medRxiv preprint doi: https://doi.org/10.1101/2021.07.20.21260855; this version posted December 4, 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 The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.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 .

Title: SARS-CoV-2 incidence, transmission and reinfection in a rural and an urban setting: results of the PHIRST-C cohort study, South Africa, 2020-2021

Authors

Cheryl Cohen 1,2 (PhD), Jackie Kleynhans 1,2 (MPH), Anne von Gottberg 1,4 (PhD), Meredith L McMorrow 3 (MD), Nicole Wolter 1,4 (PhD), Jinal N. Bhiman 1,4 (PhD), Jocelyn Moyes 1,2 (MD), Mignon du Plessis 1,4 (PhD), Maimuna Carrim 1,4 (MSc), Amelia Buys 1 (MTech), Neil A Martinson 5, 6,7 (PhD), Kathleen Kahn 8 (PhD), Stephen Tollman 8 (PhD), Limakatso Lebina 5,9 (PhD), Floidy Wafawanaka 8 (MPharm), Jacques du Toit 8 (MD), Francesc Xavier Gómez-Olivé 8 (PhD), Fatimah S. Dawood 3 (MD), Thulisa Mkhencele 1 (MSc), Kaiyun Sun (PhD) 10, Cécile Viboud 10 (PhD), for the PHIRST group, Stefano Tempia 1,2,3 (PhD)

Affiliations

1 Centre for Respiratory Diseases and Meningitis, National Institute for Communicable Diseases of the National Health Laboratory Service, Johannesburg, South Africa

2 School of Public Health, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa

3 Influenza Division, Centers for Disease Control and Prevention, Atlanta, Georgia, United States of America

4 School of Pathology, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa

5 Perinatal HIV Research Unit, MRC Soweto Matlosana Collaborating Centre for HIV/AIDS and TB, University of the Witwatersrand, South Africa

6 DST/NRF Centre of Excellence for Biomedical Tuberculosis Research, University of the Witwatersrand, Johannesburg, South Africa

7 Johns Hopkins University Center for TB Research, Baltimore, Maryland, United States of America

8 MRC/Wits Rural Public Health and Health Transitions Research Unit (Agincourt), Faculty of Health Sciences, School of Public Health, University of the Witwatersrand, Johannesburg, South Africa

9 Africa Health Research Institute, KwaZulu-Natal, South Africa

10 Division of International Epidemiology and Population Studies, Fogarty International Center, National Institutes of Health, Bethesda, MD, USA.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.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/10.1101/2021.07.20.21260855; this version posted December 4, 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 .

Corresponding author: Cheryl Cohen, Centre for Respiratory Diseases and Meningitis, National Institute for Communicable Diseases, Private Bag X4, Sandringham, 2131, Gauteng, South Africa, Telephone: 27 11 386 6593. Fax: 27 11 882 9979. E-mail: cherylc@nicd.ac.za

Keywords: SARS-CoV-2, burden, transmission, household, South Africa, HIV, rural, urban

Running title: SARS-CoV-2 community burden and transmission South Africa

Word count

Abstract:300

Text: 4468

Tables: 4

Figures: 2

medRxiv preprint doi: https://doi.org/10.1101/2021.07.20.21260855; this version posted December 4, 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 .

Research in context

Evidence before this study

Previous studies have generated wide-ranging estimates of the proportion of SARS-CoV-2 infections which are asymptomatic. A recent systematic review found that 20% (95% CI 3%-67%) of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infections remained asymptomatic throughout infection and that transmission from asymptomatic individuals was reduced. A systematic review and meta-analysis of 87 household transmission studies of SARS-CoV-2 found an estimated secondary attack rate of 19% (95% CI 16-22). The review also found that household secondary attack rates were increased from symptomatic index cases and that adults were more likely to acquire infection. As of December 2021, South Africa experienced three waves of SARS-CoV-2 infections; the second and third waves were associated with circulation of Beta and Delta variants respectively. SARS-CoV-2 vaccines became available in February 2021, but uptake was low in study sites reaching 5% fully vaccinated at the end of follow up. Studies to quantify the burden of asymptomatic infections, symptomatic fraction, reinfection frequency, duration of shedding and household transmission of SARS-CoV-2 from asymptomatically infected individuals have mostly been conducted as part of outbreak investigations or in specific settings. Comprehensive systematic community studies of SARS-CoV-2 burden and transmission including for the Beta and Delta variants are lacking, especially in low vaccination settings.

Added value of this study

We conducted a unique detailed COVID-19 household cohort study over a 13 month period in South Africa, with real time reverse transcriptase polymerase chain reaction (rRT-PCR) testing twice a week irrespective of symptoms and bimonthly serology. By the end of the study in August 2021, 749 (62%) of 1200 individuals from 222 randomly sampled households in a rural and an urban community in South Africa had at least one confirmed SARS-CoV-2 infection, detected on rRT-PCR and/or serology, and 12% (87/749) experienced reinfection. Symptom data were analysed for 662 rRT-PCR-confirmed infection episodes that occurred >14 days after the start of follow-up (of a total of 718 rRT-PCR-confirmed episodes), of these, 15% (n=97) were associated with one or more symptoms. Among symptomatic indvidiausl, 9% (n=9) were hospitalised and 2% (n=2) died. Ninety percent (200/222) of included households, had one or more individual infected with SARS-CoV-2 on rRT-PCR and/or serology within the household. SARS-CoV-2 infected index cases transmitted the infection to 25% (213/856) of susceptible household contacts. Index case ribonucleic acid (RNA) viral load proxied by rRT-PCR cycle threshold value was strongly predictive of household transmission. Presence of

medRxiv preprint doi: https://doi.org/10.1101/2021.07.20.21260855; this version posted December 4, 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 .

symptoms in the index case was not associated with household transmission. Household transmission was four times greater from index cases infected with Beta variant and fifteen times greater from index cases infected with Delta variant compared to wild-type infection. Attack rates were highest in individuals aged 13-18 years and individuals in this age group were more likely to experience repeat infections and to acquire SARS-CoV-2 infection within households. People living with HIV (PLHIV) who were not virally supressed were more likely to develop symptomatic illness when infected with SARS-CoV-2, and shed SARS-CoV-2 for longer when compared to HIV-uninfected individuals.

Implications of all the available evidence

We found a high rate of SARS-CoV-2 infection in households in a rural community and an urban community in South Africa, with the majority of infections being asymptomatic in individuals of all ages. Asymptomatic individuals transmitted SARS-CoV-2 at similar levels to symptomatic individuals suggesting that interventions targeting symptomatic individuals such as symptom-based testing and contact tracing of individuals tested because they report symptoms may have a limited impact as control measures. Increased household transmission of Beta and Delta variants, likely contributed to recurrent waves of COVID-19, with >60% of individuals infected by the end of follow-up. Higher attack rates, reinfection and acquisition in adolescents and prolonged SARS-CoV-2 shedding in PLHIV who were not virally suppressed suggests that prioritised vaccination of individuals in these groups could impact community transmission.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.20.21260855; this version posted December 4, 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 .

Abstract

Word count: 300 Target: 300

Background

By August 2021, South Africa experienced three SARS-CoV-2 waves; the second and third associated with emergence of Beta and Delta variants respectively.

Methods

We conducted a prospective cohort study during July 2020-August 2021 in one rural and one urban community. Mid-turbinate nasal swabs were collected twice-weekly from household members irrespective of symptoms and tested for SARS-CoV-2 using real-time reverse transcription polymerase chain reaction (rRT-PCR). Serum was collected every two months and tested for anti-SARS-CoV-2 antibodies.

Results

Among 115,759 nasal specimens from 1,200 members (follow-up rate 93%), 1976 (2%) were SARS-CoV-2-positive. By rRT-PCR and serology combined, 62% (749/1200) of individuals experienced ≥1 SARS-CoV-2 infection episode, and 12% (87/749) experienced reinfection. Of 662 PCR-confirmed episodes with available data, 15% (n=97) were associated with ≥1 symptom. Among 222 households, 200 (90%) had ≥1 SARS-CoV-2-positive individual. Household cumulative infection risk (HCIR) was 25% (213/856). On multivariable analysis, accounting for age and sex, index case lower cycle threshold value (OR 3.9, 95%CI 1.7-8.8), urban community (OR 2.0,95%CI 1.1-3.9), Beta (OR 4.2, 95%CI 1.7-10.1) and Delta (OR 14.6, 95%CI 5.7-37.5) variant infection were associated with increased HCIR. HCIR was similar for symptomatic (21/110, 19%) and asymptomatic (195/775, 25%) index cases (p=0.165). Attack rates were highest in individuals aged 13-18 years and individuals in this age group were more likely to experience repeat infections and to acquire SARS-CoV-2 infection. People living with HIV who were not virally supressed were more likely to develop symptomatic illness, and shed SARS-CoV-2 for longer compared to HIV-uninfected individuals.

Conclusions

In this study, 85% of SARS-CoV-2 infections were asymptomatic and index case symptom status did not affect HCIR, suggesting a limited role for control measures targeting symptomatic individuals. Increased household transmission of Beta and Delta variants, likely contributed to successive waves, with >60% of individuals infected by the end of follow-up.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.20.21260855; this version posted December 4, 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

Many low- and middle-income countries have experienced large numbers of hospitalisations and deaths related to COVID-19. However, reported levels of illness are not always proportional to the high levels of infection implied by serologic studies, suggesting reduced access to laboratory testing, or differences in transmission or susceptibility to developing SARS-CoV-2-related illness in these settings.[1,2,3] Few detailed SARS-CoV-2 cohort studies are available from middle and low income settings, where vaccination rates remain suboptimal and immunity comes primarily from natural infections, as is the case in South Africa. South Africa has a relatively young population with <5% of the population aged >65 years.[4] In 2017, the national HIV prevalence among individuals of all ages was 14%, with 7,9 million people living with HIV.[5]

Following the initial detection of SARS-CoV-2 in South Africa in March 2020, a hard lockdown was implemented with restrictions on international travel, school closures, halting of non-essential business and confining people to their homes. Subsequent to this hard lockdown there was a progressive relaxation of restrictions, beginning on 1 May 2020.[6] South Africa experienced three SARS-CoV-2 waves through August 2021; the first wave peaking in August 2020, the second, associated with the emergence of the Beta variant of SARS-CoV-2, peaking in January 2021 and the third associated with the Delta variant peaking in June 2021.[7] SARS-CoV-2 restrictions were increased moderately including school closures around the peak of the second and third waves and subsequently relaxed when case numbers decreased. Both Beta and Delta variants have been shown to escape from immunity from previous infection, be more transmissible and may be associated with more severe disease, although the epidemiological consequences of each of these parameters remain debated.[8,9,10,11]

Studies to quantify the burden of asymptomatic infections, symptomatic fraction, duration of shedding and household transmission of SARS-CoV-2 from asymptomatically infected individuals have mostly been conducted as part of outbreak investigations or in specific settings.[12,13,14,15] Comprehensive community studies of asymptomatic infection and robust individual-level epidemiologic data about infection with Beta and Delta variants are limited.

