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

PHIRST-C - A Prospective Household cohort study of Influenza, Respiratory Syncytial virus and other respiratory pathogens community burden and Transmission dynamics in South Africa – COVID-19 version

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 (MD), 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) 9, Cecile Viboud 9 (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 Division of International Epidemiology and Population Studies, Fogarty International Center, National Institutes of Health, Bethesda, MD, USA.

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 Target 300

Text: 4213 Target 3500

Tables: 4

Figures: 2

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 was lower from asymptomatic individuals.[1] A systematic review and meta-analysis of 54 household transmission studies of SARS-CoV-2 found an estimated secondary attack rate of 17% (95% CI 14-19).[2] The review also found that household secondary attack rates were increased from symptomatic index cases and that adults were more likely to acquire infection. South Africa experienced two waves of SARS-CoV-2 infections; the second wave was larger and associated with emergence of 501Y.V2 lineage. Modelling studies have suggested that SARS-CoV-2 variant 501Y.V2 may be more transmissible than other variants or that past exposure may provide limited protection, however, they were not able to establish which of these may be more important. A community study from Denmark and two healthcare worker cohorts estimated protection conferred by previous SARS-CoV-2 infection against reinfection to be between 81% and 89%. Data on protection conferred against infection with the 501Y.V2 variant are limited. 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. Comprehensive systematic community studies and studies of 501Y.V2 are lacking.

Added value of this study

We found that from the start of the SARS-CoV-2 pandemic in March 2020 through March 2021 in South Africa, 406 (34%) of 1189 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 3% (12/406) experienced reinfection. Individuals with evidence of infection identified by PCR and/or serology before the onset of the second wave (14 December 2020) experienced 84% (95% CI 65%-93%) protection against re-infection compared to individuals without evidence of previous infection. Symptom data were analysed for 254 rRT-PCR-confirmed infection episodes that occurred >14 days after the start of follow-up (of a total of 310 episodes), of these, 17% (n=43) were associated with one or more symptom, of which 14% (n=6) were hospitalised and 2% (n=1) died. Seventy three percent (161/222) of included households, had one or more individual infected with SARS-CoV-2 within the household. SARS-CoV-2 infected index cases

transmitted the infection to 16% (66/411) of susceptible household contacts. Index case ribonucleic acid (RNA) viral load estimated through cycle threshold value as proxy was strongly predictive of household transmission. Index case or household contact age group and presence of symptoms in the index case were not associated with household transmission. Household transmission was four times greater from index cases infected with Beta variant compared to non-variant infection.

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. The household cumulative infection risk was 16% and did not differ by index case or contact age or presence of symptoms but was higher in households where the index case had potentially higher viral RNA and during the second wave of SARS-CoV-2 infection. The majority of infections were asymptomatic. Asymptomatic individuals transmitted SARS-CoV-2 at similar levels to symptomatic individuals suggesting that interventions targeting symptomatic individuals such as contact tracing of individuals tested because they report symptoms may have a limited impact as a control measure. The high level of heterologous protection from natural immunity evidenced in our study also informs our understanding of possible effects of population-level immunity on SARS-CoV-2 burden and transmission, although key questions remain regarding the durability of protection in the face of continued viral evolution.

Abstract

Word count: 300 Target: 300

Background

By March 2021, South Africa experienced two waves of SARS-CoV-2 infections; the second associated with emergence of Beta variant. We estimated the burden and transmission of SARS-CoV-2 over the two waves.

Methods

We conducted a prospective cohort study during July 2020-March 2021 in one rural and one urban community. Mid-turbinate nasal swabs were collected twice-weekly from consenting 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. Household cumulative infection risk (HCIR) was defined as the proportion of household members with infection following SARS-CoV-2 introduction.

