RESEARCH ARTICLE

REVISED Household serial interval of COVID-19 and the effect of Variant B.1.1.7: analyses from prospective community cohort study (Virus Watch) [version 2; peer review: 2 approved]

Previously titled: 'Serial interval of COVID-19 and the effect of Variant B.1.1.7: analyses from prospective community cohort study (Virus Watch)'

Cyril Geismar[1], Ellen Fragaszy[2,3], Vincent Nguyen[1,2], Wing Lam Erica Fong[2], Madhumita Shrotri[2], Sarah Beale[1,2], Alison Rodger[4], Vasileios Lampos[5], Thomas Byrne[2], Jana Kovar[1], Annalan M.D. Navaratnam[1,2], Parth Patel[2], Robert W. Aldridge[2], Andrew Hayward[1], on behalf of Virus Watch Collaborative

[1]Institute of Epidemiology and Health Care, University College London, London, UK
[2]Centre for Public Health Data Science, Institute of Health Informatics, University College London, London, UK
[3]Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK
[4]Institute for Global Health, University College London, London, UK
[5]Department of Computer Science, University College London, London, UK

v2 **First published:** 09 Sep 2021, **6**:224
https://doi.org/10.12688/wellcomeopenres.16974.1
Latest published: 21 Dec 2021, **6**:224
https://doi.org/10.12688/wellcomeopenres.16974.2

Abstract

Introduction: Increased transmissibility of B.1.1.7 variant of concern (VOC) in the UK may explain its rapid emergence and global spread. We analysed data from putative household infector - infectee pairs in the Virus Watch Community cohort study to assess the serial interval of COVID-19 and whether this was affected by emergence of the B.1.1.7 variant.

Methods: The Virus Watch study is an online, prospective, community cohort study following up entire households in England and Wales during the COVID-19 pandemic. Putative household infector-infectee pairs were identified where more than one person in the household had a positive swab matched to an illness episode. Data on whether or not individual infections were caused by the B.1.1.7 variant were not available. We therefore developed a classification system based on the percentage of cases estimated to be due to B.1.1.7 in national surveillance data for different English regions and study weeks.

Results: Out of 24,887 illnesses reported, 915 tested positive for SARS-CoV-2 and 186 likely 'infector-infectee' pairs in 186 households amongst 372 individuals were identified. The mean COVID-19 serial interval was 3.18 (95%CI: 2.55-3.81, sd=4.36) days. There was no

Open Peer Review

Approval Status ✓ ✓

	1	2
version 2 (revision) 21 Dec 2021	✓ view	✓ view
version 1 09 Sep 2021	✓ view	✓ view

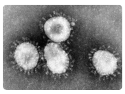
1. **Vítor H. Nascimento**, University of São Paulo, São Paulo, Brazil

2. **Benjamin J. Cowling**, The University of Hong Kong, Hong Kong, China

Any reports and responses or comments on the article can be found at the end of the article.

significant difference (p=0.267) between the mean serial interval for VOC hotspots (mean = 3.64 days, (95%CI: 2.55 – 4.73)) days and non-VOC hotspots, (mean = 2.72 days, (95%CI: 1.48 – 3.96)).
Conclusions: Our estimates of the average serial interval of COVID-19 are broadly similar to estimates from previous studies and we find no evidence that B.1.1.7 is associated with a change in serial intervals. Alternative explanations such as increased viral load, longer period of viral shedding or improved receptor binding may instead explain the increased transmissibility and rapid spread and should undergo further investigation.

Keywords
COVID-19, SARS-CoV-2, variant of concern, B.1.1.7, serial interval, epidemiology, UK



This article is included in the Coronavirus (COVID-19) collection.

Corresponding author: Cyril Geismar (cyril.geismar.16@ucl.ac.uk)

Author roles: Geismar C: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – Original Draft Preparation; **Fragaszy E**: Data Curation, Investigation, Validation, Writing – Review & Editing; **Nguyen V**: Methodology, Software; **Fong WLE**: Writing – Review & Editing; **Shrotri M**: Data Curation; **Beale S**: Data Curation; **Rodger A**: Writing – Review & Editing; **Lampos V**: Writing – Review & Editing; **Byrne T**: Data Curation; **Kovar J**: Project Administration; **Navaratnam AMD**: Data Curation; **Patel P**: Writing – Review & Editing; **Aldridge RW**: Data Curation, Funding Acquisition, Project Administration, Resources, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing; **Hayward A**: Funding Acquisition, Investigation, Project Administration, Resources, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing;

Competing interests: ACH serves on the UK New and Emerging Respiratory Virus Threats Advisory Group. AMJ was a Governor of Wellcome Trust from 2011-18 and is Chair of the Committee for Strategic Coordination for Health of the Public Research.

