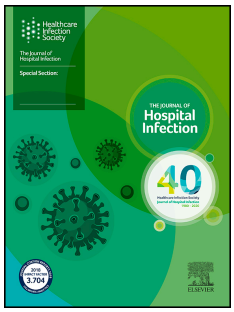
Northumbria University
NEWCASTLE

University**Library**

Journal Pre-proof



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

Hayley Colton, Matthew D. Parker, Oliver Stirrup, James Blackstone, Matthew Loose, C. Patrick McClure, Sunando Roy, Charlotte Williams, Julie McLeod, Darren Smith, Yusri Taha, Peijun Zhang, Sharon Nienyun Hsu, Beatrix Kele, Kathryn Harris, Fiona Mapp, Rachel Williams, The COG-UK HOCI Investigators, The COVID-19 Genomics UK (COG-UK) Consortium, Paul Flowers, Judith Breuer, David G. Partridge, Thushan I. de Silva

PII: S0195-6701(22)00318-8

DOI: https://doi.org/10.1016/j.jhin.2022.09.022

Reference: YJHIN 6767

To appear in: *Journal of Hospital Infection*

Received Date: 21 June 2022

Revised Date: 13 September 2022

Accepted Date: 22 September 2022

Please cite this article as: Colton H, Parker MD, Stirrup O, Blackstone J, Loose M, McClure CP, Roy S, Williams C, McLeod J, Smith D, Taha Y, Zhang P, Hsu SN, Kele B, Harris K, Mapp F, Williams R, The COG-UK HOCI Investigators, The COVID-19 Genomics UK (COG-UK) Consortium, Flowers P, Breuer J, Partridge DG, de Silva TI, 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, *Journal of Hospital Infection*, https://doi.org/10.1016/j.jhin.2022.09.022.

Title: 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.

Authors:

Hayley Colton[1,2], Matthew D. Parker[3,4], Oliver Stirrup[5], James Blackstone[6], Matthew Loose[7], C. Patrick McClure[7], Sunando Roy[8], Charlotte Williams[8], Julie McLeod[9], Darren Smith[10], Yusri Taha[11], Peijun Zhang[1], Sharon Nienyun Hsu[1,4], Beatrix Kele[12], Kathryn Harris[12], Fiona Mapp[5], Rachel Williams[8], The COG-UK HOCI Investigators, The COVID-19 Genomics UK (COG-UK) Consortium*, Paul Flowers[9], Judith Breuer[8], David G. Partridge[2], Thushan I. de Silva[1]

1. Department of Infection, Immunity and Cardiovascular Disease, University of Sheffield, Sheffield, UK
2. Directorate of Laboratory Medicine, Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, UK
3. Sheffield Biomedical Research Centre, The University of Sheffield, Sheffield, UK
4. Sheffield Bioinformatics Core, The University of Sheffield, Sheffield, UK
5. Institute for Global Health, University College London, London, UK
6. The Comprehensive Clinical Trials Unit, University College London, London, UK
7. School of Life Sciences, University of Nottingham, Nottingham, UK
8. Department of Infection, Immunity and Inflammation, Institute of Child Health, University College London, London, UK
9. School of Psychological Sciences & Health, University of Strathclyde, Glasgow, UK
10. Department of Applied Biology, Cellular and Molecular Sciences/Microbiology Group, Northumbria University, Newcastle, UK
11. Department of Infection and Tropical Medicine, Newcastle upon Tyne Hospitals NHS Foundation Trust, Royal Victoria Infirmary, Newcastle upon Tyne, UK.
12. Virology Department, East and South East London Pathology Partnership, Royal London Hospital, Barts Health NHS Trust, London, UK

* Full list of consortium names and affiliations are in the appendix

Corresponding author: Hayley Colton (h.colton@sheffield.ac.uk)

Alternative corresponding author: Thushan de Silva (t.desilva@sheffield.ac.uk)

40-word summary: SARS-CoV-2 sequencing for infection control use relies on fast turnaround time. Time between the diagnostic PCR result and arrival at the sequencing laboratory was the most prominent delay seen. Integration of pathogen sequencing into diagnostic laboratories may optimise turnaround time.

Keywords: Infection control; sequencing; SARS-CoV-2; turnaround time

Funding: COG-UK is supported by funding from the Medical Research Council (MRC) part of UK Research & Innovation (UKRI), the National Institute of Health Research (NIHR) [grant MC_PC_19027], and Genome Research Limited, operating as the Wellcome Sanger Institute. The authors acknowledge use of data generated through the COVID-19 Genomics Programme funded by the Department of Health and Social Care. The views expressed are those of the author and not necessarily those of the Department of Health and Social Care or the UK Health Security Agency.

*Conflicts of Interest***:** Since the completion of this work, MP is now an employee of Oxford Nanopore Technologies plc.

Summary:

Background: Barriers to rapid return of sequencing results can affect the utility of sequence data for infection prevention and control decisions.

Aim: To undertake a mixed-methods analysis to identify challenges sites faced in achieving a rapid turnaround time (TAT) in the COG-UK Hospital-Onset COVID-19 Infection (COG-UK HOCI) study.

Methods: For the quantitative analysis, timepoints relating to different stages of the sequencing process were extracted from both the COG-UK HOCI dataset and surveys of study sites. Qualitative data relating to the barriers and facilitators to achieving rapid TAT were included from thematic analysis.

Findings: The overall TAT, from sample collection to receipt of sequence report by infection control teams, varied between sites (median 5.1 days, range 3.0 – 29.0 days). Most variation was seen between reporting of a positive COVID-19 PCR result to sequence report generation (median 4.0 days, range 2.3 – 27.0 days). On deeper analysis, most of this variability was accounted for by differences in the delay between the COVID-19 PCR result and arrival of the sample at the sequencing laboratory (median 20.8 hours, 16.0 – 88.7 hours). Qualitative analyses suggest closer proximity of sequencing labs to diagnostic labs, increased staff flexibility and regular transport times facilitated a shorter TAT.