Findings

Among 71,759 nasal specimens from 1,189 members (follow-up rate 93%), 834 (1%) were SARS-CoV-2-positive. By PCR detection and serology combined, 34% (406/1189) of individuals experienced ≥1 SARS-CoV-2 infection episode, and 3% (12/406) experienced reinfection. Infection by PCR and/or serology before the second wave was 84% (95% confidence interval (CI) 65%-93%) protective against re-infection. Of 254 PCR-confirmed episodes with available data, 17% (n=43) were associated with ≥1 symptom, of which 21% (9/43) were medically attended. Among 222 included households, 161 (73%) had ≥1 SARS-CoV-2-positive individual. HCIR was 16% (66/411). On multivariable analysis, index case lower cycle threshold value (OR 5.8, 95% CI 1.8-19.1), urban community (OR 3.1, 95% CI 1.5-6.2) and infection with Beta variant (OR 3.7, 95% CI 1.6-8.4) were associated with increased HCIR. HCIR was similar for symptomatic (8/67,12%) and asymptomatic (61/373, 16%) index cases (p-0.302).

Interpretation

In this study, 83% 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. Previous infection was protective against SARS-CoV-2 infection in the second wave although household transmission increased following the emergence of Beta variant.

Funding

US Centers for Disease Control and Prevention

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.[3,4,5]

South Africa has a relatively young population with <5% of the population aged >65 years.[6] In 2017, the national HIV prevalence among individuals of all ages was 14%, with 7,9 million people living with HIV.[7] South Africa experienced two SARS-CoV-2 waves through March 2021; the first wave peaking in August 2020 and the second peaking in January 2021. The second wave in South Africa was associated with the emergence of the Beta variant of SARS-CoV-2.[9] This variant has been shown to escape from immunity to previous infection.[10] Preliminary data suggest that it may be more transmissible, and may be associated with more severe disease.[11,12]

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 all non-essential business and confining people to their homes. Subsequent to this hard lockdown there was a progressive relaxation of restrictions to movement beginning on 1 May 2020.[13] Restrictions were again increased moderately on 29 December 2020 around the peak of the second wave, decreasing again on 1 March 2021.

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.[14,15,16,17] Comprehensive community studies of asymptomatic infection and robust individual-level epidemiologic data about infection with Beta variant are lacking.

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 infection, the duration of viral RNA shedding and the household cumulative infection risk (HCIR) from symptomatic and asymptomatic index cases of different ages. We also evaluated the frequency of SARS-CoV-2 reinfections and protection conferred by previous infection against reinfection.

Methods

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). Data included in this analysis are from the start of the study on 16 July and 27 July 2020 at the rural and urban site respectively, to 31 March 2021, including 37 and 35 weeks of follow-up and five serum collections. 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. The Prospective Household cohort study of Influenza, Respiratory Syncytial virus and other respiratory pathogens community burden and Transmission dynamics in South Africa – COVID version (PHIRST-C) was based on a previously conducted study (PHIRST) at the same sites from 2016-2018.[21,22]

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 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-March 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)[24] and had refresher training on specimen collection and the identification of respiratory signs and symptoms each week. 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).

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 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 2020, 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 RdRp genes) was considered positive. All specimens testing rRT-PCR-positive were confirmed by re-extraction of a second aliquot, and PCR testing in duplicate. Specimens testing positive on at least one duplicate rRT-PCR were considered positive. If a specimen was confirmed positive after re-extraction, 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.[25] 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.

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).

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.

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.

An rRT-PCR-confirmed household cluster was composed of all rRT-PCR-confirmed infection episodes within a household within an interval between the first rRT-PCR positive tests of any infection episode pairs of ≤14 days (representing ≤2 mean serial intervals).[27] 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 cases (two individuals rRT-PCR positive on the same visit) were excluded from the analysis of HCIR. The generation interval was calculated as the difference between the date of the first positive rRT-PCR in the index and the secondary infection. Following examination of the distribution of calculated generation intervals, we included all secondary infections with rRT-PCR positivity <17 days after the index case onset as potential secondary cases. 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.

A variant was allocated to each episode of infection according to a hierarchical process that accounted for known lineages (i.e., wild-type or Beta variant) within episode or household clusters or occurrence of the episode in wave 1 or 2 (see supplement for details).