Grant information: The research costs for the study have been supported by the MRC Grant Ref: MC_PC 19070 awarded to UCL on 30 March 2020 and MRC Grant Ref: MR/V028375/1 awarded on 17 August 2020. The study also received $15,000 of Facebook advertising credit to support a pilot social media recruitment campaign on 18th August 2020. This study was supported by the Wellcome Trust through a Wellcome Clinical Research Career Development Fellowship to RA (206602, https://doi.org/10.35802/206602]).
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

How to cite this article: Geismar C, Fragaszy E, Nguyen V *et al.* **Household serial interval of COVID-19 and the effect of Variant B.1.1.7: analyses from prospective community cohort study (Virus Watch) [version 2; peer review: 2 approved]** Wellcome Open Research 2021, **6**:224 https://doi.org/10.12688/wellcomeopenres.16974.2

First published: 09 Sep 2021, **6**:224 https://doi.org/10.12688/wellcomeopenres.16974.1

REVISED **Amendments from Version 1**

The distinction between the generation time and the serial interval and their relation to the epidemic doubling time was added in the introduction.

Any further responses from the reviewers can be found at the end of the article

Introduction

The serial interval is defined as "the period of time between analogous phases of an infectious illness in successive cases of a chain of infection that is spread person to person" (Feinleib, 2001). Serial interval is often measured as the duration between symptom onset of a primary case and symptom onset of its secondary cases. This is a key epidemiological measure because it can allow investigation of epidemiological links between cases, and it is an important parameter in infection transmission models used to inform infection control strategies. The doubling time of epidemic infections is in part dependent on both the generation time (time between infections regardless of the symptomatic status) and the R number (the average number of secondary infections each infection produces). Diseases with shorter generation time but similar R values will have shorter doubling times. Since the generation time is seldom observable, in practice the serial interval is used as a proxy for it. Mean serial intervals vary widely for different respiratory infections and have been estimated at 2.2 days for influenza A H3N2, 2.8 days for pandemic influenza A(H1N1)pdm09, 7.5 days for respiratory syncytial virus, 11.7 days for measles, 14 days for varicella, 17.7 days smallpox, 18.0 days for mumps, 18.3 days for rubella and 22.8 days for pertussis (Vink *et al.*, 2014).

Published estimates of the serial interval of coronavirus disease 2019 (COVID-19) are largely from Asian countries prior to the emergence of variants of concern (VOC). A meta-analysis of serial interval estimates for COVID-19 found mean serial intervals ranged from 4.2 to 7.5 days with a pooled mean of 5.2 (95%CI: 4.9–5.5) (Alene *et al.*, 2021). A more recent concerning feature of COVID-19 epidemiology has been the emergence of a range of SARS-CoV-2 variants with mutations that may increase transmissibility, reduce the protective effect of immunity acquired from natural infection or vaccination and/or increase clinical severity (Alene *et al.*, 2021). These include B.1.1.7 (first described in England), 501Y.V2 (first described in South Africa) and P.1 (B.1.1.28.1 - first described in Brazil). Each of these variants rapidly became dominant in the country in which they were first described. For the B.1.1.7 variant, increased transmissibility is thought to explain the rapid emergence and global spread. Since either increased R or decreased serial interval could potentially explain more rapid emergence of B.1.1.7, it is important to understand whether serial interval differs. To date, however, there are no published comparisons of the serial interval for B.1.1.7 and previously circulating strains. We analysed data from putative household infector - infectee pairs in the Virus Watch Community cohort study to assess the serial interval of COVID-19 and whether this was affected by emergence of the B.1.1.7 variant.

Methods

This study has been approved by the Hampstead NHS Health Research Authority Ethics Committee. Ethics approval number - 20/HRA/2320. All members of participating households provided informed consent for themselves and, where relevant, for children that they were responsible for. This was electronically collected during registration. All necessary patient/participant consent has been obtained and the appropriate institutional forms have been archived.

The Virus Watch study is an online, prospective, community cohort study following up entire households in England and Wales during the COVID-19 pandemic. Participants were recruited into the Virus Watch study using a range of methods including by post, social media, SMS messages or personalised letters with incentives from General Practices. Participants were eligible if all household members agreed to take part, if they had access to the internet (Wi-Fi, fixed or on a mobile phone) and an email address. At least one household member had to be able to read English to complete the surveys. Participants were not eligible if their household was larger than 6 people (due to limitations of the online survey infrastructure). The full study design and methodology has been described elsewhere (Hayward *et al.*, 2021). Study data were collected and managed using REDCap electronic data capture tools hosted at University College London (Harris *et al.*, 2009). Data collection began on 24 June 2020 and is ongoing. As of 11th April 2021, 49,149 people across England and Wales have joined the study. Participants prospectively complete detailed daily symptom diaries recording the presence and severity of any symptoms of acute respiratory, gastrointestinal and other illnesses. At the end of each week participants are emailed a link to complete a weekly online survey where they report any symptoms from that previous week as well as the dates and outcomes of any COVID-19 swabbing conducted as part of NHS Test and Trace, work-based testing schemes, and other research studies.