Conclusion: Integration of pathogen sequencing into diagnostic laboratories may help improve sequencing TAT to allow sequence data to be of tangible value to infection control practice. Adding a quality control step upstream to increase capacity further down the workflow may also optimise TAT if lower quality samples are removed earlier on.

Introduction

The SARS-CoV-2 pandemic has highlighted the utility of large-scale genomic sequencing to influence infection prevention and control (IPC) decisions[1,2]. While the technology to rapidly sequence pathogens using next-generation sequencing has been available for some time, its use has primarily been limited to genomic surveillance and retrospective transmission studies, often done in large, centralised reference laboratories.

During the pandemic, the COVID-19 Genomics UK (COG-UK) consortium established a network of sequencing hubs, pioneering a de-centralised and distributed model of SARS-CoV-2 sequencing from National Health Service (NHS) hospitals[3]. The COG-UK Hospital-Onset COVID-19 Infection (COG-UK HOCI) study was nested within the COG-UK network, with an aim to assess the impact of sequencing and its turnaround time (TAT) on several IPC outcomes[4]. We recently reported that the likelihood of SARS-CoV-2 sequencing informing the IPC response to hospital-onset COVID-19 infections was dependent on the return of results within five days[5].

The time taken to return a potentially actionable sequencing report to the IPC team is dependent on a variety of factors. For this paper, we further interrogate data from the COG-UK HOCI study alongside additional datapoints in a post-hoc mixed-methods analysis, aiming to identify barriers in achieving a rapid sequencing TAT.

Methods

COG-UK HOCI Background and Design

COG-UK HOCI was a prospective non-randomised trial to evaluate the implementation and impact of SARS-CoV-2 sequencing on IPC practice. The study was approved by National Research Ethics Service Committee – Cambridge South: REC 20/EE/0118[4]. The study ran from December 2020 – April 2021 across fourteen UK acute NHS hospital groups. The recruiting sites were linked to one of eleven sequencing laboratories where the genomic sequencing took place.

COG-UK HOCI was split into baseline, rapid and longer turnaround phases to evaluate whether rapid sequencing (i.e., ≤48 hours) could improve IPC decision making, in comparison to a longer TAT of 5-10 days, akin to using a centralised sequencing laboratory. Possible HOCIs were identified, and the respective samples were sent to the designated sequencing laboratory. A bespoke sequencing report tool (SRT) was used to communicate the result to the IPC team for prospective action[6]. The SRT integrates genomic and epidemiological data from HOCIs to provide a one-page report identifying closely matched sequences within the hospital and at ward level, and assigns a probability estimate for nosocomial infection. Samples with genomic coverage of less than 90% could not be used to generate an SRT[5,6].

We performed parallel independent quantitative and qualitative data collection and analysis with the subsequent integration of findings.

Quantitative Data Extraction and Analysis

For each sample, the dates ± times were extracted from the COG-UK HOCI dataset for the following timepoints: 1) "*COVID-19 sample [taken] to confirm diagnosis*", 2) "*COVID-19 result reported*", 3) "*Sequence report generation*", and 4) "*Receipt of sequence report by IPC team*" (Figure 1a), in addition to patient study identifier, COG-UK ID, study site, and reason sequence was not returned within expected timeframes. Genomic coverage was extracted from the Cloud Infrastructure for Microbial Bioinformatics (CLIMB) and matched to each sample by COG-UK ID, where available.

Of the 2170 samples in the extract, we only evaluated samples from the COG-UK HOCI rapid phase, during which, sites attempted to return an SRT within 48 hours of sample collection (n=947, Figure 1b). Just under half of these samples had an SRT returned to the IPC team during the study (n=429/947, 45.3%). Reasons for failure to return an SRT included issues with sample quality/sequence failures (n=194/518, 37.5%), sequence report generation (n=130/518, 25.1%), delays of processing (n=43/518, 8.3%), insufficient sample available (n=35/518, 6.8%), or unknown (n=116/518, 22.4%). For some timepoints, several samples had only the date available with the time missing: "*COVID-19 sample [taken] to confirm diagnosis"* (n=64/947, 6.8%), *"COVID-19 result reported*" (n=18/947, 1.9%), "*Sequence report generation*" (n=429/429, 100.0%) and *"Receipt of sequence report by IPC team*" (n=146/429, 34.0%). Missing times for "*Sequence report generation*" were estimated based on the

"*Receipt of sequence report by IPC team*" timepoint. If "*Receipt of sequence report by IPC team*" was on the same date as "*Sequence report generation*", corresponding missing "*Sequence report generation*" times were replaced with either 00:00 if "*Receipt of sequence report by IPC team*" was before 02:00 (n=4/429, 0.9%), or 07:00 if "*Receipt of sequence report by IPC team*" was after 02:00 (n=288/429, 67.1%). All of the other missing times, including for the other timepoints, were replaced with 12:00 providing the dates were available, with sensitivity analyses undertaken to assess the impact of missing data imputation (Supplementary Figure 1). Samples where duration timepoints were unfeasible were either excluded from the analysis of their respective phase (n=67/947, 7.1%), or corrected where additional data were available from the site survey (n=7/947, 0.7%).

Additional data were requested from COG-UK HOCI sites by email invitation to complete a survey. Information requested included the type of sample received from diagnostic laboratory (i.e., fresh unextracted or residual nucleic acid), frequency and method of transport between diagnostic and sequencing laboratories, the number of sequencing runs per day, the sequencing platform used, and availability of additional timepoints during the sequencing process (dates ± times for when each sample arrived at the sequencing laboratory, when the sample was put on the sequencer, and when analysis of raw sequence data was commenced to generate a consensus SARS-CoV-2 sequence from each sample) (Figure 1a).

