This is a repository copy of *Characterising within-hospital SARS-CoV-2 transmission events: a retrospective analysis integrating epidemiological and viral genomic data from a UK tertiary care setting across two pandemic waves*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/198938/

Version: Submitted Version

Preprint:
Lindsey, B.B. orcid.org/0000-0003-4227-2592, Villabona-Arenas, C.J. orcid.org/0000-0001-9928-3968, Campbell, F. et al. (22 more authors) (Submitted: 2021) Characterising within-hospital SARS-CoV-2 transmission events: a retrospective analysis integrating epidemiological and viral genomic data from a UK tertiary care setting across two pandemic waves. [Preprint - medRxiv] (Submitted)

https://doi.org/10.1101/2021.07.15.21260537

© 2021 The Author(s). This preprint is made available under a CC-BY-NC-ND 4.0 International license. (http://creativecommons.org/licenses/by-nc-nd/4.0/)

Reuse
This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) licence. This licence only allows you to download this work and share it with others as long as you credit the authors, but you can't change the article in any way or use it commercially. More information and the full terms of the licence here: https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

medRxiv preprint doi: https://doi.org/10.1101/2021.07.15.21260537; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Characterising within-hospital SARS-CoV-2 transmission events: a retrospective analysis integrating epidemiological and viral genomic data from a UK tertiary care setting across two pandemic waves

Benjamin B. Lindsey[1,2]*, Ch. Julián Villabona-Arenas[3,4]*, Finlay Campbell[5]*, Alexander J. Keeley[1,2], Matthew D. Parker[6,7,8], Dhruv R Shah[1], Helena Parsons[1,2], Peijun Zhang[1], Nishchay Kakkar[2], Marta Gallis[1], Benjamin H. Foulkes[1], Paige Wolverson[1], Stavroula F Louka[1], Stella Christou[1], Amy State[2], Katie Johnson[2], Mohammad Raza[1,2], Sharon Hsu[1,7], Thibaut Jombart[3,4,9], Anne Cori[9], Sheffield COVID-19 Genomics Group^, The COVID-19 Genomics UK (COG-UK) consortium[10]^, CMMID COVID-19 working group^, Cariad M. Evans[1,2], David G. Partridge[1,2], Katherine E. Atkins[3,4,11#], Stéphane Hué[3,4#] & Thushan I. de Silva[1,2,12#]

1. The Florey Institute for Host-Pathogen Interactions & Department of Infection, Immunity and Cardiovascular Disease, Medical School, University of Sheffield, Sheffield, UK.
2. Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, UK.
3. Centre for Mathematical Modelling of Infectious Diseases, London School of Hygiene and Tropical Medicine, London, UK.
4. Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK.
5. Health Emergencies Programme, World Health Organization, Geneva, Switzerland.
6. Sheffield Biomedical Research Centre, The University of Sheffield, Sheffield, UK.
7. Sheffield Bioinformatics Core, The University of Sheffield, Sheffield, UK.
8. The Department of Neuroscience/Neuroscience Institute, The University of Sheffield, Sheffield, UK.
9. MRC Centre for Global Infectious Disease Analysis, Department of Infectious Disease Epidemiology, School of Public Health, Imperial College London.
10. https://www.cogconsortium.uk
11. Usher Institute, The University of Edinburgh, Edinburgh, UK.
12. MRC Unit The Gambia at the London School of Hygiene and Tropical Medicine, Fajara, The Gambia.

* Contributed equally

[#] Contributed equally

^ Full list of consortium names and affiliations located in the Appendix

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.15.21260537; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Structured abstract

Objectives - To characterise within-hospital SARS-CoV-2 transmission across two waves of the COVID-19 pandemic.

Design - A retrospective Bayesian modelling study to reconstruct transmission chains amongst 2181 patients and healthcare workers using combined viral genomic and epidemiological data.

Setting - A large UK NHS Trust with over 1400 beds and employing approximately 17,000 staff.

Participants - 780 patients and 522 staff testing SARS-CoV-2 positive between 1st March 2020 and 25th July 2020 (Wave 1); and 580 patients and 299 staff testing SARS-CoV-2 positive between 30th November 2020 and 24th January 2021 (Wave 2).

Main outcome measures - Transmission pairs including who-infected-whom; location of transmission events in hospital; number of secondary cases from each individual, including differences in onward transmission from community and hospital onset patient cases.

Results - Staff-to-staff transmission was estimated to be the most frequent transmission type during Wave 1 (31.6% of observed hospital-acquired infections; 95% CI 26.9 to 35.8%), decreasing to 12.9% (95% CI 9.5 to 15.9%) in Wave 2. Patient-to-patient transmissions increased from 27.1% in Wave 1 (95% CI 23.3 to 31.4%) to 52.1% (95% CI 48.0 to 57.1%) in Wave 2, to become the predominant transmission type. Over 50% of hospital-acquired infections were concentrated in 8/120 locations in Wave 1 and 10/93 locations in Wave 2. Approximately 40% to 50% of hospital-onset patient cases resulted in onward transmission compared to less than 4% of definite community-acquired cases.

Conclusions - Prevention and control measures that evolved during the COVID-19 pandemic may have had a significant impact on reducing infections between healthcare workers, but were insufficient during the second wave to prevent a high number of patient-to-patient transmissions. As hospital-acquired cases appeared to drive most onward transmissions, more frequent and rapid identification and isolation of these cases will be required to break hospital transmission chains in subsequent pandemic waves.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.15.21260537; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has resulted in multiple hospital outbreaks, exposing healthcare workers (HCWs) and non-COVID-19 patients to SARS-CoV-2 infection.[1–3] At least 32,307 patients are thought to have been infected in hospitals in the UK during the pandemic so far and an estimated 414 HCWs have died.[4] To safely continue routine and elective activities in hospitals during times of high SARS-CoV-2 incidence, it is important to discern factors that drive hospital-acquired infections. This greater understanding can be used to protect staff and patients, as well as informing further efforts to contain hospital outbreaks.

Identifying within-hospital SARS-CoV-2 transmission events using epidemiological data remains challenging for two reasons. Firstly, the high variability in the viral incubation period means it is often difficult to determine whether hospital onset cases are community- or hospital-acquired. Secondly, at least 33% of SARS-CoV-2 infections in adults are thought to be asymptomatic[5], therefore identifying all patients and HCWs contributing to transmission is challenging. In the limited instances where viral genomic data were analysed, this information was used to confirm or complement a purely epidemiological approach.[6–9] Elucidating the source of transmission events on the basis of viral genetic relatedness alone also entails considerable uncertainty due to the slow evolutionary rate of SARS-CoV-2.[10] In the time scale of an outbreak, a large proportion of individuals are infected by viruses too genetically similar to each other to distinguish genuine transmission events from unrelated infections. Furthermore, data from HCWs have rarely been included in previous

medRxiv preprint doi: https://doi.org/10.1101/2021.07.15.21260537; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

analyses[11] and the relative role that patients and HCWs have played in fuelling hospital outbreaks in the UK remains largely unknown.

The integration of genomic, epidemiological and location data into a statistical inference framework offers a possible route to more accurate estimates of within-hospital transmission. Under such an approach, a transmission event between a pair of individuals is supported if their symptom onset times are compatible with the serial interval distribution SARS-CoV-2, if the individuals are in the same hospital location at the time of a suspected transmission event and if their viral genomes exhibit a high degree of relatedness.