Symptom data were extracted and grouped into illness episodes using Stata (StataCorp, 2019). The start date of an illness episode was defined as the first day any symptoms were reported, and the end date was the final day of reported symptoms. A seven-day washout period where no symptoms were reported was used to define the end of one illness episode and the start of a new illness episode. Swab results were matched to illnesses that were within 14 days of each other. Putative household infector-infectee pairs were identified where more than one person in the household had a positive swab matched to an illness episode. Although negative serial intervals are possible, in practice it is not possible to assess the direction of transmission between pairs, so it was assumed that the minimum interval was zero days. According to the World Health Organization report on SARS-CoV-2 transmission, the estimated latest transmission can occur up to nine days after the infector's symptom onset & the incubation period for the infectee can be up to 14 days (WHO, 2020). Thus, the longest time interval between an infector's and an infectee's onset of symptoms was considered at 23 days. Our analysis conducted using R version 4.0.3 (R Core Team, 2020), considered pairs of cases with symptom onset occurring between 0 and 23 days apart

in households as possible transmission pairs (Geismar, 2021). Serial interval was calculated as the number of days between symptom onset of the pairs of cases. Figure 1 presents a hypothetical household with four confirmed COVID-19 cases and their respective symptom onset date. Any case with a symptom onset date within 23 days of a previous case will be paired. Where there were multiple potential infectors for the same infectee, these pairs were excluded from the analysis. 'Person #4' has two potential infectors as her symptom onset date is within 23 days of 'Person #2' and 'Person #3' 's symptom onset dates. Pairs containing 'Person #4' as an infectee will be removed since we cannot determine her "true" infector. Thus, we only retain the two most likely transmission pairs: 'Person #1' to 'Person #2' and 'Person #2' to 'Person #3'.

Data on whether or not individual infections were caused by the B.1.1.7 variant were not available. We therefore developed a classification system based on the percentage of cases estimated to be due to B.1.1.7 in national surveillance data for different English regions and study weeks. These surveillance data utilise a proxy indicator of B.1.1.7 known as Spike-gene target failure (SGTF) which can be picked up on most polymerase chain reaction (PCR) assays used in English community testing programmes. Infections in regions and weeks when >75% of strains were SGTF were classified as occurring in B.1.1.7 "hotspots". Infections in regions and weeks when <25% of strains were SGTF were classified as occurring in "non-hotspots". Infections in regions and weeks when 25% to 75% of strains were SGTF were classified as "undetermined" since no significant threshold was reached. Mean serial interval and 95% confidence intervals were compared in hotspot and non-hotspot areas using Welch two-sample t-tests.

Results

Out of 24,887 illnesses reported by 14,986 individuals within 9,991 households between 22/06/2020 and 07/03/2021, 7,304 were tested for SARS-CoV-2. Amongst the swabbed illnesses, 915 tested positive for SARS-CoV-2. 287 possible 'infector-infectee' pairs in 194 households amongst 424 individuals were identified. Non-unique infectees (infectees with multiple potential infectors) were then excluded as their infectors could not be determined, resulting in 186 likely pairs in 186 households amongst 372 individuals between 03/09/2020 and 22/02/2021. 43 'infector-infectee' pairs were identified in non-hotspot areas between 03/09/2020 – 05/12/2020, 69 were in hotspot areas between 14/12/2020 – 22/02/2021 and 74 in areas not determined between 09/09/2020 – 25/01/2021 (Table 1 and Table 2).

Figure 2 shows the distribution of serial intervals. The distribution peaks between 0 and 1 day and 90% of all observations lie between 0 and 8.5 days, with values spanning up to 21 days. The mean COVID-19 serial interval was 3.18 (95%CI: 2.55 - 3.81) days and its median was 2 days (Table 2).

Figure 3 compares the distribution of serial intervals for B.1.1.7 hotspot areas and non-hotspot areas. There was no significant difference (p=0.267) between the mean serial interval for VOC hotspots (mean = 3.64 days (95%CI: 2.55 – 4.73), median = 2 days) days and non-VOC hotspots, (mean = 2.72 days (95%CI: 1.48 – 3.96), median = 1 day) (Table 2).

Discussion

Our estimate of the mean serial interval of COVID-19 (3.18 days (95%CI: 2.55 - 3.81)) is within the range of previous studies reviewed by Griffin *et al.* (2020), but slightly lower than pooled estimates from meta-analysis of data from international studies in the first few months of the pandemic (5.2 (95%CI: 4.9 - 5.5)) (Alene *et al.*, 2021). Differences in populations, social contact, and timeframes may explain the range of estimates reported. The implementation of control measures, regular testing, isolation and improved knowledge of SARS-CoV-2's transmission since the start of the pandemic, may have reduced the potential for an infected person to transmit the disease over a long period of time. Multiple studies observed and attributed the decrease of the serial interval to increased control measures (Bi *et al.*, 2020; Lavezzo *et al.*, 2020; Zhao *et al.*, 2020). Ali *et al.* (2020) modelled the serial interval over time accounting for timeliness of cases' isolation and found that "serial intervals are positively associated with isolation delay". Another potential explanation of shorter serial intervals may be due the frequent and close contact among household members. This could lead to transmissions occurring earlier in the course of infection, which would result in shorter serial intervals.

