed/multiple ethnic groups	5 (2.9%, 0.96%-6.73%)	9 (0.5%)
Asian/Asian British	5 (2.9%, 0.96%-6.73%)	88 (4.9%)
Black/African/ \Caribbean/Black British	1 (0.6%, 0.01%-3.23%)	26 (1.5%)
Other ethnic groups	0 (0.0%, 0.00%- 2.15%)	14 (0.8%)
Prefer not to say	3 (1.8%, 0.37%-5.07%)	–
No response	0.00 (0%, 0.00%-2.15%)	0 (0.0%)
Nation	N=170[†]	N=1785
England	139 (81.8%, 75.13%-87.26%)	1491 (83.5%)
Northern Ireland	3 (1.8%, 0.37%-5.07%)	79 (4.4%)
Scotland	18 (10.6%, 6.40%-16.22%)	114 (6.4%)
Wales	10 (5.9%, 2.86%-10.55%)	101 (5.7%)

[*] Main occupation of participant recorded as 'delivery drivers and couriers'.
[†] Question required a response from the participant.
[‡] Age group did not meet study inclusion criteria.
N is the number of participants who provided a response to the question.

colleague during their last working week, of which, 56.2% (95%CI 29.9% to 80.2%) reported sharing the vehicle with the same colleague throughout the week.

The number of contacts made per shift, including both customer and depot interactions, was positively correlated with the number of deliveries made per shift; Figure 3.1A. Participants who reported typically delivering only large items (eg, large appliances, furniture) had the greatest number of customer and depot contacts per shift, making on average more customer contacts than deliveries per shift; 3.1B. While most drivers delivering large items only reported a one-to-one ratio of customer contacts and deliveries, one individual reported four times the number of customer contacts than deliveries.

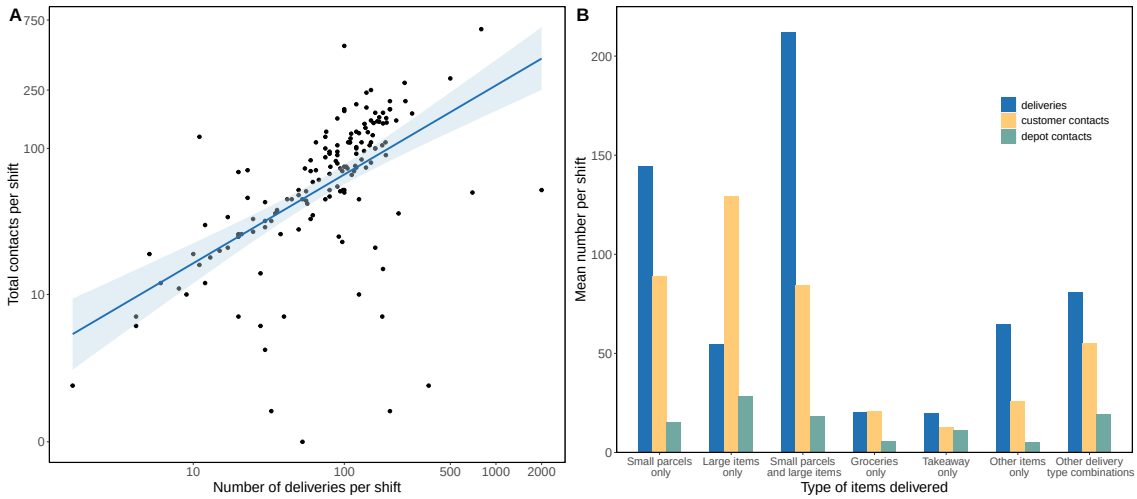

Figure 3.1: (A) Number of total contacts and deliveries made per shift. Note, x-axis and y-axis on log10-scale. Line and shaded area are a linear regression model. (B) Mean number of deliveries and contacts per shift by delivery type.

3.4.4 Frequency and type of deliveries

We found that the mean number of deliveries per shift was 121.8 (95%CI 97.9 to 152.3). Approximately half of participants (51.0%, 95%CI 42.8% to 59.1%) reported that the furthest distance they travelled from a collection point to a delivery address during their last working week was under 20 miles. The majority of delivery drivers

surveyed (52.5%, 95%CI 44.5% to 60.4%) reported that they typically delivered small parcels (including letters and mail). We found that drivers delivering small parcels and large items had the highest mean number of deliveries per shift, while takeaway and grocery delivery drivers had the lowest; Figure 3.1B.

3.4.5 Predictors of customer contacts

A negative binomial model was fitted to the number of face-to-face customer interactions per shift. The variance inflation factor, a measure of multicollinearity between a variable and remaining explanatory variables, was less than five for all model coefficients indicating multicollinearity was not present. Participants aged 18-29 (adjusted incidence rate ratio (aIRR) 1.65, 95%CI 1.07 to 2.60) and aged 40-49 (aIRR 1.64, 95%CI 1.15 to 2.34) had a higher number of customer contacts per shift than those aged 50-59; Figure 3.2, Appendix Table B.2. We found that delivery drivers who were employed by one company full time had a lower number of customer contacts per shift than those self-employed and independent (aIRR 0.66, 95%CI 0.47 to 0.94). Participants who usually deliver only groceries (aIRR 0.34, 95%CI 0.18 to 0.64), deliver only takeaways (aIRR 0.19, 95%CI 0.11 to 0.37), deliver other unlisted items (aIRR 0.43, 95%CI 0.24 to 0.80) and those who deliver other combinations of items listed (aIRR 0.58, 95%CI 0.38 to 0.93) had fewer customer contacts than those that usually deliver only small parcels.

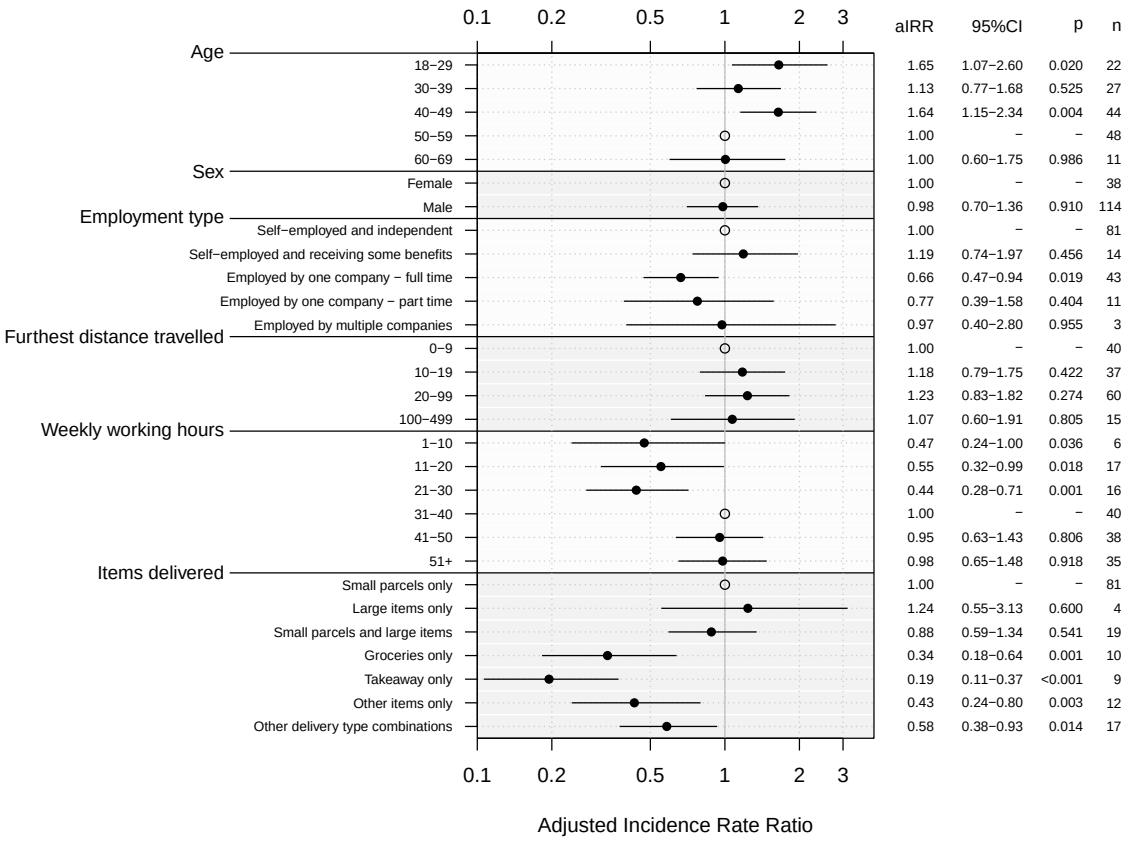

Figure 3.2: Adjusted incidence rate ratios (aIRR) for mean number of customer contacts reported for selected variables. Explanatory variables included in the model: age, sex, employment type, furthest distance travelled from the collection point to a delivery, weekly working hours and the type of items delivered. Open circles represent the reference group for each variable.

3.4.6 COVID-19 infection, self-isolation and presenteeism

We asked respondents about their COVID-19 infection status to date, as well as their behaviour following infection or potential exposure; we examined these aspects independently. We found that 3.0% (95%CI 1.0% to 6.9%) of delivery drivers surveyed reported that they had tested positive for COVID-19 since the start of the pandemic. In addition, 16.8% (95%CI 11.4% to 23.3%) of delivery drivers had self-isolated since the start of the pandemic due to a suspected or confirmed case of COVID-19. Approximately 1 in 20 drivers (5.3%, 95%CI 2.3% to 10.2%)

reported that they have continued to work while either being ill with COVID-19 symptoms or with a member of their household having a suspected or confirmed case of COVID-19. In this situation, financial reasons were most often cited as a reason for continuing to work (85.7%, 95%CI 42.1% to 99.6%).

3.4.7 Protective measures

We found that 68.3% (95%CI 60.0% to 75.7%) of participants were provided with some PPE items by their employers or contracting companies. However, less than half of participants (48.3%, 95%CI 39.9% to 56.7%) felt that the PPE provided effective protection. Face masks (82.4%, 95%CI 75.4% to 88.0%) and hand sanitiser (83.7%, 95%CI 76.8% to 89.1%) were the protective items most commonly used by delivery drivers on their last shift. Participants who shared a vehicle during their last working week most often reported using hand sanitiser to prevent infection when sharing a vehicle (81.2%, 95%CI 54.4% to 96.0%), with 50.0% (95%CI 24.7% to 75.3%) of participants reporting wearing a face mask and 50.0% (95%CI 24.7% to 75.3%) reporting keeping a window open.

3.5 Discussion

We found that delivery drivers made a large number of contacts per shift both at their depot (15.0 per shift) and with customers (71.6 per shift). In comparison, Jarvis et al. found that the mean number of contacts among the general population attending their workplace between January 2021 and March 2021 was between 3 and 10 contacts per day; this included contacts made outside of the workplace [22]. This suggests delivery drivers have a very large number of contacts compared with the general workforce at this time, which may lead to a higher risk of SARS-CoV-2 infection. The importance of contact duration in respiratory virus transmission has

been widely documented [23–27]. Face-to-face interactions between delivery drivers and customers are likely to take place outside and to be very short in duration, with only 5.4% of drivers reporting any prolonged contact (more than 5 minutes) with customers during their last working shift. Therefore, while delivery drivers have a large number of contacts, this may pose only a small risk in terms of exposure and transmission of SARS-CoV-2. Similarly, sharing a vehicle with a colleague for deliveries may be a type of high-risk contact. Nevertheless, as we found most work-related vehicle sharing was fixed-pairing (pair that share a vehicle is fixed) the risk of multiple transmission events is likely to be largely reduced. The duration of contacts made at the depot was not recorded.

The use of protective measures in the workplace was common. Most delivery drivers reported being able to maintain physical distance with customers and at the depot at least some of the time during their last shift, however, most drivers reported not being able to maintain distance at all times particularly when at the depot. Face masks and hand sanitiser were commonly used by drivers during their shift. While face masks offer varying levels of protection to the wearer, they help to prevent transmission from an infected individual to others [28]. The majority of drivers received some protective items from their employers, however, less than half of drivers felt that they provided effective protection.

By 7 December 2020, there had been 1,770,619 confirmed cases of COVID-19 in the UK, accounting for approximately 2.6% of the UK population [11, 29]. We found 3.0% of delivery drivers surveyed had tested positive for SARS-CoV-2 since the start of the pandemic, slightly higher than the national estimate, and over a sixth of delivery drivers reported having to self-isolate due to a suspected or confirmed case of SARS-CoV-2. A small proportion of delivery drivers reported working with symptoms of COVID-19 or while a member of their household had a confirmed or suspected SARS-CoV-2 infection. The majority of drivers reported continuing to work in this situation due to financial reasons, this may be associated with statutory

sick pay being unavailable for most drivers. Lack of access to paid sick leave is one of the main risk factors for respiratory infectious disease-related presenteeism [30]. Providing wider access to paid sick leave may, therefore, help to improve adherence to public health measures such as self-isolation. Delivery drivers were critical to the pandemic response, ensuring supplies were available and providing a service to clinically vulnerable individuals shielding at home, highlighting the importance of minimising exposure risk for workers and customers. Further consideration is needed on how key workers can be best supported and protected in future public health emergencies.

Participants self-reported how many face-to-face interactions they had with customers and at the depot on their last working shift as a delivery driver. For most participants, their last working shift was during the same week as completing the survey, but a small proportion (approximately 2.5%) were recalling from a shift over a month ago. There is some risk of uncertainty in recall particularly with the small proportion of participants recalling from a less recent shift. This study may suffer from recruitment bias, the survey was conducted online only without a strict recruitment process. Mean number of contacts were calculated per shift and therefore cannot be directly compared with other contact studies which calculate the number of contacts per hour or per day. Our definition of a contact is an adaptation of the definitions used by other social contact studies, rather than using a definition set by an international public health agency, such as WHO [14, 25, 31–33]. To reduce participant burden when reporting a large number of contacts, we kept the definition of a contact relatively simple. One motivation of this study was to quantify behaviours thought to be associated with transmission risk of SARS-CoV-2 to inform modelling studies, we hence used a similar definition to other pandemic contact studies to allow for comparisons across studies [14, 33, 34].

Questions pertaining to SARS-CoV-2 infection and self-isolation referred to the time period from the start of the pandemic to completing the survey. Participants

reported on their use of specific PPE items, however, we did not ask for any further details such as type of face mask worn or duration of use. Vaccination status of participants was not collected due to the timing of the study, which was released 5 days after the initial roll-out of the COVID-19 vaccine in the UK. As the vaccine was only available to clinically vulnerable individuals at this time, it was unlikely to be offered to participants during the study period [35]. Delivery driver occupation was self-reported and not confirmed. However, this occupation was reported at the time of recruitment and before beginning the survey. We did not collect data on how long participants had been working as delivery drivers, therefore, estimations of the prevalence of SARS-CoV-2 infection and self-isolation among participants may not be an accurate representation of all delivery drivers.

Data availability statement

Data are available in a public, open access repository. The survey data are available from Lancaster University's research directory at: `https://doi.org/10.17635/lancaster/researchdata/553`. License: Creative Commons Attribution licence (CC BY). The code to reproduce the statistical analyses is available at: `https://doi.org/10.5281/zenodo.7517541`.

Ethics approval

This study involves human participants and was approved by Faculty of Health and Medicine Ethics Committee at Lancaster University, reference FHMREC20040. Participants gave informed consent to participate in the study before taking part

Acknowledgments

We would like to thank the participants of the study for their time and information.

References

[1] Office for National Statistics. *Impact of the coronavirus (COVID-19) pandemic on retail sales in 2020*. en. https://www.ons.gov.uk/economy/grossdomesticproductgdp/articles/impactofthecoronaviruscovid19pandemiconretailsalesin2020/2021-01-28. Accessed: 2021. Feb. 2021.

[2] GOV UK. *Children of critical workers and vulnerable children who can access schools or educational settings*. en. https://www.gov.uk/government/publications/coronavirus-covid-19-maintaining-educational-provision/guidance-for-schools-colleges-and-local-authorities-on-maintaining-educational-provision. Accessed: 2022.

[3] Office for National Statistics. *Coronavirus and clinically extremely vulnerable people in England - Office for National Statistics*. en. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/coronavirusandclinicallyextremelyvulnerablepeopleinengland/11octoberto16october2021. Accessed: 2022. Nov. 2021.

[4] Office for National Statistics. *Coronavirus: how people and businesses have adapted to lockdowns*. en. https://www.ons.gov.uk/economy/economicoutputandproductivity/output/articles/coronavirushowpeopleandbusinesseshaveadaptedtolockdowns/2021-03-19. Accessed: 2022-1-27. Mar. 2021.

[5] World Health Organization. *Transmission of SARS-CoV-2: implications for infection prevention precautions*. en. Accessed: 2021. July 2020. URL: https://www.who.int/news-room/commentaries/detail/transmission-of-sars-cov-2-implications-for-infection-prevention-precautions.

[6] GOV UK. *COVID-19: epidemiology, virology and clinical features*. en. Accessed: 2022. 2022. URL: https://www.gov.uk/government/publications/wuhan-novel-coronavirus-background-information/wuhan-novel-coronavirus-epidemiology-virology-and-clinical-features.

[7] Office for National Statistics. *Which occupations have the highest potential exposure to the coronavirus (COVID-19)?* Accessed: 2020. 2020. URL: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/articles/whichoccupationshavethehighestpotentialexposuretothecoronaviruscovid19/2020-05-11.

[8] European Centre for Disease Prevention and Control. *COVID-19 clusters and outbreaks in occupational settings in the EU/EEA and the UK*. Accessed: 2021. 2020. URL: https://www.ecdc.europa.eu/en/publications-data/covid-19-clusters-and-outbreaks-occupational-settings-eueea-and-uk.

[9] Health and Safety Executive. *Working safely during the coronavirus (COVID-19) pandemic*. en. https://www.hse.gov.uk/coronavirus/working-safely/index.htm. Accessed: 2022.

[10] Wei, H., Daniels, S., Whitfield, C. A., et al. "Agility and Sustainability: A Qualitative Evaluation of COVID-19 Non-pharmaceutical Interventions in the UK Logistics Sector". en. In: *Front Public Health* 10 (June 2022), p. 864506.

[11] Public Health England. *Coronavirus (COVID-19) in the UK*. en. https://coronavirus.data.gov.uk. 2020.

[12] Public Health England. *Investigation of SARS-CoV-2 variants of concern in England: technical briefing 6*. Tech. rep. Feb. 2021.

[13] GOV UK. *Prime Minister's address to the nation: 4 January 2021*. en. https://www.gov.uk/government/speeches/prime-ministers-address-to-the-nation-4-january-2021. Accessed: 2022. Jan. 2021.

[14] Bridgen, J. R., Jewell, C., and Read, J. M. "Social mixing patterns in the UK following the relaxation of COVID-19 pandemic restrictions, July-August 2020: a cross-sectional online survey". en. In: *BMJ Open* 12.12 (Dec. 2022), e059231.

[15] Bridgen, J. R. E., Wei, H., Whitfield, C., et al. *CoCoNet Home Delivery Driver study data*. 2022. URL: https://www.research.lancs.ac.uk/portal/en/datasets/coconet-home-delivery-driver-study-data(702adb78-33a2-4b80-8e1c-3075a92137b4).html.

[16] Northern Ireland Statistics and Research Agency (NISRA), Office for National Statistics, Social Survey Division. *Quarterly Labour Force Survey, November, 2020 - January, 2021*. Title of the publication associated with this dataset: {UK Data Service}. Mar. 2021.

[17] Office for National Statistics, Social Survey Division, Northern Ireland Statistics and Research Agency, Central Survey Unit. *Quarterly Labour Force Survey, February - April, 2021*. Title of the publication associated with this dataset: {UK Data Service}. 2021.

[18] Office for National Statistics, Northern Ireland Statistics and Research Agency. *Quarterly Labour Force Survey, May - July, 2021*. Title of the publication associated with this dataset: {UK Data Service}. 2021.

[19] Office for National Statistics, Northern Ireland Statistics and Research Agency. *Quarterly Labour Force Survey, July - September, 2021*. Title of the publication associated with this dataset: UK Data Service{UK Data Service}. 2021.

[20] Office for National Statistics, Northern Ireland Statistics and Research Agency. *Quarterly Labour Force Survey, September - November 2021*. Title of the publication associated with this dataset: {UK Data Service}. 2022.

[21] Northern Ireland Statistics and Research Agency (NISRA), Office for National Statistics, Social Survey Division. *Quarterly Labour Force Survey, November, 2021 - January, 2022*. Title of the publication associated with this dataset: {UK Data Service}. 2022.

[22] Jarvis, C., Edmunds, J., and CMMID COVID-19 Working Group. *Social contacts in workplace the UK from the CoMix social contact survey*. Tech. rep. London School of Hygiene and Tropical Medicine, June 2021.

[23] Read, J. M., Eames, K. T. D., and Edmunds, W. J. "Dynamic social networks and the implications for the spread of infectious disease". en. In: *J. R. Soc. Interface* 5.26 (Sept. 2008), pp. 1001–1007.

[24] Smieszek, T. "A mechanistic model of infection: why duration and intensity of contacts should be included in models of disease spread". en. In: *Theor. Biol. Med. Model.* 6 (Nov. 2009), p. 25.

[25] Danon, L., Read, J. M., House, T. A., Vernon, M. C., and Keeling, M. J. "Social encounter networks: characterizing Great Britain". en. In: *Proc. Biol. Sci.* 280.1765 (Aug. 2013), p. 20131037.

[26] Thompson, H. A., Mousa, A., Dighe, A., et al. "Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Setting-specific Transmission Rates: A Systematic Review and Meta-analysis". en. In: *Clin. Infect. Dis.* 73.3 (Aug. 2021), e754–e764.

[27] De Cao, E., Zagheni, E., Manfredi, P., and Melegaro, A. "The relative importance of frequency of contacts and duration of exposure for the spread of directly transmitted infections". en. In: *Biostatistics* 15.3 (July 2014), pp. 470–483.

[28] CDC. *Masks and respirators*. en. Accessed: 2022-7-13. May 2022. URL: https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/types-of-masks.html.

[29] Office for National Statistics. *Population estimates*. en. Accessed: 2022. June 2021. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates.

[30] Daniels, S., Wei, H., Han, Y., et al. "Risk factors associated with respiratory infectious disease-related presenteeism: a rapid review". en. In: *BMC Public Health* 21.1 (Oct. 2021), p. 1955.

[31] Mossong, J., Hens, N., Jit, M., et al. "Social contacts and mixing patterns relevant to the spread of infectious diseases". en. In: *PLoS Med.* 5.3 (Mar. 2008), e74.

[32] Klepac, P., Kucharski, A. J., Conlan, A. J. K., Kissler, S., Tang, M., Fry, H., and Gog, J. R. "Contacts in context: large-scale setting-specific social mixing matrices from the BBC Pandemic project". Feb. 2020.

[33] Jarvis, C. I., Van Zandvoort, K., Gimma, A., et al. "Quantifying the impact of physical distance measures on the transmission of COVID-19 in the UK". en. In: *BMC Med.* 18.1 (May 2020), p. 124.

[34] Whitfield, C. A., Tongeren, M. van, Han, Y., et al. "Modelling the impact of non-pharmaceutical interventions on workplace transmission of SARS-CoV-2 in the home-delivery sector". en. In: *PLoS One* 18.5 (May 2023), e0284805.

[35] GOV UK. *UK COVID-19 vaccines delivery plan*. en. Accessed: 2023-2-28. Jan. 2021. URL: `https://www.gov.uk/government/publications/uk-covid-19-vaccines-delivery-plan/uk-covid-19-vaccines-delivery-plan`.

Chapter 4

Paper 3. A Bayesian approach to identifying the role of hospital structure and staff interactions in nosocomial transmission of SARS-CoV-2

Jessica R.E. Bridgen[1], Joseph M. Lewis[2,3,4], Stacy Todd[2], Miriam Taegtmeyer[2,5], Jonathan M. Read[1], Chris P. Jewell[6]

[1] Centre for Health Informatics, Computing, and Statistics, Lancaster Medical School, Lancaster University, Lancaster, UK

[2] Tropical and Infectious Diseases Unit, Liverpool University Hospitals NHS Foundation Trust, Liverpool, UK

[3] Department of Clinical Infection, Microbiology and Immunology, University of Liverpool, Liverpool, UK

[4] Department of Clinical Sciences, Liverpool School of Tropical Medicine, Liverpool, UK

[5] Department of International Public Health, Liverpool School of Tropical Medicine, Liverpool, UK

[6] Department of Mathematics and Statistics, Lancaster University, Lancaster, UK

4.1 Abstract

Nosocomial infections threaten patient safety, and were widely reported during the COVID-19 pandemic. Effective hospital infection control requires a detailed understanding of the role of different transmission pathways, yet these are poorly quantified. Using patient and staff data from a large UK hospital we demonstrate a method to infer unobserved epidemiological event times efficiently and disentangle the infection dynamics by ward. A stochastic individual-level, continuous-time state-transition model was constructed to model transmission of SARS-CoV-2, incorporating a dynamic staff-patient contact network as time-varying parameters. A Metropolis-within-Gibbs MCMC algorithm was used to estimate transmission rate parameters associated with each possible source of infection, and the unobserved infection and recovery times. We found that the total infectious pressure exerted on an individual in a ward varied over time, as did the primary inferred source of transmission. There was marked heterogeneity between wards; each ward experienced unique infectious pressure dynamics over time in terms of magnitude and primary source of transmission. Hospital infection control should consider the role of between-ward movement of staff as a key infectious source of nosocomial infection for SARS-CoV-2. With further development, this method could be implemented routinely for real-time monitoring of nosocomial transmission and to evaluate interventions.

4.2 Introduction

Healthcare-associated infections are a significant burden to health systems worldwide, and are associated with increased morbidity and mortality [1]. Transmission of SARS-CoV-2 in healthcare settings was widely reported during the first wave of the COVID-19 pandemic [2–5]. Hospital patients are considered to be especially vulnerable to severe outcomes of COVID-19 and healthcare workers have been shown to be at an increased risk of infection [6–8]. Identifying nosocomial transmission routes of SARS-CoV-2 is therefore critical to patient and staff safety. During the initial stages of the pandemic in the UK, universal testing of patients and staff was not available to rapidly identify and isolate individuals. A triage system was recommended in hospitals to isolate and test patients with suspected COVID-19 [9]. Although hospitals employed a range of infection prevention control (IPC) protocols to minimise transmission, high levels of hospital infections were reported [2, 10].