Seven of the eleven sequencing laboratories responded, and six agreed to provide additional timepoints where available. The sites who responded, and their median overall TATs, were sites E (5.7 days), L (3.0 days), J (5.4 days), M (5.0 days), K (6.0 days), I (4.1 days) and A (11.9 days) (Supplementary Table I). These sequencing laboratories processed 444 of the COG-UK HOCI rapid phase samples (n=444/947, 46.9%), and just over half had an SRT returned (n=240/444, 54.1%) (Figure 1b). Reliable timepoints were available for *"Arrival at Sequencing Lab"* and *"Time started on sequencer"* for all six laboratories, which allowed deeper analysis of the 'Sequencing' phase from the COG-UK HOCI dataset (Figure 1A). The sequencing laboratory for sites E, J and K was only able to provide the dates for both of these timepoints, so times were estimated based on their standard practice. Although site H responded to the survey, they were excluded from the analysis as they didn't successfully sequence a sample within the COG-UK HOCI rapid phase. If the duration between timepoints was illogical i.e., <0 hours, the sample was excluded from analysis of that specific segment: PCR result to Arrival at Sequencing Lab (n=3/444, 0.7%) and Analysis (n=40/240, 16.7%).

ANOVA was used to calculate statistical significance between sites for each of the durations in Figure 1a. Analysis was performed in R Studio (2021.09.1+372 "Ghost Orchid" Release).

Qualitative Design and analyses

Using a purposive sub-sample of five heterogenous study sites, 39 diverse professional-participants, all directly involved in implementing the COG-UK HOCI study, took part in semi-structured interviews between December 23rd 2020 and June 2nd 2021. Data collection focussed on their HOCI experiences. A balance of deductive and inductive thematic analysis was then conducted by a team of trained qualitative analysts. Here, we report only the findings that relate to the barriers and facilitators to achieving rapid TAT. Our main results are presented by integrating both the quantitative data on turnaround times, with qualitative findings integrated where appropriate.

Results

Overall TAT

The overall TAT was calculated from "*COVID-19 sample [taken] to confirm diagnosis*" to "*Receipt of sequence report by IPC team*". The TAT varied considerably between sites for the 429 samples which had an SRT returned to the IPC team (median 5.1 days, range 3.0 – 29.0 days, p <0.0001) (Figure 2, Supplementary Table I). Sites self-reported if samples met their 'expected' TAT (n=209/429, 48.7%). However, the COG-UK HOCI data revealed only a fraction of SRTs were returned within 48 hours of sample collection (n=9/429, 2.1%), with slightly more within 72 hours (n=83/429, 19.3%) or 120 hours of sample collection (n=204/429, 47.6%) (Figure 1b). The greatest inter-site variation was seen within the 'Sequencing' phase (median 4.0 days, range 2.3 – 27.0 days, p <0.0001), with less variation in the diagnostic (median 0.7 days, range 0.5 – 1.3 days, p <0.0001) and reporting phases (median 0.4 days, 0.2 – 1.0 days, p <0.0001).

Qualitative analyses illuminated potential reasons for the low rates of meeting expected TAT, highlighting the fragility of the whole rapid SRT pipeline, where problems in any single step had consequences for others "*you only need one thing to go wrong and it sort of snowballs really"*. In this way, TAT was vulnerable to the effects of COVID, although automated processes and effective communication could help. Table I details qualitative barriers and facilitators to meeting rapid TAT at each phase. Many rate-limiting factors in the diagnostic phase related to the impacts of COVID at the time of data collection and reported teething troubles with diagnostic processes. Facilitators to the diagnostic phase focussed on the efficient transporting of swabs to the diagnostic lab and automated systems to pick up HOCIs. For the reporting phase, barriers and facilitators related to both the generation and then dissemination of the report. Peer learning across sites generating reports and automated report generation were also notable facilitators.

Detailed breakdown of 'Sequencing' phase

We broke down the 'Sequencing' phase using the additional timepoints from the sites who responded to the survey, in order to further analyse the sequencing laboratory activity (Figure 3).

COVID-19 result reported to arrival of sample at sequencing lab: Large variability was seen between sites for the median duration between *"COVID-19 result reported*" and *"Arrival at Sequencing Lab"* (median 20.8 hours, 16.0 – 88.7 hours, p <0.0001, Supplementary Table II). We explored whether the relative location of diagnostic and sequencing laboratories could explain this variability. Site K was the furthest distance away from its sequencing lab (~205km), and had the longest median time of 88.7 hours. Sites I and L had the shortest median times of 16.0 hours and 19.0 hours respectively, and their diagnostic and sequencing laboratories were much closer, plus site L had increased transport frequency (Supplementary Table III). Qualitative analyses suggest the proximity of laboratories reduced delays, as did regular scheduled transport and dedicated pick-up times (Table I). Sequence-lab staff's ability to work flexibly also enabled more rapid TAT.

Pre-sequencing: Pre-sequencing activity was calculated from "*Arrival at Sequencing Lab*" through to "*Started on sequencer*" and would include extraction (if required), PCR, library preparation, and time between each process. All sequencing laboratories surveyed, except site M, received fresh unextracted sample rather than residual nucleic acid. DNA quantification and normalisation was performed as a quality control (QC) step prior to library

preparation at sites L, M, I and A. Sites L, I and A performed DNA quantification on all samples, whilst site M selected representative samples and their controls for testing. Sites I and A performed additional QC steps. Site I screened cycle threshold (CT) values from the diagnostic PCR, and based on their experience, only proceeded if CT <35. Site A checked PCR amplicons on gel prior to DNA quantification and only processed if a visible band was present. The median pre-sequencing activity was relatively consistent between sites (22.3 hours, range 21.2 – 32.0 hours, p <0.0001) (Supplementary Table II). Accurate start times for extraction, PCR and library preparation were unavailable, however sites L, M and A self-reported times based on typical practice (Supplementary Table IV). Site L was the only site who reported performing two library preparations per day. Only minimal differences were seen between the estimated library preparation times by sequencing platform (Supplementary Table II).