In randomly selected households from a rural and an urban community in South Africa, we estimated the cumulative incidence of SARS-CoV-2 infection using serial real-time reverse transcription PCR (rRT-PCR) and serology. We estimated the symptomatic fraction of SARS-CoV-2

medRxiv preprint doi: https://doi.org/10.1101/2021.07.20.21260855; this version posted December 4, 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, the duration of viral RNA shedding and the household cumulative infection risk (HCIR) from symptomatic and asymptomatic index cases of different ages.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.20.21260855; this version posted December 4, 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

The Prospective Household study of SARS-CoV-2, Influenza and Respiratory Syncytial virus community burden, Transmission dynamics and viral interaction in South Africa (PHIRST-C) was based on a previously conducted study (PHIRST) at the same sites from 2016-2018.[16,17] We implemented a prospective household cohort study in a rural and an urban community of South Africa with twice weekly collection of mid-turbinate nasal swabs, symptom, and health-seeking data and serum collection every two months to measure SARS-CoV-2 antibodies (Supplementary figure 1). The study included 58 weeks of follow-up at the rural site (16 July 2020 through 28 August 2021) and 56 weeks at the urban site (27 July 2020 through 28 August 2021) with seven serum collections at each site. Nasal swab collection began prior to the first wave peak in the district where the rural site was located and during the peak of the wave in the urban site.[18]

The rural site in Mpumalanga Province is nested within a health and socio-demographic surveillance system (HDSS) run by the Medical Research Council/University of Witwatersrand Rural Public Health and Health Transitions Research Unit, Agincourt.[19,20] The urban site, Jouberton Township in Klerksdorp, is located in the North West Province.

We aimed to enroll a total of 1000 individuals of all ages. Assuming an average household size of 5 individuals and loss to follow up of 10%, we planned to enroll approximately 110 households from each site (additional information on the sample size calculation is provided in the supplement). Households were randomly selected, from the HDSS database in the rural site and using Global Positioning System (GPS) coordinates in the urban site. Households with >2 members and where ≥80% of eligible members consented to participate were eligible. Details on household selection, enrolment and data collection are provided in the supplement. In brief, we first approached households previously enrolled in PHIRST, and then prospectively approached new potentially-eligible households using the site-specific sampling frame used for PHIRST until the required number of households were enrolled.

Data collection

We collected individual baseline data, including demographics, HIV status, and history of underlying illness. Study staff visited participating households twice weekly (Monday-Wednesday and Thursday-Saturday) during July 2020-August 2021 to collect mid-turbinate nasal swabs from participants and information about symptoms, absenteeism, and health system contact. Different symptoms were captured among individuals aged <5 years and ≥5 years (Supplementary table 1). Study staff entered data in the field on tablets using REDCap (Research Electronic Data Capture)[21] and had refresher training on specimen collection and the identification of respiratory signs and symptoms each week.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.20.21260855; this version posted December 4, 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 .

Study staff collected blood specimens from participants at enrollment (20 July–17 September 2020, blood draw (BD) 1), and every two months thereafter (21 September–10 October, BD2; 23 November–12 December 2020, BD3; and 25 January–21 February 2021, BD4; 22 March–11 April, BD5; 20 May – 9 June 2021, BD6; 19 July – 5 August 2021, BD7). Serology data from the first 5 blood draws have been published.[22]

Laboratory methods

Nasal specimens were collected using nasopharyngeal nylon flocked swabs, placed in Universal Transport Medium (UTM) and transported daily on ice packs to the National Institute for Communicable Diseases (NICD) in Johannesburg, South Africa, for testing. Nucleic acids were extracted from 200µl of UTM using the Microlab NIMBUS Instrument (Hamilton, Nevada, USA) with the STARMag Universal Cartridge extraction kit (Seegene Inc., Seoul, Korea) according to manufacturer instructions. Specimens were initially tested for the presence of SARS-CoV-2 nucleic acids by rRT-PCR using the Allplex™ 2019-nCoV kit (Seegene Inc., Seoul, Korea) and a BioRad CFX96 thermal cycler, according to manufacturer instructions. From March 2021, samples were tested using the Allplex™ SARS-CoV-2/FluA/FluB/RSV kit (Seegene Inc., Seoul, Korea). A cycle threshold (C_t) value of <40 on ≥1 of 3 SARS-CoV-2 PCR targets (E, N and S or RdRp genes) was considered positive. All specimens testing SARS-CoV-2-rRT-PCR-positive were confirmed by repeat testing of a second aliquot, and PCR testing in duplicate. Specimens testing positive on at least one duplicate were considered positive. If a specimen was confirmed positive after repeat testing, the results [C_t value and targets testing positive] from the first positive test were included in the analysis. A lower Ct-value on rRT-PCR (using the lowest Ct value for any target during the episode) was used as a proxy for higher RNA viral load.[23] All confirmed positive samples were tested to identify variants of concern using the Allplex™ SARS-CoV-2 Variants I assay (Seegene Inc., Seoul, Korea). This assay targets the RdRp gene, HV69/70 deletion, N501Y and E484K mutations, thus identifying the B.1.351/P1 (Beta/Gamma) and B.1.1.7 (Alpha) variants. From May 2021 SARS-CoV-2-positive samples were also tested using the Allplex™ Variants II assay (Seegene Inc., Seoul, Korea) which detects the L452R mutation (Delta) and differentiates Beta (K417N) from Gamma (K417T). SARS-CoV-2 sequencing methods are described in the supplement.

Serologic evidence of SARS-CoV-2 infection was determined using the Roche Elecsys® Anti-SARS-CoV-2 assay (Roche Diagnostics, Rotkreuz, Switzerland), using recombinant protein representing the nucleocapsid (N) antigen. The assay was performed on the Cobas e601 instrument, and a cut-off index of ≥1.0 was considered an indication of infection (seropositivity).

medRxiv preprint doi: https://doi.org/10.1101/2021.07.20.21260855; this version posted December 4, 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 .

HIV status was obtained from patient medical records if a participant reported being HIV-infected, or by nurse-administered rapid HIV test with pre- and post-test counselling for participants with unknown, or self-reported HIV-negative status. Patients newly diagnosed with HIV were referred to the nearest primary health care facility for assessment and initiation of antiretroviral treatment.

Individuals were considered fully vaccinated for SARS-CoV-2 ≥14 days after they had received a single dose of the Johnson and Johnson (J&J) vaccine or two doses of the Pfizer BioNTech (Pfizer) vaccine. They were considered partially vaccinated if they had received any vaccine dose but not meeting the above criteria. Only the J&J and Pfizer vaccines were available during the study period.

Definitions and statistical analyses

We included individuals with ≥10 completed follow-up visits. We defined a SARS-CoV-2 serology-confirmed infected individual as at least one instance of SARS-CoV-2 antibody seropositivity. We defined a SARS-CoV-2 rRT-PCR-confirmed infection episode as at least one nasal swab rRT-PCR positive for SARS-CoV-2. Infection episode duration was estimated from the first to the last day of SARS-CoV-2 rRT-PCR positivity. Details are provided in the supplement (Supplementary figure 2). An illness episode was defined as an episode with ≥1 symptom reported from one visit before, to one visit after the SARS-CoV-2 infection episode. People living with HIV (PLHIV) were deemed to be significantly immunocompromised if their CD4 T-lymphocyte count was <200 cells/µl, and not HIV virally suppressed if their HIV viral load measured ≥400 copies/ml.[24]

An rRT-PCR-confirmed household cluster was composed of all rRT-PCR-confirmed infection episodes within a household within an interval between the rRT-PCR positive tests of any infection episode pairs of ≤14 days (representing ≤2 mean serial intervals).[25] Cluster duration was estimated as the interval from the first day of rRT-PCR positivity of the first individual in a cluster to the last day of rRT-PCR positivity of the last individual in that cluster. HCIR was defined as the proportion of household members with subsequent infection following SARS-CoV-2 introduction and estimated by dividing the number of subsequent individuals with confirmed infection within a household cluster following SARS-CoV-2 introduction by the number of susceptible (no evidence of previous infection on rRT-PCR or serology) household members. The primary/index case was defined as the first individual testing positive within a cluster on rRT-PCR. Households with co-primary index cases (two individuals rRT-PCR positive for the first time on the same visit) were excluded from the analysis of HCIR.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.20.21260855; this version posted December 4, 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 generation interval was calculated as the difference between the date of the first positive rRT-PCR in the index and the secondary infection within an rRT-PCR-confirmed household cluster. Following examination of the distribution of calculated generation intervals, we included all secondary infections with rRT-PCR positivity ≤21 days after the index case onset as potential secondary cases for analysis of factors associated with generation interval. Using these definitions, it was possible for a household to experience >1 cluster of infections. Based on epidemic timing in the two communities, first wave episodes or clusters were defined as having onset before 19 December 2020 at both sites, second wave as having onset before 9 June 2021 in Agincourt and 23 June 2021 in Klerksdorp, and third wave as onset up to 28 August at both sites when intense follow-up ended.
A variant was allocated to each episode of infection according to a hierarchical process that accounted for known lineages (i.e., wild-type, Alpha, Beta or Delta variant) within episode or household clusters or occurrence of the episode in wave 1, 2 or 3 (see supplement for details).
To assess SARS-CoV-2 reinfection, we defined possible reinfection as >28 to 90 days between rRT-PCR -positive specimens (no sequence/variant data available) or between first seropositive specimen and rRT-PCR -positive specimen; probable reinfection as >90 days between rRT-PCR -positive specimens (no sequence/variant data available) or between first seropositive specimen and rRT-PCR -positive specimen; and confirmed reinfection as distinct Nextstrain clades on sequencing or variant PCR between rRT-PCR-positive specimens meeting the temporal criteria for possible or probable.[26]
The proportion of reinfections was calculated as the number of individuals with re-infection divided by the total number of individuals with evidence of prior infection.