To asses SARS-CoV-2 reinfection, we defined possible reinfection as >28 to 90 days between rRT-PCR -positive specimens (no sequence data available) or between first seropositive specimen and rRT-PCR -positive specimen; probable reinfection as >90 days between rRT-PCR -positive specimens (no sequence 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] For this analysis sequencing data were not available but sequencing is planned. 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 time 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 time). We used logistic regression for the analysis of factors associated with binary outcomes (symptomatic fraction, HCIR, index case vs other participants). We used Poisson regression for the analysis of factors associated with SARS-CoV-2 infection and we included all individuals with evidence of infection on rRT-PCR 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 estimation of protection conferred by prior infection against re-infection in the second wave, we classified all individuals who had evidence of previous infection on rRT-PCR or serology prior to 19 December 2020 as previously infected and estimated the subsequent cumulative incidence in previously infected and previously uninfected individuals. Protection was estimated as the difference in attack rate between previously uninfected and previously infected individuals divided by the attack rate in previously uninfected individuals.

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). Participants received grocery store vouchers of USD 3 per visit to compensate for time required for specimen collection and interview.

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 1,249 eligible household members, 1,189 (95%) 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 start of followup 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 76,940 potential individual follow-up visits, we collected and tested 71,759 (93%) mid-turbinate nasal swabs, of which 834 (1%) tested positive for SARS-CoV-2 on rRT-PCR (Figure 1 and Supplementary figure 5 and 6). During the study, 73% (161/222) of households had at least one individual testing SARS-CoV-2 positive on rRT-PCR or serology, with an average of 2.5 (range 1-10) infected individuals per infected household.

During the follow-up period, 34% (406/1189) of individuals experienced at least one episode of SARS-CoV-2 infection on rRT-PCR and/or serology and 3% (12/406) experienced a repeat infection. The highest proportion infected was in the 13-59 years age group (Supplementary figure 4). Of 12 repeat infection episodes, 6 (50%) were classified as possible and 5 (42%) as probable and 1 (8%) confirmed (Figure 2 and supplement). Documented infection on rRT-PCR or serology prior to the start of the second wave was associated with 84% protection against infection in the second wave (relative risk (RR) 0.16, 95% CI 0.07-0.35, attack rate in individuals with previous infection was 3% [6/211] vs 18% [177/978] in individuals without previous infection).

Among 197 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, 177 (90%) seroconverted after the episode. There were no factors signifcantly associated with seroconversion. Among 265 individuals who were seronegative at baseline and subsequently became seropositive, 231 (87%) had evidence of a rRT-PCR-confirmed infection.

On multivariable analysis, factors associated with SARS-CoV-2 cumulative incidence were age 13-18 years, 19-39 years or 40-59 years vs 5-12 years, residing in the urban community vs rural community, and being obese vs normal weight (Table 1).

Of 310 rRT-PCR-confirmed episodes, 40% (n=124) occurred in the first wave and 60% (n=186) in the second wave (Supplementary figure 5 and 6). Of those with known variant type, 8% (6/80) in the first wave and 95% (142/150) in the second wave were Beta variant. Of 254 episodes, that occurred >14 days after the start of follow-up, 17% (43/254) of individuals reported ≥1 symptom (details in supplement). Among 43 symptomatic individuals, 3 (7%) attended an outpatient clinic, 6 (14%) were hospitalised and 1 (2%) died. Among 7 symptomatic individuals who were employed or attended school, 43% (n=3) reported absenteeism. In multivariable analysis symptoms were more common in individuals aged 40-59 years and ≥60 years compared to 5-12 years (Table 2).

The mean duration of rRT-PCR positivity was 12.5 days (median 11 days, range 4-60 days) and 22% (56/254) of episodes were rRT-PCR positive at only one visit. On multivariable analysis, individuals aged ≥60 years, infected with Beta variant and with minimum Ct value <30 and ≥2 rRT-PCR targets positive shed SARS-CoV-2 RNA for longer (Table 3).