Strengths of the study include the relatively large number of pairs compared to most studies, the prospective daily recording of symptoms and weekly reporting of swab test results in a large household cohort, and our ability to assess whether a variant with apparent increases in transmissibility has an altered serial interval. Limitations of our analysis include reliance on samples taken during the national symptomatic testing programme to assess infection, meaning we are likely to have missed some infections and household transmission events. Pooled asymptomatic proportion of SARS-CoV-2 infections is estimated at 23% (95% CI 16%-30%) and we cannot assess serial intervals when either case is asymptomatic (Beale *et al.*, 2020). We can also not assess the possibility of negative serial intervals which may arise when transmission occurs prior to symptom onset and incubation period is short. Missing negative serial interval values could lead to an overestimation of our mean serial interval estimate. Finally, we do not have

Figure 1. Diagram of a hypothetical SARS-CoV-2 outbreak within a household with individual illnesses start dates.

Table 1. **Demographic characteristics of the individuals included in the analysis.**

VOC status (Number of individuals)	*Overall (n = 372)*		*Hotspot (n = 143)*		*Non-hotspot (n = 85)*		*Not determined (n = 144)*	
	n	**%**	**n**	**%**	**n**	**%**	**n**	**%**
Sex (at birth)								
Female	179	48.1	71	49.7	41	48.2	67	46.5
Male	170	45.6	68	47.6	42	49.4	60	41.7
Missing	23	6.3	4	2.7	2	2.4	17	11.8
Age (years)								
0–15	25	6.7	11	7.7	5	5.9	9	6.2
16–44	114	30.6	45	31.5	26	30.6	43	29.9
45–64	165	44.4	63	44.0	29	34.1	73	50.7
65 +	68	18.3	24	16.8	25	29.4	19	13.2
Region								
East Midlands	46	12.4	8	5.6	20	23.5	18	12.5
East of England	56	15.1	36	25.2	2	2.3	18	12.5
London	62	16.7	44	30.7	9	10.6	9	6.2
North East	20	5.4	6	4.2	6	7.1	8	5.6
North West	54	14.5	11	7.7	16	18.8	27	18.7
South East	40	10.7	28	19.6	6	7.1	6	4.2
South West	16	4.3	2	1.4	6	7.1	8	5.6
Wales	10	2.7	0	0.0	0	0.0	10	6.9
West Midlands	18	4.8	6	4.2	4	4.7	8	5.6
Yorkshire and The Humber	28	7.5	2	1.4	16	18.8	10	6.9
Missing	22	5.9	0	0.0	0	0.0	22	15.3
Number of household members								
2	214	57.5	71	49.6	55	64.7	88	61.1
3	84	22.6	41	28.7	12	14.1	31	21.5
4	54	14.5	25	17.5	12	14.1	17	11.8
5	16	4.3	2	1.4	6	7.1	8	5.6
6	4	1.1	4	2.8	0	0.0	0	0.0

VOC, variant of concern.

Table 2. **Serial Interval by variant of concern (VOC) status.**

VOC status (Number of transmissions)	**Overall (n = 186)**	**Hotspot (n = 69)**	**Non-hotspot (n = 43)**	**Not determined (n = 74)**
Median serial interval (IQR)	2.0 (0.0 – 4.0)	2.0 (0.0 – 5.0)	1.0 (0.0 – 3.5)	2.0 (0.0 – 3.0)
Mean serial interval (95%CI)(sd)	3.2 (2.5 – 3.8) (4.4)	3.6 (2.5 – 4.7)(4.5)	2.7 (1.5 – 4.0)(4.0)	3.0 (2.0 – 4.0)(4.4)