For analyses of symptomatic fraction, infection episode duration, HCIR and generation interval we only included incident episodes defined as those occurring with onset >14 days after the start of follow-up. This was because individuals tested positive at the start of follow-up (n=7 and n=32 at the rural and urban site respectively), and we did not know how long they had been shedding SARS-CoV-2, if they had symptoms previously, or who the index case was.

Proportions were compared using the Chi-squared or Fisher's exact test. We used Weibull accelerated failure time regression, for the analysis of factors associated with time-to-event outcomes (duration of shedding and generation interval). We used logistic regression for the analysis of factors associated with binary outcomes (symptomatic vs. asymptomatic, HCIR, index case vs other participants, reinfection). We used Poisson regression for the analysis of factors associated with at least one SARS-CoV-2 infection episode (cumulative incidence) including all individuals with

medRxiv preprint doi: https://doi.org/10.1101/2021.07.20.21260855; this version posted December 4, 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 .

evidence of infection on rRT-PCR and/or serology. For all analyses we accounted for within-household clustering using random effects regression models. For each multivariable model, we considered all a-priori likely biologically associated factors with the outcome of interest for which we had available data. We examined factors associated with several different outcomes, therefore the selected predictors varied across models. For analyses by age group, in each analysis we chose as reference the age group with the lowest prevalence of the outcome of interest.

Pairwise interactions were assessed graphically and by inclusion of product terms for all variables remaining in the final multivariable additive model. We conducted all statistical analyses using STATA version 14.1 (Stata Corp LP, College Station, Texas, USA). For each univariate analysis, we used all available case information. P values <0.05 were considered statistically significant.

Ethics

The PHIRST protocol was approved by the University of Witwatersrand Human Research Ethics Committee (Reference 150808) and amended to include enrollment and testing for COVID-19 on 24 June 2020 and was registered on clinicaltrials.gov on 6 August 2015 and updated on 30 December 2020 (https://clinicaltrials.gov/ct2/show/NCT02519803). Informed consent was obtained from all participants prior to data collection. Participants received grocery store vouchers of USD 3 per visit to compensate for time required for specimen collection and interview.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.20.21260855; this version posted December 4, 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 .

Results

We approached 537 households, of which 236 (52%) agreed to participate in the study, and of these 222 (94%) were included in the analysis. Of the 1,251 eligible household members, 1,200 (96%) were included. Reasons for non-inclusion are shown in Supplementary figure 3. Among the 222 households in the analysis, the median number of household members was 5 (interquartile range (IQR) 4-7), median sleeping rooms was 3 (IQR 2-4), and 49% had at least one child aged <5 years, with a higher proportion among households in the rural community (Supplementary table 2). Individuals from the rural community were younger, had a lower level of formal education and were less likely to be employed. Underlying illness was more common in the urban site, but HIV prevalence was similar between sites (14% in the rural and 17% in the urban site, p=0.173). At the end of follow up 5% of individuals were fully vaccinated againast SARS-CoV-2.

At the start of follow up in July 2020, 1% (5/443) and 15% (73/498) of individuals with available data had serologic evidence of previous SARS-CoV-2 infection at the rural and the urban site, respectively. Of 125,088 potential individual follow-up visits, we collected and tested 115,759 (93%) mid-turbinate nasal swabs, of which 1976 (2%) tested positive for SARS-CoV-2 on rRT-PCR (Figure 1 and Supplementary figure 5 and 6). During the study, 90% (200/222) of households had at least one individual testing SARS-CoV-2 positive on rRT-PCR or serology, with an average of 3.7 (range 1-10) infected individuals per infected household.

During the follow-up period, 62% (749/1200) of individuals experienced at least one episode of SARS-CoV-2 infection on rRT-PCR and/or serology and 12% (87/749) experienced a repeat infection, including one individual who experienced two repeat infections. Of 88 repeat infection episodes, 10 (11%) were classified as possible, 21 (24%) as probable and 57 (65%) as confirmed (Figure 2 and supplement). The highest proportion infected was in the 13-18 years age group (Supplementary figure 4). Repeat infection was more common in individuals aged 13-18 years and in the urban site (Supplementary Table 3).

Among 294 individuals with a positive rRT-PCR during follow-up who had a negative serology preceding the episode and available serology data >14 days after the start of the episode, 267 (91%) seroconverted after the episode. Failure to develop a serologic response was more common in individuals aged <5, 19-39 and 40-59 years (compared to 5-12 years) and episodes with a single rRT-PCR positive swab and with Ct value>30 (Supplementary table 4). Among 447 individuals who were

medRxiv preprint doi: https://doi.org/10.1101/2021.07.20.21260855; this version posted December 4, 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 .

seronegative at baseline and subsequently became seropositive, 404 (90%) had evidence of a rRT-PCR-confirmed infection.

On multivariable analysis, factors associated with SARS-CoV-2 cumulative incidence were age 5-12 and 13-18 years vs <5 years, residing in the urban community vs rural community, and being overweight vs normal weight (Table 1).

Of 718 rRT-PCR-confirmed episodes, 17% (n=124) occurred in the first wave, 55% (n=270) in the second wave and 45% (n=324) in the third wave (Supplementary figure 5 and 6). Proportionately more children and adolescents were infected with successive waves with 43% (53/124) of cases in wave one aged <19 years compared to 49% (133/270) in the second and 66% (213/324) in the third wave respectively (p<0.001). Of 662 rRT-PCR-confirmed episodes, that occurred >14 days after the start of follow-up, 15% (97/662) of individuals reported ≥1 symptom (details in supplement). Among 97 symptomatic individuals, 6 (6%) attended an outpatient clinic, 9 (9%) were hospitalised and 2 (2%) died (infection fatality ratio 0.3% (95% confidence interval (CI) 0.03%-1%)). Among 20 symptomatic individuals who were employed or attended school, 35% (n=7) reported absenteeism. In multivariable analysis symptoms were more common in individuals aged ≥19 years compared to <5 years, in PLHIV with HIV viral load ≥400 copies/ml, in obese individuals and in episodes with minimum Ct value <30 and caused by Delta variant (Table 2). Positive serology before the episode was associated with decreased symptoms.

The mean duration of rRT-PCR positivity was 11.6 days (median 11 days, range 4-137 days) and 21% (138/662) of episodes were rRT-PCR positive at only one visit. On multivariable analysis, symptomatic individuals, PLHIV with HIV viral load ≥400 copies/ml and individuals with minimum Ct value <30 shed SARS-CoV-2 RNA for longer (Table 3). Positive serology before the episode was associated with decreased duration of shedding.

Among 195 households with at least one SARS-CoV-2 infection cluster detected on rRT-PCR, 72 (37%) had one cluster, 83 (43%) had 2 clusters, 31 (16%) had 3 clusters and 9 (5%) had >3 clusters (total of 369 clusters). The average cluster duration among 336 clusters starting >14 days after the start of follow up was 15.6 days (range 4-137 days). We included 184 clusters starting >14 days after the start of follow up and with a single index case from 103 households for analysis of HCIR. In this subset of households the HCIR was 25% (213/856, 95% CI 22%-28%). HCIR was 19% (21/110) for symptomatic and 25% (195/775) for asymptomatic index cases (p=0.165) (Table 4). On multivariable

medRxiv preprint doi: https://doi.org/10.1101/2021.07.20.21260855; this version posted December 4, 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 .

analysis, low Ct value (proxy for high viral RNA load) of the index case, index case female sex, index case age 40-59 years and household contact age between 5 and 18 years, urban community and infection with Beta and Delta variant were associated with increased HCIR.

The mean generation interval was 7.5 days (range 2-21 days) (Supplementary figure 7). On multivariable analysis, generation interval was shorter if the index case was symptomatic and longer if the index case shed for longer or was infected with Alpha, Beta or Delta variant (Supplementary table 4).

medRxiv preprint doi: https://doi.org/10.1101/2021.07.20.21260855; this version posted December 4, 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 .

Discussion

Using intensive systematic repeated sampling among household cohorts in two largely unvaccinated South African communities, we found that by 18 August 2021 almost two-thirds of individuals had been infected with SARS-CoV-2. Approximately 12% of individuals experienced at least one repeat episode of infection within 13 months of follow up. Only 15% of infections were associated with symptoms; of these, 9% were hospitalised and 2% died. SARS-CoV-2 Delta variant infections were more likely to be symptomatic compared to non-variant infection. In households with at least one SARS-CoV-2 infection, SARS-CoV-2 was transmitted to 25% of household contacts irrespective of symptoms in the index case. Index and household contact age, infecting SARS-CoV-2 variant and index case RNA viral load were the main predictors of onward transmission. Household transmission was increased fourfold with the Beta variant and fifteenfold with the Delta variant.

Previous studies have generated wide-ranging estimates of the proportion of SARS-CoV-2 infections which are asymptomatic. A recent systematic review found that 20% (95% CI 3%-67%) of SARS-CoV-2 infections remained asymptomatic throughout infection and that transmission was lower from asymptomatic individuals.[27] We found that 85% of infections were asymptomatic despite active twice-a-week symptom evaluation. Three US-based cohort studies implementing weekly PCR testing for SARS-CoV-2, irrespective of symptoms, found estimates of 31% asymptomatic in a community-based study including children, 35% in pregnant women and 11% among healthcare workers, lower than estimates in our study. [28–30] The lower symptomatic fraction in these studies could be because twice weekly-swabbing allows identification of more transient symptomatic illness or because there could be truly lower symptomatic fraction in the populations included in our study. Compared to other studies, the population included in our study was relatively young with only 9% of individuals aged ≥60 years, reflecting the general South African population.[4] Despite the low overall symptomatic fraction, symptomatic fraction increased with age, in line with prior studies.[15,27]

A systematic review and meta -analysis of 87 household transmission studies of SARS-CoV-2 reported an estimated secondary attack rate of 18.9% (95% CI 16.2-22.0)[31] slightly lower than our estimate of 25%. The review also reported that household secondary attack rates were higher from symptomatic index cases, that adults were more likely than children to acquire infection and that sex was not associated with transmission. Previous studies have found that viral load and duration of SARS-CoV-2 shedding are lower among mild and asymptomatically

medRxiv preprint doi: https://doi.org/10.1101/2021.07.20.21260855; this version posted December 4, 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 .

infected individuals compared to individuals with severe illness.[32] While we did not find that symptom profile was associated with risk of infection or transmission, we did find that higher index case viral RNA load (estimated through the proxy of cycle threshold value) was associated with more frequent transmission. This is similar to a study from Catalonia where viral load was carefully quantified in adult transmitters.[33] We also found females were more likely to transmit SARS-CoV-2, potentially reflecting closer contact with other household members in this population. Several studies have demonstrated increased transmissibility of the Beta and Delta SARS-CoV-2 variants.[9,34] We found that household transmission of SARS-CoV-2 was approximately 4 times higher with Beta and 15 times higher with Delta variant compared to non-variant infections. It is possible that participants changed their behaviour once informed that they were infected with SARS-CoV-2, however data on behaviour following a SARS-CoV-2 diagnosis were not available.