On sequencer: The time spent on sequencer was calculated from the *"Started on sequencer"* timepoint through to "*Primary Analysis*" timepoint, as time the sequencing run was stopped was not readily available. The "*Primary Analysis*" timepoint was not available for sites E, J and K, so the "*Sequence Report Generation*" timepoint was used in lieu. The median time spent on sequencer for samples where the SRT was returned was 17.1 hours (range 5.7 – 62.6 hours, p <0.0001) (Supplementary Table II). Site I had the shortest median time on sequencer, and reported starting sequencing their samples around 12:00-16:00, to start primary analysis the same evening. Site A had the longest median time spent on sequencer, and reported having too many other samples to process, or having issues with reporting tools CLIMB and GLUE (n=9/11, 81.8%)[7].

Analysis: The analysis duration was calculated from "*Primary Analysis*" (or "*Sequence Report Generation*" for sites E, J and K as previously mentioned) through to the "*Receipt of sequence report by IPC team*". The median analysis duration across all sites was 4.6 hours, with a range of 2.9 – 135.8 hours (p <0.0001, Supplementary Table II). The sites who reported using DNA quantification as a QC step prior to library preparation had higher percentages of SRT return within five days than sites who didn't have a QC step (Figure 3E).

Discussion

The COG-UK HOCI trial found returning an SRT within five days changed the actions of the IPC teams in around 20% of HOCIs[5]. Through our mixed-methods analysis, we found many sites did not manage to return any of their SRTs within five days, and identified some of the challenges sites faced.

As the greatest inter-site variability was seen in the time between the diagnostic PCR result and the arrival of the sample at the sequencing laboratory, an obvious factor to optimise TAT would be to reduce the distance and/or increase the transport frequency between diagnostic and sequencing laboratories as described in the qualitative data. Integrating sequencing into diagnostic laboratories could be an ideal solution, and would also facilitate the transfer of patient level data including current location and prior ward movements from patient administration systems and provide the geotemporal data required for easy and rapid interpretation of sequence reports. Integrated laboratories have also been reported to increase regional and national processing power for the surveillance of antimicrobial resistance[8]. Where integration is not possible, reducing the distance between laboratories and regular dedicated transport times would allow laboratories to plan their workflows in order to optimise TAT and reduce the likelihood of missing samples.

The second greatest variability was seen in the median durations between the start of primary analysis and receipt of the SRT by the IPC team. As sites reported they were overwhelmed with processing other samples during the COVID-19 pandemic, adding a QC step may increase capacity further down the workflow, however caution should be applied if CT values are used, given significant variability between laboratories has been reported[9]. Additionally, if sites were able to run samples on the sequencer earlier in the day, the sequencing process could be stopped and its output analysed within the same day, allowing a shorter sequencing and analysis time.

Outside of the COVID-19 response, there have been successful reports of using rapid sequencing to influence the IPC response for other pathogens of interest, such as methicillin-resistant *Staphylococcus aureus* (MRSA), *Clostridium difficile* and vancomycin resistant Enterococcus faecium (VRE)[10,11]. For each pathogen, laboratories would have to consider pathogen-specific factors which could affect TAT, such as the time

required for culturing bacteria or the frequency of performing other tests such as immunoassays. The pressure on the microbiology laboratory workload would also vary at different points of the year, and outside of a pandemic, would likely also have an impact on TAT.

We were limited by missing and erroneous times within the datasets, however we were able to partially mitigate this through either correcting, estimating or excluding timepoints. In addition, we recognise the potential of volunteer bias within our survey data, as sites who had shorter TATs were more likely to respond and provide additional data. Although we were unable conclude whether genomic coverage was affected by either the sample type received by the sequencing laboratory, or time between sample collection and extraction, it is well reported that RNA is at risk of degradation if samples are not processed promptly[12].

Conclusions

IPC interventions in response to presumed nosocomial transmission events are often resource intensive in terms of human, financial and operational impact, and thus practice developments which confirm or refute case linkage within a clinically meaningful time-scale have the potential to be of great benefit to healthcare services.

Our results present evidence supporting the integration of pathogen sequencing into diagnostic laboratories, in order for sequence data to be of tangible value to IPC practice.

Laboratories using rapid sequencing for IPC purposes may be able to utilise our findings to streamline and optimise their own workflows for SARS-CoV-2 and other pathogens. The challenges and optimal TAT of integrated sequencing for IPC use on a larger scale needs further analysis for other pathogens, including whether challenges sites face would be similar outside of a pandemic if a short TAT is desired.

Acknowledgements

Our thanks to the COG-UK HOCI Principal Investigators: Dr Anu Chawla, Dr Maria-Teresa Cutino-Moguel, Prof Alison Holmes, Dr Nick Machin, Dr Nikunj Mahida, Dr Taibtha Mahungu, Dr Gaia Nebbia, Dr Cassie Pope, Dr Kordo Saeed, Dr Tranprit Saluja, Dr Gee Yen Shin, Dr Thushan de Silva, Dr Yusri Taha, Prof Emma Thomson and their teams for their time and support. We also acknowledge the support of the independent members of the Joint Trial Steering Committee and Data Monitoring Committee (Prof Marion Koopmans, Prof Walter Zingg, Prof Colm Bergin, Prof Karla Hemming, Prof Katherine Fielding), as well as their non-independent members (Prof Nick Lemoine, Prof Sharon Peacock). We would also thank members of COG-UK who have directly supported the COG-UK HOCI study: Dr Ewan Harrison, Dr Katerina Galai, Dr Francesc Coll, Dr Michael Chapman, Prof Thomas Connor, Prof Nick Loman and their teams. We also thank the COG-UK Consortium, and the NIHR Clinical Research Network.

References

[1] Walker A, Houwaart T, Finzer P, Ehlkes L, Tyshaieva A, Damagnez M, et al. Characterization of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Infection Clusters Based on Integrated Genomic Surveillance, Outbreak Analysis and Contact Tracing in an Urban Setting. Clin Infect Dis 2022;74:1039–46. https://doi.org/10.1093/cid/ciab588.