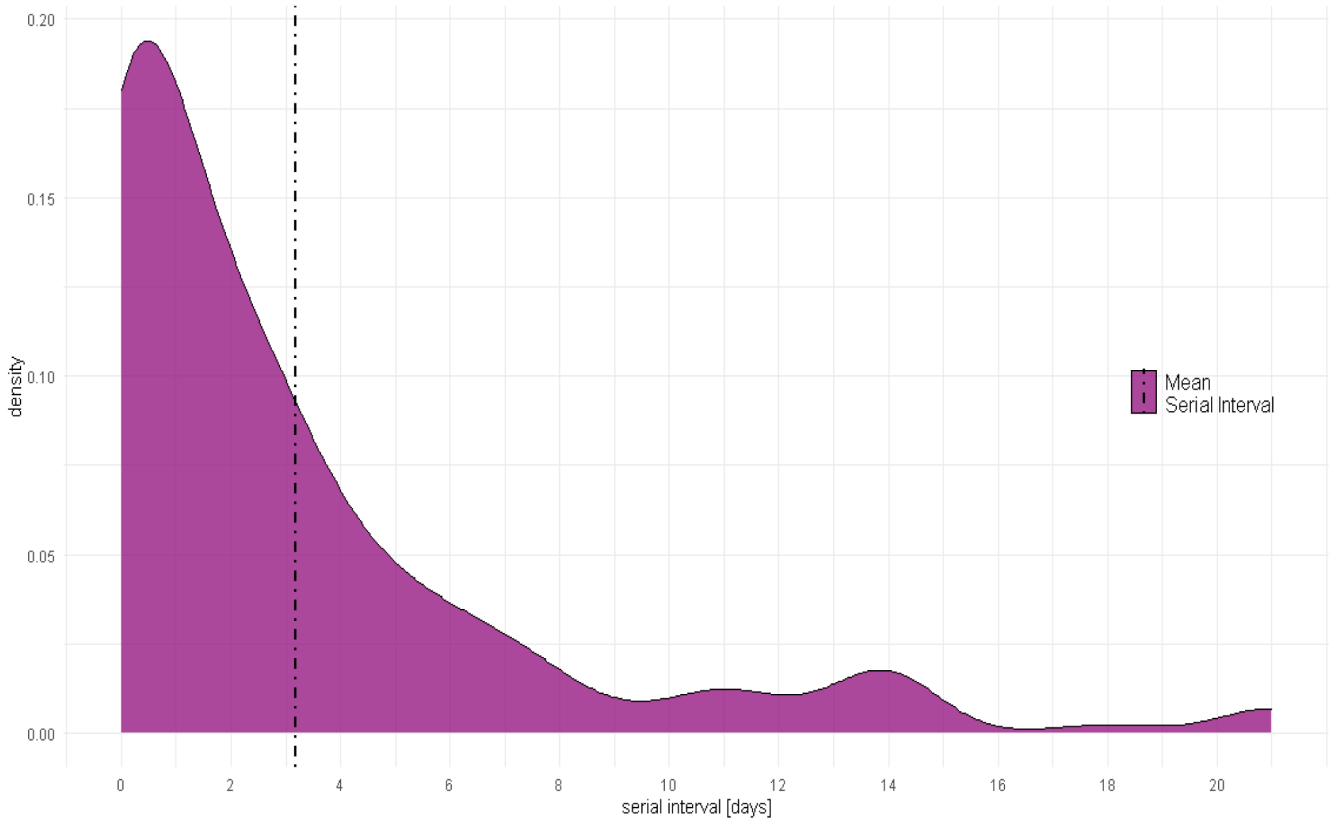


Figure 2. Serial Interval distribution and mean.

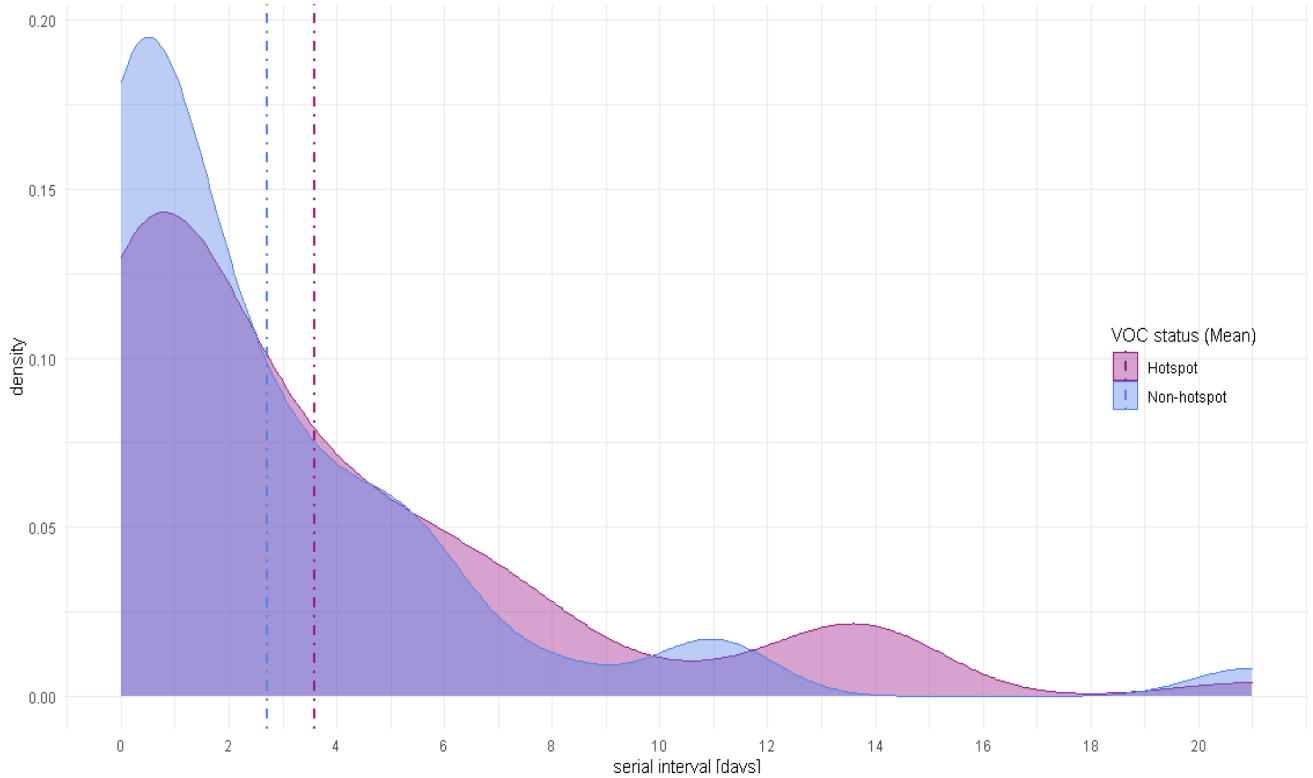


Figure 3. Serial interval density distribution for B.1.1.7 hotspot areas and non-hotspot areas (Mean). VOC, variant of concern.