Symptomatic fraction was lowest in children and adolescents aged <18 years. While attack rates were lower in children aged <5 years, they were highest in adolescents aged 13-18 years. We also found that indviduals aged 13-18 years were more likely to acquire infection within the household and more likely to experience reinfection with SARS-CoV-2. Previous studies have found lower attack rates and symptomatic fraction in children (although relatively increased in age group 13-18 years, similar to our finding) and also that children are less likely to transmit SARS-CoV-2 and have reduced susceptibility to infection.[35–37,38,39] The Delta variant has been associated with higher attack rates in children and adolescents in South Africa and elsewhere. In several countries, this has in part been attributed to a shift in the age distribution of cases as vaccination expanded to adult age groups.[40] In South Africa, where vaccination remains low, infection rates were higher among adults in the first two waves, potentially contributing to the higher attack rate in children and adolescents with Delta variant. Differences in circulating variants over time and geography may have contributed to differences in the contribution of adolescents to transmission in previous studies.[41] In addition, most previous studies did not include systematic longitudinal rRT-PCR testing irrespective of symptoms in children, adolescents and adults, potentially biasing against detection of minimally symptomatic infections in children.

Previous studies have found that PLHIV are more likely to be hospitalised and that hospitalised PLHIV without HIV viral suppression shed SARS-CoV-2 at a high viral load for longer, but data from community settings are lacking.[42–44] We found that PLHIV who were not virally suppressed

medRxiv preprint doi: https://doi.org/10.1101/2021.07.20.21260855; this version posted December 4, 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 .

were more likely to develop symptomatic illness when infected with SARS-CoV-2, and shed SARS-CoV-2 for longer when compared to HIV-uninfected individuals, potentially contributing to the evolution of novel variants of SARS-CoV-2.

Our study had several strengths. Participating households were randomly sampled from a rural and an urban South African community and followed up for 13 months through the second half of the first wave and the second and third waves of SARS-CoV-2 infection in South Africa. Participating individuals were sampled twice weekly, irrespective of symptoms, allowing for more accurate ascertainment of the burden of SARS-CoV-2 infections as well as the symptomatic fraction and transmission from asymptomatic individuals. Our unique study design combining frequent rRT-PCR and serological testing allowed for more complete ascertainment of infection burden.

Our study also had several limitations. We included two communities and households with >2 members, potentially limiting generalisability of study findings. The finding of different attack rates in the two communities suggests substantial heterogeneity in disease transmission in different geographic areas. Participants were sampled using mid-turbinate nasal swabs because of potential SARS-CoV-2 transmission risk with a collection of more sensitive nasopharyngeal swab specimens. This could have led to some missed infections. However, the strong association between rRT-PCR- and serology-confirmed infection in individuals with both specimen types available suggests that the majority of infections were detected. Repeated questioning on symptoms twice weekly may be associated with participant fatigue and under-reporting. Participants may have been informed of their SARS-CoV-2 infection status before developing symptoms, potentially affecting reporting. We implemented several measures to reduce this potential bias including weekly retraining of field workers on symptom collection and regular field supervisory visits to evaluate data collection and symptom recording. A study of influenza infection in the same population with similar study design found that 56% of individuals infected with influenza were symptomatic, suggesting the robustness of our data.[17] We did not quantify viral RNA load but instead this was inferred using Ct values as proxy.

In conclusion, we found a high rate of SARS-CoV-2 infection in households in a rural and an urban South African community with the majority of infections being asymptomatic in individuals of all ages. Individuals aged 13-18 years had the highest attack rates and were more likely to acquire infection. The household cumulative infection risk was 25% and did not differ by

medRxiv preprint doi: https://doi.org/10.1101/2021.07.20.21260855; this version posted December 4, 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 .

presence of symptoms in index case but was higher in households where the index case had a higher RNA viral load and with Beta and Delta variant infection. Asymptomatic individuals transmitted SARS-CoV-2 at similar levels as symptomatic individuals suggesting that interventions targeting symptomatic individuals, such as promotion of community testing and contact tracing of individuals tested because they report symptoms, may have limited impact in this setting.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.20.21260855; this version posted December 4, 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 .

Acknowledgements

The authors would like to thank all the individuals who kindly agreed to participate in the study as well as the many field and laboratory staff who worked tirelessly to implement the study. The MRC/Wits Rural Public Health and Health Transitions Research Unit and Agincourt Health and Socio-Demographic Surveillance System, a node of the South African Population Research Infrastructure Network (SAPRIN), is supported by the Department of Science and Innovation, the University of the Witwatersrand, the Medical Research Council, South Africa, and the Wellcome Trust, UK (grants 058893/Z/99/A; 069683/Z/02/Z; 085477/Z/08/Z; 085477/B/08/Z).

Author's contributions

Conception and design of study: CC, JK, AvG, MLM, NW, JM, NAM, KK, LL, ST

Data collection and laboratory processing: CC, JK, AvG, MLM, NW, JNB, JM, MdP, MC, AB, NAM, KK, ST, LL, FW, JdT, FXG, FSD, TMK, ST

Analysis and interpretation: CC, JK, AvG, MLM, NW, JNB, JM, MdP, MC, AB, NAM, KK, STo, LL, FW, JdT, FXG, FSD, TMK, KS, CV, ST

CC, JK, and ST accessed and verified the underlying data. CC and ST drafted the Article. All authors critically reviewed the Article. All authors had access to all the data reported in the study.

Funding. This work was supported by the National Institute for Communicable Diseases of the National Health Laboratory Service and the U.S. Centers for Disease Control and Prevention [co-operative agreement number: 6 U01IP001048-04-02] and Wellcome Trust (grant number 221003/Z/20/Z) in collaboration with the Foreign, Commonwealth and Development Office, United Kingdom.

Disclaimer. The findings and conclusions in this paper are those of the authors and do not necessarily represent the official position of the funding agencies.

Ethics

The PHIRST-C protocol was approved by the University of Witwatersrand Human Research Ethics Committee (Reference 150808) and the U.S. Centers for Disease Control and Prevention's Institutional Review Board relied on the local review (#6840). The protocol was registered on clinicaltrials.gov on 6 August 2015 and updated on 30 December 2020 (https://clinicaltrials.gov/ct2/show/NCT02519803).

medRxiv preprint doi: https://doi.org/10.1101/2021.07.20.21260855; this version posted December 4, 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 .

Potential conflicts of interest

CC has received grant support from Sanofi Pasteur, Advanced Vaccine Initiative, and payment of travel costs from Parexel. AvG has received grant support from Sanofi Pasteur, Pfizer related to pneumococcal vaccine, CDC and the Bill & Melinda Gates Foundation. NW report grants from Sanofi Pasteur and the Bill & Melinda Gates Foundation. NAM has received a grant to his institution from Pfizer to conduct research in patients with pneumonia and from Roche to collect specimens to assess a novel TB assay.

The full study protocol can be found on the National Institute for Communicable Diseases Website at the following link:

https://www.nicd.ac.za/centres/centre-for-respiratory-disease-and-meningitis/

medRxiv preprint doi: https://doi.org/10.1101/2021.07.20.21260855; this version posted December 4, 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 .

References

1. Hsiao M, Davies MA, Kalk E, et al. SARS-COV-2 SEROPREVALENCE IN THE CAPE TOWN METROPOLITAN SUB-DISTRICTS AFTER THE PEAK OF INFECTIONS. *NICD Public Heal Surveill Bull*. 2020;18(Supplementary Issue 5):2-9.
2. Arora RK, Joseph A, Van Wyk J, et al. SeroTracker: a global SARS-CoV-2 seroprevalence dashboard. *Lancet Infect Dis*. August 2020. doi:10.1016/S1473-3099(20)30631-9
3. Hallal PC, Hartwig FP, Horta BL, et al. SARS-CoV-2 antibody prevalence in Brazil: results from two successive nationwide serological household surveys. *Lancet Glob Heal*. 2020;8(11):e1390-e1398. doi:10.1016/S2214-109X(20)30387-9
4. Statistics South Africa, StatsSA. Statistical Release P0302: Mid-year population estimates 2020. *Stats SA*. 2020;(July):1-22. http://www.statssa.gov.za/publications/P0302/P03022017.pdf.
5. Simbayi L, Zuma K, Zungu N, et al. *South African National HIV Prevalence, Incidence, Behaviour and Communication Survey, 2017*. Cape Town, South Africa; 2017.
6. Government of South Africa. Regulations and guidelines – coronavirus COVID-19. https://www.gov.za/covid-19/resources/regulations-and-guidelines-coronavirus-covid-19#. Published 2021.
7. Tegally H, Wilkinson E, Giovanetti M, et al. Emergence of a SARS-CoV-2 variant of concern with mutations in spike glycoprotein. *Nature*. March 2021. doi:10.1038/s41586-021-03402-9
8. Wibmer CK, Ayres F, Hermanus T, et al. SARS-CoV-2 501Y.V2 escapes neutralization by South African COVID-19 donor plasma. *Nat Med*. March 2021. doi:10.1038/s41591-021-01285-x
9. Tao K, Tzou PL, Nouhin J, et al. The biological and clinical significance of emerging SARS-CoV-2 variants. *Nat Rev Genet*. 2021;22(12):757-773. doi:10.1038/s41576-021-00408-x
10. Pearson C, Russell T, Davies N, et al. *Estimates of Severity and Transmissibility of Novel South Africa SARS-CoV-2 Variant 501Y.V2.*; 2021. https://cmmid.github.io/topics/covid19/sa-novel-variant.html.
11. Jassat W, Mudara C, Ozougwu L, Tempia S, Blumberg LH. Increased mortality among individuals hospitalised with COVID-19 during the second wave in South Africa. *MedXRiV*. 2021.
12. Lei H, Xu X, Xiao S, Wu X, Shu Y. Household transmission of COVID-19-a systematic review and meta-analysis. *J Infect*. 2020;81(6):979-997. doi:10.1016/j.jinf.2020.08.033
13. Meyerowitz EA, Richterman A, Gandhi RT, Sax PE. Transmission of SARS-CoV-2: A Review of Viral, Host, and Environmental Factors. *Ann Intern Med*. 2021;174(1):69-79. doi:10.7326/M20-5008

medRxiv preprint doi: https://doi.org/10.1101/2021.07.20.21260855; this version posted December 4, 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 .