In this study, we reconstructed SARS-CoV-2 outbreaks in a large NHS teaching hospital trust in England during the first two UK epidemic waves. We integrated over 2,000 viral genomic sequences, patient and staff locations, and routinely-available epidemiological information in a Bayesian framework that incorporates prior knowledge on the relative contributions of within-ward and between-ward transmission, as well as the proportion of unsampled cases.[12,13] Using this approach, we characterised the dynamics of SARS-CoV-2 transmission within a hospital setting, identifying key differences across the two pandemic waves, as well as the relative contribution of different groups and hospital locations to within-hospital transmission.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.15.21260537; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Methods

Study population

All cases in the study were patients or staff who tested positive for SARS-CoV-2 at Sheffield Teaching Hospitals NHS Foundation Trust (STHNFT), Sheffield, UK, between 1st March 2020 and 24th January 2021. STHNFT is a large UK NHS hospital Trust which includes five hospitals, has an average bed occupancy of 1400, and employs approximately 17,000 staff. SARS-CoV-2 nucleic acid amplification tests (NAAT) were performed on nose and/or throat swabs throughout the pandemic in line with contemporaneous UK Department of Health and Social Care guidance[14], using Hologic Panther or an in-house dual E/RdRp gene real time PCR assay.[15,16]

Patients were included in the analysis if they tested positive for SARS-CoV-2 at or during admission. Staff were included if they tested positive for SARS-CoV-2 and had worked in a clinical area in the 14 days prior to a positive test. Information on symptom onset of patients and their ward movements, together with place of work for staff, were extracted from STHNFT electronic records, when available.

Sample Preparation, ARTIC Network PCR and Nanopore Sequencing

Sequencing was attempted on all available residual samples collected for routine diagnostic testing from STHNFT throughout the study period, with fluctuation in the proportion of positive samples sequenced due to multiple factors, including laboratory capacity and availability of stored samples. The first positive sample from each individual was selected for

medRxiv preprint doi: https://doi.org/10.1101/2021.07.15.21260537; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

sequencing. RNA was extracted from viral transport medium and subject to the ARTIC network tiled amplicon protocol[17], followed by sequencing on an Oxford Nanopore GridION X5. Base calling was performed using a high accuracy model and the default basecaller in MinKNOW (currently guppy v4). Reads were filtered based on quality and length (400bp to 700bp) and mapped to the Wuhan reference genome (GenBank accession number NC_045512). Reads were downsampled to 200x coverage in each direction and variants called using nanopolish[18] to determine changes from the reference, followed by consensus sequence generation. Samples with over 90% genome coverage were included for further analysis. Viral genomic sequences were classified into PANGO lineages using the *Phylogenetic Assignment of Named Global Outbreak LINeages* (PANGOLIN)[19] version 2.4.2 and a multiple sequence alignment built using MAFFT[20] with 10 iterative refinements. All alignment positions flagged as problematic for phylogenetic inference were removed, including highly homoplasic positions and 3′ and 5′ ends [21].

Hospital outbreak reconstruction model

To identify likely transmission events between individuals, we extended the most recent implementation of the *outbreaker2* model to capture ward-level transmission[12,13]. Our Bayesian model calculates the likelihood of a transmission event from case i to case j at a putative transmission time given the time of symptom onset for case i and j, the Hamming distance between the corresponding virus genetic sequences, and the ward that i and j were on at the time of infection . To formulate the likelihood, we used information on the generation time distribution (the delay between the infection of a primary case and the infection of a secondary case), the incubation period distribution (the delay between the

medRxiv preprint doi: https://doi.org/10.1101/2021.07.15.21260537; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

infection and symptom onset) and the assumption that an infection happened within a ward rather than between individuals on different wards.[22,23] We defined the ascertainment probability as the proportion of cases that were captured in our dataset and calculated this as the product of (i) the proportion of known cases with high quality sequence, (ii) the proportion of these cases where the ward location was known, and (iii) the proportion of all cases that were likely detected before discharge via testing. In our base case analysis, we used a global sensitivity method to capture uncertainty in (i) the proportion of cases that were imported community-acquired infections, and (ii) the symptom onset date for individuals for whom this information was unavailable and (iii) multiple work locations for some staff. We also used a point estimate for our ascertainment probability but varied this estimate in a one-way sensitivity analysis (**Full details in File S2**). A final posterior distribution of 10,000 transmission networks was inferred by integrating over the uncertainty across imputed datasets of symptom onset times, multiple staff work locations and the number (and identity) of community-acquired cases. Full details of the model, including model fitting and prior distributions, are provided in **Table S1**. All code is available at https://github.com/Chjulian/sheffield_HT.

Ethical approval

Approval for the study was obtained from the UK Health Research Authority (IRAS 281918), with sequencing performed according to The COVID-19 Genomics UK (COG-UK) study protocol approved by the Public Health England Research Ethics Governance Group (R&D NR0195).

Role of the funding source

medRxiv preprint doi: https://doi.org/10.1101/2021.07.15.21260537; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

The funder had no role in study design, data collection, data analysis, data interpretation or writing of the report.

Results

Study population, SARS-CoV-2 testing and infection control measures

During the first wave of the UK epidemic (Wave 1; defined as 1st March to 25th July 2020 for our analysis), 886/1,184 (74.8%) patients and 842/1,104 (76.3%) HCWs at STHNFT who tested positive for SARS-CoV-2 had sequence data available with over 90% genome coverage (**Figure 1**). During the second wave (Wave 2; defined here as 30th November 2020 to 24th January 2021) 669/1,183 (56.6%) SARS-CoV-2 positive patients and 651/838 (77.7%) SARS-CoV-2 positive HCWs had sequence data available with over 90% genome coverage. Cases were excluded if they were outpatients, non-clinical staff, non-STHNFT or community-based staff, household contacts of staff members, and staff who had missing ward location data, leaving 1,302 individuals in Wave 1 and 879 individuals in Wave 2 for the analysis (**Figure 1B**, **Table 1**).

SARS-CoV-2 testing policy and infection prevention and control (IPC) measures evolved throughout the pandemic (**Figure 1A**). Testing was performed in symptomatic patients with suspected SARS-CoV-2 infection on admission throughout the study period, with testing offered to symptomatic staff from 17th March 2020.[24] Testing of all admissions regardless of symptoms commenced on 25th April 2020 and screening of all asymptomatic patients and staff on wards with outbreaks from 18th May 2020. In addition to screening on admission, all patients were routinely tested on day 5 of admission from 1st Sep 2020. Routine twice weekly

medRxiv preprint doi: https://doi.org/10.1101/2021.07.15.21260537; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

testing using lateral flow devices, followed by confirmatory NAAT, was offered to staff in all clinical areas from 8th December 2020. Level 2 personal protective equipment (PPE; aprons, gloves, eye protection and fluid resistant surgical face mask) was used by staff only for seeing suspected COVID-19 cases from 17th March 2020, and for all patient contact from 8th April 2020. HCWs were mandated to wear surgical face masks in all areas of the hospital from 15th June 2020. The SARS-CoV-2 staff vaccination programme commenced on 10th December 2020.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.15.21260537; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