individual information on whether strains were B.1.1.7 and used a proxy measure of this based on levels of SGTF in different areas and times. These limitations reduce the likelihood of observing differences between B.1.1.7 other wild-type strain types at the time of the study.

Our analysis does not provide evidence to suggest that changes in serial interval explain the rapid emergence of B.1.1.7. Alternative explanations such as increased viral load (Kidd *et al.*, 2021) or improved receptor binding may instead explain the increased transmissibility and rapid spread and should undergo further investigation.

Data availability

Underlying data

We aim to share aggregate data from this project on our website and via a "Findings so far" section on our website - https://ucl-virus-watch.net/. We will also be sharing individual record level data on a research data sharing service such as the Office of National Statistics Secure Research Service. In sharing the data we will work within the principles set out in the UKRI Guidance on best practice in the management of research data. Access to use of the data whilst research is being conducted will be managed by the Chief Investigators (ACH and RWA) in accordance with the principles set out in the UKRI guidance on best practice in the management of research data. We will put analysis code on publicly available repositories to enable their reuse. Given the content of our dataset (information on infector-infectee pairs per geographic regions) for this study, we currently cannot release the data at the individual level. Data access requests to data can be made to the Virus Watch chief investigators (ACH or RWA) at the following email address: viruswatch@ucl.ac.uk.

Extended data

Analysis code available from: https://github.com/UCL-Public-Health-Data-Science/VirusWatch_COVID_Serial_Interval

Archived analysis code at time of publication: https://doi.org/10.5281/zenodo.5172106 (Geismar, 2021)

Acknowledgements

The Virus Watch Collaborative authors cited below have agreed for their names and affiliations to be included in this publication.

Virus Watch Collaborative:
Pia Hardelid[1], Isobel Braithwaite[2], Linda Wijlaars[1], Eleni Nastouli[1,3], Moira Spyer[1,3], Ben Killingley[4,5], Ingemar Cox[6], Rachel A McKendry[7], Tao Cheng[8], Yunzhe Liu[8], Anne M Johnson[9], Richard Gilson[9], Jo Gibbs[9].

1. Department of Population, Policy and Practice, UCL Great Ormond Street Institute of Child Health, London, UK

2. Centre for Public Health Data Science, Institute of Health Informatics, University College London, London, UK.

3. Francis Crick Institute, London, UK.

4. Health Protection and Influenza Research Group, Division of Epidemiology and Public Health, University of Nottingham School of Medicine, Nottingham, United Kingdom

5. University College London Hospital, London, United Kingdom.

6. Department of Computer Science, University College London, London, UK.

7. London Centre for Nanotechnology and Division of Medicine, University College London, London, UK.

8. SpaceTimeLab, Department of Civil, Environmental and Geomatic Engineering, University College London, London, UK.

9. Institute for Global Health, University College London, London, UK.

References

Alene M, Yismaw L, Assemie MA, *et al.*: **Serial interval and incubation period of COVID-19: a systematic review and meta-analysis.** *BMC Infect Dis.* 2021; **21**(1): 257.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Ali ST, Wang L, Lau EHY, *et al.*: **Serial interval of SARS-CoV-2 was shortened over time by nonpharmaceutical interventions.** *Science.* 2020; **369**(6507): 1106–1109.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Beale S, Hayward A, Shallcross L, *et al.*: **A rapid review and meta-analysis of the asymptomatic proportion of PCR-confirmed SARS-CoV-2 infections in community settings [version 1; peer review: 1 approved with reservations].** *Wellcome Open Res*. 2020; **5**: 266.
Publisher Full Text

Bi Q, Wu Y, Mei S, *et al.*: **Epidemiology and Transmission of COVID-19 in Shenzhen China: Analysis of 391 cases and 1,286 of their close contacts.** *MedRxiv.* 2020.
Publisher Full Text

Feinleib M: **A Dictionary of Epidemiology, Fourth Edition - Edited by John M. Last, Robert A. Spasoff, and Susan S. Harris.** *Am J Epidemiol.* 2001; **154**(1): 93–94.
Publisher Full Text

Geismar C: **UCL-Public-Health-Data-Science/VirusWatch-Serial_Interval: v1.0.0** (Version v1.0.0). *Zenodo.* 2021.
http://www.doi.org/10.5281/zenodo.5101611