14. Goldstein E, Lipsitch M, Cevik M. On the Effect of Age on the Transmission of SARS-CoV-2 in Households, Schools, and the Community. *J Infect Dis*. 2021;223(3):362-369. doi:10.1093/infdis/jiaa691

15. Poletti P, Tirani M, Cereda D, et al. Association of Age With Likelihood of Developing Symptoms and Critical Disease Among Close Contacts Exposed to Patients With Confirmed SARS-CoV-2 Infection in Italy. *JAMA Netw open*. 2021;4(3):e211085. doi:10.1001/jamanetworkopen.2021.1085

16. Cohen C, McMorrow ML, Martinson NA, et al. Cohort profile: A Prospective Household cohort study of Influenza, Respiratory syncytial virus and other respiratory pathogens community burden and Transmission dynamics in South Africa, 2016-2018. *Influenza Other Respi Viruses*. 2021;15(6):789-803. doi:10.1111/irv.12881

17. Cohen C, Kleynhans J, Moyes J, McMorrow ML, Treurnicht FK. Asymptomatic transmission and high community burden of seasonal influenza in an urban and a rural community in South Africa, 2017–18 (PHIRST): a population cohort study. *Lancet Glob Heal*. 2021;In press.

18. Kleynhans J, Wolter N, Tempia S, von Gottberg A, Bhiman JN. SARS-CoV-2 seroprevalence in a rural and urban community in South Africa, July 2020-February 2021: Implications for infection to hospitalisation and mortality ratios. *Submitted*. 2021.

19. Kahn K, Collinson MA, Xavier Gómez-olivé F, et al. Profile: Agincourt health and socio-demographic surveillance system. *Int J Epidemiol*. 2012;41(4):988-1001. doi:10.1093/ije/dys115

20. Garenne M, Collinson MA, Kabudula CW, Gómez-Olivé FX, Kahn K, Tollman S. Completeness of birth and death registration in a rural area of South Africa: the Agincourt health and demographic surveillance, 1992-2014. *Glob Health Action*. 2016;9(1):32795. doi:10.3402/gha.v9.32795

21. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: Building an international community of software platform partners. *J Biomed Inform*. 2019;95:103208. doi:10.1016/j.jbi.2019.103208

22. Kleynhans J, Tempia S, Wolter N, et al. SARS-CoV-2 Seroprevalence in a Rural and Urban Household Cohort during First and Second Waves of Infections, South Africa, July 2020-March 2021. *Emerg Infect Dis*. 2021;27(12). doi:10.3201/eid2712.211465

23. Rao SN, Manissero D, Steele VR, Pareja J. A Systematic Review of the Clinical Utility of Cycle Threshold Values in the Context of COVID-19. *Infect Dis Ther*. 2020;9(3):573-586. doi:10.1007/s40121-020-00324-3

24. World Health Organization. WHO Case Definitions of HIV. *World Heal Organ*. 2007.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.20.21260855; this version posted December 4, 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 .

25. Alene M, Yismaw L, Assemie MA, Ketema DB, Gietaneh W, Birhan TY. Serial interval and incubation period of COVID-19: a systematic review and meta-analysis. *BMC Infect Dis*. 2021;21(1):257. doi:10.1186/s12879-021-05950-x

26. Hadfield J, Megill C, Bell SM, et al. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics*. 2018;34(23):4121-4123. doi:10.1093/bioinformatics/bty407

27. Buitrago-Garcia D, Egli-Gany D, Counotte MJ, et al. Occurrence and transmission potential of asymptomatic and presymptomatic SARS-CoV-2 infections: A living systematic review and meta-analysis. *PLoS Med*. 2020;17(9):e1003346. doi:10.1371/journal.pmed.1003346

28. Thompson MG, Burgess JL, Naleway AL, et al. Prevention and Attenuation of Covid-19 with the BNT162b2 and mRNA-1273 Vaccines. *N Engl J Med*. 2021;385(4):320-329. doi:10.1056/NEJMoa2107058

29. Dawood FS, Porucznik CA, Veguilla V, et al. Incidence Rates, Household Infection Risk, and Clinical Characteristics of SARS-CoV-2 Infection Among Children and Adults in Utah and New York City, New York. *JAMA Pediatr*. October 2021. doi:10.1001/jamapediatrics.2021.4217

30. Dawood FS, Varner M, Tita A, et al. Incidence, Clinical Characteristics, and Risk Factors of SARS-CoV-2 Infection among Pregnant Individuals in the United States. *Clin Infect Dis*. August 2021. doi:10.1093/cid/ciab713

31. Madewell ZJ, Yang Y, Longini IM, Halloran ME, Dean NE. Factors Associated With Household Transmission of SARS-CoV-2: An Updated Systematic Review and Meta-analysis. *JAMA Netw open*. 2021;4(8):e2122240. doi:10.1001/jamanetworkopen.2021.22240

32. Liu Y, Yan L-M, Wan L, et al. Viral dynamics in mild and severe cases of COVID-19. *Lancet Infect Dis*. 2020;20(6):656-657. doi:10.1016/S1473-3099(20)30232-2

33. Marks M, Millat-Martinez P, Ouchi D, et al. Transmission of COVID-19 in 282 clusters in Catalonia, Spain: a cohort study. *Lancet Infect Dis*. February 2021. doi:10.1016/S1473-3099(20)30985-3

34. Allen H, Vusirikala A, Flannagan J, et al. Household transmission of COVID-19 cases associated with SARS-CoV-2 delta variant (B.1.617.2): national case-control study. *Lancet Reg Heal Eur*. October 2021:100252. doi:10.1016/j.lanepe.2021.100252

35. Zhu Y, Bloxham CJ, Hulme KD, et al. A meta-analysis on the role of children in SARS-CoV-2 in household transmission clusters. *Clin Infect Dis*. December 2020. doi:10.1093/cid/ciaa1825

36. Goldstein E, Lipsitch M, Cevik M. On the Effect of Age on the Transmission of SARS-CoV-2 in Households, Schools, and the Community. *J Infect Dis*. 2021;223(3):362-369. doi:10.1093/infdis/jiaa691

37. Viner RM, Mytton OT, Bonell C, et al. Susceptibility to SARS-CoV-2 Infection Among Children

medRxiv preprint doi: https://doi.org/10.1101/2021.07.20.21260855; this version posted December 4, 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 .

and Adolescents Compared With Adults: A Systematic Review and Meta-analysis. *JAMA Pediatr*. 2021;175(2):143-156. doi:10.1001/jamapediatrics.2020.4573

38. Tönshoff B, Müller B, Elling R, et al. Prevalence of SARS-CoV-2 Infection in Children and Their Parents in Southwest Germany. *JAMA Pediatr*. January 2021. doi:10.1001/jamapediatrics.2021.0001

39. Madewell ZJ, Yang Y, Longini IM, Halloran ME, Dean NE. Household Transmission of SARS-CoV-2: A Systematic Review and Meta-analysis. *JAMA Netw open*. 2020;3(12):e2031756. doi:10.1001/jamanetworkopen.2020.31756

40. Twohig KA, Nyberg T, Zaidi A, et al. Hospital admission and emergency care attendance risk for SARS-CoV-2 delta (B.1.617.2) compared with alpha (B.1.1.7) variants of concern: a cohort study. *Lancet Infect Dis*. August 2021. doi:10.1016/S1473-3099(21)00475-8

41. The National Institute for Communicable Diseases of the National Health Laboratory S. Epidemiology and clinical characteristics of laboratory-confirmed COVID-19 among individuals aged <=19 years, South Africa, 1 March 2020-28 August 2021. *Natl Inst Commun Dis*. 2021. https://www.nicd.ac.za/wp-content/uploads/2021/09/COVID-19-in-children-surveillance-report.pdf.

42. Meiring S, Tempia S, Bhiman JN, Buys A. Prolonged shedding of SARS-CoV-2 at low real-time RT-PCR cycle-threshold values amongst a subset of hospitalised immunocompromised persons living with HIV (PLHIV), South Africa. *Submitted*. 2021.

43. Boulle A, Davies M-A, Hussey H, et al. Risk factors for COVID-19 death in a population cohort study from the Western Cape Province, South Africa. *Clin Infect Dis*. August 2020. doi:10.1093/cid/ciaa1198

44. Jassat W, Mudara C, Ozougwu L, et al. Difference in mortality among individuals admitted to hospital with COVID-19 during the first and second waves in South Africa: a cohort study. *Lancet Glob Heal*. 2021;9(9):e1216-e1225. doi:10.1016/S2214-109X(21)00289-8

medRxiv preprint doi: https://doi.org/10.1101/2021.07.20.21260855; this version posted December 4, 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 .