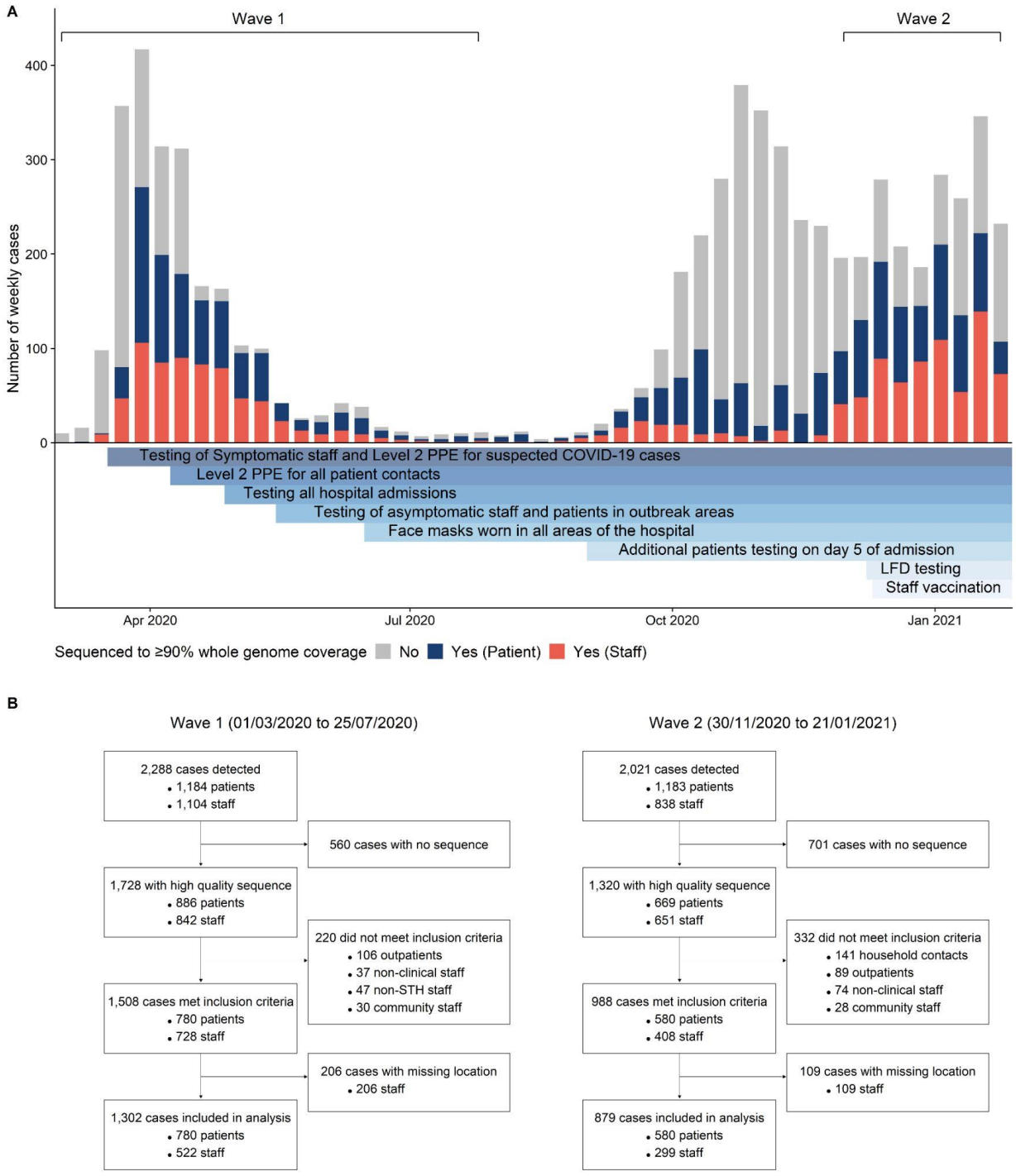


***Figure* 1. SARS-CoV-2 positive staff and healthcare worker samples included in the study**

(A) SARS-CoV-2 *positive cases in patients and staff at Sheffield Teaching Hospitals NHS*

medRxiv preprint doi: https://doi.org/10.1101/2021.07.15.21260537; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Foundation Trust (STH) over time, and implementation of testing, prevention and control interventions. All SARS-CoV-2 nucleic acid amplification tests from STH patients and staff found to be positive in the hospital (pillar 1) diagnostic laboratory are shown in grey bars. Sequences with over 90% genome coverage for samples from healthcare workers (red) and patients (blue) are shown. These relate to samples with high quality sequences (1,728 in Wave 1 and 1,320 in Wave 2) shown in ***Figure 1B****. Level 2 positive protective equipment (PPE) - aprons, gloves, eye protection and fluid resistant surgical face mask. LFD testing - Lateral flow device testing for staff two times per week. Grey bars represent SARS-CoV-2 cases tested (staff and patients) testing positive in the STH pillar 1 diagnostic laboratory with no high quality sequence available. Wave 1 and Wave 2 denote periods and samples included in the study. (B) Details of SARS-CoV-2 positive cases from patients and healthcare workers sequenced and included in the study.*

Viral genomes were classified into 64 different PANGO lineages for Wave 1 and 24 lineages for Wave 2. Lineages B.1.1.1 (471/1,302, 36.2%), B.1.1.119 (180/1,302, 13.8%) and B.1 (110/1,302, 8.4%) predominated during Wave 1, while lineages B.1.177 (293/879, 33.3%), B.1.1.7 (263/879, 29.9%) and B.1.177.4 (112/879, 12.7%) predominated during Wave 2 (**Figure S1)**.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.15.21260537; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

	Wave 1	Wave 2
Dates	01/03/2020 to 25/07/2020	30/11/2020 to 24/01/2021
Total cases	1302	879
Patients	780 (59.9%)	580 (66.0%)
Staff	522 (40.1%)	299 (34.0%)
Number of hospital locations	120	93
Global PANGO Lineages	64	24
Fraction of patients with inferred symptom onset date	30.1%	22.0%
Median ward movements per patient (range)	3 (1-9)	2 (1-9)
Patient classification		
Community onset - community associated	486 (62.3%*)	236 (40.7%*)
Community onset - suspected healthcare associated	79 (10.1%*)	67 (11.6%*)
Hospital onset - healthcare associated	80 (10.3%*)	100 (17.2%*)
Hospital onset - suspected healthcare associated	74 (9.5%*)	88 (15.2%*)
Hospital onset - indeterminate healthcare associated	61 (7.8%*)	89 (15.3%*)

* Percentage of patient cases

Table 1. Summary of study cohort. *Number of hospital locations includes wards and non-clinical areas. Ward movements refers to between ward movements and does not include bed movements within the same ward. Classification of patient cases according to likely source of infection (community or hospital-acquired) is based on SAGE criteria[25]. Community onset-community associated = positive test up to 14 days before or within 2 days after hospital admission; Community onset-suspected healthcare associated = positive test up to 14 days before or within 2 days after admission, with discharge from hospital within 14 days before test; Hospital onset- intermediate healthcare associated = positive test 3-7 days after hospital admission; Hospital onset-suspected healthcare associated = positive test 8-14 days after admission or 3-14 days after admission with discharge from hospital in 14 days before test; Hospital onset-healthcare associated = positive test 15 or more days after hospital admission.*

medRxiv preprint doi: https://doi.org/10.1101/2021.07.15.21260537; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Quantifying hospital-acquired infections

Using admission dates, symptom onset dates, SARS-CoV-2 positive test dates, ward location of cases and comparison between consensus viral genome sequences, our model inferred likely hospital transmissions between individuals, together with (i) the time of those events, (ii) the ward on which the transmission occurred, and (iii) whether the infector was a sampled or unsampled case.