To determine nosocomial transmission routes of a respiratory pathogen, the network of transmission opportunities within the hospital, hereon called the contact network must be considered. The routine running of a hospital in the UK, comprising wards, bays and side rooms, can give rise to additional structural and network transmission opportunities. Partitioning of patients into wards may be by diagnosis, severity of illness, or by infection status during an outbreak. Patients may have direct contact with staff, visitors and patients in the same ward, but also indirect contact with other patients and staff through shared equipment and objects, relevant to fomite transmission, airflow, relevant to airborne transmission, and staff acting as vectors for infectious diseases. Modelling studies have been conducted to investigate nosocomial transmission routes of numerous pathogens, most commonly *Methicillin-resistant Staphylococcus aureus* (MRSA) and *Vancomycin-resistant Enterococcus* (VRE) [11]. While transmission from patients to patients and healthcare workers to patients has been identified, there remains a lack of consensus as to the primary routes of nosocomial infection [7, 12, 13].

One of the challenges of identifying transmission routes in a hospital is distinguishing between hospital-acquired and community-acquired infections. Knight et al. report large uncertainty in the classification of nosocomial and community-acquired infections of SARS-CoV-2 during the first part of 2020 [14]. Routinely collected hospital data tends to record a timestamp of when a patient tested positive for an infection. For COVID-19, this provides a marker as to when a patient was most likely to be infectious. However, the time that a patient became infected is unobserved, and the time a patient recovered from infection and ceased to be infectious is also often unobserved. This makes the inference for nosocomial models much more difficult and computationally intensive, as these unobserved epidemiological events need to be mapped on to the structured landscape given by the composition of the hospital.

Generally, surveillance data only captures a partially observed epidemic process, and therefore requires a form of data augmentation to estimate unobserved epidemiological event times. O'Neill et al. and Gibson and Renshaw detail Bayesian data augmentation approaches to inference of general stochastic epidemic models, where unobserved event times are treated as parameters to be estimated [15, 16]. This has proved to be a popular method for conducting inference on partially observed epidemic models [17, 18]. A drawback of this method is that repeatedly calculating the likelihood can become extremely costly computationally and its use is therefore limited by population size. As an alternative method, McKinley et al. demonstrate using approximate likelihood ratios to conduct inference on partially observed epidemics, though this relies on repeatedly simulating the epidemic [19]. Previous studies which model nosocomial transmission have used data augmentation to handle unobserved infection times [20, 21]. However, these studies do not include a time-varying contact network parameter to model interactions between staff and patients.

Here we use routinely collected data from a large acute-care hospital in the UK to quantify the temporal and network dynamics of nosocomial transmission of SARS-

CoV-2. We demonstrate a Bayesian approach to conducting inference with a time-varying covariate - a fine-scale patient-staff contact network - to estimate unobserved epidemiological event times and provide insight into within-hospital transmission dynamics.

4.3 Covariate Data

Patient testing data from a large acute-care hospital in the UK was used in conjunction with staff rota data to identify routes of nosocomial transmission of SARS-CoV-2 during the first wave of the COVID-19 pandemic. A patient pathway was developed with a colour coded ward system to separate patients based on their SARS-CoV-2 infection status; Figure 4.1. Universal testing for SARS-CoV-2 was not available for patient admissions or staff during our study period. Patient diagnosis was therefore based on clinical suspicion of COVID-19 followed by a confirmatory test. Some wards (e.g. the isolation ward) could be assigned different colour areas within a single ward based on the availability of side rooms (enclosed patient rooms), which might be treated differently than communal areas. Ward colours could also change over time in view of capacity demands. PPE guidance for staff differed by ward colour. Our study period covered a four week time span from 12 April 2020 to 10 May 2020. During this time, 3,816 staff worked at least one shift at the hospital, consisting of 2,948 healthcare staff, 232 medical doctors and 636 ancillary staff. There were N = 2,981 patients admitted (including day attenders) in our study period to p = 55 wards.

Anonymised patient location, staff work patterns and SARS-CoV-2 testing results were extracted from routine trust electronic records. The study received HRA approval via IRAS and trust approval as a research project, accessing only routinely collected anonymised data (IRAS ID 288257).

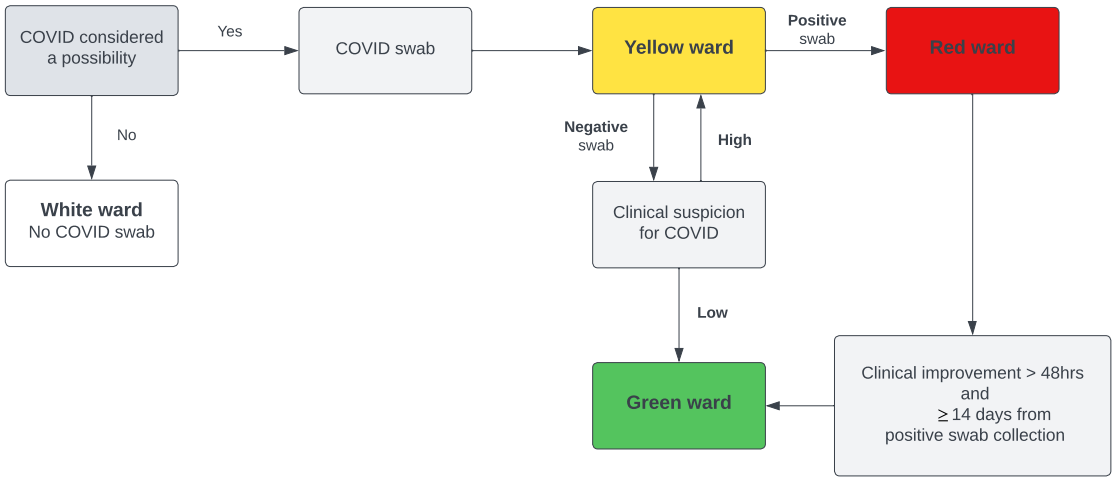

Figure 4.1: Patient pathway from initial admission to ward allocation, by suspected and confirmed SARS-CoV-2 infection status, during the study period.

4.3.1 Hospital Network

Patient movements (ward transfers, admission and discharges) and staff shift times were recorded continuously in time. However, the recorded time of these movements were not likely to be exact in practice, and the volume of changes to the patient and staff structure leads to prohibitively large data structures in RAM. Hence, patient and staff movements were aggregated to one hour time intervals to form the basis of our discrete-time contact network, comprised of T = 681 time steps.

The hospital contact network is represented in three distinct ways. Firstly, a weighted ward connectivity tensor, C, of shape $[p \times p \times T]$, where element c_{qrt} is the connectivity between ward q and r at time t, in units of number of staff working across multiple wards. This primarily consists of doctors who are assigned to a group of wards per shift. Similarly, we define a spatial adjacency tensor of shape $[p \times p \times T]$ denoted W, where w_{qrt} is the number of staff using the kitchen *if* the wards share a kitchen, and 0 otherwise. For example, in Figure 4.2, Wards A and B share kitchen A, so w_{abt} equals the total number of staff allocated to side

A (Wards A-F) on that floor of the hospital. Lastly, a membership tensor of shape $[N \times p \times T]$ denoted M, where $m_{iqt} = 1$ if individual i is a member of ward q at time t and zero otherwise.Visualisations of the contact network can be found in the supplementary material; Figure 4.7.

We introduce the dot subscript notation to indicate a slice of a tensor. For example, $C_{q\cdot t}$ refers to a slice of tensor C along the second axis, which can be thought of as a vector representing the connectivity between ward q and all hospital wards ($\forall r$) at time t.

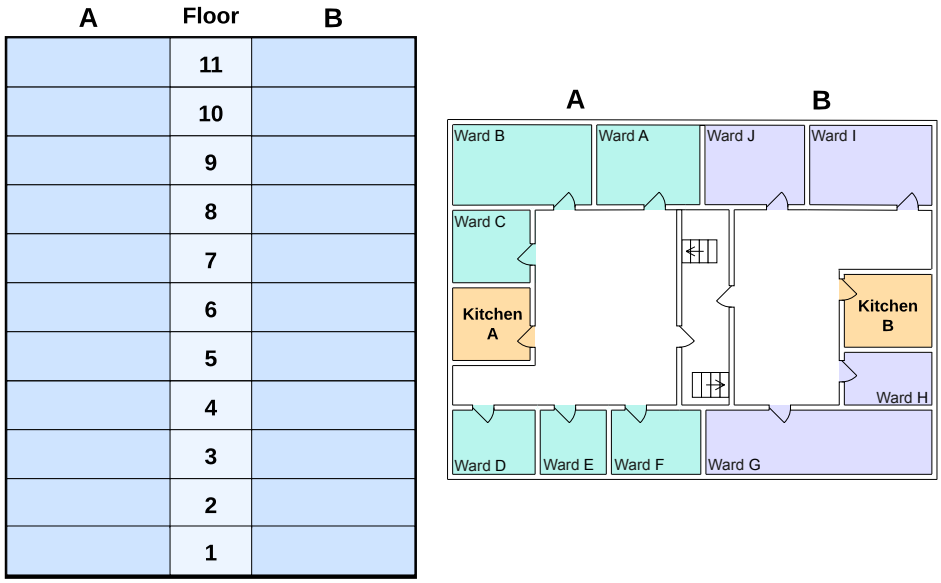

Figure 4.2: A schematic illustration of the hospital layout. Each floor of the hospital has two kitchens, one on side A and one on side B. The spatial adjacency tensor W, as defined previously, would consider all wards on the the same floor and side of the hospital as connected.

4.4 Modelling

In this section we first describe the model structure in terms of the data generating process, before outlining our approach to Bayesian inference with the appropriate likelihood function, posterior distribution and MCMC algorithm.

4.4.1 Model Structure

Transmission of SARS-CoV-2 is represented here by a stochastic individual-level, continuous-time state-transition model. At any point in time, patients are considered to belong to one of four mutually exclusive epidemiological states: *susceptible to infection (*S*)*, *infected but not infectious (*E*)*, *infected and infectious (*I*)*, or *recovered/removed from the population (*R*)*.We denote the sets of individuals in each state at time t as *S(t)*, *E(t)*, *I(t)*, *R(t)* respectively, and for convenience write $X(t) = \{S(t), E(t), I(t), R(t)\}$. Any particular patient, i is assumed to progress between the states according to the transitions [SE], [EI], and [IR] (where [PQ] denotes a transition from state P to state Q) with transition rates $\lambda_i^{\text{SE}}(t)$, $\lambda_i^{\text{EI}}(t)$, and $\lambda_i^{\text{IR}}(t)$ respectively, as defined below.

At time t, the rate at which patient $i = 1, \ldots, N$ becomes infected, i.e. $\lambda_i^{\text{SE}}(t)$, is assumed to be a function of the time-evolving infectious landscape. Within the hospital, we assume that i experiences infectious pressure from four separate sources: other patients in the same ward; patients in other wards connected by staff being assigned to more than one ward; the spatial structure of wards connected by adjoining kitchens and constant 'background' infection, representing sources of infection not explicitly modelled by the hospital structure. Within-hospital routes of infection are represented by our patient-staff contact networks (C, W, M). Within the continuous-time model, changes to the patient-ward and staff-ward structures (either via patient movements or staff-ward allocation) are assumed to occur at discrete intervals of one hour, aggregating the precisely time-stamped patient movement and staff shift data to the hour. Continuous-time was used to model the epidemic process to account for the the fluidity of events in the hospital through time, thus avoiding the necessary act of choosing a time-step. In addition, this leads to more efficient sampling from the posterior distribution of censored event times, avoiding large jumps in posterior density that otherwise occur in discrete-time systems.

The rate at which an individual transitions from state S to E can be described as a time-dependent infectious pressure. We assume that infections occur at points of a right-continuous time inhomogeneous Poisson process at a rate equal to the sum of infectious pressure on susceptibles immediately before that time point. The infectious pressure is density dependent and is defined at the individual-level. An individual can either be present in the hospital H ($i \in H$) and admitted to ward q ($i \in q$), or in the community ($i \notin H$). The infectious pressure on a susceptible individual i ($i \in S(t)$) at time t can be defined as:

$$\lambda_i^{\text{SE}}(t) = \begin{cases} \beta_1 I_q(t) + (\beta_2 \vec{C}_{q\cdot t} + \beta_3 \vec{W}_{q\cdot t})^\intercal \cdot \vec{I}(t) + \beta_4 & \text{if } i \in S(t), i \in q, i \in H \\ \beta_5 & \text{if } i \in S(t), i \notin H \\ 0 & \text{otherwise} \end{cases} \tag{4.1}$$

where β_1 denotes the transmission rate for within-ward mixing and $I_q(t)$ represents the number of infected individuals on ward q at time t. β_2 and β_3 are transmission rates for between-ward mixing, with $\vec{C}_{q\cdot t}$ denoting the connectivity between ward q and all other wards at time t, and $\vec{W}_{q\cdot t}$ describing the connectivity between ward q and all other wards by ward proximity at time t. $\vec{I}(t)$ denotes a vector whose q_{th} entry is equal to $I_q(t)$ and thus represents the number of infected individuals on each ward at time t. β_4 is a background hospital transmission rate. To allow for individuals to be infected before and between hospital admissions, we define a constant infectious pressure β_5 that is exerted on to susceptibles in the community.

Similarly, we can define an individual's [EI] transition rate as follows:

$$\lambda_i^{\text{EI}}(t) = \begin{cases} \alpha & \text{if } i \in E(t) \\ 0 & \text{otherwise} \end{cases} \tag{4.2}$$

where $i \in E(t)$ represents an individual i residing in the E state at time t.

Likewise, an individual's [IR] transition rate is defined as:

$$\lambda_i^{\text{IR}}(t) = \begin{cases} \gamma & \text{if } i \in I(t) \\ 0 & \text{otherwise} \end{cases} \tag{4.3}$$

where $i \in I(t)$ represents an individual i in the I state at time t.

Thus the [EI] and [IR] transition rates are assumed to be constant across individuals and time, and for identifiability reasons we fix $\alpha = 1/4\,\text{day}^{-1}$ and $\gamma = 1/5\,\text{day}^{-1}$ respectively [22].

The epidemic process is assumed to be Markovian. The data generating process is outlined in Algorithm 2.

Algorithm 2: Gillespie's Direct algorithm

Input: initial state $X(0) = \{S(0), E(0), I(0), R(0)\}$, wallclock time $t = 0$, contact matrix change time $s = 0$, transition rate functions $\lambda^{\text{SE}}(t), \lambda^{\text{EI}}(t), \lambda^{\text{IR}}(t)$

while $\sum_{i=1}^{N} [\lambda_i^{\text{SE}}(t) + \lambda_i^{\text{EI}}(t) + \lambda_i^{\text{IR}}(t)] > 0$ **do** `// transition rates are greater than 0`

- $t^\star \sim \text{Exponential}\left(\sum_{i=1}^{N} [\lambda_i^{\text{SE}}(t) + \lambda_i^{\text{EI}}(t) + \lambda_i^{\text{IR}}(t)]\right)$ `// draw time to next transmission event` ;
- **if** $t + t^\star > s$ **then**
  - Set $t := s$;
  - Set $s := s + \frac{1}{24}$; `// advance` s `by 1 hour`
  - Set $k := 3N + 1$;
- **else**
  - $k \sim \text{Discrete}\left(\frac{\vec{h}(t)}{\sum_{i=1}^{3N} h_i(t)}\right)$ `// Choose event`
  - where $\vec{h}(t) = (\lambda_1^{\text{SE}}(t), \ldots, \lambda_N^{\text{SE}}(t), \lambda_1^{\text{EI}}(t), \ldots, \lambda_N^{\text{EI}}(t), \lambda_1^{\text{IR}}(t), \ldots, \lambda_N^{\text{IR}}(t))^T$;
  - Let $j = \lfloor k/N \rfloor$, $i = (k \bmod N)$; `// enumerate transition and individual`
  - **switch** j **do**
    - **case** 0 **do** `// S->E`
      - $S(t + t^*) = S(t) - \{i\}$;
      - $E(t + t^*) = E(t) + \{i\}$;
    - **case** 1 **do** `// E->I`
      - $E(t + t^*) = E(t) - \{i\}$;
      - $I(t + t^*) = I(t) + \{i\}$;
    - **case** 2 **do** `// I->R`
      - $I(t + t^*) = I(t) - \{i\}$;
      - $R(t + t^*) = R(t) + \{i\}$;
  - **end**
  - $t := t + t^\star$;
- **end**
- Record t, k ;

end

Result: The ordered lists **t**, **k**

4.4.2 Bayesian Inference

A Metropolis-within-Gibbs Markov chain Monte Carlo (MCMC) algorithm is used to estimate the transmission rate parameters $\theta = \{\beta_1, \beta_2, \beta_3, \beta_4, \beta_5\}$ and the unobserved [SE] and [IR] state transition times for each infection event, denoted t_{SE} and t_{IR} respectively, according to the method of Jewell et al. [18]. Each infection transition event is associated with an observed [EI] transition time, where an individual's [EI] transition time is assumed to be two days before the collection of their first positive swab. A sensitivity analysis was conducted to assess the robustness of the findings given different delays between a patient's first positive test and their [EI] transition time. The inclusion of a time-varying covariate, the contact networks, in the continuous-time model adds to the complexity of computing the likelihood. We define the likelihood of observing $(\mathbf{t}, \mathbf{k})$ transitions, in terms of an ordered event list, where an event is considered to be a transition event (transitioning between states: S$\rightarrow$ E, E$\rightarrow$ I or I$\rightarrow$ R) or a hospital network update, as follows:

$$f(\mathbf{t}, \mathbf{k} \big| X(0), \theta, C, W, M, t_0 = 0) = \prod_{l=1}^{|\mathbf{t}|} \left[h_{k_l}(t_l) e^{-\sum_{i=1}^{3N} h_i(t_l)(t_l - t_{l-1})} \right] \tag{4.4}$$

where $\mathbf{t}$ and $\mathbf{k}$ are the (time-ordered) lists of event times and indices as in Algorithm 2, $X(0)$ denotes the initial conditions, θ the parameters as above, and covariate data C, W and M. $\vec{h}(t_k)$ is a vector of length 3N+1 of hazard rates for the 3 transitions, with N = 2,981 individuals and a covariate marker to indicate a hospital network update. That is, $\vec{h}(t) = (\lambda_1^{\text{SE}}(t), \ldots, \lambda_N^{\text{SE}}(t), \lambda_1^{\text{EI}}(t), \ldots, \lambda_N^{\text{EI}}(t), \lambda_1^{\text{IR}}(t), \ldots, \lambda_N^{\text{IR}}(t), 1)^\intercal$ similar to Algorithm 2. Thus $h_{k_l}(t_l)$ denotes the hazard rate for the k_lth transition at time t_l. Note that the trailing 1 corresponds to a contact matrix change (a hospital network update), such that for a timepoint at which the contact matrices change the likelihood term represents the probability of surviving an epidemiological state transition in the preceding time interval.

The following prior distributions were chosen for all transmission parameters

$(\beta_1, \beta_2, \beta_3, \beta_4, \beta_5)$ denoted by $f(\theta)$. The prior distributions were chosen to favour small beta values which were still bound by zero.

$$\beta_i \overset{iid}{\sim} \text{Gamma}(1.1, 1000) \tag{4.5}$$

The joint conditional posterior can therefore be defined as:

$$\pi(\theta, \mathbf{t}_z, \mathbf{k}_z \big| X(0), \mathbf{t}_{-z}, \mathbf{k}_{-z}, C, W, M, t_0 = 0) \propto f(\mathbf{t}, \mathbf{k} \big| X(0), \theta, C, W, M, t_0 = 0) f(\theta) \tag{4.6}$$

where $z = ([\text{SE}], [\text{IR}])$ and $\mathbf{t}_z$ and $\mathbf{t}_{-z}$ denote partitioning of elements of the lists $\mathbf{t}$ into unobserved and observed events respectively (and similarly for $\mathbf{k}$).

As the likelihood is intractable to integration, a Metropolis-within-Gibbs MCMC algorithm is used to sample from the joint posterior. Full details of the MCMC algorithm used can be found in the supplementary materials; Algorithm 3. To evaluate chain convergence, the Gelman and Rubin potential scale reduction statistic is calculated for three chains with differing starting values drawn from the prior distribution [23]. The posterior predictive distribution is analysed visually to assess how well our modelled estimates fit the observed data.

4.4.3 Code Implementation

The model and MCMC were implemented in Python v.3.9 using TensorFlow and TensorFlow Probability for GPU acceleration [24, 25]. The model code is available at: `https://zenodo.org/doi/10.5281/zenodo.10567172`.

4.4.4 Infection Hazard Attributable Fraction

Having computed the joint posterior distribution, we are able to investigate how within-ward and between-ward dynamics contribute to nosocomial transmission. We calculate the attributable fraction, denoted AF, per infection for each transmission type; within-ward transmission β_1; between-ward transmission rates β_2 and β_3; background hospital transmission rate β_4 and community transmission rate β_5, i.e. the proportional contribution of each transmission type to each infection. For example, the attributable fraction for infected individual i from between-ward transmission on ward q can be defined as follows:

$$P(AF_{i_{bw}}) = \frac{(\beta_2 \vec{C_{q \cdot t}} + \beta_3 \vec{W_{q \cdot t}})^\intercal \cdot \vec{I}(\widetilde{t})}{\lambda_i^{\text{SE}}(\widetilde{t})},\ \widetilde{t} = \mathbf{t}_{\text{idx}(\mathbf{k}=i)} \tag{4.7}$$

where idx($\mathbf{k} = i$) denotes the index of the element of $\mathbf{k}$ which is equal to i (i.e. corresponding to the [SE] event for individual i).

The attributable fractions are aggregated for each set of sampled parameter estimates to compute the mean attributable fraction and 95% Credible Intervals (95%CrI) for each transmission type per infection. Similarly, using the sampled parameter estimates we can identify which wards and associated ward colours nosocomial infections take place in.

4.5 Results

The results presented are of SARS-CoV-2 infections recorded in a UK hospital one month in to the first wave of the pandemic. In a population size of 2,981 patients from 12 April 2020 to 10 May 2020, we have identified 131 infection events, [EI] transitions, which are a combination of community-acquired and nosocomially-acquired infections. Historic testing data, 3 days prior to our study start date,

was used to identify the 58 patients which are considered to be initially infectious, residing in the I state, in our model.

4.5.1 Parameter Estimation

To estimate our parameters of interest ($\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, t_{SE}, t_{IR}$), we ran a Metropolis-within-Gibbs MCMC algorithm for 11,000 iterations removing the first 1000 samples as burn-in. Convergence across all parameters is seen which is confirmed by the potential scale reduction statistic computed for three independent chains; supplementary materials Figure 4.8, Table 4.2. The pairwise correlation of transmission parameters was also assessed; supplementary materials Figure 4.9. In Figure 4.3, we present the kernel density estimates for each of the transmission rate parameters and for a random sample of transition event times. A constraint when drawing event times is that an individual who tested positive in hospital must have been admitted to hospital before their [IR] transition time. As the observed [EI] transition times are set to two days prior to a patient's first positive hospital test, there may be a period of time pre-admission which is unexplored by the MCMC algorithm, as seen with sample 1 and 6 of Figure 4.3B.

The posterior predictive distribution is used to evaluate how well our model fits the observed data. An epidemic is simulated forward using the data generating model (Algorithm 2) for each of the 10,000 sets of parameter estimates sampled by the MCMC algorithm. We compare the number of [EI] transitions occurring each day from the simulations with the observed data; Figure 4.4. With the exception of one peak of [EI] transitions on day 12, the observed data sits comfortably within the 95%CrI of the aggregated simulation data. Figure 4.4 shows that in a given simulation, the peaks and troughs of the number of [EI] transitions per day may similarly sit outside of the credible interval.

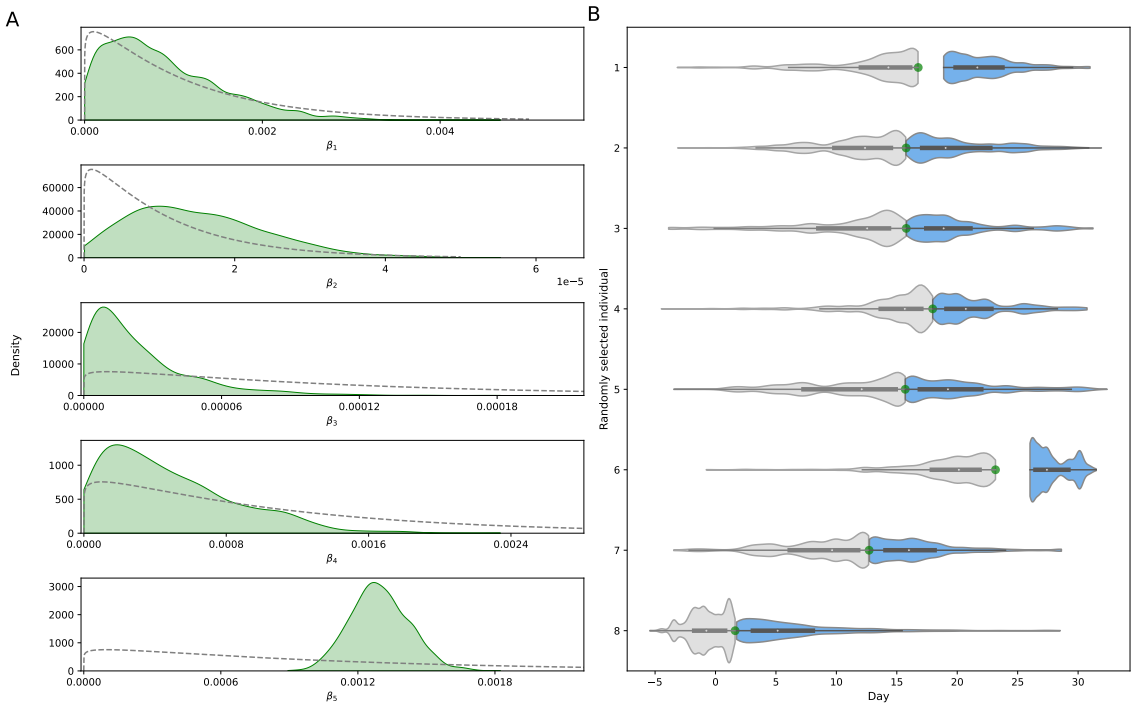

Figure 4.3: (A) Kernel density estimates for each transmission rate parameter. Dashed lines represent the associated prior distribution.(B) Density estimates of [SE] and [IR] transition times for a random sample of eight observed [EI] events. Distributions for the associated [SE] and [IR] transition times are shown in grey and blue respectively. The observed [EI] transition time for each randomly selected individual is represented as a green circle positioned between the [SE] and [IR] distributions.