Among 139 households with at least one SARS-CoV-2 infection cluster detected on rRT-PCR, 108 (77%) had one cluster, 30 (21%) had 2 clusters and 1 had 3 clusters (total of 171 clusters). The average cluster duration among 138 clusters starting >14 days after the start of follow up was 15.8 days (range 4-60 days). We included 97 clusters starting >14 days after the start of follow up and with a single index case from 81 households for analysis of HCIR. In this subset of households the HCIR was 16% (66/411). HCIR was 12% (8/67) for symptomatic and 16% (61/373) for asymptomatic index cases (p = 0.302) (Table 4). On multivariable analysis, low Ct value (proxy for high viral RNA load) of the index case, urban community and infection with Beta variant were associated with increased HCIR. HCIR did not differ by age of the index case or household contact. When compared to non-index cases, index cases within household clusters were less likely to be aged <5 years and 5-12 years compared to 40-59 years and more likely to be obese compared to normal weight (Supplementary table 3).

The mean generation interval was 7.0 days (range 2-16 days) (Supplementary figure 7). On multivariable analysis, accounting for age, generation interval was longer if the index case shed for

longer or was infected with Beta variant and shorter if the index case Ct value was ≤30 (Supplementary table 4).

Discussion

Using intensive repeated sampling in two South African communities, we found that by 31 March 2021 more than one-third of individuals had been infected with SARS-CoV-2. Notably, only 16% of infections were associated with symptoms; of these, 14% were hospitalised and 2% died. SARS-CoV-2 was transmitted to approximately 16% of household contacts. Transmission did not depend on the index case age or presence of symptoms. Instead, index case RNA viral load was the main predictor of onward transmission. Approximately 3% of individuals experienced at least one repeat episode of infection within 9 months of follow up, and infection in the first wave was 84% protective against infection in the second wave. Infection with SARS-CoV-2 Beta variant, was associated with a similar symptomatic fraction to non-variant infection.[9] However, duration of shedding was longer and household transmission was increased fourfold with the Beta 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.[1] We found that 16% of infections were symptomatic despite active twice-a-week symptom evaluation. We are not aware of comparable community-based studies with systematic regular RT-PCR testing and symptom ascertainment, limiting our ability to discern the contribution of the study design to the observed differences or to assess whether there may be truly lower symptomatic fraction in the populations studied. 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.[6] Despite the low overall symptomatic fraction, symptomatic fraction increased with age, in line with prior studies.[1,17]

A systematic review and meta analysis of 54 household transmission studies of SARS-CoV-2 reported an estimated secondary attack rate of 16.6% (95% CI 14.0-19.3)[2] similar to our estimate of 16%. The review also reported that household secondary attack rates were higher from symptomatic index cases and that adults were more likely than children to acquire infection. Previous studies have found that viral load and duration of SARS-CoV-2 shedding are lower among mild and asymptomatically infected individuals compared to individuals with severe illness.[28] While we did not find that factors like age or symptom profile were associated with risk of infection transmission or acquisition, 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.[29] 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.

Modelling studies have suggested that SARS-CoV-2 Beta variant may be more transmissible than other variants or that past exposure may provide limited protection, however, they were not able to establish which of these may be more important.[11] We found that household transmission of SARS-CoV-2 was approximately 4 times higher with Beta variant compared to non-variant infections. In vitro studies have found that serum from individuals previously infected with prior lineages of SARS-CoV-2 provides limited cross-protection against Beta variant.[10] In contrast, we found, that previous infection in the first wave conferred 84% protection against infection in the second wave when Beta variant predominated in South Africa.[9] This estimate was similar to estimates of 81% in a community study from Denmark[30] and 84% and 89% protection against infection seen in two different cohorts of UK healthcare workers in the absence of the Beta variant.[31,32] In contrast to our study, an analysis of the placebo arm of a clinical trial in South Africa, where most infections were due to Beta variant, found a similar attack rate in individuals with and without evidence of previous SARS-CoV-2 infection.[33]