Our model identified 473 (95% credible interval (CI) 310 to 688) hospital-acquired infections among our study population during Wave 1, of which 85 (95% CI 21 to 192) were unsampled. During Wave 2, there were 402 (95% CI 293 to 538) estimated hospital-acquired infections, of which 52 (95% CI 11 to 120) were unsampled. In Wave 1, patient cases comprised 40.9% (95% CI 36.1 to 45.5%) of all sampled hospital-acquired cases, compared to 65.1% (95% CI 60.3 to 69.4%) of all sampled hospital-acquired cases in Wave 2. Our model estimates showed good agreement with previous *a priori* inpatient epidemiological definitions of community or hospital onset categories (**Table 1 and Table S2**).[25] Specifically, no 'community onset-community associated' patient cases were identified as hospital-acquired by our model during either wave, and the majority of 'hospital onset-hospital acquired' and 'hospital onset-suspected hospital acquired' cases were identified as likely hospital-acquired by our model.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.15.21260537; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Transmission chain reconstruction

We identified 95 (95% CI 82 to 109) transmission chains (defined as contiguous transmission events between 2 or more cases) in Wave 1 and 72 (95% CI to 61-84) transmission chains in Wave 2 (**Figure S2**). The median number of cases per transmission chain was 3 (95% CI 3 to 4) in Wave 1 and 4 (95% CI 3 to 5) in Wave 2. A staff member was identified as the index case in 50.6% (95% CI 42.0 to 58.0%) of transmission chains in Wave 1 and in 31.3% (95% CI 23.1 to 39.7%) of transmission chains in Wave 2. Forty different PANGO lineages were involved in transmission chains in Wave 1 and 13 were found in Wave 2 chains.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.15.21260537; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

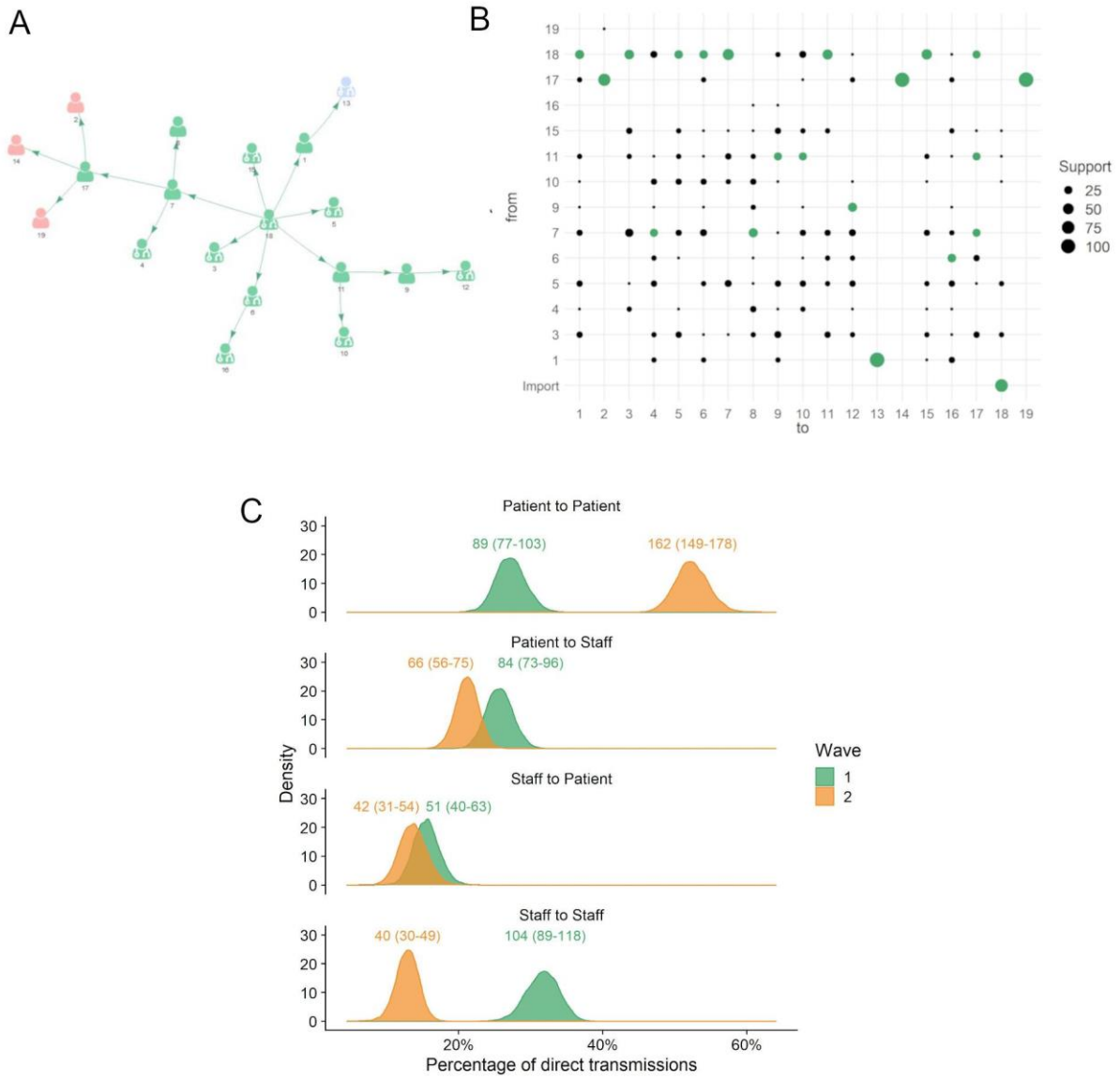


Figure 2. Within-hospital transmission chains and estimated infections within and between patients and healthcare workers (A) *An example transmission chain showing transmission between staff (icons with stethoscope) and patients (icons without stethoscope). Icons are labeled with their case identification number. Each color represents a separate ward where the infection occurred.* (B) *Uncertainty plot representing the support for each potential transmission pair shown in the example network in panel 2A. Support represents the*

medRxiv preprint doi: https://doi.org/10.1101/2021.07.15.21260537; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

percentage of networks across all those sampled, where a given infector is assigned to each case. Numbers on the plot correspond to the case identification numbers in Fig 2A and the green circles correspond to the transmission pairs displayed. Import = likely community-acquired infection imported into hospital. From = infector, to = infectee. (C) Comparison of the percentage of each transmission type between the two waves. The distributions display the percentages throughout the 10,000 plausible networks. *The numbers above each distribution are the absolute numbers of each transmission pair with 95% credible intervals shown within the brackets.*

Of the transmissions between sampled cases in Wave 1, 31.6% (104/329, 95% CI 26.9 to 35.8%) were staff-to-staff events, 27.1% (89/329, 95% CI 23.3 to 31.4%) were patient-to-patient, 25.5% (84/329, 95% CI 22.1 to 29.3%) were patient-to-staff and 15.5% (51/329, 95% CI 12.2 to 19.1%) were staff-to-patient (**Figure 2C**). By contrast, during Wave 2, the majority of transmission events (162/311; 52.1%, 95% CI 48.0 to 57.1%) between sampled cases were patient-to-patient events, with 21.2% (66/311, 95% CI 18.0 to 24.1%) patient-to-staff, 13.5% (42/311, 95% CI 10.1 to 17.5%) staff-to-patient and 12.9% (40/311, 95% CI 9.5 to 15.9%) staff-to-staff transmission events (**Figure 2C**).