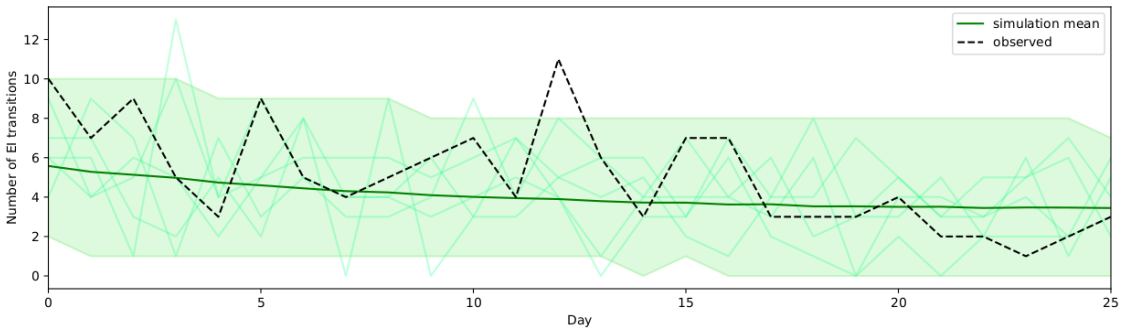

Figure 4.4: Posterior predictive distribution of the number of EI transitions each day formed from 10,000 stochastic simulations over the joint posterior. Mean simulated number of [EI] transitions in dark green with 95% Credible Interval as the shaded area. Five individual simulations are displayed in the faint green lines. The observed number of [EI] transitions per day is shown by the dashed line.

4.5.2 Nosocomial Transmission Routes

One of the parameters of interest is the associated [SE] transition time for each infection. Estimating this enables us to identify which infections were most likely to be contracted nosocomially rather than in the community. Moreover, for each infection we analyse how the different types of transmission contribute to the infectious pressure exerted on an individual directly before their infection event. Hospital-acquired infections are likely for 15.3% (20/131) of detected infections. These infections have a mean community attributable fraction of less than 0.5, with the majority (15/20) having a mean community attributable fraction of less than 0.1; Figure 4.5. Between-ward transmission is the highest mean attributable fraction for 13 of the 20 infections identified as nosocomial. Infection events of patient 4 and patient 15 have the highest mean attributable fractions for between-ward transmission of 0.72 (95%CrI 0.12 to 0.99) and 0.83 (95%CrI 0.29 to 1.00); Figure 4.5. Four nosocomial infections have a mean within-ward attributable fraction of zero, indicating that these infections occurred when there were no infectious individuals admitted to the individual's ward.

To assess the plausibility of our model output, we further explored the hospital data associated with the 15 individuals who were identified as most likely to have hospital-acquired infections. One individual appeared to have been tested on the day they were admitted to hospital; on further inspection, this individual had been discharged from a 15-day hospital spell two days prior to readmission which was captured within our study period. The other 14 individuals identified had been admitted to hospital for an average of 13.3 days (95%CrI 6.62 to 22.76) before their positive test sample was collected.

Using our modelling framework we are able to identify the wards in which nosocomial infections were likely to occur. Nosocomial infections are identified based on whether an individual's [SE] transition time is during a hospital admission spell. For each

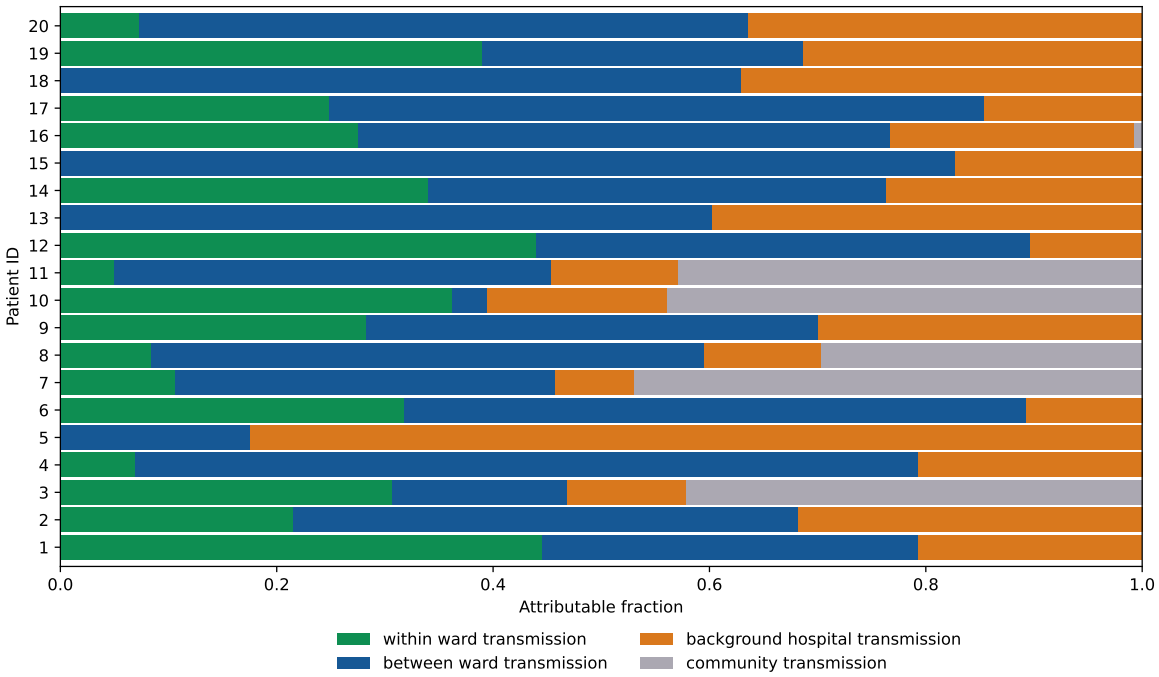

Figure 4.5: Mean attributable fraction for each transmission type per infection event for 10,000 posterior samples. Infection events are displayed if they have a mean community transmission attributable fraction less than 0.5

set of sampled parameters, the percentage distribution of nosocomial infections by ward is calculated and aggregated for the 10,000 samples. We find that four wards (<10% of the wards) account for the locations of 63.1% of nosocomially-acquired infections, with wards 13 and 23 accounting for 21.1% (95%CrI 11.76 to 30.00) and 17.5% (95%CrI 10.00 to 25.00) respectively. Similarly, we compute the percentage distribution of nosocomial infections by ward type; Table 4.1. Nosocomial infections are recorded most frequently in green wards (40.0%, 95%CrI 33.33 to 60.00) and wards which are assigned three colours of green, red and yellow (13.6%, 95%CrI 6.67 to 25.00).

The modelling suggests that the infectious pressure exerted on an individual in a ward changes considerably over time; this may be explained by the dynamic contact network. The within-ward infectious pressure will increase with the number of infectious individuals on the ward. Similarly, if there is an increase of infectious

individuals on wards considered to be connected, the between-ward infectious pressure will increase. Interestingly, the source of infectious pressure exerted on individuals in the four wards which recorded the highest number of nosocomial infections was also dynamic; Figure 4.6. For the first 23 days of our study period, each of these four wards were classified as green wards with the exception of ward 23 which was a mixed ward, either classified as green-red-yellow or green-red ward. We find infectious pressure dynamics varied by ward and ward colour; Figure 4.6.

Individuals on wards 13 and 22 would have experienced clear peaks of within-ward infectious pressure (these are visualised in Figure 4.6 by the peaks of green dots, within-ward pressure, on days 12 to 15 and days 1 to 4 respectively). Whereas an individual on ward 33 would have experienced an infectious pressure driven by between-ward dynamics. This can be seen in Figure 4.6 with between-ward infectious pressure, shown by the blue dots, consistently contributing most to the infectious pressure. Similarly, individuals on ward 23 would have primarily experienced infectious pressure driven by between-ward dynamics. However, we can see the interchange of between-ward infectious pressure and within-ward pressure (shown by the blue and green dots in Figure 4.6) towards the end of the study period.

We conducted a sensitivity analysis of the delay between an individual's [EI] transition time and an individual's first positive test in hospital to assess the robustness of our results; supplementary materials Figures 3-5. Delays of one day and three days were considered in addition to the two days used for our primary analysis. A byproduct of altering the [EI] transition times is that the initial conditions of our model change, with a different number of individuals initially residing in the I state; supplementary materials Table 3. The number of nosocomial infections identified therefore varied. However, between-ward dynamics were found to be the key driver of infectious pressure in each analysis, and the same four wards were identified as housing the most nosocomial infections.

Table 4.1: Percentage of total nosocomial infections by designated ward colour.

Ward colour	Percentage of nosocomial infections (95%CrI)
green	40.04 (33.33-60.00)
green-red-yellow	13.59 (6.67-25.00)
white	10.29 (5.56-23.53)
yellow	10.00 (5.56-20.00)
white-yellow	9.20 (5.56-20.00)
red	5.68 (5.00-9.09)
green-red	5.61 (5.00-9.09)
red-yellow	5.61 (5.26-9.09)

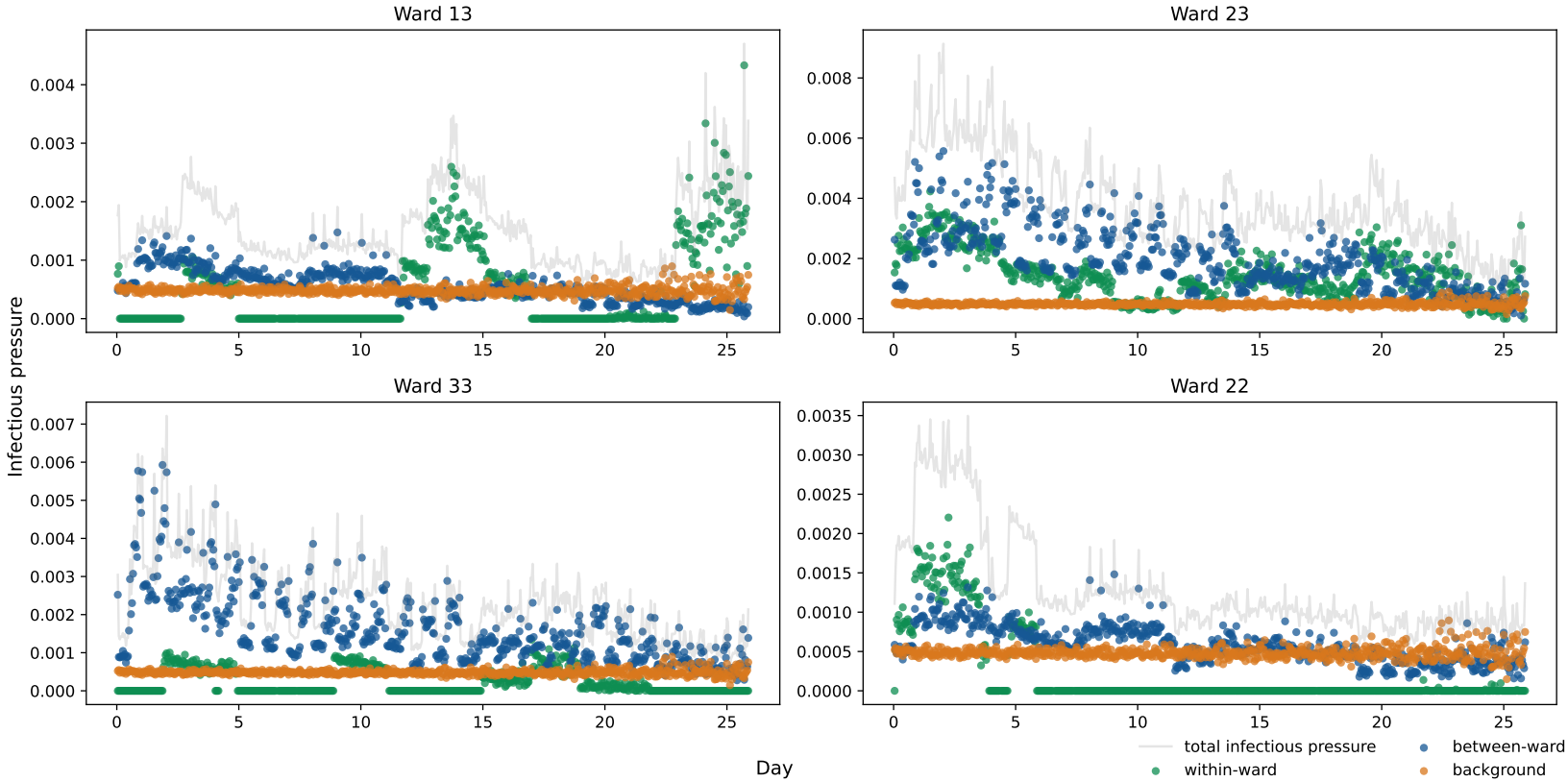

Figure 4.6: Mean infectious pressure for an individual on the stated ward at each [SE] transition, calculated for 10,000 posterior samples. The wards displayed are those with the highest mean number of nosocomial infections.

4.6 Discussion

This study provides a statistical framework for conducting inference on hospital outbreak dynamics to quantify the relative contribution of transmission routes for nosocomial infections. We have demonstrated that transmission parameters and unobserved event times can be inferred by incorporating a discrete time-varying patient-staff contact network and testing data into a continuous-time stochastic epidemic model. By estimating these parameters we are able to disentangle the different components of infectious pressure and identify routes of nosocomial transmission. While computing the likelihood for these models can be computationally intensive, we found that with GPU acceleration (1 x NVIDIA V100 32GB), 11,000 iterations of our MCMC took approximately 61 minutes. A Bayesian approach enables us to assess uncertainty in our parameter estimates by simulating forwards stochastically from the joint posterior.

We estimate that 15.3% of identified SARS-CoV-2 infections in patients identified in hospital were nosocomially-acquired, other studies which examined hospital infections during the first wave of the pandemic have reported similar findings [2, 6, 12, 26]. Nonetheless, we expect that this is an underrepresentation of the full extent of within-hospital transmission. Asymptomatic infections are not captured in this study as universal testing of admissions had not yet been introduced. Similarly, if a patient had been discharged while unknowingly infected, their infection would be unrecorded. Additionally, we have not accounted for infections which may have been missed due to individuals not being tested due to unobserved symptoms, or due to imperfect sensitivity and specificity of SARS-CoV-2 tests. A simple method which is used to identify nosocomial infections is to consider the time interval between admission and symptom onset. However, this can lead to an infected patient with multiple hospital spells in quick succession being misclassified as a community-acquired infection [14]. The modelling approach presented here enabled us to identify an individual who was discharged from a long hospital stay and re-admitted two days

later, as having a likely nosocomially-acquired infection.

We found that the total infectious pressure exerted on an individual in a ward changes over time, as does the primary source of transmission. When comparing wards which housed nosocomial infections, it was clear that infectious pressure dynamics varied greatly by ward. Moreover, these dynamics varied across wards which were designated with the same ward colour. For most nosocomial infections, the most likely source of infection was captured by between-ward dynamics, suggesting that the patient pathway implemented was successful in separating susceptible patients from infectious patients. This finding is supported by several other studies which found that healthcare workers were a likely source of nosocomial infection [13, 27, 28]. Evans et al. found that indirect transmission from infected patients was the most likely route for nosocomial transmission, where indirect transmission may happen through healthcare workers acting as vectors for transmission or through fomite transmission [7]. It is difficult to pinpoint the exact cause of between-ward transmission without further infection data. If universal testing had been implemented at this point, we could include staff as individuals in the model which may help to further disentangle transmission types. Between-ward transmission in our model is driven by the staff-patient contact network, which indicates that the hospital contact network is a key route for nosocomial transmission.

Nosocomial infections were most likely to be contracted within four wards. These wards tended to be wards designated as green on the patient pathway. Patients would be placed in green wards after receiving a negative test result for SARS-CoV-2 and if they were also of low clinical suspicion of having COVID-19, or if they were being stepped down from a red ward. This suggests that infected individuals may have been missed due to test sensitivity or remaining infectious after being stepped down from a red ward. Separating patients who were not suspected of having COVID-19 on admission appears to have been effective, with

fewer nosocomial infections occurring on white wards than green wards. However, undetected asymptomatic transmission may be more likely to occur on white wards as patients were not tested on admission.

There are several limitations to our approach. Firstly, in order to conduct the inference we fix an individual's [EI] transition time to be two days prior to their first positive test. This allows for a delay between patients becoming infectious and being tested [29]. The effect of this assumption was tested in our sensitivity analysis (supplementary materials), and although we note that the initial conditions, β_1, β_2 and β_3 do change in response to changing the delay, the overall conclusion of which wards presented the highest risk is robust. A feasible innovation to our inference algorithm would be to estimate this interval, along with formal inference on the initial conditions, as part of the MCMC. We do not consider reinfections due to the short length of our study period and the low risk of reinfection for the wild-type virus.[30] Additionally, we assume that individuals progress from exposed to infectious and infectious to recovered at constant rates. As our study period is during the initial phase of the pandemic, vaccination status and virus variants were not considered. The data augmentation methodology used could be developed further to better cope with unknown disease status on admission and occult infections at the end of the time window. Furthermore, the force of infection exerted on the community could be based on prevalence estimates. Community prevalence was not well known for our study period, with estimates often based on hospital admission data, as SARS-CoV-2 tests were not readily available to the public. Nevertheless, this should be considered for future studies. We also did not explore different model structures, using the standard SEIR structure often used for COVID-19 models. Computationally, calculating the likelihood in continuous-time is expensive, it would be worthwhile to investigate the potential gains and losses of a fully discrete model and the impact of time aggregation of the movement data (aggregating to less or more than an hour).

Our findings allow us to draw some important conclusions regarding effectiveness of infection prevention and control measures in this hospital early in the pandemic. At this time, universal SARS-CoV-2 testing was not logistically possible, and a lack of isolation facilities meant that patients had to be cohorted based on COVID-19 risk. Firstly, we conclude that stratifying patient risk of SARS-CoV2 infection based on clinical assessment (primarily presence of respiratory symptoms) was successful, in that few nosocomial infections likely occurred in white (low risk) and yellow (possible COVID-19) wards. Secondly, most nosocomial infections likely occurred in green wards, and were primarily driven by between-ward transmission and hence the staff-patient contact network. These findings can inform future planning for outbreaks of respiratory pathogens and provide support for strategies to reduce staff-patient transmission, such as asymptomatic staff testing which was implemented in the NHS later in the pandemic [31].

We have presented a Bayesian approach to quantifying routes of nosocomial transmission of SARS-CoV-2 which could be applied to other respiratory infections and extended to outbreaks of bacterial infections. To our knowledge, inference methods such as the ones presented here have not been used for real-time IPC monitoring due to the computational complexity and impractical runtimes. We have demonstrated an efficient method to infer epidemiological event times and nosocomial transmission routes while accounting for the intricacies of the hospital network. This could be used to retrospectively evaluate and simulate interventions. Additionally, with further development and the appropriate infrastructure in place, we believe that methods such as these could be implemented at a similar scale during a prolonged hospital outbreak to alert IPC teams to potential hotspots of transmission and assess effectiveness of interventions.

4.7 Supplementary Material

4.7.1 Hospital Network

The hospital staff-patient contact network, as defined in the manuscript, is represented by weighted tensors in our model. Figure 1 describes the mean ward connectivity by staff working across multiple wards, C, for our study period in panel A and the mean ward connectivity by ward proximity, W, for our study period in panel B. We found that fewer wards were connected by ward proximity C, however wards were connected more strongly, with a higher weighting between connected wards when compared to W.

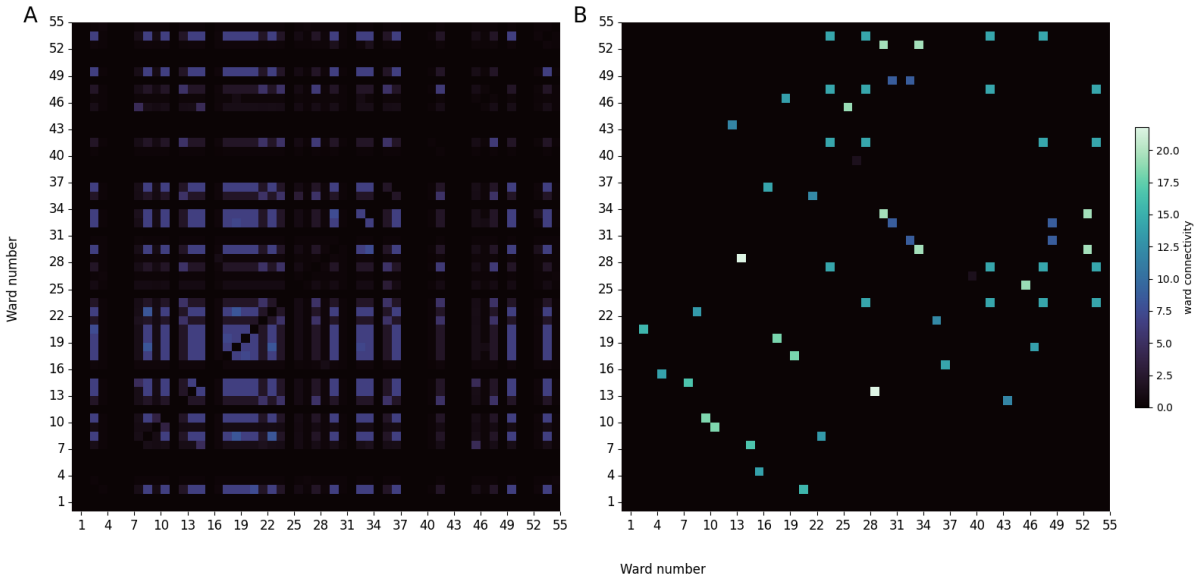

Figure 4.7: (A) Mean ward connectivity C for the study period, weighted by staff assigned to multiple wards. (B) Mean ward proximity W, weighted by the number of staff assigned to each ward.

4.7.2 MCMC algorithm

Using the Metropolis-within-Gibbs algorithm, we iteratively sample from the conditional posterior of the transmission parameters given the missing and observed

data, and the conditional posterior of the missing data given the observed data and the transmission parameters, to ultimately sample from the joint posterior of the parameters and missing data given the observed data. For each of the 11,000 iterations, we sample from the conditional posteriors for the transmission parameters, $\beta = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5)^\intercal$ in turn and approximately 10% of the unobserved state transition times [32], defined as a row matrix $U = (T_{SE}, T_{IR})$ with dimensions $dim(U) = [1, 320]$. We sample 33 $(320/10 + 1)$ event times per MCMC iteration. In algorithm 3 the notation β_i refers to the ith element of β, where β_{-i} denotes all elements other than the ith of β, similarly with the jth element of U_j. A log-normal proposal distribution is used for the β transmission parameters. The notation Normal(μ, σ) denotes the normal distribution with mean μ and standard deviation σ. The proposal variances were tuned manually through pilot runs of the MCMC algorithm.

Algorithm 3: Metropolis-within-Gibbs algorithm

Input: β, U and tuning constants: $\sigma_\beta = (\sigma_{\beta_1}, \sigma_{\beta_2}, \sigma_{\beta_3}, \sigma_{\beta_4}, \sigma_{\beta_5})$, σ_U
Output: β, U
Initialise: $\beta^{(0)}$, $U^{(0)}$, $n = 0$, $x = 0$

while $n < 11,000$ **do**
  for *i in 1 ... 5* **do**
    $\beta^* \sim Normal(\beta_i, \sigma_{\beta_i})$
    Compute
      $\alpha(\beta_i, \beta^*) = 1 \ \wedge \frac{\pi(\beta^*|\beta_{-i}, U)}{\pi(\beta_i|\beta_{-i}, U)} \cdot \frac{\beta^*}{\beta_i}$
    if $Uniform(0,1) < \alpha(\beta_i, \beta^*)$ **then**
      set $\beta_i = \beta^*$
    end
  end
  while $x < 33$ **do**
    Choose an event time to move, U_j, where
     $j = \text{DiscreteUniform}\big[0,\ dim(U)\big)$
    $U^* \sim Normal(U_j, \sigma_U)$
    Compute
      $\alpha(U_j, U^*) = 1 \ \wedge \frac{\pi(U^*|U_{-j}, \beta)}{\pi(U_j|U_{-j}, \beta)}$
    if $Uniform(0,1) < \alpha(U_j, U^*)$ **then**
      set $U_j = U^*$
    end
    $x = x + 1$
  end
  $n = n + 1$
end

4.7.3 Model Fit

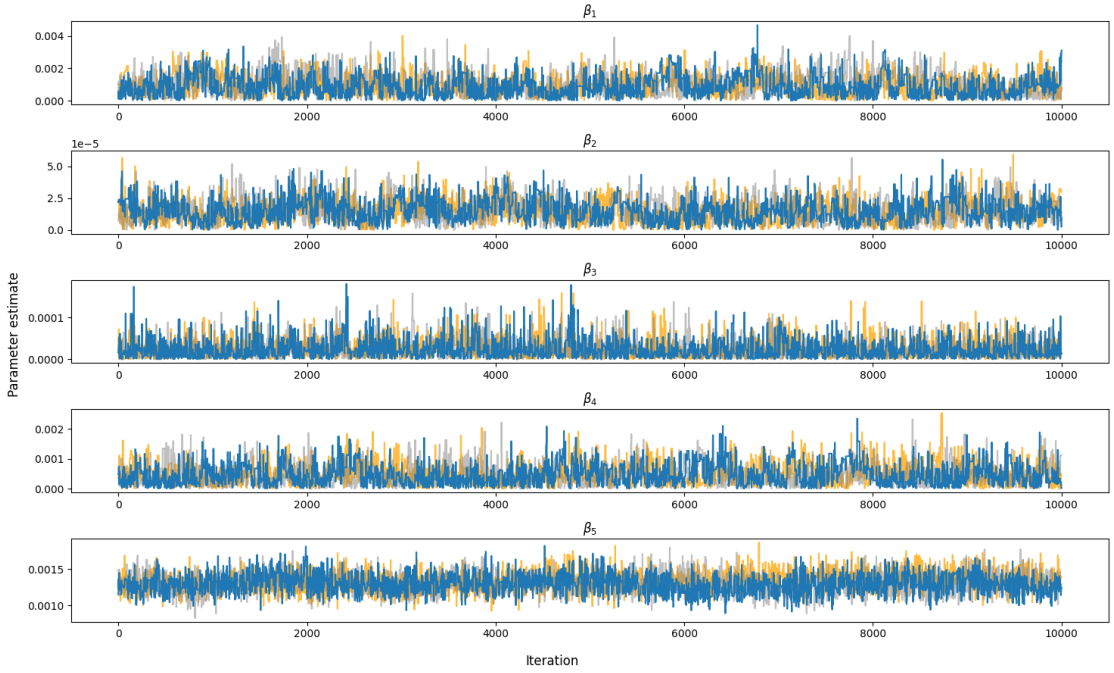

Figure 4.8: MCMC trace plots of three independent chains.