Tables and Figures

Table 1: Factors associated with cumulative incidence of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection on real-time reverse transcription polymerase chain reaction (rRT-PCR) and/or serology in a rural and an urban community, South Africa, 2020-2021

Variable		SARS-CoV-2 infection n/N (%)	Univariate RR[e] (95% CI)	Multivariable aRR[e] (95% CI)
Site	Rural	368/643 (57)	Reference	Reference
	Urban	381/557 (68)	1.9 (1.2-3.0)	1.7 (1.1-2.7)
Age group (years)	<5	75/154 (49)	Reference	Reference
	5-12	205/340 (60)	2.0 (1.2-3.2)	1.8 (1.1-2.9)
	13-18	132/170 (78)	5.1 (2.8-9.2)	4.4 (2.4-8.1)
	19-39	165/265 (62)	1.9 (1.1-3.0)	1.5 (0.9-2.6)
	40-59	115/168 (68)	2.7 (1.5-4.7)	2.1 (0.9-4.2)
	≥60	57/103 (55)	1.5 (0.8-2.8)	1.2 (0.6-2.4)
Sex	Female	454/717 (63)	1.1 (0.8-1.5)	
	Male	295/483 (61)	Reference	
HIV and viral load copies/ml[a]	Uninfected	608/971 (63)	Reference	
	Infected <400	87/136 (64)	1.1 (0.7-1.7)	
	Infected ≥400	22/31 (71)	1.5 (0.6-3.8)	
	HIV and/or viral load unknown	32/62 (52)	0.5 (0.3-1.1)	
HIV and CD4+ T cell count/ml[b]	Uninfected	608/971 (63)	Reference	
	Infected ≥200	99/151 (66)	1.1 (0.7-1.7)	
	Infected <200	8/14 (57)	1.0 (0.3-3.7)	
	HIV and/or CD4+ unknown	34/64 (53)	0.6 (0.3-1.2)	
Other underlying illness[c]	Absent	672/1075 (63)	Reference	
	Present	77/125 (62)	0.8 (0.5-1.3)	
BMI[d]	Underweight	55/85 (65)	1.3 (0.7-2.3)	1.1 (0.6-2.1)
	Normal weight	371/642 (58)	Reference	Reference
	Overweight	150/219 (68)	1.7 (1.1-2.5)	1.6 (1.1-2.6)
	Obese	171/252 (68)	1.5 (1.0-2.1)	1.4 (0.9-2.2)
Number of individuals in household	3-5	311/511 (61)	Reference	
	6-10	369/571 (65)	1.3 (0.8-2.1)	
	≥11	69/118 (58)	1.0 (0.3-2.9)	
Crowding (>2 people/sleeping room)	No	381/640 (60)	Reference	
	Yes	368/560 (66)	1.5 (0.9-2.4)	

HIV – Human immunodeficiency virus, BMI – Body mass index, RR – relative risk

Additional variables evaluated but not found to be significant on univariate or multivariable analysis: use of alcohol, current or previous smoking, current or previous tuberculosis, household income, fuel used for cooking, main water source.

[a]HIV data available for 1147 of 1200 (96%) individuals. Among 176 people living with HIV (PLHIV), 167 (95%) had available data on HIV viral load [b]Among 176 PLHIV, 165 (93%) had available data on CD4+ T cell count [c]Self-reported history of asthma, lung disease, heart disease, stroke, spinal cord injury, epilepsy, organ transplant, immunosuppressive therapy, organ transplantation, cancer, liver disease, renal disease or diabetes [d]BMI=body mass index calculated using the formula (weight in

medRxiv preprint doi: https://doi.org/10.1101/2021.07.20.21260855; this version posted December 4, 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 .

kilograms)/(height in metres squared). We defined BMI categories as follows: underweight - age <18 years weight for age or BMI <-2 standard deviations of the World Health Organization (WHO) Child Growth Standards, age ≥18 years BMI <18.5kg/m2; overweight - age <18 years BMI >+1 and ≤+2 standard deviations of the WHO growth standards, age ≥18 years BMI ≥25 and <30kg/m2, obese – age <18 years BMI >+2 standard deviations of the WHO growth standards, age ≥18 years BMI ≥30 kg/m2 [e]Estimated using Poisson regression adjusted for clustering by site and household

medRxiv preprint doi: https://doi.org/10.1101/2021.07.20.21260855; this version posted December 4, 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 .

Table 2: Factors associated with symptomatic illness[a] among real-time reverse transcription polymerase chain reaction (rRT-PCR)-confirmed severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)-infected individuals in a rural and an urban community, South Africa, 2020-2021

Variable		Symptomatic illness n/N (%)	Univariate OR[d] (95% CI)	Multivariable aOR[d] (95% CI)
Age group (years)	<5	6/65 (9)	1.6 (0.6-4.5)	2.0 (0.7-5.8)
	5-12	11/184 (6)	Reference	Reference
	13-18	11/124 (9)	1.5 (0.6-3.6)	1.9 (0.8-4.7)
	19-39	26/152 (17)	3.2 (1.5-6.8)	4.1 (1.9-8.9)
	40-59	24/83 (30)	6.4 (3.0-13.9)	6.8 (2.8-16.3)
	≥60	19/54 (35)	8.5 (3.7-19.5)	12.5 (4.7-33.4)
Sex	Female	65/403 (16)	1.3 (0.8-2.1)	
	Male	32/259 (12)	Reference	
HIV and viral load copies/ml[e]	Uninfected	75/535 (14)	Reference	Reference
	Infected <400	9/68 (13)	1.0 (0.5-2.0)	0.5 (0.2-1.1)
	Infected ≥400	10/25 (40)	4.2 (1.8-9.6)	3.3 (1.3-8.4)
	HIV and/or viral load unknown	3/34 (9)	0.6 (0.2-2.0)	0.5 (0.1-1.7)
HIV and CD4+ T cell count/ml[e]	Uninfected	75/535 (14)	Reference	
	Infected ≥200	16/83 (19)	1.5 (0.8-2.7)	
	Infected <200	3/8 (38)	3.7 (0.9-16.0)	
	HIV and/or CD4+ unknown	3/36 (8)	0.6 (0.2-1.9)	
Other underlying illness[b]	Absent	78/602 (13)	Reference	
	Present	19/60 (32)	3.1 (1.6-5.3)	
BMI[c]	Underweight	4/53 (8)	0.7 (0.3-2.2)	0.6 (0.2-2.0)
	Normal weight	34/341 (10)	Reference	Reference
	Overweight	22/124 (18)	1.9 (1.1-3.5)	1.1 (0.6-2.3)
	Obese	37/141 (26)	3.2 (1.9-5.4)	2.2 (1.2-4.1)
Duration of viral RNA shedding (days)	≤4	10/138 (7)	Reference	
	>4	87/524 (17)	2.5 (1.3-5.0)	
Minimum Ct value	≤30	89/547 (16)	2.6 (1.2-5.5)	2.6 (1.1-5.8)
	>30	8/115 (7)	Reference	Reference
Seropositive before the episode	No	89/552 (16)	Reference	Reference
	Yes	8/102 (8)	0.4 (0.2-0.9)	0.4 (0.2-0.9)
SARS-CoV-2 vaccine status	Unvaccinated	82/609 (13)	Reference	Reference
	Partially vaccinated	4/9 (44)	5.1 (1.4-19.5)	0.9 (0.2-4.1)
	Fully vaccinated	11/44 (25)	2.1 (1.1-4.4)	0.6 (0.2-1.5)
Epidemic wave	1	7/69 (10)	Reference	
	2	43/269 (16)	1.7 (0.7-3.9)	
	3	47/324 (15)	1.5 (0.6-3.4)	
Variant	Wild type	7/67 (10)	Reference	Reference
	Beta	43/263 (16)	1.7 (0.7-3.9)	1.8 (0.7-4.6)
	Alpha	0/7 (0)	Not estimated	Not estimated
	Delta	47/325 (14)	1.4 (0.1-3.3)	2.6 (1.1-6.6)

OR – Odds ratio, a – adjusted, CI – confidence interval, HIV – Human immunodeficiency virus, Ct – cycle threshold, BMI – Body mass index, n- number, NE – not estimated

medRxiv preprint doi: https://doi.org/10.1101/2021.07.20.21260855; this version posted December 4, 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 .

[a]≥1 symptom vs no symptom reported, analysis restricted to 662 episodes of PCR confirmed infection with onset >14 days after the start of follow up [b]Self-reported history of asthma, lung disease, heart disease, stroke, spinal cord injury, epilepsy, organ transplant, immunosuppressive therapy, organ transplantation, cancer, liver disease, renal disease or diabetes [c]BMI=body mass index calculated using the formula (weight in kilograms)/(height in metres squared). We defined BMI categories as follows: underweight - age <18 years weight for age or BMI <-2 standard deviations of the World Health Organization (WHO) Child Growth Standards, age ≥18 years BMI <18.5kg/m2; overweight - age <18 years BMI >+1 and ≤+2 standard deviations of the WHO growth standards, age ≥18 years BMI ≥25 and <30kg/m2, obese – age <18 years BMI >+2 standard deviations of the WHO growth standards, age ≥18 years BMI ≥30 kg/m2 [d]Estimated using logistic regression adjusted for clustering by site and household [e]HIV data available for 629 of 662 (95%) individuals eligible for this analysis. Among 94 PLHIV eligible for this analysis, 93 (99%) had available data on HIV viral load and 91 (97%) had available data on CD4+ T cell count.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.20.21260855; this version posted December 4, 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 .