In Wave 1, 55.3% (104/188, 95% CI 48.9 to 61.2%) of staff infections resulted from another staff case, which decreased to 37.7% (40/106, 95% CI 29.3 to 45.4%) in Wave 2. In Wave 1, 63.6% (89/140, 95% CI 56.1 to 71.0%) of patient infections resulted from another patient case which increased to 79.4% (162/204, 95% CI 73.7 to 84.6%) in Wave 2.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.15.21260537; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Ward and Bay level transmission

Identified transmission events were not evenly distributed across the 132 hospital locations included, but isolated to 38 wards in 3 of the 5 hospitals within STHNFT. The 8 wards with the highest number of infections in Wave 1 accounted for 51.0% (95% CI 39.7 to 63.4%) of all transmissions, indicating the presence of transmission hot spots. A similar finding was observed during Wave 2, where 10 wards accounted for 50.1% (95% CI 40.6 to 60.5%) of all transmission events (**Figure 3**). We found evidence that the relative importance of specific wards in contributing to overall transmission was maintained across the two waves (Spearman's Rank correlation Rho 0.54, $P<0.0001$, Ranked by mean number of transmissions per ward). However, there was considerable variability between waves, and several wards that were transmission hotspots in Wave 1 did not make up the 10 wards accounting for >50% of transmissions in Wave 2 (**Figure 3A-B**). Equally, several wards with no transmission event during Wave 1 were identified as transmission hotspots in Wave 2. Considerable variability was also seen between wards in the proportion of infectors and infectees made up by patients and staff (**Figure 3C-F**).

The highest number of infections on a single ward was 50 (95% CI 40 to 58) in Wave 1 but decreased to 30 (95% CI 23 to 36) in Wave 2. The highest number of separate transmission chains on a single ward was 8 (95% CI 5 to 11) for Wave 1 and 5 (95% CI 3 to 7) for Wave 2.

Wards comprise a combination of multi-bed bays with shared bathroom facilities and individual en-suite side rooms. We used a post hoc analysis to evaluate the contribution of bay-level transmission between patients to the outbreak. We identified 38.3% (95% CI 29.9

medRxiv preprint doi: https://doi.org/10.1101/2021.07.15.21260537; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

to 47.1%) of patient-patient transmissions in Wave 1 and 33.8% (95% CI 27.9 to 39.6%) in Wave 2 were between patients who shared a bay at some point during their stay. We estimated an increased risk of transmission between individuals who shared a bay compared with those who shared a ward as 2.8 (95% CI 2.2 to 3.5) times higher in Wave 1 and a 2.5 (95% CI 2.1 to 2.9) times higher in Wave 2.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.15.21260537; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

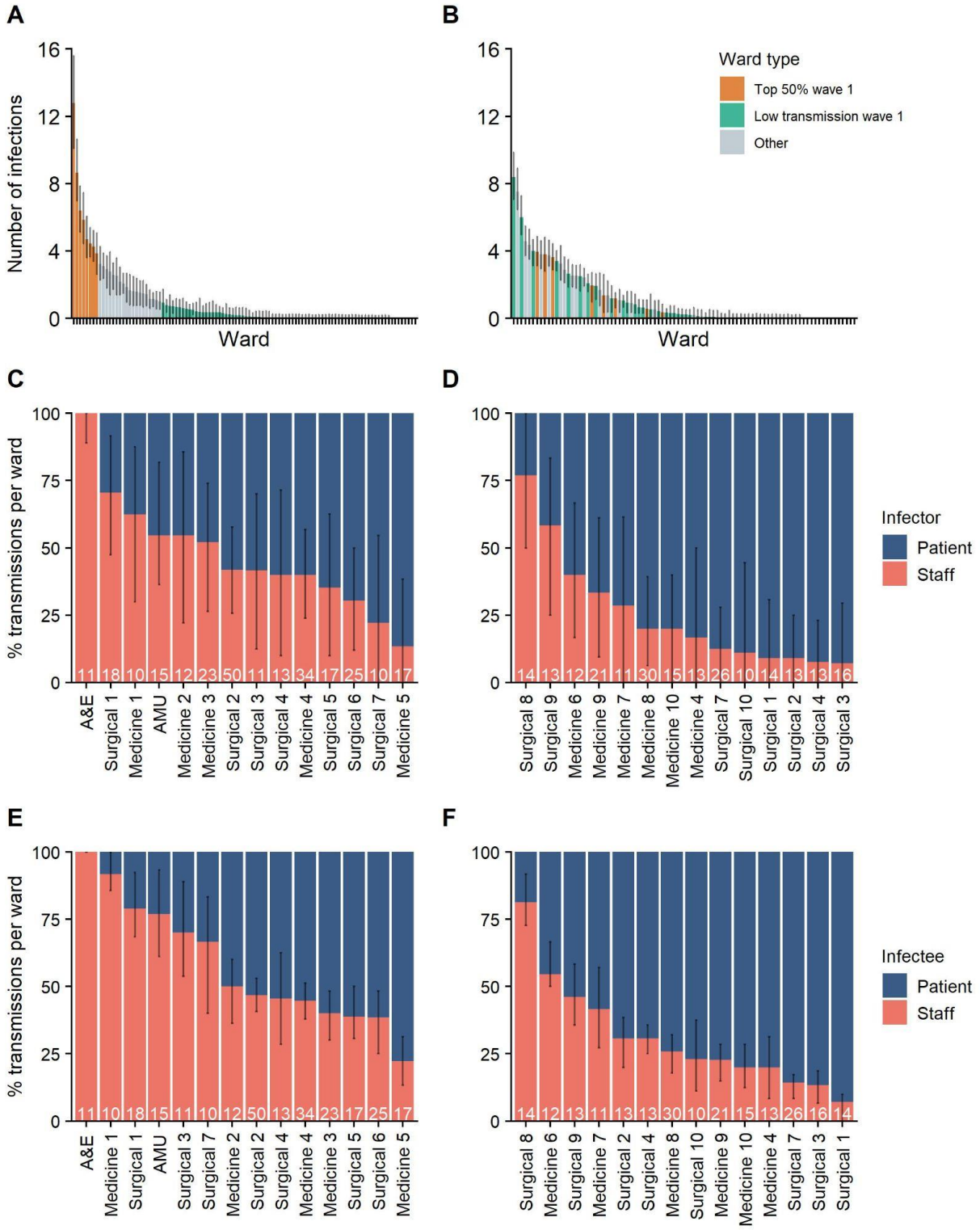


Figure 3. Hospital locations contributing to transmission events. *Wards are ordered by the percentage of total transmission events identified in Wave 1 (A) and Wave 2 (B). Wards*