Table 4.2: Potential Scale Reduction Statistic.

Parameter	Potential Scale Reduction Statistic
β_1	1.02
β_2	1.01
β_3	1.00
β_4	1.00
β_5	1.04

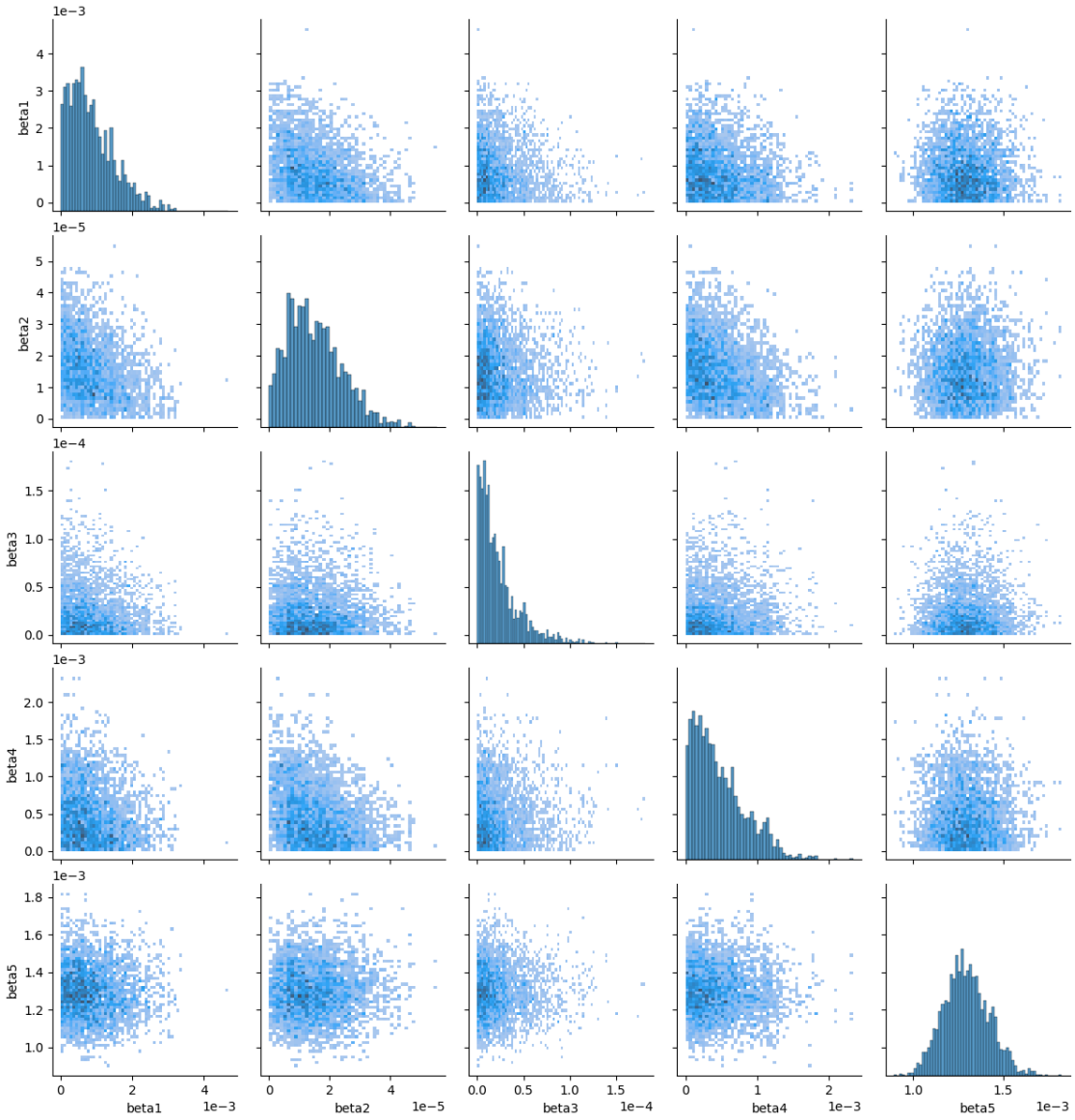

Figure 4.9: Pairwise scatter plots of transmission rate parameters with histograms for the marginal plots along the diagonal.

4.7.4 Sensitivity Analysis

We assessed how robust our results were to the choice of our observed [EI] transition times. Our primary analysis considers an individual's [EI] transition time to be two days prior to their first recorded positive test result, a 2 day delay. We subsequently conducted our analysis with a 1 day delay and a 3 day delay. Our initial conditions are sensitive to the delay chosen; Table 4.3.

Figure 4.10 presents the kernel density estimates for each of the transmission rate parameters for the three time delays. We find that the number of nosocomial infections identified varies with the length of the delay between an individual's [EI] transition time and their first positive swab, most likely due to the change in the initial conditions. We identified 26 nosocomial infections when considering a delay of one day and 17 nosocomial infections with a delay of three days. In Figure 4.11, we compare the mean attributable fraction associated with each identified nosocomial infection across the different time delays. The same four wards are identified as housing the most nosocomial infections in each analysis; Figure 4.12. The dynamics within a ward do not change considerably when different time delays are considered.

Table 4.3: Initial conditions for varying time delays between an individual's observed [EI] transition time and their first positive test result.

[EI] delay	Number of initially infected
1 day	50
2 day	58
3 day	68

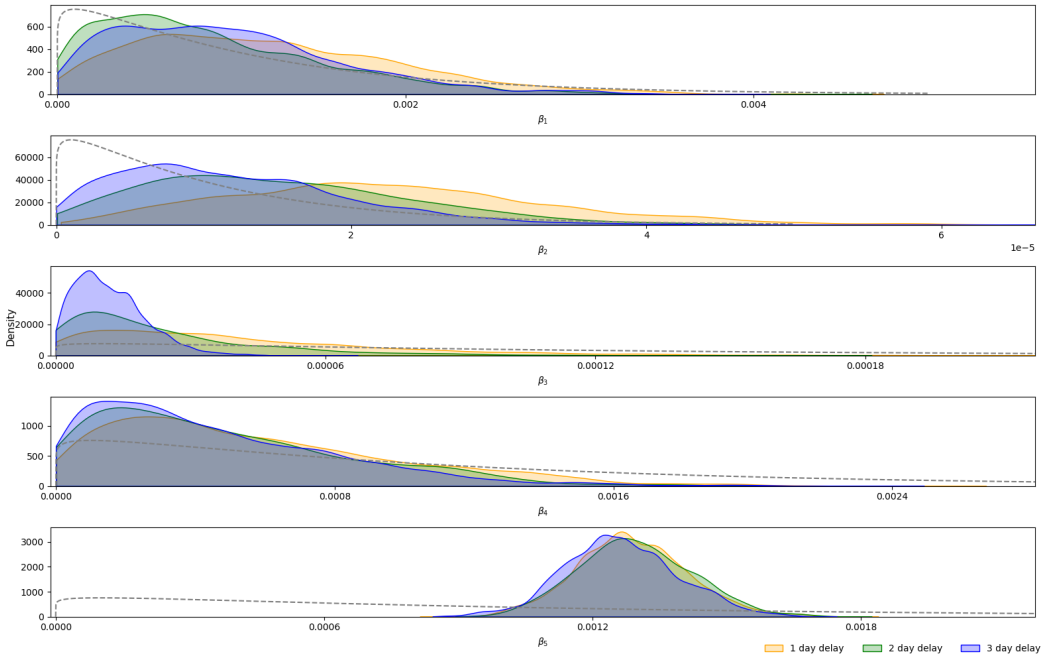

Figure 4.10: Kernel density estimates for each transmission rate parameter for varying time delays between an individual's observed [EI] transition time and their first positive test result. Dashed lines represent the associated prior distribution.

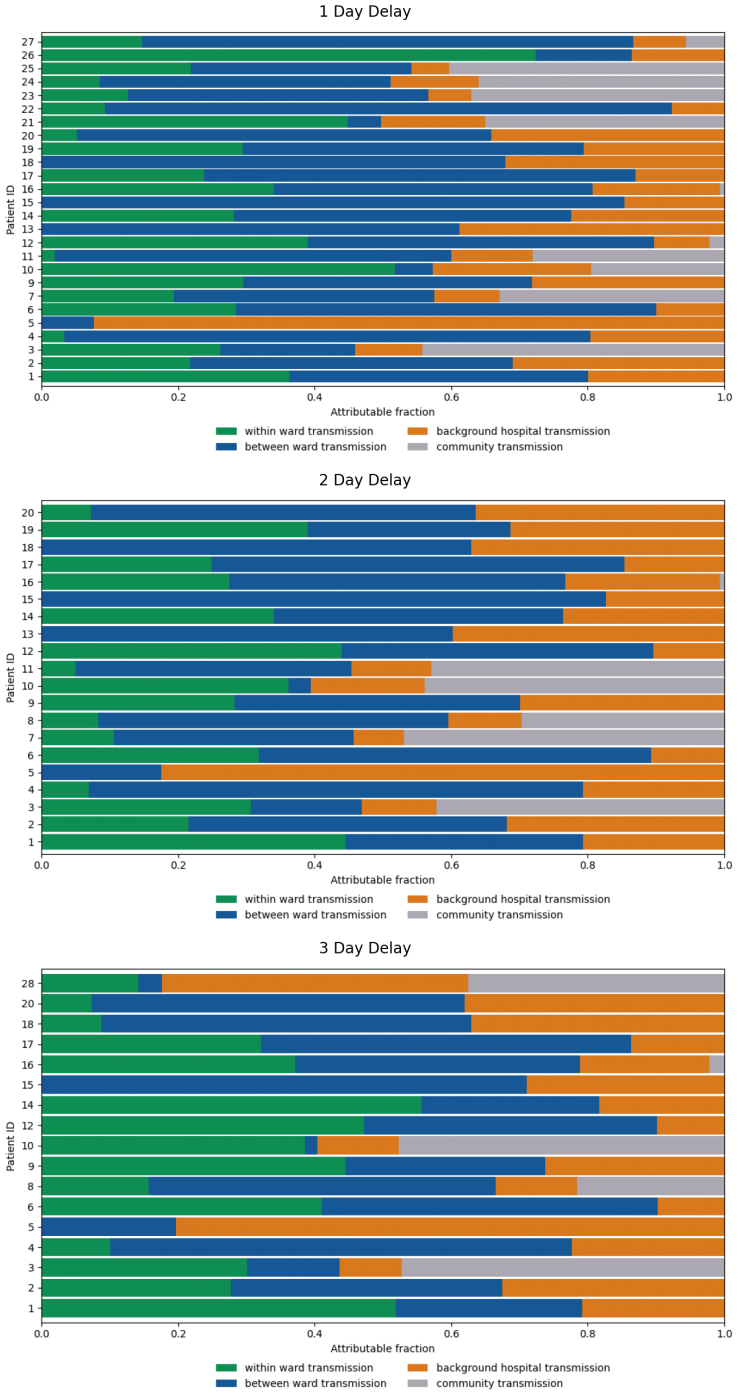

Figure 4.11: Nosocomial infections identified for varied time delays between an individual's observed [EI] transition time and their first positive test result. Mean attributable fraction for each transmission type per infection event for 10,000 posterior samples. Infection events are displayed if they have a mean community transmission attributable fraction less than 0.5. Each nosocomial infection is numbered according to the original manuscript figure (2 day delay) to enable comparison between individual infections in each analysis.

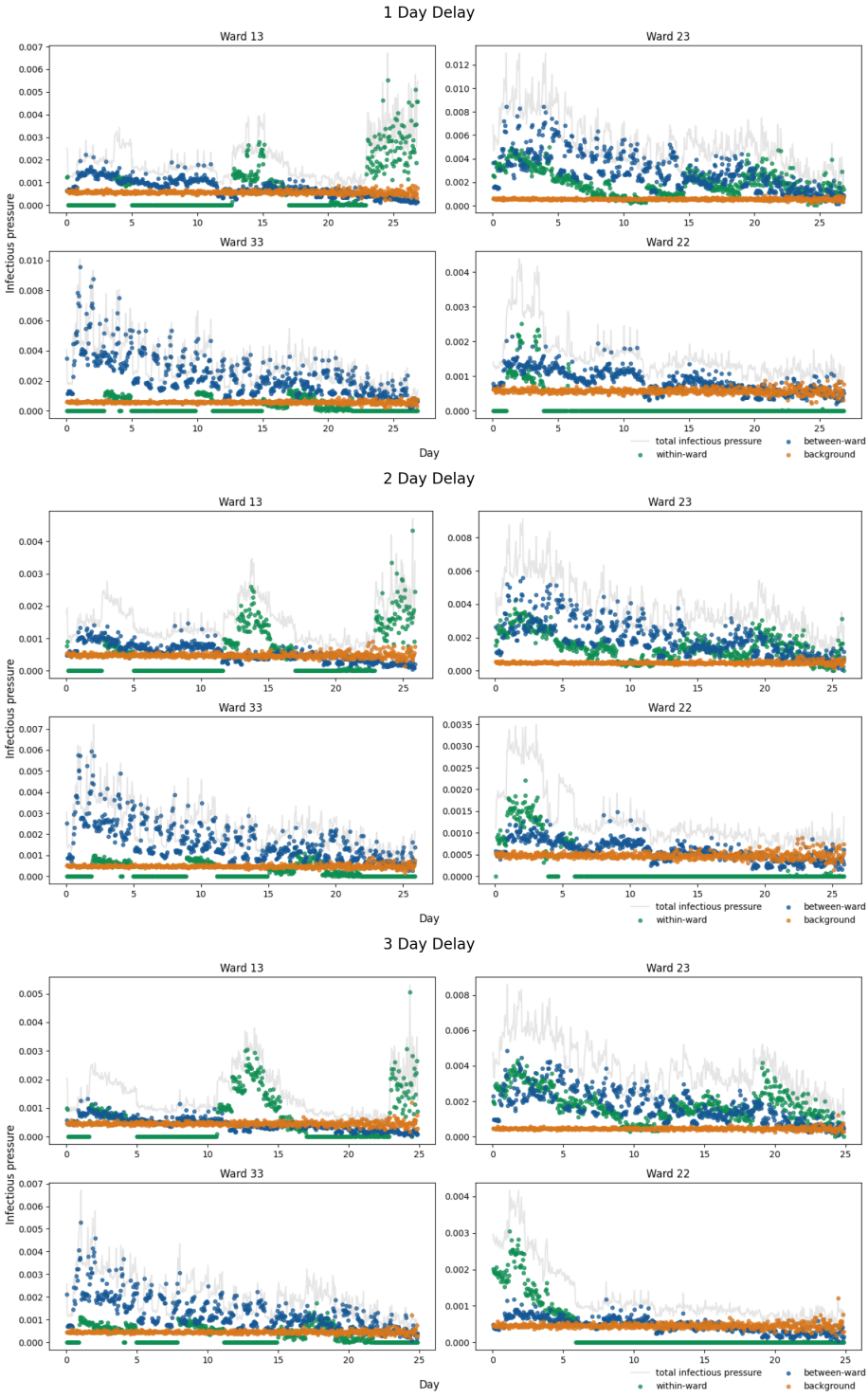

Figure 4.12: Ward dynamics for the four wards with the highest mean number of nosocomial infections shown for varied time delays between an individual's observed [EI] transition time and their first positive test result. Mean infectious pressure for an individual on the stated ward at each [SE] transition, calculated for 10,000 posterior samples.

References

[1] World Health Organization. *Report on the Burden of Endemic Health Care-Associated Infection Worldwide*. Tech. rep. 2011.

[2] Read, J. M., Green, C. A., Harrison, E. M., et al. "Hospital-acquired SARS-CoV-2 infection in the UK's first COVID-19 pandemic wave". en. In: *Lancet* 398.10305 (Sept. 2021), pp. 1037–1038.

[3] Abbas, M., Robalo Nunes, T., Martischang, R., Zingg, W., Iten, A., Pittet, D., and Harbarth, S. "Nosocomial transmission and outbreaks of coronavirus disease 2019: the need to protect both patients and healthcare workers". In: *Antimicrob. Resist. Infect. Control* 10.1 (2021).

[4] Wu, Z. and McGoogan, J. M. "Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72314 Cases From the Chinese Center for Disease Control and Prevention". In: *JAMA* (2020).

[5] Suárez-García, I., Martínez de Aramayona López, M. J., Sáez Vicente, A., and Lobo Abascal, P. "SARS-CoV-2 infection among healthcare workers in a hospital in Madrid, Spain". en. In: *J. Hosp. Infect.* 106.2 (Oct. 2020), pp. 357–363.

[6] Rickman, H. M., Rampling, T., Shaw, K., et al. "Nosocomial Transmission of Coronavirus Disease 2019: A Retrospective Study of 66 Hospital-acquired Cases in a London Teaching Hospital". en. In: *Clin. Infect. Dis.* 72.4 (Feb. 2021), pp. 690–693.

[7] Evans, S., Stimson, J., Pople, D., Bhattacharya, A., Hope, R., White, P. J., and Robotham, J. V. "Quantifying the contribution of pathways of nosocomial acquisition of COVID-19 in English hospitals". en. In: *Int. J. Epidemiol.* 51.2 (May 2022), pp. 393–403.

[8] Ward, H., Atchison, C., Whitaker, M., et al. "SARS-CoV-2 antibody prevalence in England following the first peak of the pandemic". en. In: *Nat. Commun.* 12.1 (Feb. 2021), p. 905.

[9] European Centre for Disease Prevention and Control. *Infection prevention and control and preparedness for COVID-19 in healthcare settings - second update*. 2020.

[10] GOV UK. *COVID-19: infection prevention and control (IPC)*. en. Jan. 2020. URL: `https://www.gov.uk/government/publications/wuhan-novel-coronavirus-infection-prevention-and-control`.

[11] Kleef, E. van, Robotham, J. V., Jit, M., Deeny, S. R., and Edmunds, W. J. "Modelling the transmission of healthcare associated infections: a systematic review". en. In: *BMC Infect. Dis.* 13 (June 2013), p. 294.

[12] Bhattacharya, A., Collin, S. M., Stimson, J., et al. "Healthcare-associated COVID-19 in England: A national data linkage study". en. In: *J. Infect.* 83.5 (Nov. 2021), pp. 565–572.

[13] Aghdassi, S. J. S., Schwab, F., Peña Diaz, L. A., et al. "Risk factors for nosocomial SARS-CoV-2 infections in patients: results from a retrospective matched case–control study in a tertiary care university center". en. In: *Antimicrob. Resist. Infect. Control* 11.1 (Jan. 2022), pp. 1–10.

[14] Knight, G. M., Pham, T. M., Stimson, J., et al. "The contribution of hospital-acquired infections to the COVID-19 epidemic in England in the first half of 2020". en. In: *BMC Infect. Dis.* 22.1 (June 2022), p. 556.

[15] O'Neill, P. D. and Roberts, G. O. "Bayesian Inference for Partially Observed Stochastic Epidemics". In: *J. R. Stat. Soc. Ser. A Stat. Soc.* 162.1 (1999), pp. 121–129.

[16] Gibson, G. J. and Renshaw, E. "Estimating parameters in stochastic compartmental models using Markov chain methods". en. In: *Math. Med. Biol.* 15.1 (Mar. 1998), pp. 19–40.

[17] Streftaris, G. and Gibson, G. J. "Bayesian inference for stochastic epidemics in closed populations". In: *Stat. Modelling* 4.1 (Apr. 2004), pp. 63–75.

[18] Jewell, C. P., Kypraios, T., Neal, P., and Roberts, G. O. "Bayesian analysis for emerging infectious diseases". en. In: {*Bayesian Analysis*} 4.3 (Sept. 2009), pp. 465–496.

[19] McKinley, T., Cook, A. R., and Deardon, R. "Inference in Epidemic Models without Likelihoods". en. In: *Int. J. Biostat.* 5.1 (July 2009).

[20] Forrester, M. L., Pettitt, A. N., and Gibson, G. J. "Bayesian inference of hospital-acquired infectious diseases and control measures given imperfect surveillance data". en. In: *Biostatistics* 8.2 (Apr. 2007), pp. 383–401.

[21] Kypraios, T., O'Neill, P. D., Huang, S. S., Rifas-Shiman, S. L., and Cooper, B. S. "Assessing the role of undetected colonization and isolation precautions in reducing methicillin-resistant Staphylococcus aureus transmission in intensive care units". en. In: *BMC Infect. Dis.* 10 (Feb. 2010), p. 29.

[22] GOV UK. *COVID-19: epidemiology, virology and clinical features*. en. Accessed: 2022. 2022. URL: https://www.gov.uk/government/publications/wuhan-novel-coronavirus-background-information/wuhan-novel-coronavirus-epidemiology-virology-and-clinical-features.

[23] Brooks, S. P. and Gelman, A. "General Methods for Monitoring Convergence of Iterative Simulations". In: *J. Comput. Graph. Stat.* 7.4 (Dec. 1998), pp. 434–455.

[24] Martín Abadi, Ashish Agarwal, Paul Barham, et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. 2015. URL: https://www.tensorflow.org/.

[25] Dillon, J. V., Langmore, I., Tran, D., et al. "TensorFlow Distributions". In: (Nov. 2017). arXiv: 1711.10604 [cs.LG].

[26] Carter, B., Collins, J. T., Barlow-Pay, F., et al. "Nosocomial COVID-19 infection: examining the risk of mortality. The COPE-Nosocomial Study (COVID in Older PEople)". en. In: *J. Hosp. Infect.* 106.2 (Oct. 2020), pp. 376–384.

[27] Borges, V., Isidro, J., Macedo, F., et al. "Nosocomial Outbreak of SARS-CoV-2 in a "Non-COVID-19" Hospital Ward: Virus Genome Sequencing as a Key Tool to Understand Cryptic Transmission". en. In: *Viruses* 13.4 (Apr. 2021).

[28] Thoma, R., Kohler, P., Haller, S., Maenner, J., Schlegel, M., and Flury, D. "Ward-level risk factors associated with nosocomial coronavirus disease 2019 (COVID-19) outbreaks: A matched case-control study". en. In: *Antimicrob Steward Healthc Epidemiol* 2.1 (Mar. 2022), e49.

[29] European Centre for Disease Prevention and Control. *Infection*. en. Accessed: April 2023. Sept. 2020. URL: https://www.ecdc.europa.eu/en/covid-19/latest-evidence/infection.

[30] Pulliam, J. R. C., Schalkwyk, C. van, Govender, N., et al. "Increased risk of SARS-CoV-2 reinfection associated with emergence of Omicron in South Africa". en. In: *Science* 376.6593 (May 2022), eabn4947.

[31] NHS England. *Standard operating procedure for rollout of lateral flow devices for asymptomatic staff testing for COVID-19 in primary care*. en. Tech. rep. 2021.

[32] Kypraios, T. "Efficient Bayesian inference for partially observed stochastic epidemics and a new class of semi-parametric time series models". en. PhD thesis. Lancaster University, 2007.

Chapter 5

Discussion

The overarching aim of this thesis was to identify the implications of social contact patterns for public health and hospital infection control in the context of the COVID-19 pandemic. Two distinct approaches were taken in this thesis. Firstly, Chapters 2 and 3 outlined two cross-sectional studies conducted during the pandemic which aimed to quantify and characterise social contact patterns at a population level and for a specific occupation. These primary data collection studies provide insight into the levels of social mixing (the contact rate) and adherence to control measures, such as social distancing, at different points of the pandemic. Secondly, Chapter 4 presented a Bayesian approach to quantifying the relative role of nosocomial transmission routes of SARS-CoV-2, accounting for a detailed staff-patient contact network. Detailed data on social contact patterns is vital for understanding the drivers of transmission risk in different settings and populations. Chapter 4 demonstrates how fine-grained contact data, as collected in Chapters 2 and 3, may be incorporated into an individual-level transmission model and the benefits of doing so, which in this case was quantifying relative routes of transmission.

5.1 Chapter Overviews

Quantifying social contact patterns at a population level is a valuable method to assess adherence to control measures and to identify sociodemographic factors associated with heterogeneities in contact rate. Shortly following the emergence of COVID-19 in the UK, a multitude of NPIs were implemented to curb community transmission of SARS-CoV-2 by reducing non-household contact. Chapter 2 of this thesis introduced a UK social contact study conducted in July 2020 when pandemic restrictions had been relaxed following the first national lockdown. This period signified the start of the second wave of SARS-CoV-2 in the UK; it is likely that the level of social mixing during our study period enabled epidemic growth. The daily contact rate remained considerably lower than that recorded by pre-pandemic studies, suggesting continued adherence to control measures during the study period. Individuals who attended a workplace in-person or were employed in healthcare were less able to maintain social distance and had a higher daily non-household contact rate, representing a potentially elevated occupational infection risk. Limited evidence was found that those shielding had reduced non-household contacts compared with non-shielders, suggesting that self-isolating or shielding individuals required greater support to follow the government guidelines and reduce their infection risk. Although this study is limited by its sample size and representativeness, it provides insight into adherence to NPIs and quantification of contact patterns at a time of epidemic growth.

Key workers, defined as those critical to the pandemic response, were found to be at an increased risk of SARS-CoV-2 infection [149]. Contact patterns are thought to be a driver of occupational variation in infection risk during respiratory outbreaks. Chapter 3 described an occupational contact study of UK home delivery drivers conducted in December 2020 to quantify the interactions of delivery drivers, and identify the protective measures they adopted during the pandemic. This cross-sectional survey captured interactions of delivery drivers during the peak and gradual

decline of the second wave of SARS-CoV-2 in the UK. This study found that delivery drivers made a large number of face-to-face contacts compared to other working adults at this time, suggesting an increased risk of exposure and infection. However, this increased risk of exposure and transmission may be alleviated as contact was reported to be of short duration and was likely to have occurred outside. Although protective equipment, such as facemasks and hand sanitiser, was commonly used by drivers, less than half of drivers felt that it provided effective protection. Furthermore, a small proportion of drivers reported having worked while ill with symptoms of COVID-19 or while a member of their household had a suspected or confirmed case of COVID-19. Most drivers reported continuing to work in this situation due to financial reasons, which may be associated with statutory sick pay being unavailable for most drivers. A motivation of this study was to provide parameter estimates of contact rates of home delivery drivers for a modelling study which aimed to assess the impact of NPIs in the home delivery sector [58]. Whitfield et al. utilised a few key results from this study: average number of contacts made per shift, the proportion of drivers that worked whilst ill or with a member of their household ill with COVID-19 symptoms, and the proportion of drivers who reported self-isolating during the pandemic. Key workers were critical to the pandemic response, especially during periods of lockdown, ensuring essential services and supplies remained available. Proactive measures should be in place to ensure key workers are protected and supported.