Griffin J, Casey M, Collins Á, *et al.*: **Rapid review of available evidence on the serial interval and generation time of COVID-19.** *BMJ Open.* 2020; **10**(11): e040263.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Harris PA, Taylor R, Thielke R, *et al.*: **Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support.** *J Biomed Inform.* 2009; **42**(2): 377–381.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Hayward A, Fragaszy E, Kovar J, *et al.*: **Risk factors, symptom reporting,**

healthcare-seeking behaviour and adherence to public health guidance: protocol for Virus Watch, a prospective community cohort study. *medRxiv.* 2021.
Publisher Full Text

Kidd M, Richter A, Best A, *et al.*: **S-Variant SARS-CoV-2 Lineage B1.1.7 Is Associated With Significantly Higher Viral Load in Samples Tested by TaqPath Polymerase Chain Reaction.** *J Infect Dis.* 2021; **223**(10): 1666–1670.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Lavezzo E, Franchin E, Ciavarella C, *et al.*: **Suppression of a SARS-CoV-2 outbreak in the Italian municipality of Vo'.** *Nature.* 2020; **584**(7821): 425–429.
PubMed Abstract | **Publisher Full Text**

R Core Team: **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2020.
Reference Source

StataCorp: **Stata Statistical Software: Release 16.** College Station, TX: StataCorp LLC, 2019.

Vink MA, Bootsma MCJ, Wallinga J: **Serial intervals of respiratory infectious diseases: a systematic review and analysis.** *Am J Epidemiol.* 2014; **180**(9): 865–875.
PubMed Abstract | **Publisher Full Text**

World Health Organization: **Transmission of SARS-CoV-2: implications for infection prevention precautions: scientific brief, 09 July 2020**. World Health Organization, License: CC BY-NC-SA 3.0 IGO. 2020.
Reference Source

Zhao S, Cao P, Chong MKC, *et al.*: **COVID-19 and gender-specific difference: Analysis of public surveillance data in Hong Kong and Shenzhen, China, from January 10 to February 15, 2020.** *Infect Control Hosp Epidemiol.* 2020; **41**(6): 750–751.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Open Peer Review

Current Peer Review Status: ✔ ✔

Version 2

Reviewer Report 22 December 2021

https://doi.org/10.21956/wellcomeopenres.19188.r47730

✔ **Benjamin J. Cowling**
WHO Collaborating Centre for Infectious Disease Epidemiology and Control, School of Public Health, The University of Hong Kong, Hong Kong, China

No further comments.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Infectious disease epidemiology

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 21 December 2021

https://doi.org/10.21956/wellcomeopenres.19188.r47729

✔ **Vítor H. Nascimento**
Department of Electronic Systems Engineering, University of São Paulo, São Paulo, Brazil

I have no further comments on the paper.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Signal processing/statistics applied to epidemiology

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 02 November 2021

https://doi.org/10.21956/wellcomeopenres.18741.r46856

Benjamin J. Cowling
WHO Collaborating Centre for Infectious Disease Epidemiology and Control, School of Public Health, The University of Hong Kong, Hong Kong, China

This is a valuable study of the household serial interval of COVID-19 based on the FluWatch cohort which includes tens of thousands of households from across the UK.

Minor comments and suggestions:

1. The title could be modified to "Household serial intervals ..." since the data reflect serial intervals in households but not other settings.

2. The exclusion of negative serial intervals might lead to overestimation of the mean serial interval? This could be clarified at the end of the second Discussion paragraph?

3. The standard deviation of serial intervals could also be presented in the abstract and also in Table 2 and the main text.

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Infectious disease epidemiology

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 22 Dec 2021

Cyril Geismar, University College London, London, UK

We thank the reviewer for their valuable insight.
All suggestions have been added to the 2nd version of the paper.

Competing Interests: No competing interests were disclosed.

Reviewer Report 27 September 2021

https://doi.org/10.21956/wellcomeopenres.18741.r45799

Vítor H. Nascimento
Department of Electronic Systems Engineering, University of São Paulo, São Paulo, Brazil

The paper compares the mean serial intervals for COVID-19 in areas with a large incidence of variant B.1.1.7, with areas in which the wild variant is dominant, concluding that there is no significant difference between the means.

This comparison presents a number of difficulties, most of which are discussed in the paper. These difficulties are inherent to estimates of the serial interval distribution, and, although should be taken into account, in my opinion, do not invalidate the conclusions, since data about different variants was obtained at about the same time, using the same procedure.

- The data is restricted to infector-infectee pairs within households, so pairs with people in different households are not taken into account. This will probably bias down the observed serial intervals since contacts within households tend to be closer.

- In the introduction, it is mentioned that serial intervals and the doubling time of the epidemic are related to the reproduction number - but to be precise, it is the generation interval (time between infections) that has this relation. Since the generation interval is seldom observable, the serial interval is used as a proxy for it. I believe this should be

acknowledged in the paper.