[2] Francis R V., Billam H, Clarke M, Yates C, Tsoleridis T, Berry L, et al. The Impact of Real-Time Whole-Genome Sequencing in Controlling Healthcare-Associated SARS-CoV-2 Outbreaks. J Infect Dis 2022;225:10–8. https://doi.org/10.1093/infdis/jiab483.

[3] The COVID-19 Genomics UK (COG-UK) consortium. An integrated national scale SARS-CoV-2 genomic surveillance network. Lancet Microbe 2020;1(3):e99–e100.

[4] Blackstone J, Stirrup O, Mapp F, Panca M, Copas A, Flowers P, et al. Protocol for the COG-UK hospital-onset COVID-19 infection (HOCI) multicentre interventional clinical study: evaluating the efficacy of rapid genome sequencing of SARS-CoV-2 in limiting the spread of COVID-19 in UK NHS hospitals. BMJ Open 2022;12:e052514. https://doi.org/10.1136/bmjopen-2021-052514.

[5] Stirrup O, Blackstone J, Mapp F, MacNeil A, Panca M, Holmes A, et al. Evaluating the effectiveness of rapid SARS-CoV-2 genome sequencing in supporting infection control teams: the COG-UK hospital-onset COVID-19 infection study. MedRxiv 2022:2022.02.10.22270799. https://doi.org/10.1101/2022.02.10.22270799.

[6] Stirrup OT, Hughes J, Parker M, Partridge DG, Shepherd JG, Blackstone J, et al. Rapid feedback on hospital onset sars-cov-2 infections combining epidemiological and sequencing data. Elife 2021;10:1–30. https://doi.org/10.7554/eLife.65828.

[7] Singer JB, Thomson EC, McLauchlan J, Hughes J, Gifford RJ. GLUE: A flexible software system for virus sequence data. BMC Bioinformatics 2018;19:1–18. https://doi.org/10.1186/s12859-018-2459-9.

[8] Kekre M, Arevalo SA, Valencia MF, Lagrada ML, Macaranas PK V., Nagaraj G, et al. Integrating Scalable Genome Sequencing into Microbiology Laboratories for Routine Antimicrobial Resistance Surveillance. Clin Infect Dis 2021;73:S258–66. https://doi.org/10.1093/cid/ciab796.

[9] Evans D, Cowen S, Kammel M, O'Sullivan DM, Stewart G, Grunert HP, et al. The Dangers of Using Cq to Quantify Nucleic Acid in Biological Samples: A Lesson From COVID-19. Clin Chem 2021;68:153–62. https://doi.org/10.1093/clinchem/hvab219.

[10] Eyre DW, Golubchik T, Gordon NC, Bowden R, Piazza P, Batty EM, et al. A pilot study of rapid benchtop sequencing of Staphylococcus aureus and Clostridium difficile for outbreak detection and surveillance. BMJ Open 2012;2:1–9. https://doi.org/10.1136/bmjopen-2012-001124.

[11] McGann P, Bunin JL, Snesrud E, Singh S, Maybank R, Ong AC, et al. Real time application of whole genome sequencing for outbreak investigation – What is an achievable turnaround time? Diagn Microbiol Infect Dis 2016;85:277–82. https://doi.org/10.1016/j.diagmicrobio.2016.04.020.

[12] Lewandowski K, Bell A, Miles R, Carne S, Wooldridge D, Manso C, et al. The Effect of Nucleic Acid Extraction Platforms and Sample Storage on the Integrity of Viral RNA for Use in Whole Genome Sequencing. J Mol Diagnostics 2017;19:303–12. https://doi.org/10.1016/j.jmoldx.2016.10.005.

[13] Xu Shuangbin, Chen Meijun, Feng Tingze, Zhan Li, Zhou Lang YG. Use ggbreak to Effectively Utilize Plotting Space to Deal With Large Datasets and Outliers. Front Genet 2021;12. https://doi.org/10.3389/fgene.2021.774846.

Barriers	Facilitators	Indicative extract
Overall turnaround time		
The volume and impact of COVID infections stalled all steps and processes within TAT The number of steps within TAT meant if only one element was problematic – it had knock on effects throughout	Flexibility of staff resource to be responsive to particular situational dynamics (ebbs and flows) reduced bottle necks in the multi-staged TAT process Close diagnostic and sequencing labs Robust workflow and automatization Skilled human resource Effective communication across the SRT pipeline	"when you have 150 new cases a day, it just, it makes everything grind to a halt, you know, the patient flow in the hospital as well as specifically for the HOCI study, like even the diagnostic lab can't cope, it has an impact on the research lab and the flow of samples. And then we have so many samples, the sequencing isn't as good, means more samples failing and you know, handling the data's a lot harder"
Diagnostic phase		
COVID demands made it difficult to collect samples (e.g., volume of patients) Volume of samples meant delays to testing (competing testing streams) in diagnostic labs Staff meeting called for each potential HOCI (at start of trial) delayed classification Untrained staff appraising clinical notes without expert knowledge made classifications difficult (at start of trial) Reliance on a clinician for classification slowed down process (at start of trial)	Having porters available to transport swabs to diagnostic lab Regular transport system between sites and labs (e.g., taxis) Automated systems to pick up HOCIs were developed once screening established Established IPC team at hospital made the classifications	"It felt was a bit like a brick wall a lot of the time unfortunately. Just the systems and the way it works and the fact that we take swabs and they have to be couriered over to [hospital name B]. It takes a bit longer over here; we don't quite get the turnround"
Sequencing phase		
Staff absences made it difficult to get samples to sequencing lab Volume of HOCI samples to sequence and the prioritisation of health care worker (HCW) samples make it difficult to generate sequence data file Samples going missing because of test validation studies made it difficult to generate the sequence data file Samples with high CT values wasted sequence capacity Volume of sequencing occurring limited sequencing	Dedicated pick-up time from diagnostic lab meant sufficient time at sequence-lab Having sequence lab close-by reduced delays High quality, committed sequence-lab teams were able to offer flexible working patterns Dedicated resource assisted with sequencing phase.	"they worked really late that night to make sure that things were still kept on track. So I think that's, I think that's an amazing effort on their part"
Reporting phase		
IT problems with memory and grid lines Problems with updating tools for lineage The tool was not designed for the amount of data Access to GLUE delayed report generation There were problems sending the PDF of the report to some email accounts (NHS particularly)	Input from bioinformatician and IT Communication and peer-learning across sites running reports Automated report generation and dissemination – to password protected HOCI slack channel/Teams [workplace messaging systems] and necessary NHS emails	"I think it's a GLUE server that's been down and over the weekend apparently the sys[tem] admin don't kind of, the CLIMB admin don't work, so we were kind of stuck where we had worked quite hard to get sequence data out and we can actually get the reporting tool out…. And that's a bit frustrating I guess when something out of your hands goes wrong and you've done everything possible to try and, yeah"