Hospitals were under extraordinary pressure during the pandemic. While various infection control measures were employed, high levels of nosocomial infections were reported during the first wave of the pandemic [80]. Chapter 4 of this thesis utilised routinely collected data from a large UK hospital during the first wave of the pandemic to identify the relative role of hypothesised transmission pathways of nosocomial infection. I developed a Bayesian framework to infer unobserved epidemiological event times and disentangle the within-hospital infectious pressure dynamics. A stochastic individual-level continuous-time state-transition model was

used to represent nosocomial transmission of SARS-CoV-2 within the hospital. A dynamic staff-patient contact network was incorporated into the model as time-varying parameters, constructed from patient and staff location data. Transmission rate parameters for each source of infection and unobserved infection and recovery times were estimated using a MCMC algorithm. Infectious pressure dynamics were found to vary over time and across wards. Between-ward dynamics, driven by the staff-patient contact network in our model, were found to be the primary source of infection.

5.2 Implications for Public Health

Chapters 2 and 3 of this thesis highlighted the inadequate levels of support and protection for vulnerable individuals during the COVID-19 pandemic. Individuals in the two key worker groups identified in these studies, healthcare workers and home delivery drivers, were found to have a higher contact-rate than other individuals and were less able to maintain social distance, suggesting an increased risk of exposure and potentially infection [150, 151]. Furthermore, Chapter 2 found that shielding individuals, those identified as the most clinically vulnerable, were less able to reduce their contact rate compared to other participants. As clinically vulnerable individuals are at an increased risk of serious outcomes of COVID-19, it is important to identify how these individuals can be better supported to reduce their risk of infection. Disparities in health outcomes during respiratory infectious disease outbreaks were well documented before the emergence of COVID-19 [152]. Proactive measures should have been in place prior to the pandemic, to enable rapid implementation of protective measures for those most at-risk. Wider access to paid sick leave may improve adherence to measures such as self-isolation during an epidemic and reduce community transmission. Studies have found that insufficient access to paid sick leave is a risk factor for non-adherence to isolation guidelines during respiratory infectious disease outbreaks [153]. From a public health

perspective, support should always be available to enable workers to isolate when ill, not limited to an outbreak scenario. Consideration at a policy level is required to ensure that adequate measures for the most vulnerable are available to be rapidly implemented during the next public health emergency.

This thesis has demonstrated the benefit of capturing social mixing data during an outbreak, and how this data can be subsequently used to identify key drivers of transmission. Chapter 2 of this thesis provided insight into adherence to NPIs during the summer of 2020, and highlighted the substantial reduction in social mixing compared to levels recorded prior to the pandemic. Large-scale population-based surveys were conducted in many countries to capture social contact patterns and assess adherence to NPIs [154]. This social contact data can be subsequently used to parametrise mathematical models of COVID-19. While this data is important to collect, it is not feasible to conduct large-scale population-based surveys during every outbreak, or each time restrictions on social mixing change. Challenges to conducting contact studies include the financial cost and the difficulty of recruiting a large representative sample. Moreover, participant burden increases with the number and detail of survey questions, and responses are subject to recall bias. It follows that alternative data sources should be considered.

Klepac et al. describe an app-based approach to passively collect movement data, reducing participant burden [38]. Modelling studies have also utilised GPS data from commercial companies such as Google and SafeGraph to detect changes in mobility patterns [101, 155–157]. One advantage of these data sources is that they enable the identification of high-risk venues for infection, utilising data on the number of visitors and the average time spent at a venue. Chang et al. developed a SARS-CoV-2 transmission model which integrated mobility data and predicted that certain venues, such as fitness centres, restaurants and religious establishments, disproportionately contribute to increasing infection rates [155]. Although anonymised mobility data is useful for modelling purposes, the

disadvantage of this data is that children and elderly individuals are under-represented [158]. Additionally, as mobility data generally lacks demographic information, it is not possible to identify whether subgroups of a population are behaving differently, which is an important consideration when evaluating the effectiveness of NPIs. Flight data was similarly used at the beginning of the pandemic to study the importation of cases [1, 159, 160]. Furthermore, many countries conducted contact tracing during the pandemic to identify and break chains of transmission in the community. Contact tracing data could be an incredibly rich data source for estimating levels of social contact and thus the network of transmission opportunities in a population. The effectiveness of contact tracing during the pandemic was dependent on capacity to rapidly trace a potentially large number of contacts [161, 162]. This has led to the development of a variety of digital contact tracing apps during the pandemic. For example, the NHS COVID-19 app, developed as part of NHS Test and Trace, informed individuals if they had been exposed to an infected individual using the close proximity of devices as a proxy for close contact. The effectiveness of digital contact tracing is contingent on its uptake within a population and has had varying levels of success during the pandemic [59, 163, 164]. As with any large-scale surveillance system, there are concerns surrounding equity and ethics that need to be addressed, such as digital exclusion and privacy [165, 166]. While there may not be one single source of routinely collected data which is entirely suitable for quantifying social contact patterns, consideration should be given to how these data sources can be repurposed and combined to capture the levels of social mixing occurring in a population during the next public health emergency.

5.3 Implications for Hospital Infection Control

In the early stages of the pandemic, universal SARS-CoV-2 testing was not available for patients or staff in UK hospitals. Chapter 4 found that a triage system, which

cohorted patients on admission based on their suspected infection status, was an effective way of reducing nosocomial infections amongst those considered to be at low risk of having COVID-19. Furthermore, Chapter 4 detailed how between-ward dynamics, driven by the staff-patient contact network, appeared to be the primary source of nosocomial SARS-CoV-2 infection at a large UK hospital during the first wave of the pandemic. Strategies to reduce staff-patient transmission should be a primary consideration for respiratory outbreaks. For example, adequate PPE may help to reduce staff-patient transmission. Moreover, early adoption of asymptomatic testing and isolation of staff may substantially reduce the number of transmission opportunities within a hospital.

Advanced statistical inference methods can be used to improve the identification of nosocomial infections and quantify relative routes of transmission. Although these methods are not currently used as part of hospital infection control, due to the computational complexity and impractical runtimes, if implemented in real time, these methods could alert hospital infection control to hotspots of transmission and enable rapid evaluation of interventions. The statistical framework demonstrated in Chapter 4 provides a basis of how this could be achieved. Existing models of nosocomial transmission most often simulate an outbreak rather than fit models to data [167]. Those that do tend to either simplify the staff-patient contact network or focus on a small population of the hospital in order to conduct the inference efficiently [91, 94, 168, 169]. Using the framework outlined in Chapter 4, I was able to infer unobserved epidemiological event times and quantify transmission routes for patients in a large UK hospital over a four week period, while accounting for the intricacies of the staff-patient contact network. Further development is needed to construct a more flexible framework which could be used for non-respiratory diseases. Different model structures would need to be available to best suit the pathogen of interest. Additionally, the data augmentation method could be improved to better address unknown disease statuses on admission and occult infections at the end of the study period. A further challenge of implementing these more computationally

intensive inference methods is the computer infrastructure currently available in UK hospitals. As HCAIs continue to pose a significant burden to the NHS, investment in computer infrastructure should be increased to integrate more sophisticated methods into hospital infection control procedures.

5.4 Future Research

Epidemic models are increasingly being used to support policy decisions [101, 170]. As social contact patterns are known to have important implications for respiratory disease transmission, heterogenous mixing tends to be incorporated into models of these diseases. At the population scale, this is often implemented using age-specific contact rates derived from the POLYMOD data collected in 2005 [40]. However, the POLYMOD study has several limitations. Firstly, the data is right-censored for Great Britain as participants were only able record a maximum of 29 contacts. Secondly, there was a relatively small sample for each country, with only 1,012 participants reporting for Great Britain. Including more detailed behavioural information into epidemic models may be useful for evaluating interventions and identifying populations at greater risk of infection. Granular data on contact patterns and other behaviours (e.g, propensity to leave home, types of transport used, location of contacts and ability to maintain social distance) was collected during the pandemic; Chapters 2 and 3. This detailed information should be incorporated into individual-level transmission models to determine the extent to which infection risk is reduced by certain behaviours. For example, Kucharski et al. developed a model of individual-level transmission, stratifying by different settings (e.g. household, school, workplace), to assess the effectiveness of testing, isolation, and physical distancing scenarios [171]. This study used the BBC pandemic dataset [38] to simulate contact events in different settings. Further, Haw et al. developed an individual-level transmission model which was integrated with a multi-sector economic model to project the spread of SARS-CoV-

2 in the workplace [172]. Heterogeneity of contact rates were considered by age and workplace sector in the model by using derived contact rates from a French contact study [173]. This study was able to calculate optimal closure strategies by workplace sector, noting that a major challenge is identifying the nature and magnitude of changes in behaviour due to a lack of available data. Studies such as these incorporate more detailed behavioural data than simply age-structured models, however with the wealth of data available since the pandemic more fine-grained data could be included in future studies. For example, detailed data on shielding behaviour, adherence to social distancing, home-working and public transport usage (as collected in Chapter 2) could be directly incorporated into such models to understand what drives infection risk. An improved understanding of infection risk would enable more effective control measure strategies to be rapidly implemented during future epidemic scenarios. These may include initially targeting NPIs at high-contact occupations, and different vaccination prioritisation strategies. The value of integrating this detailed individual-level information into epidemic models should be evaluated to increase understanding of the data granularity that is needed to identify the drivers of transmission risk in different populations and settings, informing future decisions on data collection.

As demonstrated in Chapter 3, targeted social contact surveys can provide valuable insight into social mixing patterns in different populations. National-scale social contact surveys can provide useful insight into the level of social mixing in a population and improve understanding of how demographic factors are associated with contact rate. However, these tend to have a relatively small sample size and cannot provide a more nuanced understanding of contact patterns within specific populations and settings [37]. Quantifying social mixing patterns of high-contact occupations and key workers may assist with the evaluation of interventions and provide insight into the cause of superspreading events in workplaces, such as those observed at meat processing plants during the COVID-19 pandemic [174]. Targeted surveys could also be used to quantify social mixing of hard-to-

reach populations, such as refugees and individuals experiencing homelessness. A refined understanding of social contact patterns can aid with the identification of individuals at high risk of infection during infectious disease outbreaks and with the evaluation of interventions. Conducting a series of social contact studies for occupations identified as high-risk during the pandemic would provide data on occupation-specific tasks and behaviours. This data could in turn be incorporated into occupational transmission models to identify drivers of transmission and thus the most effective methods to break chains of transmission in these settings, in preparation for future infectious disease outbreaks.

The methodology presented in Chapter 4 only required routinely collected patient and staff data, avoiding any additional burden of data collection during an outbreak. However, the model would benefit from additional data where available, to improve understanding of the hospital structure and transmission pathways. Studies have shown the benefit of combining epidemiological data with genomic data when analysing hospital outbreaks [175–177]. The median turnaround time for genome sequencing was 5.1 days during the pandemic [178]. While this delay is not ideal for real-time monitoring of nosocomial transmission, sequencing data would be useful when modelling prolonged outbreaks. Moreover, this turnaround time is likely to decrease as technology improves. Chapter 4 identified the staff-patient contact network as a key source of nosocomial infection, however more detailed staff movement data and staff testing data would be needed to determine the underlying dynamics of staff-patient transmission. Detailed data on the duration and proximity of contact between staff and patients could be captured with real-time locating systems [179, 180]. A comprehensive study of nosocomial transmission capturing airflow in wards, staff and patient interactions with real-time locating systems, environmental samples, and staff and patient infection data during an outbreak, would be an extremely valuable resource to inform current modelling assumptions.

5.5 Conclusion

This thesis demonstrated how contemporary social contact data collected during the pandemic can provide insight into adherence to NPIs, be used to identify subgroups of the population which may be at a greater risk of infection, and aid in the identification of transmission routes in high-risk settings. Chapters 2 and 3 of this thesis quantified and characterised social contact patterns for different populations. At the national scale, the average daily non-household contact rate was found to be considerably lower in July 2020 compared to that recorded prior to the pandemic, suggesting adherence to NPIs. Subgroups of the population were identified which had a high daily non-household contact rate and were less able to maintain social distance, which could have led to an elevated infection risk. These studies demonstrated how contact rates vary with demographic characteristics and how different behaviours (e.g. shielding, social distancing) were associated with these. In turn, these patterns can assist with identifying population subgroups which may require greater support and protection during infectious disease outbreaks. Further, these insights can provide critical parameter estimates for transmission models of respiratory infectious diseases. An example of incorporating granular contact data, as collected in Chapters 2 and 3, into an individual-level model to identify key drivers of transmission is provided in Chapter 4 of this thesis. Chapter 4 demonstrates how routinely collected patient infection data coupled with information on the staff-patient contact network can be used to quantify relative routes of nosocomial transmission. This study highlighted that the patient-staff contact route was a key source of nosocomial infection during the fist-wave of the pandemic. With further development, the statistical framework presented in Chapter 4 could be used for real-time monitoring of nosocomial transmission and the evaluation of hospital infection control interventions. Contact patterns are central in driving epidemic dynamics of respiratory pathogens and are likely to form a critical component in future modelling and intervention assessment.

Appendix A

Paper 1: Supplemental Tables and Study Material

A.1 Tables

Table A.1: Comparison of non-household contact rates across different UK contact surveys.

	Mean contact rate, d^{-1} (25th and 75th percentile)				
Study, contact type and geography of sample	CoCoNet Number of participants	CoCoNet non-household contacts UK	CoMix non-household contacts UK [45]	Social Contact Survey non-household contacts Great Britain[42]	POLYMOD non-household contacts Great Britain[40]
Sampling period		28 July 2020 to 14 August 2020 Non-lockdown period	24 March 2020 to 27 March 2020 Lockdown period	2009 Pre-pandemic	2005-2006 Pre-pandemic
All participants	5,037	2.9 (0, 3)	1.4 (0, 1)	25.9 (5, 23)	9.6 (4, 13)
Age group					
0-9	5	1.0 (0, 1)	-	29.8 (7, 46)	8.9 (3, 13)
10-19	37	3.6 (0, 3)	-	43.1 (8, 40)	12.3 (5, 18)
20-29	250	3.2 (0, 3)	1.1 (0, 1)*	29.0 (7, 27)	9.7 (5, 13)
30-39	594	2.6 (0, 3)	1.4 (0, 1)	25.4 (6, 25)	8.9 (4, 12)
40-49	1167	3.1 (0, 4)	1.4 (0, 1)	30.8 (6, 29)	9.8 (4, 13)
50-59	1617	3.3 (0, 3)	1.6 (0, 2)	28.6 (6, 26)	8.2 (3, 11)
60-69	1056	2.3 (0, 3)	1.4 (0, 2)	23.0 (4, 18)	8.2 (4, 11)
70-79	290	1.9 (0, 3)	1.1 (1, 1)**	19.1 (3, 17)	6.8 (3, 11)
80+	21	1.7 (0, 2)	-	13.2 (1,10)	-
Sex†					
Female	3978	2.8 (0, 3)	1.4 (0, 1)	27.5 (5, 26)	10.2 (4, 14)
Male	1037	2.9 (0, 3)	1.3 (0, 1)	22.6 (4, 19)	9.0 (4, 13)
Prefer not to say	22	1.5 (0, 2)	-	-	-
Household size					

1	875	2.8 (0, 3)	1.6 (1, 2)	24.3 (4, 20)	7.5 (3, 11)
2	1902	2.6 (0, 3)	1.5 (0, 2)	23.7 (5, 21)	9.2 (4, 12)
3	979	3.2 (0, 3)	1.2 (0, 1)	24.6 (5, 25)	9.6 (4, 14)
4	896	2.7 (0, 3)	1.4 (0, 1)	33.4 (6, 33)	10.0 (4, 14)
5	284	3.2 (0, 4)	1.1 (0, 1)	30.7 (7, 30)	10.9 (5, 15)
6+	101	3.8 (0, 4)	1.1 (0, 1)	45.6 (8, 36)	10.1 (5, 15)

[*] CoMix age group 18-29.
[**] CoMix age group 70+.
[†] Comix and POLYMOD report participants' gender rather than sex.

Table A.2: Adjusted incidence rate ratios for number of daily non-household contacts by select variables. Intercept of 0.41 (95 %CI 0.35-0.48). Dispersion parameter of 1.07 (95%CI 1.00-1.14).

	Multivariable analysis[1]	
	aIRR (95%CI)	**p-value**
Age		
0-9	0.70 (0.13-3.69)	0.673
10-19	0.90 (0.50-1.67)	0.724
20-29	1.16 (0.97-1.39)	0.109
30-39	0.86 (0.76-0.97)	0.016
40-49	0.90 (0.82-1.00)	0.039
50-59	1.00	-
60-69	0.89 (0.79-1.00)	0.046
70-79	0.94 (0.78-1.14)	0.542
80+	1.23 (0.69-2.25)	0.458
Sex		
Female	1.00	-
Male	1.01 (0.93-1.10)	0.772
Prefer not to say	0.53 (0.27-1.05)	0.063
Ethnicity		
White	1.00)	-
Mixed/Multiple ethnic groups	1.09 (0.77-1.58)	0.625
Asian/Asian British	0.54 (0.36-0.82)	0.004
Black/African/Caribbean/Black British	0.46 (0.17-1.23)	0.136
Other ethnic group	0.54 (0.18-1.80)	0.308
Nation		
England	1.00	-
Northern Ireland	0.92 (0.59-1.47)	0.713
Scotland	0.80 (0.68-0.95)	0.009
Wales	1.22 (0.99-1.50)	0.060
Household size		
1	1.00	-
2	0.92 (0.83-1.02)	0.105
3	0.99 (0.88-1.12)	0.878
4	0.84 (0.74-0.96)	0.008
5	0.97 (0.81-1.16)	0.717
6+	0.85 (0.65-1.12)	0.247
Day of the week (contacts recorded)		
Weekday	1.00	-
Weekend	1.14 (0.99-1.31)	0.066
Left home yesterday		

Table A.2 continued from previous page

	Multivariable analysis[1] aIRR (95%CI)	p-value
No	1.00	-
Yes	5.58 (4.92-6.33)	<0.001
Dwelling type		
House or bungalow	1.00	-
Flat, maisonette or apartment	0.95 (0.85-1.07)	0.435
Mobile or temporary structure	1.13 (0.53-2.65)	0.760
Assisted living facility	2.25 (0.39-37.68)	0.442
Care home	2.84 (0.42-50.22)	0.345
Other	1.28 (0.68-2.54)	0.461
School or work situation		
Employed - working from home	1.00	-
School pupil - studying at home	0.83 (0.28-2.46)	0.739
School pupil - attending school	2.90 (0.95-10.02)	0.080
College or University student	0.84 (0.63-1.12)	0.229
Employed - going to place of work	3.33 (3.02-3.66)	<0.001
Self employed	1.63 (1.43-1.87)	<0.001
Healthcare professional	5.10 (4.29-6.10)	<0.001
Unemployed	1.10 (0.89-1.36)	0.399
Furloughed	1.20 (0.98-1.48)	0.074
Unable to work	0.96 (0.71-1.31)	0.815
Retired	1.24 (1.09-1.42)	0.001
COVID-19 circumstance		
Not self isolating or shielding	1.00	-
Self isolating - I have symptoms*	4.05 (1.94-9.72)	0.001
Self isolating - Someone in my household has symptoms	1.31 (0.51-3.63)	0.604
Self isolating - Someone in my support bubble has symptoms	0.81 (0.21-3.32)	0.771
Self isolating - precaution/told to by Test and Trace	0.58 (0.43-0.79)	<0.001
Shielding - I am a vulnerable individual	0.82 (0.66-1.01)	0.056
Shielding - I live with a vulnerable individual	0.79 (0.62-1.02)	0.065
Not sure	0.56 (0.39-0.79)	0.001

[1] Adjusted for age, sex, ethnicity, nation, household size, day of the week, left home, dwelling type, school or work situation and COVID-19 circumstance.

* This increased contact rate is due to one participant who was self-isolating with symptoms reporting a large number of contacts (see results).

Table A.3: Adjusted incidence rate ratios for number of daily non-household contacts by select variables. Variables identified through a forward stepwise model selection process. Intercept of 0.41 (95 %CI 0.35-0.48). Dispersion parameter of 1.07 (95%CI 1.00-1.14).

	Multivariable analysis[1]	
	aIRR (95%CI)	**p-value**
Age		
0-9	0.69 (0.13-3.65)	0.663
10-19	0.88 (0.49-1.64)	0.688
20-29	1.14 (0.96-1.36)	0.142
30-39	0.86 (0.76-0.97)	0.012
40-49	0.90 (0.82-0.99)	0.032
50-59	1.00	-
60-69	0.89 (0.79-1.00)	0.047
70-79	0.95 (0.79-1.15)	0.598
80+	1.24 (0.69-2.27)	0.446
Sex		
Female	1.00	-
Male	1.01 (0.93-1.11)	0.762
Prefer not to say	0.53 (0.27-1.05)	0.061
Ethnicity		
White	1.00	-
Mixed/Multiple ethnic groups	1.08 (0.76-1.56)	0.667
Asian/Asian British	0.53 (0.35-0.81)	0.003
Black/African/Carribean/Black British	0.46 (0.17-1.23)	0.135
Other ethnic group	0.53 (0.17-1.77)	0.290
Nation		
England	1.00	-
Northern Ireland	0.92 (0.59-1.47)	0.708
Scotland	0.79 (0.67-0.94)	0.006
Wales	1.23 (1.00-1.51)	0.052
Household size		
1	1.00	-
2	0.92 (0.83-1.02)	0.133
3	1.00 (0.89-1.13)	0.981
4	0.85 (0.75-0.96)	0.010
5	0.98 (0.83-1.17)	0.839
6+	0.88 (0.68-1.16)	0.357
Day of the week (contacts recorded)		
Weekday	1.00	-
Weekend	1.14 (0.99-1.31)	0.074

Table A.3 continued from previous page

	Multivariable analysis[1] aIRR (95%CI)	p-value
Left home yesterday		
No	1.00	
Yes	5.60 (4.94-6.36)	<0.001
School or work situation		
Employed - working from home	1.00	-
School pupil - studying at home	0.84 (0.29-2.48)	0.752
School pupil - attending school	2.96 (0.97-10.23)	0.074
College or University student	0.86 (0.65-1.14)	0.287
Employed - going to place of work	3.32 (3.02-3.66)	<0.001
Self employed	1.63 (1.43-1.87)	<0.001
Healthcare professional	5.05 (4.25-6.03)	<0.001
Unemployed	1.11 (0.90-1.37)	0.354
Furloughed	1.20 (0.98-1.48)	0.074
Unable to work	0.96 (0.71-1.30)	0.798
Retired	1.24 (1.09-1.42)	0.001
COVID-19 circumstance		
Not self-isolating or shielding	1.00	-
Self isolating - I have symptoms*	4.07 (1.96-9.79)	0.001
Self isolating - Someone in my household has symptoms	1.30 (0.51-3.60)	0.614
Self-isolating - Someone in my support bubble has symptoms	0.82 (0.21-3.34)	0.780
Self isolating - precaution/told to do so by Test and Trace	0.58 (0.43-0.79)	<0.001
Shielding - I am a vulnerable individual	0.81 (0.66-1.01)	0.050
Shielding - I live with a vulnerable individual	0.79 (0.62-1.01)	0.063
Not sure	0.55 (0.39-0.78)	0.001

[1] Adjusted for age, sex, ethnicity, nation, household size, day of the week, left home, school or work situation and COVID-19 circumstance.

* This increased contact rate is due to one participant who was self-isolating with symptoms reporting a large number of contacts (see results).

Table A.4: Characteristics of participants who reported 'Self isolating' or 'Shielding' as their COVID circumstance. N is the number of participants who provided a response to the question.

	Number of self-isolating participants (%)	**Number of shielding participants (%)**	**Number of participants not self-isolating or shielding (%)**
Age group	N = 136	N = 353	N = 4,511
0-9	0 (0.0%)	0 (0.0%)	
10-19	1 (0.7%)	4 (1.1%)	33 (0.7%)
20-29	7 (5.1%)	12 (3.4%)	233 (5.2%)
30-39	11 (8.1%)	28 (7.9%)	549 (12.2%)
40-49	17 (12.5%)	64 (18.1%)	1093 (24.2%)
50-59	41 (30.1%)	95 (26.9%)	1465 (32.5%)
60-69	32 (23.5%)	104 (29.5%)	905 (20.1%)
70-79	22 (16.2%)	43 (12.2%)	215 (4.8%)
80+	5 (3.7%)	3 (0.8%)	13 (0.3%)
No response	0 (0.0%)	0 (0.0%)	0 (0.0%)
Sex	N = 136	N = 353	N = 4,511
Female	111 (81.6%)	286 (81.0%)	3548 (78.7%)
Male	25 (18.4%)	65 (18.4%)	943 (20.9%)
Prefer not to say	0 (0.0%)	2 (0.6%)	20 (0.4%)
No response	0 (0.0%)	0 (0.0%)	0 (0.0%)
Ethnicity	N = 136	N = 353	N = 4,511
White	126 (92.6%)	334 (94.6%)	4336 (96.1%)
Mixed/Multiple ethnic groups	5 (3.7%)	8 (2.3%)	36 (0.8%)
Asian/Asian British	3 (2.2%)	2 (0.6%)	43 (1.0%)
Black/African/Caribbean/ Black British	0 (0.0%)	2 (0.6%)	9 (0.2%)
Other ethnic group	1 (0.7%)	0 (0.0%)	6 (0.1%)
Prefer not to say	0 (0.0%)	3 (0.8%)	13 (0.3%)
No response	1 (0.7%)	4 (1.1%)	68 (1.5%)
Left home yesterday	N = 135	N = 350	N = 4,495
No	64 (47.4%)	145 (41.4%)	778 (17.3%)
Yes	71 (52.6%)	205 (58.6%)	3717 (82.7%)
No response	0 (0.0%)	0 (0.0%)	0 (0.0%)
Part of a support bubble	N = 136	N= 352	N = 4,505
No	87 (64.0%)	238 (67.6%)	2664 (59.1%)
Yes	49 (36.0%)	114 (32.4%)	1841 (40.9%)
No response	0 (0.0%)	0 (0.0%)	0 (0.0%)

Table A.4 continued from previous page

	Number of self-isolating participants (%)	**Number of shielding participants (%)**	**Number of participants not self-isolating or shielding (%)**
Social distancing	N = 53	N= 167	N = 2,989
Yes, all of the time	39 (73.6%)	117 (70.1%)	1728 (57.8%)
More than half of the time	10 (18.9%)	36 (21.6%)	877 (29.3%)
Less than half of the time	4 (7.5%)	11 (6.6%)	278 (9.3%)
No, none of the time	0 (0.0%)	3 (1.8%)	86 (2.9%)
Not sure	0 (0.0%)	0 (0.0%)	20 (0.7%)
No response	0 (0.0%)	0 (0.0%)	0 (0.0%)
	Mean daily non-household contact rate (IQR)		
	Sef-isolating participants	Shielding participants	Participants not shielding or self-isolating
	N = 134	N = 348	N = 4,484
All participants	1.2 (0, 2)	1.3 (0, 2)	3.1 (0, 3)

Table A.5: Adjusted incidence rate ratios for number of daily non-household contacts by select variables. Self-isolating individual with large number of contacts removed for this analysis (see results). Intercept of 0.41 (95%CI 0.35-0.48). Dispersion parameter of 1.07 (95%CI 1.00-1.14).