Table 3: Factors associated with duration of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) viral RNA positivity in a rural and an urban community, South Africa, 2020-2021[a]

		Viral RNA shedding duration (days)	Univariate	Multivariable
Variable		**Mean±SD (Range)**	**HR[c] (95% CI)**	**aHR[c] (95% CI)**
Age group (years)	<5	11.6±6.8 (4-35)	Reference	
	5-12	10.9±7.4 (4-52)	1.1 (0.8-1.4)	
	13-18	11.5±7.8 (4-60)	1.0 (0.7-1.3)	
	19-39	11.8±12.7 (4-137)	0.8 (0.6-0.9)	
	40-59	11.8±7.1 (4-43)	1.0 (0.7-1.3)	
	≥60	13.2±9.0 (4-52)	0.7 (0.5-0.9)	
Sex	Female	12.2±10.1 (4-137)	0.8 (0.7-0.9)	
	Male	10.6±6.7 (4-52)	Reference	
HIV and viral load copies/ml[d]	Uninfected	11.3±7.5 (4-60)	Reference	Reference
	Infected <400	11.7±7.3 (4-52)	1.0 (0.7-1.2)	0.9 (0.7-1.1)
	Infected ≥400	18.5±26.6 (4-137)	0.3 (0.2-0.5)	0.4 (0.3-0.6)
	HIV and/or viral load unknown	8.9±5.6 (4-33)	1.5 (1.1-2.2)	1.3 (0.9-1.9)
HIV and CD4+ T cell count/ml[d]	Uninfected	11.3±7.5 (4-60)	Reference	
	Infected ≥200	11.9±8.4 (4-52)	0.9 (0.7-1.1)	
	Infected <200	30.2±43.5 (4-137)	0.1 (0.1-0.3)	
	HIV and/or CD4+ unknown	9.2 (4-33)	1.5 (1.1-2.1)	
Other underlying illness	Absent	11.6±9.1 (4-137)	Reference	
	Present	12.0±7.0 (4-34)	1.0 (0.8-1.4)	
BMI[b]	Underweight	12.9±10.8 (4-60)	0.8 (0.6-1.1)	
	Normal weight	11.2±9.6 (4-137)	Reference	
	Overweight	12.3±7.6 (4-46)	1.0 (0.8-1.2)	
	Obese	11.6±7.5 (4-43)	1.0 (0.8-1.2)	
Symptoms	Absent	11.0±7.5 (4-60)	Reference	Reference
	Present	14.8±14.7 (4-137)	0.6 (0.5-0.7)	0.7 (0.6-0.9)
Minimum Ct value	≤30	12.8±9.3 (4-137)	0.3 (0.3-0.4)	0.3 (0.3-0.4)
	>30	6.0±4.3 (4-28)	Reference	Reference
Seropositive before episode	No	12.3±9.3 (4-137)	Reference	Reference
	Yes	7.9±5.5 (4-36)	2.0 (1.6-2.5)	1.3 (1.1-1.7)
SARS-CoV-2 vaccine status	Unvaccinated	11.5±9.1 (4-137)	Reference	
	Partially vaccinated	15.3±7.9 (4-29)	0.7 (0.4-1.4)	
	Fully vaccinated	12.1±6.7 (4-31)	1.0 (0.7-1.3)	
Variant	Wild type	10.8±7.5 (4-43)	Reference	Reference
	Beta	13.0±11.5 (4-137)	0.7 (0.5-0.9)	1.0 (0.5-2.2)
	Alpha	12.0±7.7 (4-22)	0.9 (0.4-1.9)	0.8 (0.6-1.1)
	Delta	10.6±6.5 (4-52)	1.1 (0.8-1.4)	1.1 (0.8-1.5)

SD – Standard deviation, HR – Hazard ratio, CI – confidence interval, Ct – cycle threshold, HIV – Human immunodeficiency virus, BMI – Body mass index, Ct – cycle threshold, rRT-PCR – real-time reverse transcription polymerase chain reaction

[a]Estimated using Weibull accelerated failure time regression adjusted for clustering by site and household. Hazard ration <1 corresponds to prolonged duration of viral RNA shedding. Samples were collected at 3 to 4 day intervals hence values of 4 days represent a single positive swab. Analysis restricted to 662 episodes of PCR confirmed infection with onset >14 days after the start of

medRxiv preprint doi: https://doi.org/10.1101/2021.07.20.21260855; this version posted December 4, 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 .

follow up [b]BMI=body mass index calculated using the formula (weight in kilograms)/(height in metres squared). We defined BMI categories as follows: underweight - age <18 years weight for age or BMI <-2 standard deviations of the World Health Organization (WHO) Child Growth Standards, age ≥18 years BMI <18.5kg/m2; overweight - age <18 years BMI >+1 and ≤+2 standard deviations of the WHO growth standards, age ≥18 years BMI ≥25 and <30kg/m2, obese – age <18 years BMI >+2 standard deviations of the WHO growth standards, age ≥18 years BMI ≥30 kg/m2 [c]Estimated using Weibull accelerated failure time regression adjusted for clustering by site and household [d]HIV data available for 629 of 662 (95%) of individuals eligible for this analysis. Among 94 PLHIV eligible for this analysis, 93 (99%) had available data on HIV viral load and 91 (97%) had available data on CD4+ T cell count.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.20.21260855; this version posted December 4, 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 .

Table 4: Factors associated with household cumulative infection risk (HCIR)[a] in a rural and an urban community, South Africa, 2020-2021

Variable		HCIR n/N (%)	Univariate OR[e] (95% CI)	Multivariable aOR[e] (95% CI)
Characteristics of the index case				
Age group (years)	<5	18/54 (33)	7.5 (2.0-27.9)	5.2 (0.3-21.7)
	5-12	47/187 (25)	3.0 (1.1-8.9)	1.8 (0.6-6.1)
	13-18	50/182 (27)	2.5 (0.9-7.5)	1.8 (0.6-5.8)
	19-39	48/247 (19)	2.4 (0.8-6.9)	2.2 (0.7-7.1)
	40-59	42/117 (36)	6.5 (2.0-20.9)	4.6 (1.3-16.3)
	≥60	8/69 (12)	Reference	Reference
Sex	Female	141/505 (28)	2.0 (1.2-3.3)	2.1 (1.2-3.6)
	Male	72/351 (21)	Reference	Reference
HIV and viral load copies/ml[c]	Uninfected	169/683 (25)	Reference	
	Infected <400	32/107 (30)	1.1 (0.6-2.2)	
	Infected ≥400	5/27 (19)	0.7 (0.2-2.8)	
	HIV and/or viral load unknown	7/39 (18)	0.9 (0.3-3.2)	
HIV and CD4+ T cell count/ml[c]	Uninfected	169/683 (25)	Reference	
	Infected ≥200	31/115 (27)	1.0 (0.5-1.9)	
	Infected <200	5/15 (33)	1.2 (0.2-6.3)	
	HIV and/or CD4+ unknown	8/43 (19)	0.9 (0.3-3.0)	
BMI[b]	Underweight	19/66 (29)	1.0 (0.4-2.4)	
	Normal weight	101/431 (23)	Reference	
	Overweight	38/173 (22)	0.9 (0.5-1.8)	
	Obese	55/186 (30)	1.9 (1.1-3.4)	
Symptoms	Absent	195/775 (25)	Reference	
	Present	21/110 (19)	0.9 (0.4-1.7)	
Duration of viral RNA shedding (days)	≤4	16/171 (9)	Reference	
	>4	200/714 (28)	4.3 (2.2-8.6)	
Minimum Ct value	≤30	200/720 (28)	3.6 (1.7-7.6)	3.9 (1.7-8.8)
	>30	16/146 (11)	Reference	Reference
Epidemic wave	1	17/162 (10)	Reference	
	2	74/350 (21)	2.4 (1.1-5.2)	
	3	125/373 (34)	8.9 (4.0-19.9)	
Variant	Wild type	16/153 (10)	Reference	Reference
	Beta	81/366 (22)	14.6 (0.6-352.3)	4.2 (1.7-10.1)
	Alpha	4/7 (57)	2.7 (1.2-6.0)	14.9 (0.7-324.7)
	Delta	115/359 (32)	9.5 (4.1-21.8)	14.6 (5.7-37.5)
Site	Rural	106/517 (21)	Reference	Reference
	Urban	107/339 (32)	1.6 (0.9-2.9)	2.0 (1.1-3.9)
Characteristics of the household contact				
Age group (years)	<5	26/112 (23)	1.4 (0.7-2.8)	1.3 (0.6-2.7)
	5-12	73/277 (26)	1.9 (1.1-3.5)	2.0 (1.1-3.8)
	13-18	39/118 (33)	2.8 (1.4-5.5)	3.2 (1.5-6.8)
	19-39	31/170 (18)	Reference	Reference

medRxiv preprint doi: https://doi.org/10.1101/2021.07.20.21260855; this version posted December 4, 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 .

	40-59	25/97 (26)	1.7 (0.8-3.5)	2.1 (0.9-4.7)
	≥60	19/82 (23)	1.4 (0.6-3.1)	1.6 (0.7-3.8)
Sex	Female	130/503 (26)	1.3 (0.9-1.9)	
	Male	83/353 (24)	Reference	
HIV and viral load copies/ml[d]	Uninfected	179/711 (25)	Reference	
	Infected <400	17/83 (20)	0.8 (0.4-1.5)	
	Infected ≥400	9/22 (41)	1.9 (0.6-5.9)	
	HIV and/or viral load unknown	8/40 (20)	0.6 (0.2-1.6)	
HIV and CD4+ T cell count/ml[d]	Uninfected	179/711 (25)	Reference	
	Infected ≥200	25/97 (26)	1.0 (0.5-1.8)	
	Infected <200	1/8 (13)	0.6 (0.1-6.5)	
	HIV and/or CD4+ unknown	8/40 (20)	0.6 (0.2-1.6)	
Other underlying illness	Absent	194/765 (25)	Reference	
	Present	19/91 (21)	0.7 (0.4-1.3)	

CI - confidence interval, HIV – Human immunodeficiency virus, BMI – Body mass index, Ct – Cycle threshold, rRT-PCR – real-time reverse transcription polymerase chain reaction

Additional factors evaluated but not found to be statistically significant include underlying tuberculosis of index or household contact, BMI of contact, alcohol or smoking of index or contact, SARS-CoV-2 vaccination status of index or contact, fuel used for cooking, water source for handwashing, number of people in household and crowding.