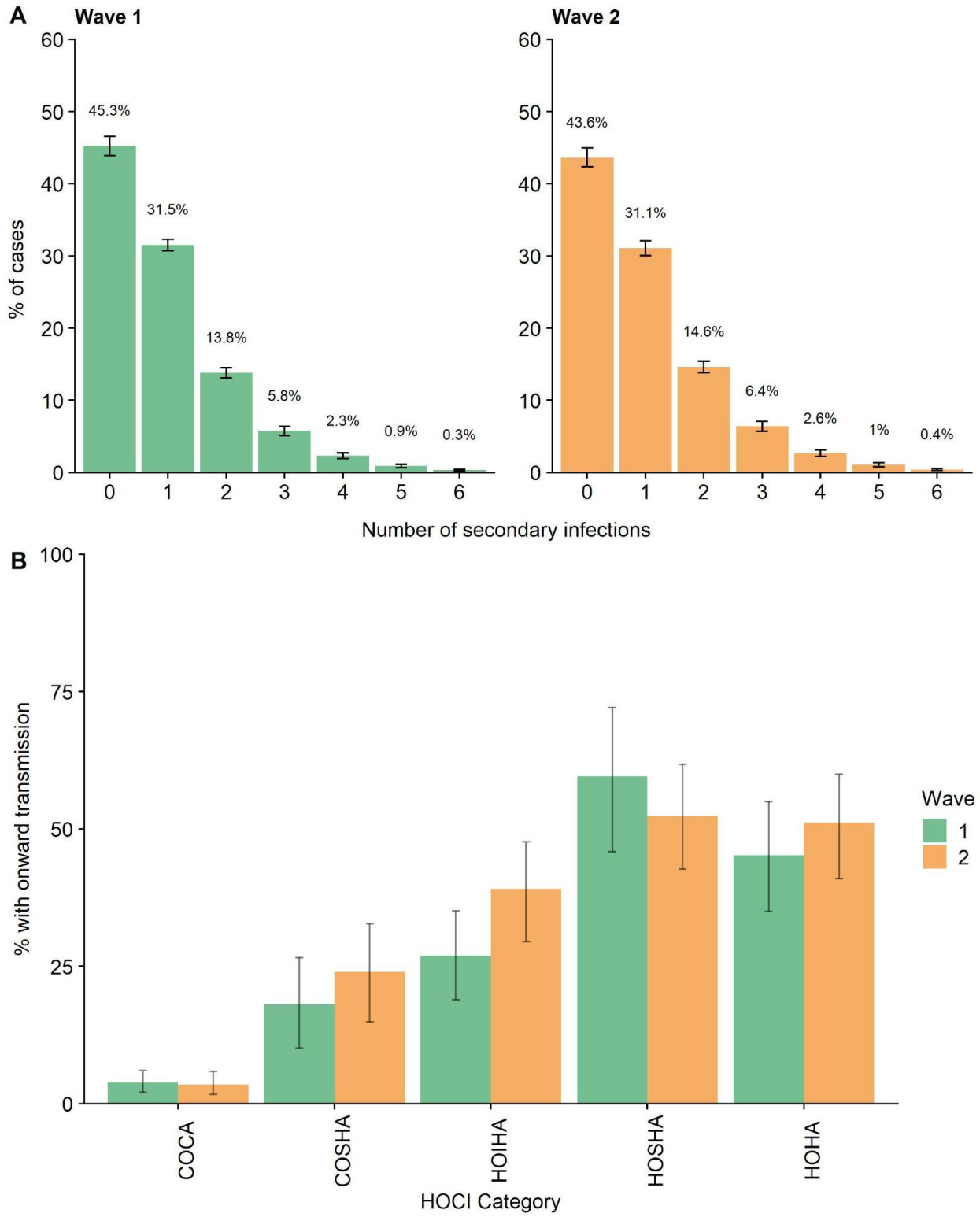
medRxiv preprint doi: https://doi.org/10.1101/2021.07.15.21260537; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

contributing to 50% of all transmission events in Wave 1 (n=8) are coloured in red in both Wave 1 and Wave 2 datasets. Wards with less than 1 transmission (mean of 10,000 networks) in Wave 1 are coloured in green in both Wave 1 and Wave 2 datasets. (C-F) Percentage of staff and patient infector and infectees per ward involved in transmission events. All wards with 10 or more transmissions in Wave 1 and Wave 2 are shown. Wards are ordered by the percentage of staff infector cases in Wave 1 (c) and Wave 2 (d) and percentage of staff infectee cases in Wave 1 (e) and Wave 2 (f). Numbers of transmission events per ward are shown at the bottom of each column. A&E = *accident and emergency department*; AMU = *acute medical unit.*

Secondary cases

The crude mean number of secondary cases was 0.30 (95% CI 0.21 to 0.38) for Wave 1 and 0.40 (95% CI 0.31 to 0.48) for Wave 2. Adjusting for unsampled cases, we estimated that 45.3% (95% CI 43.9 to 46.6%) of infections in Wave 1 and 43.6% (95% 42.3 to 45.0%) of infections in Wave 2 resulted in no onward transmission. Only 0.5% (95% CI 0.3 to 0.6%) infections in Wave 1 and 0.6% (95% CI 0.4 to 0.9%) infections in Wave 2 resulted in more than 5 secondary cases (**Figure 4**). Fewer patients classified as having community onset-community associated infections gave rise to secondary cases within the hospital (3.7%, 95% CI 1.8 to 5.6 for Wave 1; 3.5%, 95% CI 1.7 to 5.9 for Wave 2) compared to those with hospital onset-hospital acquired infections (45.5%, 95% CI 33.8 to 55.0 for Wave 1; 51.2%, 95% CI 41.0 to 60.0 for Wave 2), or other categories of hospital-onset cases (**Figure 4b**, **Table S2**). All findings were consistent across sensitivity analyses (**Figures S3-4**)

medRxiv preprint doi: https://doi.org/10.1101/2021.07.15.21260537; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .



medRxiv preprint doi: https://doi.org/10.1101/2021.07.15.21260537; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Figure 4. Secondary cases distributions and onward infections by HOCI category. (A) *the percentage of cases with each number of secondary cases after adjusting for unsampled cases.* (B) *The percentage of each HOCI category case with at least 1 onward infection.* COCA = *Community onset community associated*, COSHA = *Community onset suspected hospital associated*, HOIHA = *Hospital onset indeterminate hospital associated*, HOSHA = *Hospital onset suspected hospital associated*, HOHA = *Hospital onset hospital associated. Classification of patient cases according to likely source of infection (community or hospital-acquired) is based on* SAGE *criteria*[25]. *Community onset-community associated = positive test up to 14 days before or within 2 days after hospital admission; Community onset-suspected healthcare associated = positive test up to 14 days before or within 2 days after admission, with discharge from hospital within 14 days before test; Hospital onset- intermediate healthcare associated = positive test 3-7 days after hospital admission; Hospital onset-suspected healthcare associated = positive test 8-14 days after admission or 3-14 days after admission with discharge from hospital in 14 days before test; Hospital onset-healthcare associated = positive test 15 or more days after hospital admission.*

Discussion

To our knowledge, our findings represent the largest collection of SARS-CoV-2 genomic and hospital epidemiology data to date used to reconstruct directional transmission networks, where we estimated hospital-acquired SARS-CoV-2 infections across two pandemic waves in the UK using a Bayesian framework. Importantly, our model accounts for unsampled events within these networks, which is crucial given the likely presence of unidentified infections or those lacking sequence data. We observed different contributions to the total

medRxiv preprint doi: https://doi.org/10.1101/2021.07.15.21260537; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

number of within-hospital transmission events from those occurring between and within staff and patients across the two waves. We identified transmission hotspots within our institution, with a relatively small proportion of locations accounting for most hospital-acquired infections in staff and patients. We also found that the majority of SARS-CoV-2 infections resulted in onward transmission, with secondary cases identified in greater than 50% of infections but relatively few so called 'superspreader' events.

While much attention has been paid to staff potentially acquiring SARS-CoV-2 infections from patients due to perceived or real deficiencies in PPE, our findings suggest that the majority of HCW infections during the first pandemic wave were acquired from other HCWs. The contribution of these staff-to-staff infections to hospital-acquired transmission reduced dramatically during the autumn 2020 wave of SARS-CoV-2. Staff were less likely to initiate hospital transmission chains in Wave 2, accounting for 31.3% of index cases compared to 50.6% in Wave 1. Infection control practice and understanding of SARS-CoV-2 transmission evolved considerably during the pandemic. Improved social distancing and wearing of face coverings in non-clinical areas may explain some of these observations. In addition, the importance of asymptomatic transmission was increasingly appreciated and twice weekly lateral flow testing for healthcare workers was introduced in December 2020. Furthermore, seroprevalence rates of over 25% have been reported in HCWs following the first pandemic wave,[26] including in our NHS Trust,[27] which may have contributed to greater protection and reduced transmission in some areas. For example, we have reported that the SARS-CoV-2 seroprevalence in staff working on our acute medical unit by June 2020

medRxiv preprint doi: https://doi.org/10.1101/2021.07.15.21260537; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

was over 40%.[27] This was an area which was a hotspot for transmissions involving staff during our Wave 1 analysis but had very few transmission events identified during Wave 2. Nevertheless, by Wave 2, most staff infections in our NHS trust were estimated to have been acquired from patients, so further efforts are required to increase protection for HCWs. Staff vaccination is anticipated to have a large impact but is unlikely to have played a significant role in our observations due to introduction towards the end of Wave 2.