	Multivariable analysis[1]	
	aIRR (95%CI)	**p-value**
Age		
0-9	0.69 (0.13-3.64)	0.662
10-19	0.89 (0.49-1.64)	0.691
20-29	1.15 (0.97-1.39)	0.114
30-39	0.86 (0.76-0.98)	0.021
40-49	0.91 (0.82-1.00)	0.044
50-59	1.00	-
60-69	0.89 (0.79-1.00)	0.047
70-79	0.94 (0.78-1.14)	0.540
80+	1.23 (0.69-2.25)	0.457
Sex		
Female	1.00	-
Male	1.02 (0.93-1.11)	0.730
Prefer not to say	0.53 (0.27-1.05)	0.063
Ethnicity		
White	1.00)	-
Mixed/Multiple ethnic groups	1.13 (0.79-1.62)	0.511
Asian/Asian British	0.54 (0.36-0.82)	0.004
Black/African/Caribbean/Black British	0.47 (0.17-1.23)	0.137
Other ethnic group	0.54 (0.18-1.79)	0.305
Nation		
England	1.00	-
Northern Ireland	0.92 (0.59-1.47)	0.713
Scotland	0.80 (0.68-0.95)	0.009
Wales	1.22 (1.00-1.50)	0.059
Household size		
1	1.00	-
2	0.92 (0.83-1.02)	0.105
3	0.98 (0.87-1.11)	0.805
4	0.84 (0.74-0.95)	0.007
5	0.97 (0.81-1.15)	0.692
6+	0.85 (0.65-1.12)	0.244
Day of the week (contacts recorded)		
Weekday	1.00	-
Weekend	1.14 (0.99-1.31)	0.068

Table A.5 continued from previous page

	Multivariable analysis[1]	
	aIRR (95%CI)	**p-value**
Left home yesterday		
No	1.00	-
Yes	5.54 (4.89-6.29)	<0.001
Dwelling type		
House or bungalow	1.00	-
Flat, maisonette or apartment	0.95 (0.85-1.07)	0.414
Mobile or temporary structure	1.21 (0.56-2.87)	0.647
Assisted living facility	2.25 (0.39-37.46)	0.441
Care home	2.78 (0.41-48.81)	0.355
Other	1.28 (0.68-2.54)	0.460
School or work situation		
Employed - working from home	1.00	-
School pupil - studying at home	0.85 (0.29-2.50)	0.763
School pupil - attending school	2.90 (0.95-10.00)	0.080
College or University student	0.86 (0.65-1.15)	0.291
Employed - going to place of work	3.34 (3.04-3.68)	<0.001
Self employed	1.64 (1.43-1.87)	<0.001
Healthcare professional	5.12 (4.31-6.12)	<0.001
Unemployed	1.10 (0.89-1.37)	0.369
Furloughed	1.21 (0.99-1.49)	0.063
Unable to work	1.00 (0.74-1.35)	0.993
Retired	1.25 (1.09-1.42)	0.001
COVID-19 circumstance		
Not self isolating or shielding	1.00	-
Self isolating - I have symptoms	0.83 (0.24-2.82)	0.775
Self isolating - Someone in my household has symptoms	1.31 (0.51-3.62)	0.606
Self isolating - Someone in my support bubble has symptoms	0.81 (0.21-3.30)	0.768
Self isolating - precaution/told to by Test and Trace	0.58 (0.43-0.79)	<0.001
Shielding - I am a vulnerable individual	0.81 (0.66-1.00)	0.050
Shielding - I live with a vulnerable individual	0.79 (0.62-1.02)	0.065
Not sure	0.56 (0.39-0.79)	0.001

[1] Adjusted for age, sex, ethnicity, nation, household size, day of the week, left home, dwelling type, school or work situation and COVID-19 circumstance.

Table A.6: Association of participant characteristics and maintaining social distancing more than half of the time with contacts (adjusted odds ratios).
N = 3058. Model intercept of 9.34 (7.24-12.20).

	Adjusted Odds Ratio (95%CI)[1]	**p-value**
Age		
10-19	0.30 (0.08-1.25)	0.077
20-29	0.62 (0.38-1.06)	0.069
30-39	0.66 (0.46-0.95)	0.024
40-49	0.88 (0.65-1.19)	0.415
50-59	1.00	-
60-69	0.88 (0.61-1.28)	0.496
70-79	0.91 (0.49 -1.79)	0.778
80+	0.28 (0.08-1.30)	0.064
School or work situation		
Employed - working from home	1.00	-
College or University student	0.68 (0.33-1.55)	0.335
Employed - going to place of work	0.71 (0.53-0.96)	0.025
Self employed	1.48 (0.92-2.51)	0.126
Healthcare professional	0.26 (0.17-0.40)	<0.001
Unemployed	0.59 (0.35-1.07)	0.068
Furloughed	1.01 (0.54-2.05)	0.979
Unable to work	1.58 (0.56-6.61)	0.453
Retired	1.40 (0.90-2.19)	0.136

[1] Adjusted for age and school or work situation.
[*] School pupils excluded from analysis due to insufficient data.

Table A.7: Association of participant characteristics with odds (adjusted odds ratios) of visiting another household (N = 4,030). Model intercept 0.11 (0.09-0.14).

	Adjusted Odds Ratio (95%CI)[1]	**p-value**
Age		
10-19	0.75 (0.18-2.20)	0.643
20-29	0.77 (0.47-1.21)	0.277
30-39	0.85 (0.62-1.15)	0.313
40-49	1.03 (0.82-1.30	0.805
50-59	1.00	-
60-69	1.09 (0.86-1.38)	0.479
70-79	1.18 (0.79-1.72)	0.413
80+	0.73 (0.11-2.73)	0.682
Sex		
Female	1.23 (0.98-1.56)	0.073
Male	1.00	-
Prefer not to say	1.07 (0.17-3.92)	0.926
Part of a support bubble		
No	1.00	-
Yes	1.92 (1.61-2.28)	<0.001

[1] Adjusted for age, sex and whether a participants was a part of a support bubble.

[*] 0-9 year olds excluded from analysis due to insufficient data.

A.2 Participant Information Sheet

PARTICIPANT INFORMATION SHEET
CoCoNet Survey

Research Project: COVID-19 Contact Network (CoCoNet) Study

Thank you for your interest in taking part in the CoCoNet Survey. Please read the information below carefully to decide whether or not you would like to take part.

What is the study about?
This is a short online survey which aims to measure people's interactions and travel under the current COVID-19 social distancing restrictions. The results of the survey will be used to improve UK predictions of COVID-19 spread, and to help identify which restrictions could be lifted safely in the future.

Can I take part?
Anyone who lives in the UK can take the survey. We are keen to get responses from people of all ages, and across all areas of the UK. For children under the age of 13 who wish to participate we ask that their parent or guardian reads through the information sheet for children with them, and that they complete the online survey together.

What will I be asked to do if I take part?
We will ask you to complete a short online survey, which should take about 5 to 10 minutes. We will ask questions on topics such as: how often you are leaving the house, the types of places you've visited, and how many people you meet during the day. **The survey is completely anonymous, so please answer questions as honestly as you can.**

Completion of the survey is voluntary at each stage. You can close the survey at any time, but any questions you have answered up to that point will be collected. As the data collected is anonymous, you will not be able to withdraw any information from the study has collected once you have started the survey.

Will my data be identifiable?
This survey is anonymous. We do ask for the first part of your home postcode. This is so we can work out which region of the UK you live in. However, we will not be able to identify you based on the information you provide. Postcode information will be removed from any data or results published by the study. All reasonable steps will be taken to protect the anonymity of the participants involved in this project.

At the end of the survey we will ask if you are happy to be contacted in the future for a follow-up survey and to receive an update on the study findings. If you choose to do so, we will ask for your email address. Your email address will be stored securely by the researchers

at Lancaster University. You can contact one of the researchers listed below, or ask to unsubscribe from the mailing list on receipt of an email, and your email address will be deleted from our records. Your email address will not be shared or used for any other reason. Your survey responses will remain anonymous.

Your personal data (email address) will be processed as a *task in the public interest* under GDPR and in accordance with the UK's Data Protection Act. For further information about how Lancaster University processes personal data for research purposes and your data rights please visit our webpage: www.lancaster.ac.uk/research/data-protection.

All data collected by the survey will be stored securely for 10 years, and only the researchers conducting the study will have access to all of the information. We may share fully anonymised data (without postcode information) with other COVID-19 researchers. Any research published may also provide unrestricted public access to the anonymised data.

What will happen to the results?
The results will be submitted as part of a PhD thesis at Lancaster University, and may also be submitted for publication in scientific journals. They may also be presented at national and international conferences. We expect our results of the study to be published in 2020.

Are there any risks?
We do not expect there to be any risk associated with participating in this study. However, if you experience any distress following participation, or are worried about catching coronavirus COVID-19, please use the resources provided at the end of this document.

Are there any benefits to taking part?
There is no direct benefit to you from taking part. However, we hope that you will find the research interesting, and we hope to use the results to help inform the epidemiology of COVID-19 outbreak.

Who has reviewed the project?
This study has been reviewed and approved by the Faculty of Health and Medicine Research Ethics Committee at Lancaster University.

Where can I obtain further information about the study if I need it?
For further information please contact one of the following researchers involved in the study:

Jessica Bridgen - PhD student, Lancaster Medical School
Email: j.bridgen@lancaster.ac.uk

Dr Jonathan Read - Senior Lecturer in Biostatistics and Epidemiology, Lancaster Medical School
Email: jonathan.read@lancaster.ac.uk

Complaints
If you wish to make a complaint or raise concerns about any aspect of this study, and do not want to speak to the researchers, you can contact:

Professor Joanne Knight +44 (0)1524 594800
Chair in Applied Data Science Email: jo.knight@lancaster.ac.uk
Faculty of Health and Medicine
(Lancaster Medical School)
Lancaster University
Lancaster
LA1 4YG

If you wish to speak to someone outside of the Lancaster Medical School Doctorate Programme, you may also contact:
Dr Laura Machin Tel: +44 (0)1524 594973
Chair of FHM REC Email: l.machin@lancaster.ac.uk
Faculty of Health and Medicine
(Lancaster Medical School)
Lancaster University
Lancaster
LA1 4YG

Thank you for taking the time to read this information sheet.

How can I get help or information related to coronavirus COVID-19?
Should you feel concerned or distressed either as a result of taking part, or in the future, the following resources may help:

If you need medical help, go to NHS 111 online.

If you're struggling because of coronavirus (COVID-19) and need support:
https://www.gov.uk/find-coronavirus-support

If you need help with your mental health or wellbeing:
https://www.nhs.uk/using-the-nhs/nhs-services/mental-health-services/dealing-with-a-mental-health-crisis-or-emergency/
or
https://www.mind.org.uk/information-support/coronavirus/coronavirus-and-your-wellbeing/#collapse838f8

For COVID-19 NHS Medical advice:
https://www.nhs.uk/conditions/coronavirus-covid-19/

For information about COVID-19 UK Government guidance:
https://www.gov.uk/coronavirus

A.3 Participant Information Sheet for Children

Health & Medicine | Lancaster University

Information sheet
CoCoNet Survey

To be read and talked through with children under 13

We would like to invite you to take part in a research study called CoCoNet.

This study is trying to understand where people go and who they talk to. This is important when thinking about the spread of a disease like coronavirus.

Can I take part?

If you live in the UK and your parent or the person that looks after you is happy to complete the survey with you then you can take part.

What do I have to do?

You will be asked to answer some questions online with the help of your parent or the person that looks after you. It will take about 5 - 10 minutes to finish.

We will ask about things like who you live with, where you go when you leave your home and who you talk to.

Health & Medicine | Lancaster University

What happens to my answers?

We will keep hold of your answers on a computer database with everyone else's answers to help us with our study. We will not know who you are from your answers, we do not ask for your name or your address.

You can choose to give us an email address so that we can ask you to complete the questions again later. We will also tell you what we have found so far. If you choose to do this, we will make sure no one else sees your email address.

Does anything else happen?

We will publish the results of the survey online, once we have studied them. Remember, no one will know which answers were yours.

If you have any questions, you can contact one of the researchers at any time.

You can email Jessica Bridgen on j.bridgen@lancaster.ac.uk
or
Dr Jonathan Read on jonathan.read@lancaster.ac.uk

Thank you for reading this information sheet.

A.4 Survey Questions

The questionnaire was presented as an online Qualtrics survey which could be completed on a computer or mobile phone.
Participants were directed to the survey through the study website: `https://www.lancaster.ac.uk/health-and-medicine/research/coconet-study/`.

Q1. Are you aged 13 or over?

- Yes
- No

Q1.a **Please make sure you agree to the following before continuing with the survey:**

You currently live in the UK;

You have read the **Participant Information Sheet** and fully understand what is expected of you within this study;

Your participation is voluntary and you are aware that you can stop the survey at any point; You understand that data submitted prior to closing the survey will be collected;

You consent to Lancaster University keeping the anonymised data for a period of 10-years after the study has finished;

If you are filling out the survey on behalf of someone else, please make sure you have their consent before continuing.

☐ **I consent to taking part in the CoCoNet study**

Q1.b **If you are under the age of 13 we do ask that a parent or guardian fills out the survey with you.**

Please take the time to read through and discuss the *Information sheet for Children* together. The parent or guardian should also read through the more detailed *Participant Information Sheet*.

Please make sure you both agree to the following before continuing with the survey:

I live in the UK;

I have read and understood the information sheet(s);

I understand I can stop the survey at any point;

I understand that my answers will be kept for 10 years after the study has finished.

	Child	Parent/Guardian
I agree to take part in the CoCoNet study/ I consent to my child taking part in the study	☐	☐

Q2. Where in the UK do you currently live?

- England
- Northern Ireland
- Scotland
- Wales
- I do not live in the UK

Q3. What is your age?

- 0 - 9 years old
- 10 - 19 years old
- 20 - 29 years old
- 30 - 39 years old
- 40 - 49 years old
- 50 - 59 years old
- 60 - 69 years old
- 70 - 79 years old
- Aged 80 or over

Q4. What is your sex?

The answer you give can be different from what is on your birth certificate.

- Female
- Male
- Prefer not to say

Q5. Which of the following best describes your ethnicity?

- English / Welsh / Scottish / Northern Irish / British
- Irish
- Gypsy or Irish Traveller
- Any other White background
- White and Black Caribbean
- White and Black African
- White and Asian
- Any other Mixed / Multiple ethnic background
- Indian
- Pakistani
- Bangladeshi
- Chinese
- Any other Asian background
- African
- Caribbean
- Any other Black / African / Caribbean background
- Arab
- Any other ethnic group

- Prefer not to say

Q6. What is the first part of your home postcode?

For example, if your home postcode was LA1 4YW then you would enter LA1.

Q7. Which type of accommodation best describes your home?

- Flat, maisonette or apartment
- House or bungalow
- Mobile or temporary structure
- Assisted living facility
- Care home
- Other

Q8. What is your current school or work situation?

- School pupil - studying at home
- School pupil - still attending school
- College or University student
- Employed - working from home
- Employed - still going to place of work
- Self Employed
- Healthcare professional
- Unemployed
- Furloughed

- Unable to work
- Retired
- Other

Q9. Currently, do you regularly meet members of the general public as part of your job?

- Yes
- No

Q10. Are you self-isolating or shielding because of COVID-19? *A vulnerable individual here refers to a clinically extremely vulnerable person.*

- I am not self-isolating or shielding
- Self Isolating - I have symptoms of COVID
- Self Isolating - Someone in my household has symptoms of COVID
- Self Isolating - Someone in my support bubble has symptoms of COVID
- Self Isolating - As a precaution / told to do so by Test and Trace
- Shielding - I am a vulnerable individual
- Shielding - I live with a vulnerable individual
- Not sure

Q11. How many other people currently live with you at home?

- 0 - I live alone
- 1
- 2
- 3

- 4
- 5 or more

Q12.How many people of each age group live with you at home?
Do not include yourself.

Drop down options of 0, 1, 2, 3, 4, 5 or more for each age group.

- 0 - 9 year olds
- 10 - 19 year olds
- 20 - 29 year olds
- 30 - 39 year olds
- 40 - 49 year olds
- 50 - 59 year olds
- 60 -69 year olds
- 70 - 79 year olds
- Aged 80 or over

Q13. Have you formed a support bubble with another household?
A single-person household can join with one other household and interact without maintaining social distance.

- Yes
- No

Q14. How many people of each age group are part of your support bubble?
Do not include your own household members.

Drop down options of 0, 1, 2, 3, 4, 5 or more for each age group.

- 0 - 9 year olds
- 10 - 19 year olds

- 20 - 29 year olds
- 30 - 39 year olds
- 40 - 49 year olds
- 50 - 59 year olds
- 60 -69 year olds
- 70 - 79 year olds
- Aged 80 or over

Q15. Thinking about the past 7 days, on how many of these days did you meet someone from your support bubble?

- None
- 1 day
- 2 days
- 3 days
- 4 days
- 5 days
- 6 days
- 7 days
- Not sure

Q16. Did you leave your home or property yesterday?
Do not include going into your private garden, but do include visits to shared or communal gardens or spaces.

- Yes
- No

Q17. Where did you go yesterday? Tick all that apply.

☐ Visited the home of someone else

☐ My school or workplace

☐ Doctor's surgery or healthcare facility

☐ Supermarket or convenience store

☐ Other shops or retail spaces (e.g. garden centre, clothing shops, drive-through food outlets)

☐ Restaurant, café or pub

☐ For a walk or exercise

☐ Other - *please do not include any identifying information*

Q18. What modes of transport did you use yesterday? Tick all that apply.

☐ I walked or cycled

☐ I travelled in a car by myself

☐ I travelled in a car with another person(s)

☐ I took a bus, tram or train

☐ I took an aeroplane or ferry

Q19. Not including those that you live with, how many people did you meet yesterday?

Only include those you had a face-to-face conversation with.

- None
- 1
- 2
- 3
- 4
- 5
- 6
- 7

- 8
- 9
- 10
- 11
- 12
- 13
- 14
- 15 or more

Q20. Please tell us about each of the people you met yesterday.
Information collected for up to 14 contacts.

How old were they? *Please estimate the person's age if you are unsure.*								Did you meet this person indoors? *For example, in a shop, office or house.*		Did anyone that you live with also meet this person yesterday?		
0 - 4	5 - 9	10 - 19	20 - 39	40 - 59	60 - 69	70 - 79	80 +	Yes	No	Yes	No	Not sure

Q21. Please tell us about each of the people you met yesterday.
Information collected for up to 14 contacts.

How old were they? *Please estimate the person's age if you are unsure.*								Did you meet this person indoors? *For example, in a shop, office or house.*	
0 - 4	5 - 9	10 - 19	20 - 39	40 - 59	60 - 69	70 - 79	80+	Yes	No

Q22. How many people of each age group did you meet yesterday?
Drop down choice of integers 1 - 19 or '20 or more' for each age group.

- 0-4 year olds
- 5-9 year olds
- 10-19 year olds
- 20-39 year olds

- 40-59 year olds
- 60-69 year olds
- 70-79 year olds
- Aged 80 or over

Q23. Did you meet these people indoors or outdoors?
For example, meeting someone indoors could be in a shop, office or house etc.

- I met everyone indoors
- I met most people indoors
- I met most people outdoors
- I met everyone outdoors
- Not sure

Q24. Were you able to maintain social distance from everyone you met yesterday?
Do not include people that you live with or those in your support bubble. Please refer to the government advice for the recommended social distance in your area.

- Yes, all of the time
- More than half of the time
- Less than half of the time
- No, none of the time
- Not sure

Q25. Of the people you live with, how many people stayed at home all day yesterday?

- None
- 1
- 2

- 3
- 4
- 5 or more
- Not sure

Q26. Thinking about the past 7 days, on how many of these days did you leave your home or property?

- None
- 1 day
- 2 days
- 3 days
- 4 days
- 5 days
- 6 days
- 7 days
- Not sure

Q27. What was the furthest distance from home you travelled over the past 7 days?

- Less than 2 miles (3 km)
- 2 - 9 miles (3 - 15 km)
- 10 - 19 miles (16 - 31 km)
- 20 - 49 miles (32 - 79 km)
- 50 miles (80 km) or more

Appendix B

Paper 2: Supplemental Tables and Study Material

B.1 Tables

Table B.1: Participants' employment information. N is the number of participants who provided a response to the question.

	Number of participants (%)
Current employment situation (N = 162)	
Self-employed and completely independent	88 (54.3%)
Self-employed but receiving certain workers benefits	14 (8.6%)
Employed by one company - full time	46 (28.4%)
Employed by one company - part time	11 (6.8%)
Employed by multiple companies	3 (1.9%)
No response	0 (0.0%)
Current weekly working hours (N = 162)† *	
0	0 (0.0%)
1 - 10	6 (3.7%)
11 - 20	17 (10.5%)
21 - 30	17 (10.5%)
31 - 40	44 (27.2%)
41 - 50	42 (25.9%)
51 or more	36 (22.2%)
In receipt of statutory sick leave pay (N = 151)†*	
Yes	38 (25.2%)
No	103 (68.2%)
Not sure	10 (6.6%)

† Question required a response from the participant.
* As a delivery driver.

Table B.2: Adjusted incidence rate ratios (aIRR) for number of customer contacts per shift by select variables.

	Multivariable analysis[1]	
	aIRR (95%CI)	**p-value**
Age		
18-29	1.65 (1.07-2.60)	0.020
30-39	1.13 (0.77-1.68)	0.525
40-49	1.64 (1.15-2.34)	0.004
50-59	1.00	-
60-69	1.00 (0.60-1.75	0.986
Sex		
Female	1.00	-
Male	0.98 (0.70-1.36)	0.910
Employment type		
Self-employed and independent	1.00	-
Self-employed and receiving some benefits	1.19 (0.74-1.97)	0.456
Employed by one company - full time	0.66 (0.47-0.94)	0.019
Employed by one company - part time	0.77 (0.39-1.58)	0.404
Employed by multiple companies	0.97 (0.40-2.80)	0.955
Furthest distance from depot to delivery (miles)		
0-9	1.00	-
10-19	1.18 (0.79-1.75)	0.422
20-99	1.23 (0.83-1.82)	0.274
100-499	1.07 (0.60-1.91)	0.805
Weekly working hours		
1-10	0.47 (0.24-1.00)	0.036
11-20	0.55(0.32-0.99)	0.018
21-30	0.44 (0.28-0.71)	0.001
31-40	1.00	-
41-50	0.95 (0.63-1.43)	0.806
51+	0.98 (0.65-1.48)	0.918
Items normally delivered		
Small parcels only	1.00	-
Large items only	1.24(0.55-3.13)	0.600
Small parcels and large items	0.88(0.59-1.34)	0.541
Groceries only	0.34(0.18-0.64)	0.001
Takeaway only	0.19(0.11-0.37)	0.001
Other items only	0.43(0.24-0.80)	0.003
Other delivery type combinations	0.58(0.38-0.93)	0.014

B.2 Participant Information Sheet

Health & Medicine | Lancaster University | MANCHESTER 1824

PARTICIPANT INFORMATION SHEET
CoCoNet: Home Delivery Driver Study

Research Project: COVID-19 Contact Network (CoCoNet): Home Delivery Driver Study

Thank you for your interest in taking part in the CoCoNet: Home Delivery Driver Survey. Please read the information below carefully to decide whether or not you would like to take part.

What is the study about?
This is a short online survey which aims to better understand how UK delivery drivers contribute to the control of COVID-19 and what measures have been put in place to protect them and their customers.

Can I take part?
You must be aged 18 or over to take part in the study and currently be working as a home delivery driver in the UK.

What will I be asked to do if I take part?
We will ask you to complete a short online survey, which should take about 5 to 10 minutes. We will ask questions on topics such as: how many people you meet during a shift at work, whether you share a vehicle with another delivery worker and whether or not you are using any personal protective equipment (PPE) at work.

The survey is completely anonymous, so please answer questions as honestly as you can.
Completion of the survey is voluntary at each stage. You can close the survey at any time, but any questions you have answered up to that point will be collected. As the data collected is anonymous, you will not be able to withdraw any information from the study once you have started the survey.

Will my data be identifiable?
This survey is anonymous. We do ask for the first part of your home postcode. This is so we can work out which region of the UK you live in. However, we will not be able to identify you based on the information you provide. Postcode information will be removed from any data or results published by the study. All reasonable steps will be taken to protect the anonymity of the participants involved in this project.

At the end of the survey we will ask if you are happy to be contacted in the future for a follow-up survey and to receive an update on the study findings. If you choose to do so, we will ask for your email address. Your email address will be stored securely by the researchers at Lancaster University. You can contact one of the researchers listed below, or ask to unsubscribe from the mailing list on receipt of an email, and your email address will be

deleted from our records. Your email address will not be shared or used for any other reason. Your survey responses will remain anonymous.

Your personal data (email address) will be processed as a *task in the public interest* under GDPR and in accordance with the UK's Data Protection Act. For further information about how Lancaster University processes personal data for research purposes and your data rights please visit our webpage: www.lancaster.ac.uk/research/data-protection.

All data collected by the survey will be stored securely for 10 years, and only the researchers conducting the study will have access to all of the information. We may share fully anonymised data (without postcode information) with other COVID-19 researchers. Any research published may also provide unrestricted public access to the anonymised data.

What will happen to the results?
The results will be submitted as part of a PhD thesis at Lancaster University, and may also be submitted for publication in scientific journals. They may also be presented at national and international conferences.

Are there any risks?
We do not expect there to be any risk associated with participating in this study. However, if you experience any distress following participation, or are worried about catching coronavirus COVID-19, please use the resources provided at the end of this document.

Are there any benefits to taking part?
There is no direct benefit to you from taking part, however, we hope you will find the research interesting.

Who has reviewed the project?
This study has been reviewed and approved by the Faculty of Health and Medicine Research Ethics Committee at Lancaster University.