Table I. The barriers and facilitators to the main phases of rapid turnaround time

CLIMB: Cloud Infrastructure for Microbial Bioinformatics; CT: cycle threshold; GLUE: Genes Linked by Underlying Evolution[7]; HOCI: Hospital-onset COVID-19 Infection; IMT: Incident Management Team; IPC: Infection Prevention and Control; NHS: National Health Service; SRT: Sequence Report Tool

Figure 1

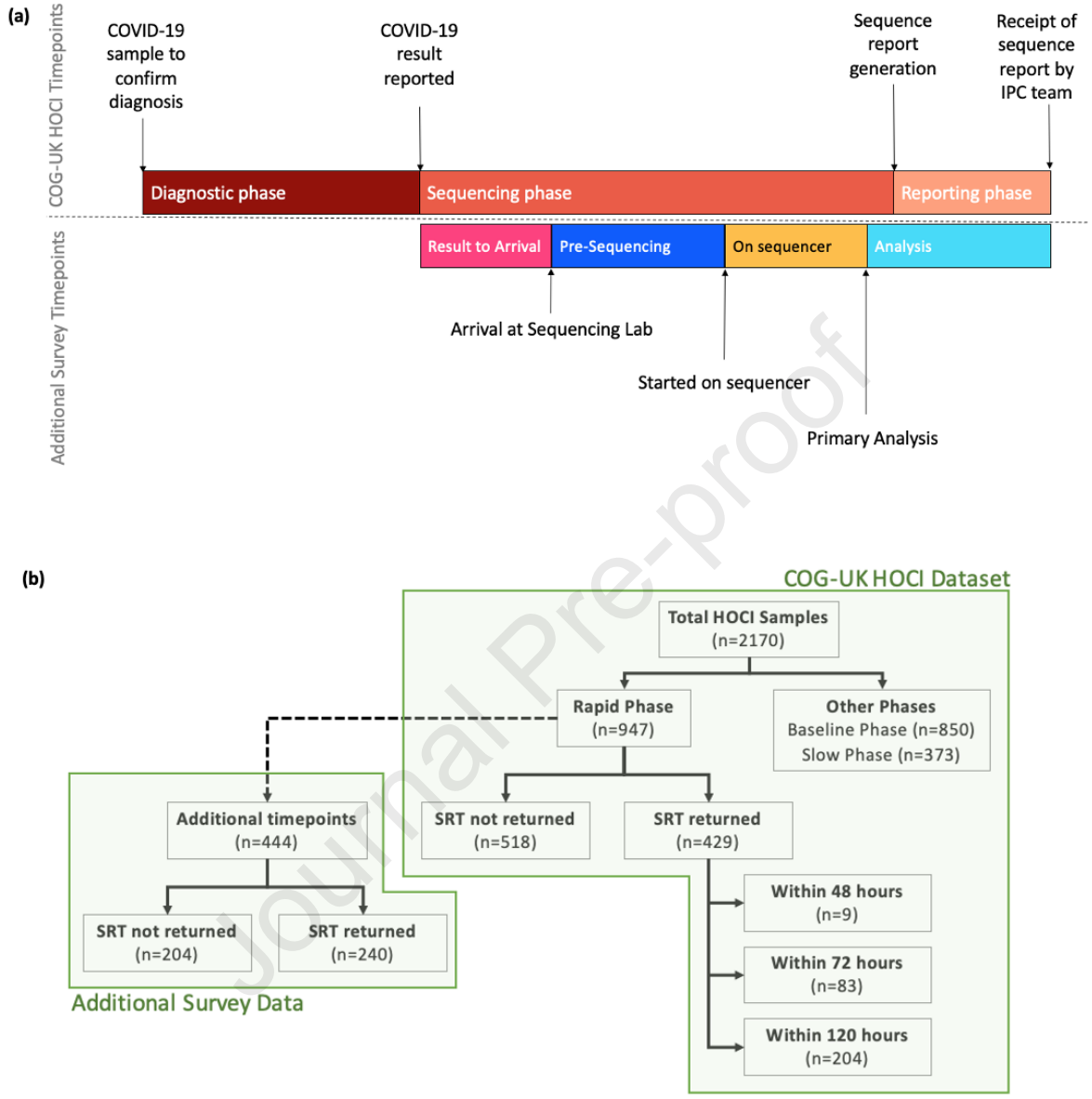
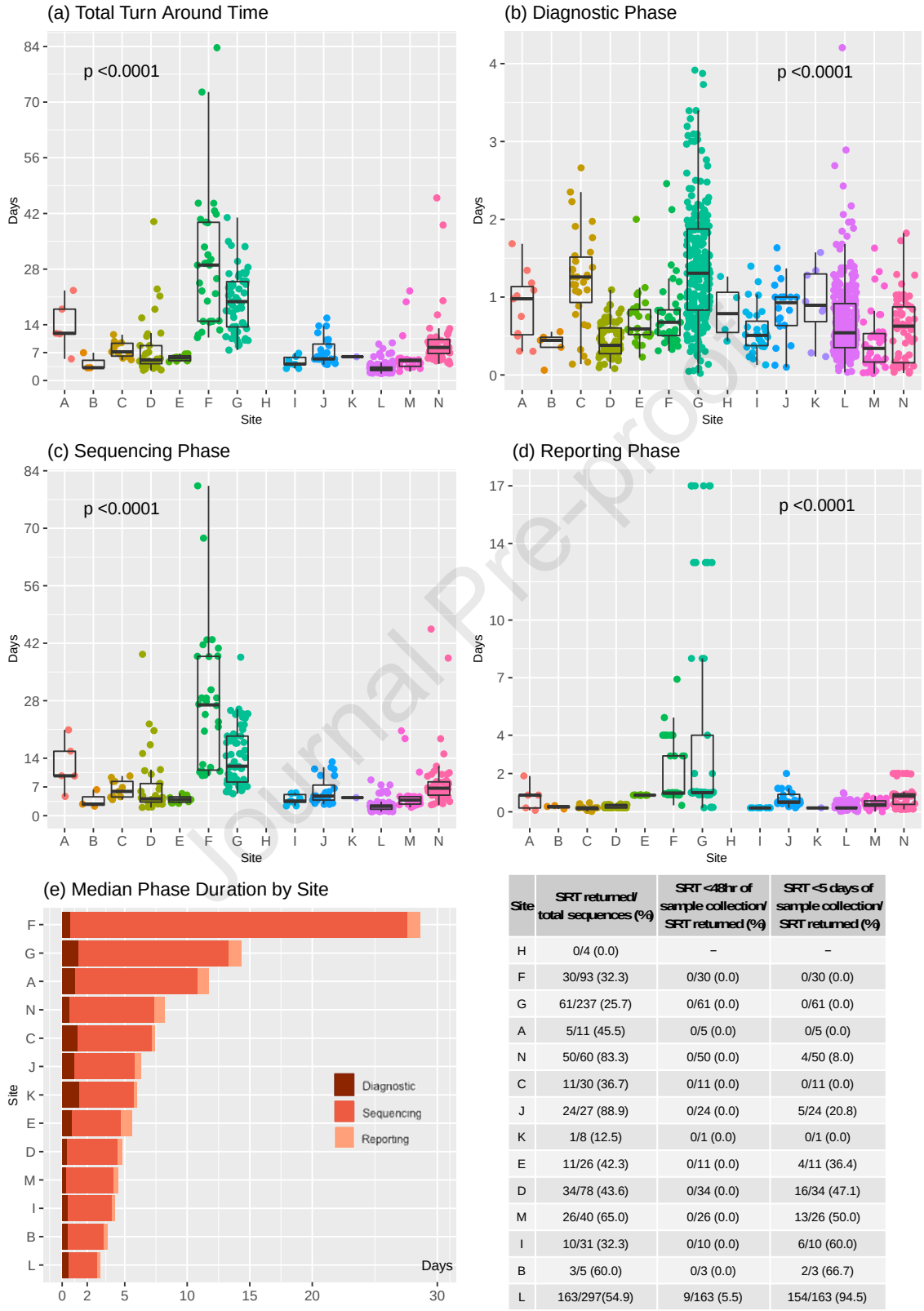
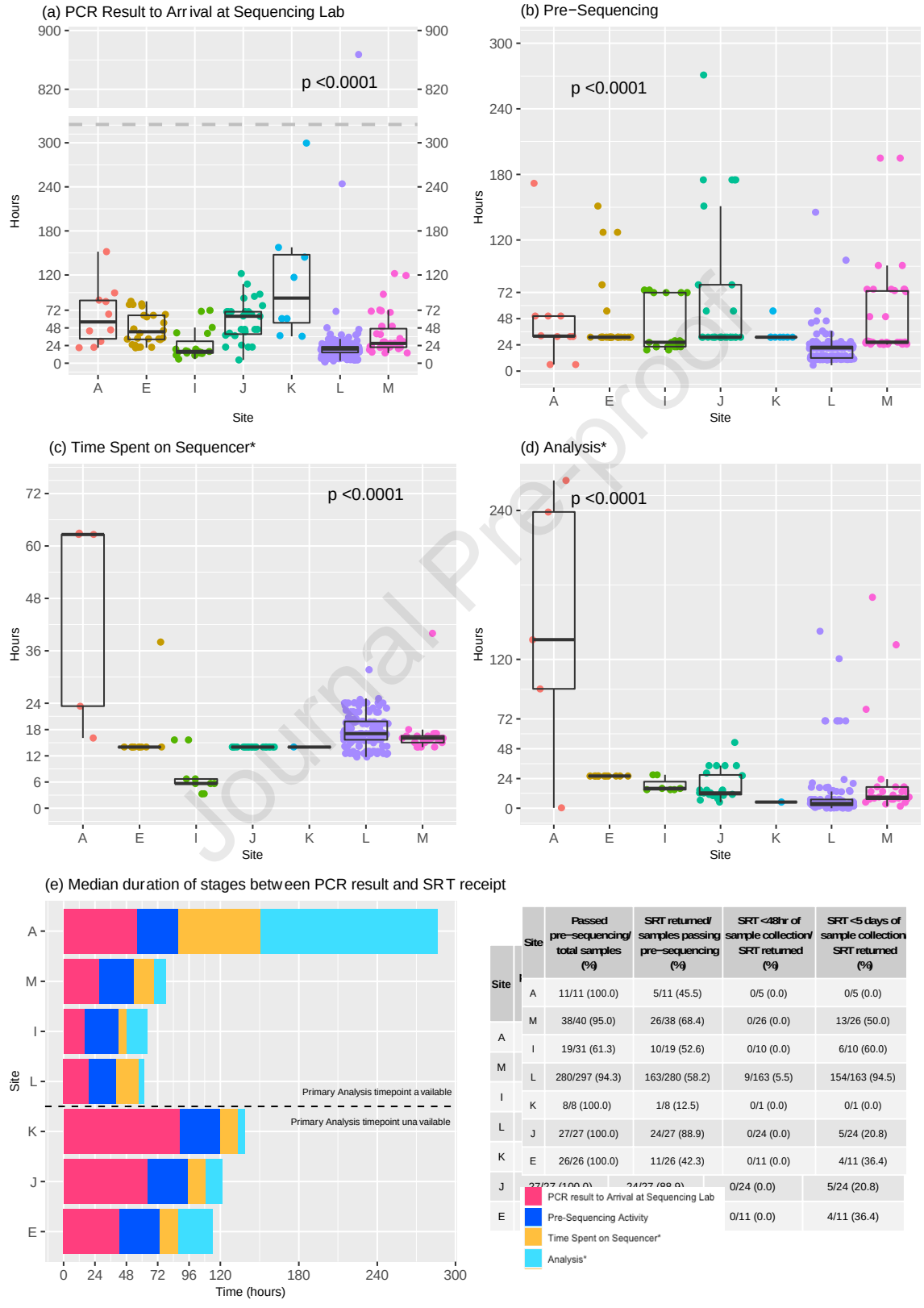