Hospital-acquired infections during Wave 2 were overwhelmingly dominated by patient-to-patient transmissions. The reasons for these events are likely to be multifactorial. UK hospitals faced significant bed pressures during this period and unlike during Wave 1, attempts were made to maintain as many routine and elective procedures for as long as possible. By this point, all patients in our hospitals were being routinely tested by NAAT for SARS-CoV-2 on admission and on day 5. Accordingly, the percentage of patients included in our dataset with asymptomatic infection increased from 10.4% during Wave 1 to 23.9% in Wave 2. The intense increase in patient-to-patient infections unfortunately occurred despite this enhanced focus on preventing asymptomatic transmission. Most of our transmission hotspots were wards built over two decades ago, with 6 – 8 beds per bay, and shared toilet facilities between every 1 to 2 bays.[28] While ventilation in these settings is in line with applicable regulations at the time of construction, none were designed with a respiratory pandemic in mind. Any contribution from these fixed estate issues will be challenging to address in a short timeframe. While viruses with greater transmissibility could also have played a role during Wave 2, circulation of the B.1.1.7/alpha variant occurred

medRxiv preprint doi: https://doi.org/10.1101/2021.07.15.21260537; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

relatively late in our region compared to many other parts of the UK, and many Wave 2 transmission events were due to other SARS-CoV-2 lineages.

We found that the distribution of secondary cases was very similar across both waves, with over 50% of SARS-CoV-2 cases resulting in onward transmission, although only 10% of all infections resulted in more than two secondary cases. Our findings are different from those in a smaller study focusing on a few large clusters in another UK hospital, where 20% of individuals caused 80% of transmission events.[11] Importantly, we find that hospital-acquired SARS-CoV-2 cases give rise to a greater number of secondary cases than community-onset community-associated cases. Cases admitted from the community already suspected as having COVID-19 will have been isolated in single cubicles or COVID-19 cohort areas more rapidly, thus limiting opportunities for onward transmission. As severe disease requiring hospitalisation often occurs later in infection, they may also be at a less infectious stage, although hospitalised cases may also shed viable virus for longer.[29] In contrast, individuals with no SARS-CoV-2 symptoms who later acquired nosocomial infection may have initially be placed in bays with other susceptible patients, all of whom tested SARS-CoV-2 negative on screening tests at admission. Given the high viral loads during the first few days of infection, including during pre-symptomatic stages,[30] our data suggest that these individuals may acquire SARS-CoV-2 in hospital and have ample opportunity for onward transmission before being detected and isolated. This finding indicates that asymptomatic testing of patients on admission and day 5 was insufficient to prevent these scenarios. Daily testing of patients in the first week of admission or more regular testing throughout admission may allow greater opportunity for intervention, as well as more recent recommendations to IPC guidance such as routine wearing of masks by all

medRxiv preprint doi: https://doi.org/10.1101/2021.07.15.21260537; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

patients in bays. Equally, rapid point-of-care testing (POCT) on admission may also reduce the window for transmission early in admission as it allows earlier isolation of asymptomatic community-acquired cases. Our Trust instituted POCT for all medical admissions in mid-January 2021.

Our study has several limitations that are important to consider. Firstly, despite the large number of individuals included, this is a single centre study and may not be generalisable across all UK hospitals given the heterogeneity in practice, building infrastructure, and patient population that exists. Our organisation had a high number of documented hospital-acquired infections in patients between March 2020 and March 2021 (n=795), but was not an outlier with 7 other NHS Trusts with higher numbers (highest n=1,463).[31] Seven of the top 10 busiest NHS Trusts (including our own) were also in the top 10 Trusts with the highest number of hospital-acquired COVID-19 infections in patients, indicating a common theme that may be a driver of nosocomial SARS-CoV-2 infections.[32] The effectiveness of various infection control measures on within-hospital transmissions over time in our setting is also likely to be generalisable to many UK hospitals, as they were based on national guidance applicable to all NHS Trusts. Many institutions will share similar issues regarding outdated infrastructure with 43% of NHS Trusts occupying estates greater than 30 years old.[33]

Although we did not have a selective sampling strategy, either for case detection or sequencing of positive cases, it is possible that there was an unobserved sampling bias. For example, as individuals with higher viral loads will be more infectious and their samples more likely to result in successful sequencing, they are more likely to have been included in our dataset. Our model attempted to account for unsampled events but would not mitigate any

medRxiv preprint doi: https://doi.org/10.1101/2021.07.15.21260537; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

bias entirely. While we had electronic records of precise location data for patients during all times of their admission, staff location data were less granular and dependent on self-reported areas of work in the 14 days prior to infection. We also had to make several assumptions regarding priors in our model, but have attempted to undertake sensitivity analyses to test any potential impact of these .

With this study, we provide evidence that the integration of clinical surveillance data, viral genomic information and modelling enhances our capacity to unravel the complex transmission dynamics of SARS-CoV-2 in times and places of high incidence. The application of such a high-resolution framework to healthcare settings offers attractive perspectives for guiding the development of a safe environment for both staff and patients, as it may have a significant impact on the reduction of SARS-CoV-2 hospital transmission in subsequent epidemic waves.

Contributions

Contributor roles were assigned as per http://credit.niso.org/. BBL, CJVA, FC, KEA, SH & TIdS were involved in the conceptualisation of the study. BBL, AJK, MDP, DRS, PZ, NK, MG, BHF, PW, SFL, SC, AS, KJ, MR, SH, HP, CME, DGP, KEA, SH and TIdS were involved in data collection and curation. BBL, CJVA, FC, MDP, SH and KEA were involved in data analysis. AC, TJ, SH, KEA and TIdS were involved in supervision of the project. BBL and CJVA were involved in data visualisation. FC was involved in software development. BBL, CJVA, FC, KEA, SH & TIdS wrote the original draft. All authors were involved in reviewing and editing the final manuscript.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.15.21260537; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Declaration of interests

The authors declare no competing interests.

Acknowledgements

We thank the Sheffield Bioinformatics Core for their thoughtful discussion. We would like to thank the members of the Sheffield Biomedical Research centre for their continued support of the SARS-CoV-2 sequencing work in Sheffield. We thank all partners of and contributors to the COG-UK consortium, who are listed at https://www.cogconsortium.uk/about/. Sequencing of SARS-CoV-2 samples was undertaken by the Sheffield COVID-19 Genomics Group as part of the COG-UK CONSORTIUM and supported by funding from the Medical Research Council (MRC) part of UK Research & Innovation (UKRI), the National Institute of Health Research (NIHR) and Genome Research Limited, operating as the Wellcome Sanger Institute. MDP and DW are funded by the NIHR Sheffield Biomedical Research Centre (BRC - IS-BRC-1215-20017). TIdS is supported by a Wellcome Trust Intermediate Clinical Fellowship (110058/Z/15/Z). CJVA and KEA were funded by an ERC Starting Grant (action number 757688). This study is partially funded by the NIHR Health Protection Research Unit in Modelling and Health Economics, a partnership between Public Health England, Imperial College London and LSHTM (grant code NIHR200908); and acknowledges funding from the MRC Centre for Global Infectious Disease Analysis (reference MR/R015600/1), jointly funded by the UK MRC and the UK Foreign, Commonwealth & Development Office (FCDO), under the MRC/FCDO Concordat agreement and is also part of the EDCTP2 programme supported by the European Union. *Disclaimer*: "*The views expressed are those of*

medRxiv preprint doi: https://doi.org/10.1101/2021.07.15.21260537; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

the author(s) and not necessarily those of the NIHR, Public Health England or the Department of Health and Social Care."