Where can I obtain further information about the study if I need it?
For further information please contact one of the following researchers involved in the study:

Jessica Bridgen - PhD student, Lancaster Medical School
Email: j.bridgen@lancaster.ac.uk

Dr Jonathan Read - Senior Lecturer in Biostatistics and Epidemiology, Lancaster Medical School
Email: jonathan.read@lancaster.ac.uk

Prof Martie Van Tongeren - Professor of Occupational and Environmental Health, Manchester University
Email: Martie.J.Van-Tongeren@manchester.ac.uk

Complaints
If you wish to make a complaint or raise concerns about any aspect of this study, and do not want to speak to the researchers, you can contact:

Professor Joanne Knight +44 (0)1524 594800
Chair in Applied Data Science Email: jo.knight@lancaster.ac.uk
Faculty of Health and Medicine
(Lancaster Medical School)
Lancaster University
Lancaster
LA1 4YG

If you wish to speak to someone outside of the Lancaster Medical School Doctorate Programme, you may also contact:
Dr Laura Machin Tel: +44 (0)1524 594973
Chair of FHM REC Email: l.machin@lancaster.ac.uk
Faculty of Health and Medicine
(Lancaster Medical School)
Lancaster University
Lancaster
LA1 4YG

Thank you for taking the time to read this information sheet.
How can I get help or information related to coronavirus COVID-19?
Should you feel concerned or distressed either as a result of taking part, or in the future, the following resources may help:

If you need medical help, go to NHS 111 online.

If you're struggling because of coronavirus (COVID-19) and need support:
https://www.gov.uk/find-coronavirus-support

If you need help with your mental health or wellbeing:
https://www.nhs.uk/using-the-nhs/nhs-services/mental-health-services/dealing-with-a-mental-health-crisis-or-emergency/
or
https://www.mind.org.uk/information-support/coronavirus/coronavirus-and-your-wellbeing/#collapse838f8

For COVID-19 NHS Medical advice:
https://www.nhs.uk/conditions/coronavirus-covid-19/

For information about COVID-19 UK Government guidance:
https://www.gov.uk/coronavirus

If you have concerns about work:
https://www.citizensadvice.org.uk/work/

B.3 Survey Questions

Survey questions were presented as an online Qualtrics survey which could be completed on a computer or mobile phone.

Q1a. Are you aged 18 or over?

- Yes
- No

Q1b. Do you work as a home delivery driver in the UK? (including mail, parcels, homeware, takeaway, groceries etc.)

- Yes
- No

Q1c.

Please make sure you agree to the following before continuing with the survey:

You currently live in the UK;

You are a home delivery driver;

You have read the **Participant Information Sheet** and fully understand what is expected of you within this study;

Your participation is voluntary, and you are aware that you can stop the survey at any point;
You consent to Lancaster University keeping the anonymised data for a period of 10-years after the study has finished;

- **I consent to taking part in the CoCoNet: Home Delivery Driver study**

Q2. Where in the UK do you currently live?

- England
- Northern Ireland
- Scotland
- Wales
- I do not live in the UK

Q3. What is your age?

- 18 - 29 years old
- 30 - 39 years old
- 40 - 49 years old
- 50 - 59 years old
- 60 - 69 years old
- 70 - 79 years old
- Aged 80 or over

Q4. What is your sex?

The answer you give can be different from what is on your birth certificate.

- Female
- Male
- Prefer not to say

Q5. Which of the following best describes your ethnicity?

- English / Welsh / Scottish / Northern Irish / British
- Irish

- Gypsy or Irish Traveller
- Any other White background
- White and Black Caribbean
- White and Black African
- White and Asian
- Any other Mixed / Multiple ethnic background
- Indian
- Pakistani
- Bangladeshi
- Chinese
- Any other Asian background
- African
- Caribbean
- Any other Black / African / Caribbean background
- Arab
- Any other ethnic group
- Prefer not to say

Q6. What is the first part of your home postcode?

For example, if your home postcode was LA1 4YW then you would enter LA1.

Q7. What is you highest level of education?

- Higher education and professional/vocational equivalents
- A levels, vocational level 3 and equivalents

- Trade apprenticeships
- GCSE/ O level grade A*-C, vocational level 2 and equivalents
- Qualifications at level 1 and below
- Other qualifications
- No qualifications
- Not sure

Q8. Are you self-isolating or shielding because of COVID-19?

A vulnerable individual here refers to a clinically extremely vulnerable person.

- I am not self-isolating or shielding
- Self Isolating - I have symptoms of COVID
- Self Isolating - Someone in my household has symptoms of COVID
- Self Isolating - Someone in my support bubble has symptoms of COVID
- Self Isolating - Someone I have been in contact with has symptoms of COVID
- Self Isolating - Told to do so by a contact tracer
- Self Isolating - Told to do so by a contact tracing phone app
- Self Isolating - In travel-related quarantine
- Shielding - I am a vulnerable individual
- Shielding - I live with a vulnerable individual
- Other - *please do not include any identifying information*
  _____
- Not sure

Q9. Have you ever had to self-isolate due to COVID-19 infection (either suspected or confirmed)?

- Yes

- No

Q10. In which month(s) have you had to self-isolate? Tick all that apply.

- ☐ March 2020
- ☐ April 2020
- ☐ May 2020
- ☐ June 2020
- ☐ July 2020
- ☐ August 2020
- ☐ September 2020
- ☐ October 2020
- ☐ November 2020
- ☐ December 2020
- ☐ January 2021
- ☐ February 2021
- ☐ March 2021

Q11. Have you ever tested positive for COVID-19?

- Yes
- No

Q12. In which month(s) did you test positive? Tick all that apply.

- ☐ March 2020
- ☐ April 2020
- ☐ May 2020
- ☐ June 2020
- ☐ July 2020
- ☐ August 2020

- ☐ September 2020
- ☐ October 2020
- ☐ November 2020
- ☐ December 2020
- ☐ January 2021
- ☐ February 2021
- ☐ March 2021

Q13. How many other people currently live with you at home?

- 0 - I live alone
- 1
- 2
- 3
- 4
- 5 or more

Q14. How many people of each age group live with you at home?

Do not include yourself.

Drop down options of 0, 1, 2, 3, 4, 5 or more for each age group.

0 - 9 year olds
10 - 19 year olds
20 - 29 year olds
30 - 39 year olds
40 - 49 year olds
50 - 59 year olds
60 -69 year olds
70 - 79 year olds
Aged 80 or over

Q15. What do you normally deliver?

- ☐ Letters and mail
- ☐ Small parcels
- ☐ Large items (e.g. large appliances, furniture)
- ☐ Takeaway food
- ☐ Groceries
- ☐ Other - please specify

  _____

Q16. What is your current employment situation?

- Self-employed and completely independent
- Self-employed but receiving certain workers benefits (e.g. holiday pay, sick pay). Please specify - do not include any identifying information

  _____
- Employed by one company - full time
- Employed by one company - part time
- Employed by multiple companies

Q17. How many hours do you currently work as a delivery driver per week?

- 0 hours
- 1 - 10 hours
- 11 - 20 hours
- 21 - 30 hours
- 31 - 40 hours
- 41 - 50 hours
- 51 hours or more

Q18. How many deliveries did you make during your last shift?

Q19. When was the last shift you worked as a delivery driver?

- This week
- Earlier this month
- Last month
- 2 months ago or more

Q20. Did you use your own vehicle to make deliveries on your last shift?

- Yes
- No

Q21. During your last shift, how many customers did you meet face-to-face?

Q22. On your last shift, which types of face-to-face contact did you make with customers? Tick all that apply.

☐ Brief face-to-face interaction with customer (less than 5 minutes)

☐ Prolonged face-to-face interaction with customer (more than 5 minutes)

☐ Entered a customer's property (e.g. for installation or to drop off a heavy item)

☐ Hand signature required from customer

☐ Other, please specify

Q23. Were you able to maintain social distance (at least 2 metres apart) from all of the customers you met on your last shift?

- Yes, all of the time
- More than half of the time
- Less than half of the time
- No, none of the time
- Not sure

Q24. During your last shift, how many people did you meet in the location where you collect items for delivery (e.g. depots or food outlets)?

Only include those you had a face-to-face conversation with.

Q25. Were you able to maintain social distance (at least 2 metres apart) from everyone you met in the location where you collect items for delivery last shift?

- Yes, all of the time
- More than half of the time
- Less than half of the time
- No, none of the time
- Not sure

Q26. During your last working week, what was the furthest distance you travelled from a collection point (at a depot or restaurant) to a delivery address?

- 0 - 9 miles
- 10 - 19 miles
- 20 - 99 miles
- 100 - 499 miles
- 500 miles or more

Q27. During your last working week, did you share a vehicle with a colleague to make deliveries? (e.g. for large items)

- Yes
- No

Q28. During your last working week, did you always share a vehicle with the same colleague?

- Yes
- No

Q29. What measures have you taken to prevent infection when sharing a vehicle? Tick all that apply.

- ☐ Keeping windows open
- ☐ Wearing a facemask
- ☐ Using hand sanitiser
- ☐ Other, please specify

Q30. On your last shift, did you use any Personal Protective Equipment (PPE)? Tick all that apply.

- ☐ Facemask
- ☐ Gloves
- ☐ Hand sanitiser
- ☐ Other, please specify

Q31. Was the PPE provided by your employers or contracting companies/platforms?

- Yes
- No
- Not sure

Q32. Do you think they are effective protection?

- Yes
- No
- Not sure

Q33. Do you think you have adequate knowledge and information about workplace COVID-19 risk?

- Yes
- No

Q34. Where do you learn about workplace COVID-19 risk? Tick all that apply.

☐ Company/platform communications
☐ Government website
☐ News and social media
☐ Other, please specify
_____

Q35. Do you receive statutory sick leave pay for your delivery occupation?

- Yes
- No
- Not sure

Q36. How would you cover your living expenses if you had to take time off work due to being ill with COVID-19 or COVID-19 symptoms? Tick all that apply.

☐ Special COVID-19 funding scheme provided by the contracting companies/platforms

☐ Government self-employment income support scheme

☐ Social benefits

☐ Other, please specify

_____

Q37. Since March 2020, have you worked whilst being ill with COVID-19 symptoms, or with a member of your household having a suspected or confirmed case of COVID-19?

COVID-19 symptoms include: a high temperature, a new continuous cough, a loss or change to your sense of smell or taste

- Yes
- No
- Not sure

Q38. What was the main reason you still went to work? Tick all that apply.

☐ Financial reasons

☐ Unable to find someone to cover my shifts

☐ Asked to or encouraged to carry on working by my employers

☐ Other, please specify

_____

References

[1] Read, J. M., Bridgen, J. R. E., Cummings, D. A. T., Ho, A., and Jewell, C. P. "Novel coronavirus 2019-nCoV COVID-19: early estimation of epidemiological parameters and epidemic size estimates". en. In: *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 376.1829 (July 2021), p. 20200265.

[2] Blomquist, P. B., Bridgen, J., Bray, N., et al. *Enhancing epidemiological surveillance of the emergence of the SARS-CoV-2 Omicron variant using spike gene target failure data, England, 15 November to 31 December 2021.* 2022.

[3] Webster, H. H., Nyberg, T., Sinnathamby, M. A., et al. "Hospitalisation and mortality risk of SARS-COV-2 variant omicron sub-lineage BA.2 compared to BA.1 in England". en. In: *Nat. Commun.* 13.1 (Oct. 2022), p. 6053.

[4] Harman, K., Nash, S. G., Webster, H. H., et al. "Comparison of the risk of hospitalisation among BA.1 and BA.2 COVID-19 cases treated with sotrovimab in the community in England". en. In: *Influenza Other Respi. Viruses* 17.5 (May 2023), e13150.

[5] Keeling, M. J. and Eames, K. T. D. "Networks and epidemic models". en. In: *J. R. Soc. Interface* 2.4 (Sept. 2005), pp. 295–307.

[6] Wallinga, J., Edmunds, W. J., and Kretzschmar, M. "Perspective: human contact patterns and the spread of airborne infectious diseases". en. In: *Trends Microbiol.* 7.9 (Sept. 1999), pp. 372–377.

[7] Read, J. M., Edmunds, W. J., Riley, S., Lessler, J., and Cummings, D. A. T. "Close encounters of the infectious kind: methods to measure social mixing behaviour". en. In: *Epidemiol. Infect.* 140.12 (Dec. 2012), pp. 2117–2130.

[8] Ferguson, N. "Capturing human behaviour". en. In: *Nature* 446.7137 (Apr. 2007), p. 733.

[9] Cabinet Office. *Staying at home and away from others (social distancing).* Accessed: 2020. 2020. URL: `https://www.gov.uk/government/publications/full-guidance-on-staying-at-home-and-away-from-others/full-guidance-on-staying-at-home-and-away-from-others`.

[10] Davies, N. G., Kucharski, A. J., Eggo, R. M., Gimma, A., Edmunds, W. J., and Mathematical Modelling of Infectious Diseases COVID-19 working group, C. for the. "Effects of non-pharmaceutical interventions on COVID-19 cases, deaths, and demand for hospital services in the UK: a modelling study". en. In: *Lancet Public Health* 5.7 (July 2020), e375–e385.

[11] Keeling, M. J., Hill, E. M., Gorsich, E. E., et al. "Predictions of COVID-19 dynamics in the UK: Short-term forecasting and analysis of potential exit strategies". en. In: *PLoS Comput. Biol.* 17.1 (Jan. 2021), e1008619.

[12] World Health Organization. *Chapter 2. Current context: the COVID-19 pandemic and continuing challenges to global health*. en. Accessed: 2022. 2022. URL: https://www.who.int/about/funding/invest-in-who/investment-case-2.0/challenges.

[13] Hale, Thomas, Noam Angrist, Beatriz Kira, Anna Petherick, Toby Phillips, Samuel Webster. "Variation in Government Responses to COVID-19". In: *Blavatnik School of Government Working Paper* Version 6.0 (2020).

[14] World Health Organization. *WHO Coronavirus (COVID-19) Dashboard*. Accessed: 2023. 2020. URL: https://covid19.who.int/data.

[15] Whittaker, C., Walker, P. G. T., Alhaffar, M., et al. "Under-reporting of deaths limits our understanding of true burden of covid-19". en. In: *BMJ* 375 (Oct. 2021), n2239.

[16] Wang, H., Paulson, K. R., Pease, S. A., et al. "Estimating excess mortality due to the COVID-19 pandemic: a systematic analysis of COVID-19-related mortality, 2020–21". en. In: *Lancet* 399.10334 (Apr. 2022), pp. 1513–1536.

[17] GOV UK. *COVID-19: epidemiology, virology and clinical features*. en. Accessed: 2022. 2022. URL: https://www.gov.uk/government/publications/wuhan-novel-coronavirus-background-information/wuhan-novel-coronavirus-epidemiology-virology-and-clinical-features.

[18] World Health Organization. *Transmission of SARS-CoV-2: implications for infection prevention precautions*. en. Accessed: 2021. July 2020. URL: https://www.who.int/news-room/commentaries/detail/transmission-of-sars-cov-2-implications-for-infection-prevention-precautions.

[19] Oran, D. P. and Topol, E. J. "The Proportion of SARS-CoV-2 Infections That Are Asymptomatic : A Systematic Review". en. In: *Ann. Intern. Med.* 174.5 (May 2021), pp. 655–662.

[20] Ma, Q., Liu, J., Liu, Q., et al. "Global Percentage of Asymptomatic SARS-CoV-2 Infections Among the Tested Population and Individuals With Confirmed COVID-19 Diagnosis: A Systematic Review and Meta-analysis". en. In: *JAMA Netw Open* 4.12 (Dec. 2021), e2137257.

[21] Buitrago-Garcia, D., Egli-Gany, D., Counotte, M. J., et al. "Occurrence and transmission potential of asymptomatic and presymptomatic SARS-CoV-2 infections: A living systematic review and meta-analysis". en. In: *PLoS Med.* 17.9 (Sept. 2020), e1003346.

[22] World Health Organisation. *Tracking SARS-CoV-2 variants*. en. Accessed: 2022. 2022. URL: https://www.who.int/activities/tracking-SARS-CoV-2-variants.

[23] GOV UK. *CMO confirms cases of coronavirus in England*. en. Accessed: 2022. 2020. URL: https://www.gov.uk/government/news/cmo-confirms-cases-of-coronavirus-in-england.

[24] World Health Organisation. *Listings of WHO's response to COVID-19*. en. Accessed: 2022. URL: https://www.who.int/news/item/29-06-2020-covidtimeline.

[25] Public Health England. *Coronavirus (COVID-19) in the UK*. Accessed: 2020. 2020. URL: https://coronavirus.data.gov.uk.

[26] GOV UK. *Major new measures to protect people at highest risk from coronavirus*. en. Accessed: 2022. Mar. 2020. URL: https://www.gov.uk/government/news/major-new-measures-to-protect-people-at-highest-risk-from-coronavirus.

[27] GOV UK. *NHS Test and Trace statistics (England): methodology*. en. Accessed: 2022. 2022. URL: https://www.gov.uk/government/publications/nhs-test-and-trace-statistics-england-methodology/nhs-test-and-trace-statistics-england-methodology.

[28] GOV.UK. *New campaign urges public to get tested twice a week*. en. Accessed: 2022. 2021. URL: https://www.gov.uk/government/news/new-campaign-urges-public-to-get-tested-twice-a-week.

[29] NHS England. *Landmark moment as first NHS patient receives COVID-19 vaccination*. Accessed: 2022. 2020. URL: https://www.england.nhs.uk/2020/12/landmark-moment-as-first-nhs-patient-receives-covid-19-vaccination/.

[30] Joint Committee on Vaccination and Immunisation. *Joint Committee on Vaccination and Immunisation: advice on priority groups for COVID-19 vaccination, 30 December 2020*. en. Accessed: 2021. 2020. URL: https://www.gov.uk/government/publications/priority-groups-for-coronavirus-covid-19-vaccination-advice-from-the-jcvi-30-december-2020/joint-committee-on-vaccination-and-immunisation-advice-on-priority-groups-for-covid-19-vaccination-30-december-2020.

[31] World Health Organization. "Ten health issues WHO will tackle this year". In: *World Health Organization* (2019).

[32] Robertson, E., Reeve, K. S., Niedzwiedz, C. L., et al. "Predictors of COVID-19 vaccine hesitancy in the UK household longitudinal study". en. In: *Brain Behav. Immun.* 94 (May 2021), pp. 41–50.

[33] Office for National Statistics. *Coronavirus and vaccine hesitancy*. en. Accessed: 2022. July 2021. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthandwellbeing/bulletins/coronavirusandvaccinehesitancygreatbritain/26mayto20june2021.

[34] Chaudhuri, K., Chakrabarti, A., Chandan, J. S., and Bandyopadhyay, S. "COVID-19 vaccine hesitancy in the UK: a longitudinal household cross-sectional study". en. In: *BMC Public Health* 22.1 (Jan. 2022), p. 104.

[35] Anderson, R. M. and May, R. M. *Infectious Diseases of Humans: Dynamics and Control*. en. OUP Oxford, 1991.

[36] Hethcote, H. W. "The Mathematics of Infectious Diseases". In: *SIAM Rev.* 42.4 (Jan. 2000), pp. 599–653.

[37] Hoang, T., Coletti, P., Melegaro, A., et al. "A Systematic Review of Social Contact Surveys to Inform Transmission Models of Close-contact Infections". en. In: *Epidemiology* 30.5 (Sept. 2019), pp. 723–736.

[38] Klepac, P., Kucharski, A. J., Conlan, A. J. K., Kissler, S., Tang, M., Fry, H., and Gog, J. R. "Contacts in context: large-scale setting-specific social mixing matrices from the BBC Pandemic project". Feb. 2020.

[39] Vanhems, P., Barrat, A., Cattuto, C., et al. "Estimating potential infection transmission routes in hospital wards using wearable proximity sensors". en. In: *PLoS One* 8.9 (Sept. 2013), e73970.

[40] Mossong, J., Hens, N., Jit, M., et al. "Social contacts and mixing patterns relevant to the spread of infectious diseases". en. In: *PLoS Med.* 5.3 (Mar. 2008), e74.

[41] Read, J. M., Lessler, J., Riley, S., et al. "Social mixing patterns in rural and urban areas of southern China". en. In: *Proc. Biol. Sci.* 281.1785 (June 2014), p. 20140268.

[42] Danon, L., Read, J. M., House, T. A., Vernon, M. C., and Keeling, M. J. "Social encounter networks: characterizing Great Britain". en. In: *Proc. Biol. Sci.* 280.1765 (Aug. 2013), p. 20131037.

[43] Latsuzbaia, A., Herold, M., Bertemes, J.-P., and Mossong, J. "Evolving social contact patterns during the COVID-19 crisis in Luxembourg". en. In: *PLoS One* 15.8 (Aug. 2020), e0237128.

[44] Eames, K. T. D., Tilston, N. L., Brooks-Pollock, E., and Edmunds, W. J. "Measured dynamic social contact patterns explain the spread of H1N1v influenza". en. In: *PLoS Comput. Biol.* 8.3 (Mar. 2012), e1002425.

[45] Jarvis, C. I., Van Zandvoort, K., Gimma, A., et al. "Quantifying the impact of physical distance measures on the transmission of COVID-19 in the UK". en. In: *BMC Med.* 18.1 (May 2020), p. 124.

[46] Gimma, A., Munday, J. D., Wong, K. L. M., et al. "Changes in social contacts in England during the COVID-19 pandemic between March 2020 and March 2021 as measured by the CoMix survey: A repeated cross-sectional study". en. In: *PLoS Med.* 19.3 (Mar. 2022), e1003907.

[47] *CoMix study - Social contact survey in the UK*. en. `https://cmmid.github.io/topics/covid19/comix-reports.html`. Accessed: 2024-2-18. Oct. 2020.

[48] Ipsos. *The power of research panels*. en. `https://www.ipsos.com/en-us/knowledge/new-services/power-research-panels`. Accessed: 2024-2-18. Jan. 2021.

[49] *Social contact patterns in the EU/EEA during the COVID-19 pandemic*. en. `https://www.ecdc.europa.eu/en/publications-data/social-contact-patterns-eueea-during-covid-19-pandemic`. Accessed: 2024-2-18. Sept. 2022.

[50] Jarvis, C. I., Gimma, A., Zandvoort, K. van, Wong, K. L. M., CMMID COVID-19 working group, and Edmunds, W. J. "The impact of local and national restrictions in response to COVID-19 on social contacts in England: a longitudinal natural experiment". en. In: *BMC Med.* 19.1 (Feb. 2021), p. 52.

[51] GOV UK. *SPI-M-O: Consensus statement on COVID-19, 30 June 2021*. en. `https://www.gov.uk/government/publications/spi-m-o-consensus-statement-on-covid-19-30-june-2021`. Accessed: 2023. July 2021.

[52] Acke, S., Couvreur, S., Bramer, W. M., Schmickler, M.-N., De Schryver, A., and Haagsma, J. A. "Global infectious disease risks associated with occupational exposure among non-healthcare workers: a systematic review of the literature". en. In: *Occup. Environ. Med.* 79.1 (Jan. 2022), pp. 63–71.

[53] Su, C.-P., Perio, M. A. de, Cummings, K. J., McCague, A.-B., Luckhaupt, S. E., and Sweeney, M. H. "Case Investigations of Infectious Diseases Occurring in Workplaces, United States, 2006-2015". en. In: *Emerg. Infect. Dis.* 25.3 (Mar. 2019), pp. 397–405.

[54] Voirin, N., Payet, C., Barrat, A., et al. "Combining high-resolution contact data with virological data to investigate influenza transmission in a tertiary care hospital". en. In: *Infect. Control Hosp. Epidemiol.* 36.3 (Mar. 2015), pp. 254–260.

[55] Jiang, L., Ng, H. L., Ho, H. J., Leo, Y. S., Prem, K., Cook, A. R., and Chen, M. I. "Contacts of healthcare workers, patients and visitors in general wards in Singapore". en. In: *Epidemiol. Infect.* 145.14 (Oct. 2017), pp. 3085–3095.

[56] Isella, L., Romano, M., Barrat, A., et al. "Close encounters in a pediatric ward: measuring face-to-face proximity and mixing patterns with wearable sensors". en. In: *PLoS One* 6.2 (Feb. 2011), e17144.

[57] Jarvis, C., Edmunds, J., and CMMID COVID-19 Working Group. *Social contacts in workplace the UK from the CoMix social contact survey*. Tech. rep. London School of Hygiene and Tropical Medicine, June 2021.

[58] Whitfield, C. A., Tongeren, M. van, Han, Y., et al. "Modelling the impact of non-pharmaceutical interventions on workplace transmission of SARS-CoV-2 in the home-delivery sector". en. In: *PLoS One* 18.5 (May 2023), e0284805.

[59] Wymant, C., Ferretti, L., Tsallis, D., et al. "The epidemiological impact of the NHS COVID-19 app". en. In: *Nature* 594.7863 (June 2021), pp. 408–412.

[60] Office for National Statistics. *Coronavirus (COVID-19) Infection Survey: methods and further information*. en. `https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/methodologies/covid19infectionsurveypilotmethodsandfurtherinformation`. Accessed: 2023.

[61] Seventer, J. M. van and Hochberg, N. S. "Principles of Infectious Diseases: Transmission, Diagnosis, Prevention, and Control". In: *International Encyclopedia of Public Health (Second Edition)*. Ed. by S. R. Quah. Oxford: Academic Press, Jan. 2017, pp. 22–39.

[62] Niveau, G. "Prevention of infectious disease transmission in correctional settings: a review". en. In: *Public Health* 120.1 (Jan. 2006), pp. 33–41.

[63] Taylor, A., Goldberg, D., Emslie, J., et al. "Outbreak of HIV infection in a Scottish prison". en. In: *BMJ* 310.6975 (Feb. 1995), pp. 289–292.

[64] Utsumi, M., Makimoto, K., Quroshi, N., and Ashida, N. "Types of infectious outbreaks and their impact in elderly care facilities: a review of the literature". en. In: *Age Ageing* 39.3 (May 2010), pp. 299–305.

[65] NHS. *Healthcare associated infections*. en. Accessed: 2023. URL: `https://www.england.nhs.uk/patient-safety/healthcare-associated-infections/`.

[66] World Health Organization. *Report on the Burden of Endemic Health Care-Associated Infection Worldwide*. Tech. rep. 2011.

[67] World Health Organization. *Global report on infection prevention and control*. en. Tech. rep. May 2022.

[68] Cassini, A., Plachouras, D., Eckmanns, T., et al. *Burden of Six Healthcare-Associated Infections on European Population Health: Estimating Incidence-Based Disability-Adjusted Life Years through a Population Prevalence-Based Modelling Study*. 2016.