- Due to limitations in the software used, households with more than 6 people were not included in the study. It would be good to include in the text how many households are in this situation. One limitation of the study is that households with many people would tend to lower down the serial interval distribution, since if more people live together, they may presumably have closer contacts than in households with less people. However, since the same kind of bias was applied to all virus variants, I believe this would not affect the final conclusion.

- Asymptomatic cases were not included (since the serial interval is not defined for asymptomatic cases). This might be a source of bias if one variant has a larger percentage of asymptomatic infections than the other.

- Negative serial intervals were also not considered (the absolute value of the serial interval is used instead). Previous studies report a rate of approximately 13% negative serial intervals (Griffin *et al.*, 2020).

- The distribution of serial intervals will depend on NPIs. Despite the fact that the study considers cases occurring in different regions at the same time, it might be important to check if different regions were not having different sets of restrictions in place (for example, in a region under lockdown, people would stay at home most of the time, potentially decreasing the within-household serial interval). This potential difference between hotspots was not discussed in the paper.

- Although the mean serial interval is particularly important, the whole distribution affects estimates of the reproduction number - so it might be worth comparing the distributions, not only their means.

The study is an important contribution to understanding the observed differences between variants, and a worthy contribution to the literature.

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Partly

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Signal processing/statistics applied to epidemiology

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 29 Oct 2021

Cyril Geismar, University College London, London, UK

1.The data is restricted to infector-infectee pairs within households, so pairs with people in different households are not taken into account. This will probably bias down the observed serial intervals since contacts within households tend to be closer.

- We thank the reviewer for their valuable insights. We agree that the clustering method we implemented may impact our results. We state this limitation in the first paragraph of the Discussion section. To consider transmissions occurring outside the household we would require data about 1) the contact between members in our cohort and 2) data from those outside our cohort. Due to ethical and security concerns, point 1 is not possible as it requires us to have access to a deeper volume of data akin to surveillance for which we do not have ethical approval for. Point 2 is not possible as our study actively requires participants to knowingly provide data; we only have this at the household level.

2. In the introduction, it is mentioned that serial intervals and the doubling time of the epidemic are related to the reproduction number - but to be precise, it is the generation interval (time between infections) that has this relation. Since the generation interval is seldom observable, the serial interval is used as a proxy for it. I believe this should be acknowledged in the paper.

- We thank the reviewer for pointing this out and will update for all audiences.

3. Due to limitations in the software used, households with more than 6 people were not included in the study. It would be good to include in the text how many households are in this situation. One limitation of the study is that households with many people would tend to lower down the serial interval distribution, since if more people live together, they may presumably have closer contacts than in households with less people. However, since the same kind of bias was applied to all virus variants, I believe this would not affect the final conclusion.

- We thank the reviewer for providing their insights into the limitation on household size. We agree that the bias applies to all virus variants and is not likely to impact the conclusion. Furthermore, according to the ONS report “Families and households in the UK: 2020”: “*The average household size in the UK is 2.4 while there were 162,900 (0.6%) households in the UK with seven or more people*''. Therefore we believe that our results are representative.
  https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/families/bulletin
  .

4. Asymptomatic cases were not included (since the serial interval is not defined for asymptomatic cases). This might be a source of bias if one variant has a larger percentage of asymptomatic infections than the other.

- We agree with the reviewer. We acknowledged in the paper that we were unable to measure asymptomatic transmissions and that the estimates provided are strictly measures of the serial interval when both the infector and infectee are symptomatic.

5. Negative serial intervals were also not considered (the absolute value of the serial interval is used instead). Previous studies report a rate of approximately 13% negative serial intervals (Griffin et al., 2020).

- We thank the reviewer for pointing this out with supporting evidence from Griffin et al. (2020)'s rapid review. We did acknowledge the possibility of COVID-19 negative serial interval and explained why we were not able to consider this:
  "Although negative serial intervals are possible, in practice it is not possible to assess the direction of transmission between pairs, so it was assumed that the minimum interval was zero days."
  Given the nature of our data (self-reported), we were not able to determine cases when transmission occurs prior to symptom onset.
  Although we acknowledge the possibility of a significant proportion of negative serial interval in COVID-19 transmissions, we do not believe that there is sufficient evidence from enough studies with significant transmission pairs to report a specific percentage.

6. The distribution of serial intervals will depend on NPIs. Despite the fact that the study considers cases occurring in different regions at the same time, it might be important to check if different regions were not having different sets of restrictions in place (for example, in a region under lockdown, people would stay at home most of the time, potentially decreasing the within-household serial interval). This potential difference between hotspots was not discussed in the paper.

- We thank the reviewer for their comments. To mitigate any variation related to the localised implementation of policies, our recruitment process was conducted using the royal mail list to recruit evenly across the UK; therefore our results are a close estimate to the average and were less likely impacted by local variation.

7.Although the mean serial interval is particularly important, the whole distribution affects estimates of the reproduction number - so it might be worth comparing the distributions, not only their means.

- We used the Welch two sample t test to compare VOC and non-VOC hotspots. This statistical test is a test of the distribution and not the mean itself as it accounts for the standard deviation which is a measure of the distribution.

Competing Interests: No competing interests were disclosed.