References

1 Zhan M, Qin Y, Xue X, *et al.* Death from Covid-19 of 23 Health Care Workers in China. N *Engl J Med* 2020;**382**:2267–8.

2 Rickman HM, 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. *Clin Infect Dis* 2020;**72**:690–3.

3 Heinzerling A, Stuckey MJ, Scheuer T, *et al.* Transmission of COVID-19 to Health Care Personnel During Exposures to a Hospitalized Patient — Solano County, California, February 2020. MMWR. Morbidity and Mortality Weekly Report. 2020;**69**:472–6. doi:10.15585/mmwr.mm6915e5

4 Nasir DBW-SA. Coronavirus (COVID-19) related deaths by occupation, England and Wales - Office for National Statistics. 2021.https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/causesofdeath/bulletins/coronaviruscovid19relateddeathsbyoccupationenglandandwales/deathsregisteredbetween9marchand28december2020 (accessed 3 Jun 2021).

5 Oran DP, Topol EJ. The Proportion of SARS-CoV-2 Infections That Are Asymptomatic : A Systematic Review. *Ann Intern Med* 2021;**174**:655–62.

6 Lucey M, Macori G, Mullane N, *et al.* Whole-genome sequencing to track SARS-CoV-2 transmission in nosocomial outbreaks. *Clin Infect Dis* Published Online First: 19 September 2020. doi:10.1093/cid/ciaa1433

7 Hamilton WL, Tonkin-Hill G, Smith ER, *et al.* Genomic epidemiology of COVID-19 in care homes in the east of England. *Elife* 2021;**10**. doi:10.7554/eLife.64618

8 Ellingford JM, George R, McDermott JH, *et al.* Genomic and healthcare dynamics of nosocomial SARS-CoV-2 transmission. *Elife* 2021;**10**. doi:10.7554/eLife.65453

9 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. *Viruses* 2021;**13**. doi:10.3390/v13040604

10 Abbas M, Nunes TR, Cori A, *et al.* Explosive Nosocomial Outbreak of SARS-CoV-2 in a Rehabilitation Clinic: The Limits of Genomics for Outbreak Reconstruction. SSRN

medRxiv preprint doi: https://doi.org/10.1101/2021.07.15.21260537; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Electronic Journal. doi:10.2139/ssrn.3800084

11 Illingworth C, Hamilton W, Warne B, *et al.* Superspreaders drive the largest outbreaks of hospital onset COVID-19 infection. 2021. doi:10.31219/osf.io/wmkn3

12 Campbell F, Didelot X, Fitzjohn R, *et al.* outbreaker2: a modular platform for outbreak reconstruction. *BMC Bioinformatics* 2018;**19**:363.

13 Campbell F, Cori A, Ferguson N, *et al.* Bayesian inference of transmission chains using timing of symptoms, pathogen genomes and contact data. *PLoS Comput Biol* 2019;**15**:e1006930.

14 Coronavirus (COVID-19): guidance and support. https://www.gov.uk/coronavirus (accessed 2 Jun 2021).

15 Tillett RL, Sevinsky JR, Hartley PD, *et al.* Genomic evidence for reinfection with SARS-CoV-2: a case study. *Lancet Infect Dis* 2021;**21**:52–8.

16 Colton H, Ankcorn M, Yavuz M, *et al.* Improved sensitivity using a dual target, E and RdRp assay for the diagnosis of SARS-CoV-2 infection: Experience at a large NHS Foundation Trust in the UK. J. Infect. 2021;**82**:159–98.

17 Artic Network. https://artic.network/ncov-2019 (accessed 2 Jun 2021).

18 Simpson J. *nanopolish*. Github https://github.com/jts/nanopolish (accessed 2 Jun 2021).

19 Rambaut A, Holmes EC, Hill V, *et al.* A dynamic nomenclature proposal for SARS-CoV-2 to assist genomic epidemiology. doi:10.1101/2020.04.17.046086

20 Katoh K, Misawa K, Kuma K-I, *et al.* MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 2002;**30**:3059–66.

21 NicolaDeMaio, n_j_loman, goldman. Issues with SARS-CoV-2 sequencing data. 2020.https://virological.org/t/issues-with-sars-cov-2-sequencing-data/473 (accessed 27 May 2021).

22 Challen R, Brooks-Pollock E, Tsaneva-Atanasova K, *et al.* Meta-analysis of the SARS-CoV-2 serial interval and the impact of parameter uncertainty on the COVID-19 reproduction number. doi:10.1101/2020.11.17.20231548

23 Lauer SA, Grantz KH, Bi Q, *et al.* The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Ann Intern Med* 2020;**172**:577–82.

24 Keeley AJ, Evans C, Colton H, *et al.* Roll-out of SARS-CoV-2 testing for healthcare workers at a large NHS Foundation Trust in the United Kingdom, March 2020. *Euro*

medRxiv preprint doi: https://doi.org/10.1101/2021.07.15.21260537; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Surveill 2020;**25**. doi:10.2807/1560-7917.ES.2020.25.14.2000433

25 S1056 Contribution of nosocomial transmission to the first wave. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/961210/S1056_Contribution_of_nosocomial_infections_to_the_first_wave.pdf (accessed 30 Jun 2021).

26 Shields A, Faustini SE, Perez-Toledo M, *et al.* SARS-CoV-2 seroprevalence and asymptomatic viral carriage in healthcare workers: a cross-sectional study. *Thorax* 2020;**75**:1089–94.

27 Hodgson D, Colton H, Hornsby H, *et al.* Risk factors for SARS-CoV-2 seroprevalence following the first pandemic wave in UK healthcare workers in a large NHS Foundation Trust. medRxiv. 2021. doi:10.1101/2021.07.07.21260151

28 Partridge DG, Evans CM, Raza M, *et al.* Lessons from a large norovirus outbreak: impact of viral load, patient age and ward design on duration of symptoms and shedding and likelihood of transmission. *J Hosp Infect* 2012;**81**:25–30.

29 van Kampen JJA, van de Vijver DAMC, Fraaij PLA, *et al.* Duration and key determinants of infectious virus shedding in hospitalized patients with coronavirus disease-2019 (COVID-19). *Nat Commun* 2021;**12**:267.

30 Jones TC, Biele G, Mühlemann B, *et al.* Estimating infectiousness throughout SARS-CoV-2 infection course. *Science* 2021;**373**. doi:10.1126/science.abi5273

31 Campbell D, Bawden A. Up to 8,700 patients died after catching Covid in English hospitals. The Guardian. 2021.http://www.theguardian.com/world/2021/may/24/up-to-8700-patients-died-after-catching-covid-in-english-hospitals (accessed 3 Jun 2021).

32 England: busiest hostpitals. https://www.statista.com/statistics/504252/leading-busy-hospitals-ranked-by-number-of-admissions-england-uk/ (accessed 14 Jul 2021).

33 Department of Health and Social Care. NHS property and estates: Naylor review. 2017.https://www.gov.uk/government/publications/nhs-property-and-estates-naylor-review (accessed 12 Jul 2021).