[69] Knight, G. M., Pham, T. M., Stimson, J., et al. "The contribution of hospital-acquired infections to the COVID-19 epidemic in England in the first half of 2020". en. In: *BMC Infect. Dis.* 22.1 (June 2022), p. 556.

[70] GOV UK. *Health and Social Care Act 2008: code of practice on the prevention and control of infections and related guidance*. en. Accessed: 2023. 2022. URL: https://www.gov.uk/government/publications/the-health-and-social-care-act-2008-code-of-practice-on-the-prevention-and-control-of-infections-and-related-guidance/health-and-social-care-act-2008-code-of-practice-on-the-prevention-and-control-of-infections-and-related-guidance.

[71] Loveday, H. P., Wilson, J. A., Pratt, R. J., et al. "epic3: national evidence-based guidelines for preventing healthcare-associated infections in NHS hospitals in England". en. In: *J. Hosp. Infect.* 86 Suppl 1 (Jan. 2014), S1–70.

[72] Cowling, B. J., Park, M., Fang, V. J., Wu, P., Leung, G. M., and Wu, J. T. "Preliminary epidemiological assessment of MERS-CoV outbreak in South Korea, May to June 2015". en. In: *Euro Surveill.* 20.25 (June 2015), pp. 7–13.

[73] Hsieh, Y.-H. "2015 Middle East Respiratory Syndrome Coronavirus (MERS-CoV) nosocomial outbreak in South Korea: insights from modeling". en. In: *PeerJ* 3 (Dec. 2015), e1505.

[74] Mousavi, E. S., Kananizadeh, N., Martinello, R. A., and Sherman, J. D. "COVID-19 Outbreak and Hospital Air Quality: A Systematic Review of Evidence on Air Filtration and Recirculation". en. In: *Environ. Sci. Technol.* 55.7 (Apr. 2021), pp. 4134–4147.

[75] World Health Organization. *MERS outbreaks*. en. Accessed: 2023. 2023. URL: https://www.emro.who.int/health-topics/mers-cov/mers-outbreaks.html.

[76] Hui, D. S., Azhar, E. I., Kim, Y.-J., Memish, Z. A., Oh, M.-D., and Zumla, A. "Middle East respiratory syndrome coronavirus: risk factors and determinants of primary, household, and nosocomial transmission". en. In: *Lancet Infect. Dis.* 18.8 (Aug. 2018), e217–e227.

[77] Chowell, G., Abdirizak, F., Lee, S., Lee, J., Jung, E., Nishiura, H., and Viboud, C. "Transmission characteristics of MERS and SARS in the healthcare setting: a comparative study". en. In: *BMC Med.* 13 (Sept. 2015), p. 210.

[78] Ki, M. "2015 MERS outbreak in Korea: hospital-to-hospital transmission". en. In: *Epidemiol. Health* 37 (July 2015), e2015033.

[79] Kim, K. H., Tandi, T. E., Choi, J. W., Moon, J. M., and Kim, M. S. "Middle East respiratory syndrome coronavirus (MERS-CoV) outbreak in South Korea, 2015: epidemiology, characteristics and public health implications". en. In: *J. Hosp. Infect.* 95.2 (Feb. 2017), pp. 207–213.

[80] Read, J. M., Green, C. A., Harrison, E. M., et al. "Hospital-acquired SARS-CoV-2 infection in the UK's first COVID-19 pandemic wave". en. In: *Lancet* 398.10305 (Sept. 2021), pp. 1037–1038.

[81] Zhou, Q., Gao, Y., Wang, X., et al. "Nosocomial infections among patients with COVID-19, SARS and MERS: a rapid review and meta-analysis". en. In: *Ann Transl Med* 8.10 (May 2020), p. 629.

[82] Wang, D., Hu, B., Hu, C., et al. "Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China". en. In: *JAMA* 323.11 (Mar. 2020), pp. 1061–1069.

[83] Rickman, H. M., Rampling, T., Shaw, K., et al. "Nosocomial Transmission of Coronavirus Disease 2019: A Retrospective Study of 66 Hospital-acquired Cases in a London Teaching Hospital". en. In: *Clin. Infect. Dis.* 72.4 (Feb. 2021), pp. 690–693.

[84] Ward, H., Atchison, C., Whitaker, M., et al. "SARS-CoV-2 antibody prevalence in England following the first peak of the pandemic". en. In: *Nat. Commun.* 12.1 (Feb. 2021), p. 905.

[85] Nguyen, L. H., Drew, D. A., Graham, M. S., et al. "Risk of COVID-19 among front-line health-care workers and the general community: a prospective cohort study". en. In: *Lancet Public Health* 5.9 (Sept. 2020), e475–e483.

[86] European Centre for Disease Prevention and Control. *Infection prevention and control and preparedness for COVID-19 in healthcare settings - second update*. 2020.

[87] Treibel, T. A., Manisty, C., Burton, M., et al. "COVID-19: PCR screening of asymptomatic health-care workers at London hospital". en. In: *Lancet* 395.10237 (May 2020), pp. 1608–1610.

[88] Sun, H., Lu, M., Chen, S., Cheng, Z., Xiong, Y., and Wang, X. "Nosocomial SARS-CoV-2 infection among nurses in Wuhan at a single centre". en. In: *J. Infect.* 80.6 (June 2020), e41–e42.

[89] Jewkes, S. V., Zhang, Y., and Nicholl, D. J. "Nosocomial spread of COVID-19: lessons learned from an audit on a stroke/neurology ward in a UK district general hospital". en. In: *Clin. Med.* 20.5 (Sept. 2020), e173–e177.

[90] National Healthcare Safety Network. *Patient Safety Component Manual*. 2023.

[91] Kypraios, T., O'Neill, P. D., Huang, S. S., Rifas-Shiman, S. L., and Cooper, B. S. "Assessing the role of undetected colonization and isolation precautions in reducing methicillin-resistant Staphylococcus aureus transmission in intensive care units". en. In: *BMC Infect. Dis.* 10 (Feb. 2010), p. 29.

[92] Kwok, K. O., Leung, G. M., Lam, W. Y., and Riley, S. "Using models to identify routes of nosocomial infection: a large hospital outbreak of SARS in Hong Kong". en. In: *Proc. Biol. Sci.* 274.1610 (Mar. 2007), pp. 611–617.

[93] Basu, S., Andrews, J. R., Poolman, E. M., et al. "Prevention of nosocomial transmission of extensively drug-resistant tuberculosis in rural South African district hospitals: an epidemiological modelling study". en. In: *Lancet* 370.9597 (Oct. 2007), pp. 1500–1507.

[94] Evans, S., Stimson, J., Pople, D., Bhattacharya, A., Hope, R., White, P. J., and Robotham, J. V. "Quantifying the contribution of pathways of nosocomial acquisition of COVID-19 in English hospitals". en. In: *Int. J. Epidemiol.* 51.2 (May 2022), pp. 393–403.

[95] Bhattacharya, A., Collin, S. M., Stimson, J., et al. "Healthcare-associated COVID-19 in England: A national data linkage study". en. In: *J. Infect.* 83.5 (Nov. 2021), pp. 565–572.

[96] Aghdassi, S. J. S., Schwab, F., Peña Diaz, L. A., et al. "Risk factors for nosocomial SARS-CoV-2 infections in patients: results from a retrospective matched case–control study in a tertiary care university center". en. In: *Antimicrob. Resist. Infect. Control* 11.1 (Jan. 2022), pp. 1–10.

[97] Heesterbeek, H., Anderson, R. M., Andreasen, V., et al. "Modeling infectious disease dynamics in the complex landscape of global health". en. In: *Science* 347.6227 (Mar. 2015), aaa4339.

[98] Brauer, F. "Mathematical epidemiology: Past, present, and future". en. In: *Infect Dis Model* 2.2 (May 2017), pp. 113–127.

[99] Bernoulli, D. and Chapelle, D. "Essai d'une nouvelle analyse de la mortalité causée par la petite vérole, et des avantages de l'inoculation pour la prévenir". fr. In: (May 2023).

[100] Keeling, M. J. and Rohani, P. *Modeling Infectious Diseases in Humans and Animals*. en. Princeton University Press, Sept. 2011.

[101] McBryde, E. S., Meehan, M. T., Adegboye, O. A., et al. "Role of modelling in COVID-19 policy development". en. In: *Paediatr. Respir. Rev.* 35 (Sept. 2020), pp. 57–60.

[102] Metcalf, C. J. E., Morris, D. H., and Park, S. W. "Mathematical models to guide pandemic response". en. In: *Science* 369.6502 (July 2020), pp. 368–369.

[103] GOV UK. *Technical report on the COVID-19 pandemic in the UK Chapter 5: modelling*. en. Tech. rep. 2023.

[104] Unkel, S., Farrington, C. P., Garthwaite, P. H., Robertson, C., and Andrews, N. "Statistical Methods for the Prospective Detection of Infectious Disease Outbreaks: A Review". en. In: *J. R. Stat. Soc. Ser. A Stat. Soc.* 175.1 (July 2011), pp. 49–82.

[105] Diggle, P. and Giorgi, E. *Model-based Geostatistics for Global Public Health: Methods and Applications*. en. CRC Press, June 2021.

[106] Becker, N. G. and Britton, T. "Statistical studies of infectious disease incidence". en. In: *J. R. Stat. Soc. Series B Stat. Methodol.* 61.2 (July 1999), pp. 287–307.

[107] Brauer, F. "Compartmental Models in Epidemiology". In: *Mathematical Epidemiology*. Ed. by F. Brauer, P. van den Driessche, and J. Wu. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 19–79.

[108] Ross, R. "An application of the theory of probabilities to the study of a priori pathometry.—Part I". In: *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 92.638 (1916), pp. 204–230.

[109] Ross, R. and Hudson, H. P. "An application of the theory of probabilities to the study of a priori pathometry.—Part II". In: *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 93.650 (1917), pp. 212–225.

[110] Ross, R. and Hudson, H. P. "An application of the theory of probabilities to the study of a priori pathometry.—Part III". In: *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 93.650 (1917), pp. 225–240.

[111] Kermack, W. O., McKendrick, A. G., and Walker, G. T. "A contribution to the mathematical theory of epidemics". In: *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 115.772 (1927), pp. 700–721.

[112] Tolles, J. and Luong, T. "Modeling Epidemics With Compartmental Models". en. In: *JAMA* 323.24 (June 2020), pp. 2515–2516.

[113] Wallinga, J., Teunis, P., and Kretzschmar, M. "Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents". en. In: *Am. J. Epidemiol.* 164.10 (Nov. 2006), pp. 936–944.

[114] Anderson, R. M. and May, R. M. "Age-related changes in the rate of disease transmission: implications for the design of vaccination programmes". en. In: *J. Hyg.* 94.3 (June 1985), pp. 365–436.

[115] Schenzle, D. "An age-structured model of pre- and post-vaccination measles transmission". en. In: *IMA J. Math. Appl. Med. Biol.* 1.2 (1984), pp. 169–191.

[116] Anderson, R. M., May, R. M., and McLean, A. R. "Possible demographic consequences of AIDS in developing countries". en. In: *Nature* 332.6161 (Mar. 1988), pp. 228–234.

[117] Andersson, H. and Britton, T. *Stochastic Epidemic Models and Their Statistical Analysis*. en. Springer New York, Aug. 2000.

[118] Gillespie, D. T. "Exact stochastic simulation of coupled chemical reactions". In: *J. Phys. Chem.* 81.25 (Dec. 1977), pp. 2340–2361.

[119] Becker, N. "The uses of epidemic models". en. In: *Biometrics* 35.1 (Mar. 1979), pp. 295–305.

[120] Neal, P. J. and Roberts, G. O. "Statistical inference and model selection for the 1861 Hagelloch measles epidemic". en. In: *Biostatistics* 5.2 (Apr. 2004), pp. 249–261.

[121] Kretzschmar, M. and Morris, M. "Measures of concurrency in networks and the spread of infectious disease". en. In: *Math. Biosci.* 133.2 (Apr. 1996), pp. 165–195.

[122] Keeling, M. J., Woolhouse, M. E., Shaw, D. J., et al. "Dynamics of the 2001 UK foot and mouth epidemic: stochastic dispersal in a heterogeneous landscape". en. In: *Science* 294.5543 (Oct. 2001), pp. 813–817.

[123] Deardon, R., Brooks, S. P., Grenfell, B. T., et al. "INFERENCE FOR INDIVIDUAL-LEVEL MODELS OF INFECTIOUS DISEASES IN LARGE POPULATIONS". en. In: *Stat. Sin.* 20.1 (Jan. 2010), pp. 239–261.

[124] Ferguson, N. M., Cummings, D. A. T., Cauchemez, S., et al. "Strategies for containing an emerging influenza pandemic in Southeast Asia". en. In: *Nature* 437.7056 (Sept. 2005), pp. 209–214.

[125] Riley, S. and Ferguson, N. M. "Smallpox transmission and control: spatial dynamics in Great Britain". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 103.33 (Aug. 2006), pp. 12637–12642.

[126] Germann, T. C., Kadau, K., Longini Jr, I. M., and Macken, C. A. "Mitigation strategies for pandemic influenza in the United States". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 103.15 (Apr. 2006), pp. 5935–5940.

[127] Ferguson, N. M., Cummings, D. A. T., Fraser, C., Cajka, J. C., Cooley, P. C., and Burke, D. S. "Strategies for mitigating an influenza pandemic". en. In: *Nature* 442.7101 (July 2006), pp. 448–452.

[128] Merler, S. and Ajelli, M. "The role of population heterogeneity and human mobility in the spread of pandemic influenza". en. In: *Proc. Biol. Sci.* 277.1681 (Feb. 2010), pp. 557–565.

[129] Mollison, D., Isham, V., and Grenfell, B. "Epidemics: models and data". en. In: *J. R. Stat. Soc. Ser. A Stat. Soc.* 157.1 (1994), pp. 115–149.

[130] Bolstad, W. M. *Understanding Computational Bayesian Statistics*. en. John Wiley & Sons, Dec. 2009.

[131] O'Neill, P. D. and Roberts, G. O. "Bayesian Inference for Partially Observed Stochastic Epidemics". In: *J. R. Stat. Soc. Ser. A Stat. Soc.* 162.1 (1999), pp. 121–129.

[132] Bayes, M. and Price, M. "LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S". la. In: *Philos. Trans. R. Soc. Lond.* 53.0 (Dec. 1763), pp. 370–418.

[133] Schoot, R. van de, Depaoli, S., King, R., et al. "Bayesian statistics and modelling". en. In: *Nature Reviews Methods Primers* 1.1 (Jan. 2021), pp. 1–26.

[134] Dempster, A. P. "A Generalization of Bayesian Inference". In: *J. R. Stat. Soc. Series B Stat. Methodol.* 30.2 (1968), pp. 205–247.

[135] McKinley, T., Cook, A. R., and Deardon, R. "Inference in Epidemic Models without Likelihoods". en. In: *Int. J. Biostat.* 5.1 (July 2009).

[136] Bisset, K. R., Chen, J., Feng, X., Kumar, V. S. A., and Marathe, M. V. "EpiFast: a fast algorithm for large scale realistic epidemic simulations on distributed memory systems". In: *Proceedings of the 23rd international conference on Supercomputing*. ICS '09. Yorktown Heights, NY, USA: Association for Computing Machinery, June 2009, pp. 430–439.

[137] Heslop, D. J., Chughtai, A. A., Bui, C. M., and MacIntyre, C. R. "Publicly available software tools for decision-makers during an emergent epidemic-Systematic evaluation of utility and usability". en. In: *Epidemics* 21 (Dec. 2017), pp. 1–12.

[138] Sellman, S., Tsao, K., Tildesley, M. J., et al. "Need for speed: An optimized gridding approach for spatially explicit disease simulations". en. In: *PLoS Comput. Biol.* 14.4 (Apr. 2018), e1006086.

[139] Pellegrino, M., Lombardo, G., Cagnoni, S., and Poggi, A. "High-Performance Computing and ABMS for High-Resolution COVID-19 Spreading Simulation". en. In: *Future Internet* 14.3 (Mar. 2022), p. 83.

[140] Jewell, C. P., Hale, A. C., Rowlingson, B. S., Suter, C., Read, J. M., and Roberts, G. O. "Bayesian inference for high-dimensional discrete-time epidemic models: spatial dynamics of the UK COVID-19 outbreak". In: (June 2023). arXiv: `2306.07987 [stat.ME]`.

[141] Whitehouse, M., Whiteley, N., and Rimella, L. "Consistent and fast inference in compartmental models of epidemics using Poisson Approximate Likelihoods". en. In: *J. R. Stat. Soc. Series B Stat. Methodol.* (July 2023), qkad065.

[142] Gilks, W. R., Richardson, S., and Spiegelhalter, D. *Markov Chain Monte Carlo in Practice*. Ed. by W. R. Gilks, S. Richardson, and D. Spiegelhalter. Chapman and Hall/CRC Interdisciplinary Statistics. Philadelphia, PA: Chapman and Hall/CRC, Dec. 1995.

[143] Gibson, G. J. and Renshaw, E. "Estimating parameters in stochastic compartmental models using Markov chain methods". en. In: *Math. Med. Biol.* 15.1 (Mar. 1998), pp. 19–40.

[144] Jewell, C. P., Kypraios, T., Neal, P., and Roberts, G. O. "Bayesian analysis for emerging infectious diseases". en. In: {*Bayesian Analysis*} 4.3 (Sept. 2009), pp. 465–496.

[145] Cauchemez, S., Temime, L., Valleron, A.-J., Varon, E., Thomas, G., Guillemot, D., and Boëlle, P.-Y. "S. pneumoniae transmission according to inclusion in conjugate vaccines: Bayesian analysis of a longitudinal follow-up in schools". en. In: *BMC Infect. Dis.* 6 (Jan. 2006), p. 14.

[146] Auranen, K., Arjas, E., Leino, T., and Takala, A. K. "Transmission of Pneumococcal Carriage in Families: A Latent Markov Process Model for Binary Longitudinal Data". In: *J. Am. Stat. Assoc.* 95.452 (Dec. 2000), pp. 1044–1053.

[147] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. "Equation of state calculations by fast computing machines". en. In: *J. Chem. Phys.* 21.6 (June 1953), pp. 1087–1092.

[148] Hastings, W. K. "Monte Carlo Sampling Methods Using Markov Chains and Their Applications". In: *Biometrika* 57.1 (1970), pp. 97–109.

[149] Topriceanu, C.-C., Wong, A., Moon, J. C., et al. "Impact of lockdown on key workers: findings from the COVID-19 survey in four UK national longitudinal studies". en. In: *J. Epidemiol. Community Health* 75.10 (Oct. 2021), pp. 955–962.

[150] European Centre for Disease Prevention and Control. *COVID-19 clusters and outbreaks in occupational settings in the EU/EEA and the UK*. Accessed: 2021. 2020. URL: https://www.ecdc.europa.eu/en/publications-data/covid-19-clusters-and-outbreaks-occupational-settings-eueea-and-uk.

[151] Industrial Injuries Advisory Council. *COVID-19 and occupation: position paper 48*. Tech. rep. Mar. 2021.

[152] Quinn, S. C. and Kumar, S. "Health inequalities and infectious disease epidemics: a challenge for global health security". en. In: *Biosecur. Bioterror.* 12.5 (2014), pp. 263–273.

[153] Daniels, S., Wei, H., Han, Y., et al. "Risk factors associated with respiratory infectious disease-related presenteeism: a rapid review". en. In: *BMC Public Health* 21.1 (Oct. 2021), p. 1955.

[154] Liu, C. Y., Berlin, J., Kiti, M. C., et al. "Rapid review of social contact patterns during the COVID-19 pandemic". In: *Epidemiology* 32.6 (2021).

[155] Chang, S., Pierson, E., Koh, P. W., Gerardin, J., Redbird, B., Grusky, D., and Leskovec, J. "Mobility network models of COVID-19 explain inequities and inform reopening". en. In: *Nature* 589.7840 (Jan. 2021), pp. 82–87.

[156] Kang, Y., Gao, S., Liang, Y., Li, M., Rao, J., and Kruse, J. "Multiscale dynamic human mobility flow dataset in the U.S. during the COVID-19 epidemic". en. In: *Sci Data* 7.1 (Nov. 2020), p. 390.

[157] Davies, N. G., Barnard, R. C., Jarvis, C. I., et al. "Association of tiered restrictions and a second lockdown with COVID-19 deaths and hospital admissions in England: a modelling study". en. In: *Lancet Infect. Dis.* 21.4 (Apr. 2021), pp. 482–492.

[158] Ma, K. C. and Lipsitch, M. "Big data and simple models used to track the spread of COVID-19 in cities". en. In: *Nature* 589.7840 (Jan. 2021), pp. 26–28.

[159] Wu, J. T., Leung, K., and Leung, G. M. "Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study". en. In: *Lancet* 395.10225 (Feb. 2020), pp. 689–697.

[160] Verity, R., Okell, L. C., Dorigatti, I., et al. "Estimates of the severity of coronavirus disease 2019: a model-based analysis". en. In: *Lancet Infect. Dis.* 20.6 (June 2020), pp. 669–677.

[161] Keeling, M. J., Hollingsworth, T. D., and Read, J. M. "Efficacy of contact tracing for the containment of the 2019 novel coronavirus (COVID-19)". en. In: *J. Epidemiol. Community Health* 74.10 (Oct. 2020), pp. 861–866.

[162] Kretzschmar, M. E., Rozhnova, G., Bootsma, M. C. J., Boven, M. van, Wijgert, J. H. H. M. van de, and Bonten, M. J. M. "Impact of delays on effectiveness of contact tracing strategies for COVID-19: a modelling study". en. In: *Lancet Public Health* 5.8 (Aug. 2020), e452–e459.

[163] Salathé, M., Althaus, C., Anderegg, N., et al. "Early evidence of effectiveness of digital contact tracing for SARS-CoV-2 in Switzerland". en. In: *Swiss Med. Wkly* 150 (Dec. 2020), w20457.

[164] Vogt, F., Haire, B., Selvey, L., Katelaris, A. L., and Kaldor, J. "Effectiveness evaluation of digital contact tracing for COVID-19 in New South Wales, Australia". en. In: *Lancet Public Health* 7.3 (Mar. 2022), e250–e258.

[165] Bengio, Y., Janda, R., Yu, Y. W., et al. "The need for privacy with public digital contact tracing during the COVID-19 pandemic". en. In: *Lancet Digit Health* 2.7 (July 2020), e342–e344.

[166] Braithwaite, I., Callender, T., Bullock, M., and Aldridge, R. W. "Automated and partly automated contact tracing: a systematic review to inform the control of COVID-19". en. In: *Lancet Digit Health* 2.11 (Nov. 2020), e607–e621.

[167] Kleef, E. van, Robotham, J. V., Jit, M., Deeny, S. R., and Edmunds, W. J. "Modelling the transmission of healthcare associated infections: a systematic review". en. In: *BMC Infect. Dis.* 13 (June 2013), p. 294.

[168] McBryde, E. S., Pettitt, A. N., Cooper, B. S., and McElwain, D. L. S. "Characterizing an outbreak of vancomycin-resistant enterococci using hidden Markov models". en. In: *J. R. Soc. Interface* 4.15 (Aug. 2007), pp. 745–754.

[169] Cassidy, R., Kypraios, T., and O'Neill, P. D. "Modelling, Bayesian inference, and model assessment for nosocomial pathogens using whole-genome-sequence data". en. In: *Stat. Med.* 39.12 (May 2020), pp. 1746–1765.

[170] Dubé, C., Ribble, C., Kelton, D., and McNab, B. "A review of network analysis terminology and its application to foot-and-mouth disease modelling and policy development". en. In: *Transbound. Emerg. Dis.* 56.3 (Apr. 2009), pp. 73–85.

[171] Kucharski, A. J., Klepac, P., Conlan, A. J. K., et al. "Effectiveness of isolation, testing, contact tracing, and physical distancing on reducing transmission of SARS-CoV-2 in different settings: a mathematical modelling study". en. In: *Lancet Infect. Dis.* 20.10 (Oct. 2020), pp. 1151–1160.

[172] Haw, D. J., Forchini, G., Doohan, P., et al. "Optimizing social and economic activity while containing SARS-CoV-2 transmission using DAEDALUS". en. In: *Nat Comput Sci* 2.4 (Apr. 2022), pp. 223–233.

[173] Béraud, G., Kazmercziak, S., Beutels, P., et al. "The French Connection: The First Large Population-Based Contact Survey in France Relevant for the Spread of Infectious Diseases". en. In: *PLoS One* 10.7 (July 2015), e0133203.

[174] Pokora, R., Kutschbach, S., Weigl, M., et al. "Investigation of superspreading COVID-19 outbreak events in meat and poultry processing plants in Germany: A cross-sectional study". en. In: *PLoS One* 16.6 (June 2021), e0242456.

[175] Eibach, D., Casalegno, J.-S., Bouscambert, M., et al. "Routes of transmission during a nosocomial influenza A(H3N2) outbreak among geriatric patients and healthcare workers". en. In: *J. Hosp. Infect.* 86.3 (Mar. 2014), pp. 188–193.

[176] Raven, K. E., Gouliouris, T., Brodrick, H., et al. "Complex Routes of Nosocomial Vancomycin-Resistant Enterococcus faecium Transmission Revealed by Genome Sequencing". en. In: *Clin. Infect. Dis.* 64.7 (Apr. 2017), pp. 886–893.

[177] Quainoo, S., Coolen, J. P. M., Hijum, S. A. F. T. van, Huynen, M. A., Melchers, W. J. G., Schaik, W. van, and Wertheim, H. F. L. "Whole-Genome Sequencing of Bacterial Pathogens: the Future of Nosocomial Outbreak Analysis". en. In: *Clin. Microbiol. Rev.* 30.4 (Oct. 2017), pp. 1015–1063.

[178] Colton, H., Parker, M. D., Stirrup, O., et al. "Factors affecting turnaround time of SARS-CoV-2 sequencing for inpatient infection prevention and control decision making: analysis of data from the COG-UK HOCI study". en. In: *J. Hosp. Infect.* 131 (Jan. 2023), pp. 34–42.

[179] Ng, G. Y. and Ong, B. C. "Contact tracing using real-time location system (RTLS): a simulation exercise in a tertiary hospital in Singapore". en. In: *BMJ Open* 12.10 (Oct. 2022), e057522.

[180] Hellmich, T. R., Clements, C. M., El-Sherif, N., et al. "Contact tracing with a real-time location system: A case study of increasing relative effectiveness in an emergency department". en. In: *Am. J. Infect. Control* 45.12 (Dec. 2017), pp. 1308–1311.