[a]HCIR was defined as the probability of secondary infections within a household following SARS-CoV-2 introduction and estimated by dividing the number of subsequent individuals with confirmed infection within a household cluster following SARS-CoV-2 introduction by the number of susceptible (no previous infection on rRT-PCR or serology) household members. An rRT-PCR-confirmed household cluster was composed of all rRT-PCR-confirmed infection episodes within a household within an interval between the rRT-PCR positive tests of any infection episode pairs of ≤14 days. [b]BMI=body mass index calculated using the formula (weight in kilograms)/(height in metres squared). We defined BMI categories as follows: underweight - age <18 years weight for age or BMI <-2 standard deviations of the World Health Organization (WHO) Child Growth Standards, age ≥18 years BMI <18.5kg/m2; overweight - age <18 years BMI >+1 and ≤+2 standard deviations of the WHO growth standards, age ≥18 years BMI ≥25 and <30kg/m2, obese – age <18 years BMI >+2 standard deviations of the WHO growth standards, age ≥18 years BMI ≥30 kg/m2 [c]HIV data available for 1401 of 1730 (81%) of index case individuals eligible for this analysis. Among 226 index case PLHIV eligible for this analysis, 224 (99%) had available data on HIV viral load and 219 (97%) had available data on CD4+ T cell count [d]HIV data available for 1670 of 1730 (97%) of household contact individuals eligible for this analysis. Among 228 household contact PLHIV eligible for this analysis, 217 (95%) had available data on HIV viral load and 214 (94%) had available data on CD4+ T cell count [e]Estimated using logistic regression adjusted for clustering by site and household

Figure 1: Top panel: Percentage testing real-time reverse transcription polymerase chain reaction (rRT-PCR)-positive per study visit and cumulative percentage with evidence of infection (attack rate) on serology only and on rRT-PCR and serology combined, a rural site and an urban site, South Africa, 2020-2021. Middle panel: Results of serology and rRT-PCR of individuals enrolled in the PHIRST-C study, a rural site and an urban site, South Africa, 2020-2021. Columns are individual follow up visits and rows are individual participants. Individuals within the same household are numbered consecutively (appear below one another). Follow up visits are coloured white if no sample was tested, light pink if the sample tested negative for SARS-CoV-2 and coloured red if the nasopharyngeal swab tested positive for SARS-CoV-2. Cells at the time of serology blood draws are coloured according to the results of serology as follows: light blue - serology negative, dark blue - serology positive. Bottom panel: Percent of rRT-PCR-positive samples typed as wild type (WT) or variant of concern (Alpha, Beta or Delta) by follow-up visit (10 visit moving average used for smoothing).

medRxiv preprint doi: https://doi.org/10.1101/2021.07.20.21260855; this version posted December 4, 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/10.1101/2021.07.20.21260855; this version posted December 4, 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 .

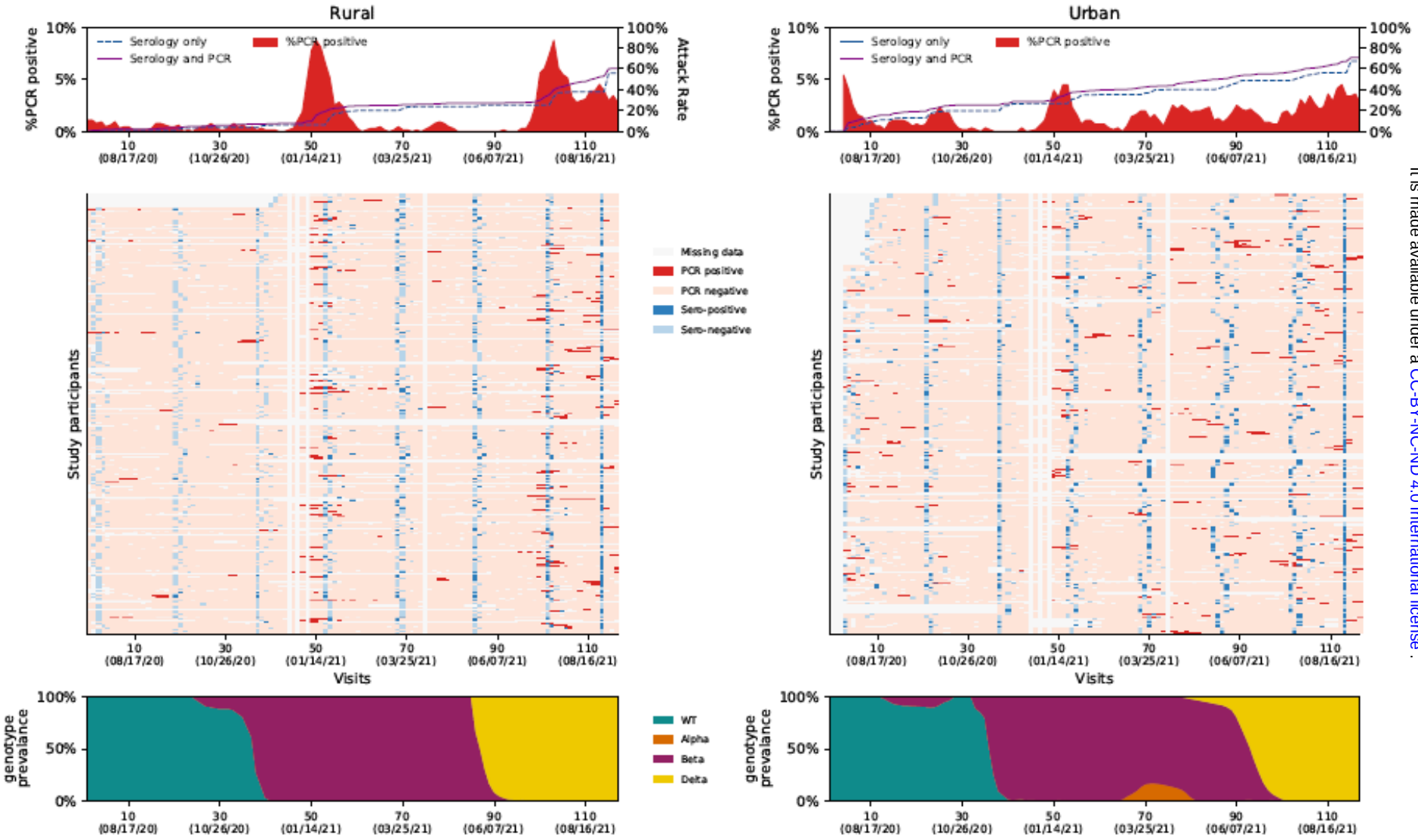
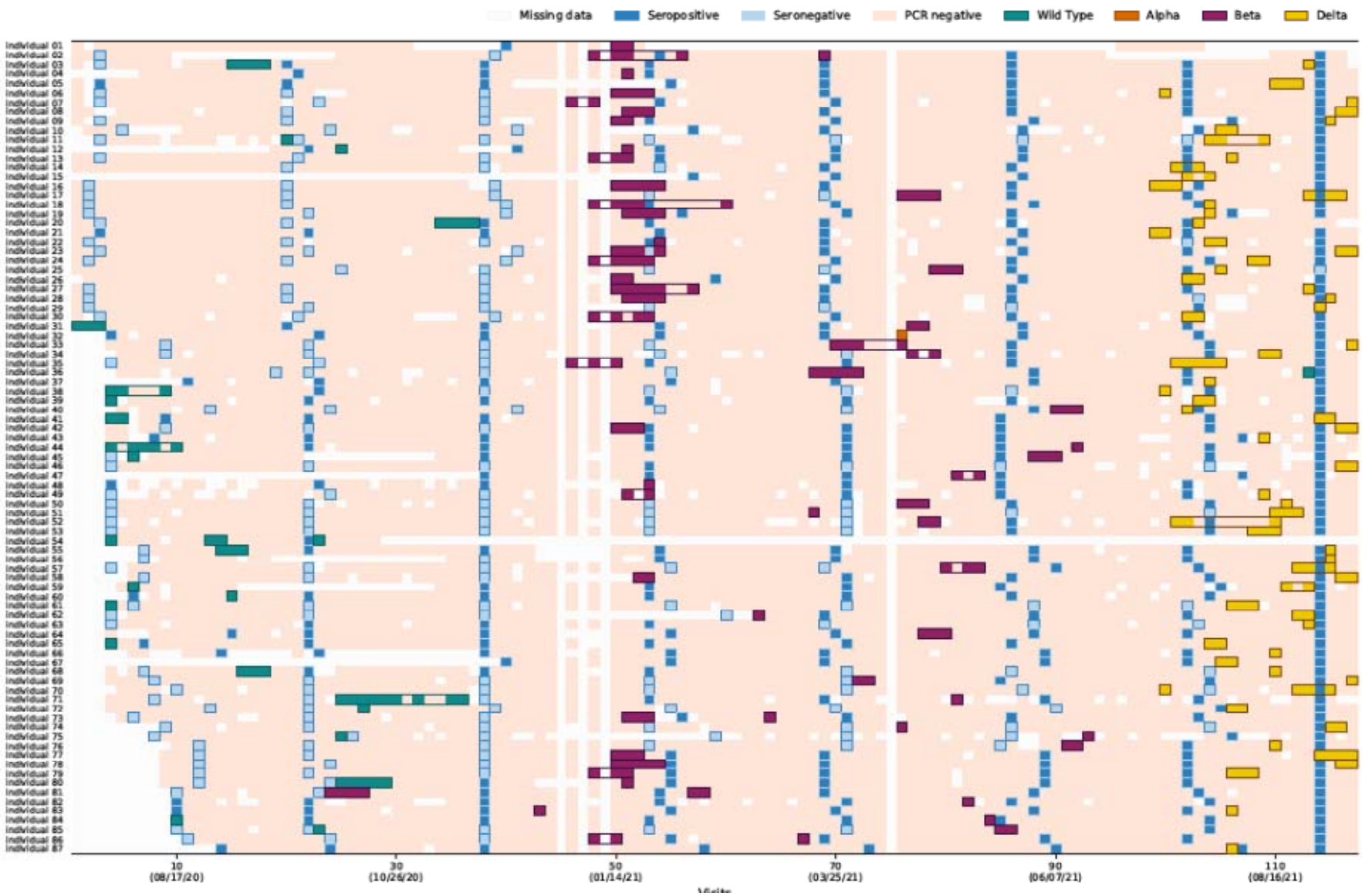


Figure 2: Timing of results of serology and real-time reverse transcription polymerase chain reaction (rRT-PCR)-of 87 individuals with definite, probable or possible reinfections, a rural site and an urban site, South Africa, 2020-2021

Columns are individual follow up visits and rows are individual participants. Follow up visits are coloured white if no sample was tested, light pink if the sample tested negative for SARS-CoV-2. PCR-positive follow up visits are coloured different colours according to infecting variant. Infection episodes are outlined in corresponding colours. Within an episode some visits may test negative or be missed, these are coloured light pink or white. Cells at the time of serology blood draws are coloured according to the results of serology as follows: light blue serology negative, dark blue serology positive.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.20.21260855; this version posted December 4, 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/10.1101/2021.07.20.21260855; this version posted December 4, 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/10.1101/2021.07.20.21260855; this version posted December 4, 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 .