Figure 2



Site	SRT returned/ total sequences (%)	SRT <48hr of sample collection/ SRT returned (%)	SRT <5 days of sample collection/ SRT returned (%)
H	0/4 (0.0)	–	–
F	30/93 (32.3)	0/30 (0.0)	0/30 (0.0)
G	61/237 (25.7)	0/61 (0.0)	0/61 (0.0)
A	5/11 (45.5)	0/5 (0.0)	0/5 (0.0)
N	50/60 (83.3)	0/50 (0.0)	4/50 (8.0)
C	11/30 (36.7)	0/11 (0.0)	0/11 (0.0)
J	24/27 (88.9)	0/24 (0.0)	5/24 (20.8)
K	1/8 (12.5)	0/1 (0.0)	0/1 (0.0)
E	11/26 (42.3)	0/11 (0.0)	4/11 (36.4)
D	34/78 (43.6)	0/34 (0.0)	16/34 (47.1)
M	26/40 (65.0)	0/26 (0.0)	13/26 (50.0)
I	10/31 (32.3)	0/10 (0.0)	6/10 (60.0)
B	3/5 (60.0)	0/3 (0.0)	2/3 (66.7)
L	163/297(54.9)	9/163 (5.5)	154/163 (94.5)

Figure 3



Site	Passed pre−sequencing/ total samples (%)	SRT returned/ samples passing pre−sequencing (%)	SRT <48hr of sample collection/ SRT returned (%)	SRT <5 days of sample collection/ SRT returned (%)
A	11/11 (100.0)	5/11 (45.5)	0/5 (0.0)	0/5 (0.0)
M	38/40 (95.0)	26/38 (68.4)	0/26 (0.0)	13/26 (50.0)
I	19/31 (61.3)	10/19 (52.6)	0/10 (0.0)	6/10 (60.0)
L	280/297 (94.3)	163/280 (58.2)	9/163 (5.5)	154/163 (94.5)
K	8/8 (100.0)	1/8 (12.5)	0/1 (0.0)	0/1 (0.0)
J	27/27 (100.0)	24/27 (88.9)	0/24 (0.0)	5/24 (20.8)
E	26/26 (100.0)	11/26 (42.3)	0/11 (0.0)	4/11 (36.4)

Figure 1. Summary of data used from the COG-UK HOCI dataset.

(a) Phases of SARS-CoV-2 sequencing and available data from different timepoints. Pre-sequencing activity includes processes such as extraction (if required), amplicon generation for Polymerase Chain Reaction (PCR), library preparation, and the time spent waiting in between each process. Primary analysis denotes processing of raw sequence data to generate a consensus sequence.

(b) Breakdown of data acquired from main COG-UK HOCI dataset and additional data acquired through survey of sites. COG-UK HOCI: COVID-19 Genomics UK Hospital-Onset COVID-19 Infection Study; IPC: Infection Prevention and Control; SRT: Sequence Report Tool

Figure 2. Scatterplots for all samples processed during the rapid phase using timepoints available within COG-UK HOCI extract for (a) total turnaround time (n=429/429), (b) diagnostic phase (n=880/947), (c) 'Sequencing' phase (n=429/429) and (d) reporting phase (n=429/429). Boxplots represent medians plus interquartile ranges (IQR25 and IQR75). 67 samples were excluded from assessment of the diagnostic phase due to apparent errors in data entry. (e) Median durations for the diagnostic, sequencing and reporting phases by COG-UK HOCI site, using timepoints for the "rapid phase" samples in COG-UK HOCI extract which had an SRT returned (n=429/947). The associated table shows the number of samples which were processed within the rapid phase per site.

Figure 3. Durations within the 'Sequencing' phase for each sample by site for (a) PCR result to Arrival at Sequencing Lab (n=406/444), (b) Pre-sequencing (n=406/444), (c) Time Spent on Sequencer* (n=239/240), and (d) Sequence analysis* (n=200/240). Boxplots represent IQR25, median and IQR75. For 3A, the y axis was broken in order to show outliers using the R package ggbreak[13]. (e) Median durations and number of samples for each stage of process from "COVID-19 result reported" onwards for the samples processed within the rapid phase from the surveyed sites (n=444/947). As the "Primary Analysis" timepoint was not available for sites E, J and K, the "Sequence Report Generation" timepoint from the COG-UK HOCI dataset was used in lieu, which corresponds to Figures 3(c)-(e).

*For time spent on sequencer and analysis, only samples which had an SRT returned were used (n=240/444), as unsuccessful sequences may artificially shorten the durations if the run was aborted.