We found a lower symptomatic fraction in children aged <18 years compared to adults. While attack rates were lower in children aged <12 years, rates in children aged 13-18 years were similar to those in adults aged 19-59 years. We did not find differences in the risk of transmission or acquisition of SARS-CoV-2 in children compared to adults, possibly related to limited statistical power, however we did find that children aged <13 years were less likely to introduce SARS-CoV-2 into the home. 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.[34–36,37,2] Differences between studies could be related to differences in study design as most previous studies did not include systematic longitudinal rRT-PCR testing irrespective of symptoms in both children and adult groups.

Our study had several limitations. We included two communities and households with >2 members, potentially limiting generalisability of study findings. The finding of very 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 disease 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.[22] Our study covered a 9-month period, and we cannot speculate on the duration of natural immunity and risk of reinfection over longer timescales. We did not quantify viral RNA load but instead this was inferred using Ct values as proxy.

Our study also had several strengths. Participating householdss were randomly sampled from a rural and an urban South African community and followed up for 9 months through the second half of the first wave and the second wave 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. The combination of rRT-PCR and serological testing allowed for more complete ascertainment of infection burden and protection conferred by previous infection.

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. The household cumulative infection risk was 16% and did not differ by index case or contact age or presence of symptoms in index case but was higher in households where the index case had a higher RNA viral load and with Beta variant infection. Asymptomatic individuals transmitted SARS-CoV-2 at similar levels as symptomatic individuals suggesting that interventions targeting symptomatic individuals, such as contact tracing of individuals tested because they report symptoms, may have limited impact in this setting. The high level of heterologous protection from natural immunity evidenced in our study also informs our thinking about the possible effects of population-level immunity on SARS-CoV-2 burden and transmission, although key questions remain regarding the durability of protection in the face of continued viral evolution.

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, and the Medical Research Council, South Africa, and previously 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, STo, 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].

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). Participants receive grocery store vouchers of ZAR50 (USD 3) per visit to compensate for time required for specimen collection and interview.

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/

References

1. 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

2. 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

3. 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.

4. 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

5. 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

6. 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.

7. Simbayi L, Zuma K, Zungu N, et al. *South African National HIV Prevalence, Incidence, Behaviour and Communication Survey, 2017*. Cape Town, South Africa; 2017.

8. The National Institute for Communicable Diseases of the National Health Laboratory S. *COVID-19 Weekly Epidemiology Brief: Week Ending 20 February 2021 (Week 7 of 2021)*.; 2021. https://www.nicd.ac.za/diseases-a-z-index/covid-19/surveillance-reports/weekly-epidemiological-brief/.

9. 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

10. 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

11. 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.

12. 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.

13. 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.

14. 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

15. 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

16. 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

17. 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

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. Cohen C, McMorrow ML, Martinson NA, Kahn K, Treurnicht FK, Moyes J. Cohort Profile: a Prospective Household cohort study of Influenza, Respiratory Syncytial virus, and other respiratory pathogens community burden and Transmission dynamics in South Africa

(PHIRST), 2016-2018. *MedXRiV*. 2021.

22. 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.

23. Mccarthy K, Tempia S, Kufa T, Kleynhans J, Wolter N. The Importation and Establishment of Community Transmission of SARS-CoV-2 During the First Eight Weeks of the South African COVID-19 Epidemic. *SSRN Prepr*. 2021. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3792114.

24. 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

25. 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

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. 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

28. 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

29. 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

30. Hansen CH, Michlmayr D, Gubbels SM, Mølbak K, Ethelberg S. Assessment of protection against reinfection with SARS-CoV-2 among 4 million PCR-tested individuals in Denmark in 2020: a population-level observational study. *Lancet (London, England)*. 2021;397(10280):1204-1212. doi:10.1016/S0140-6736(21)00575-4

31. Hall VJ, Foulkes S, Charlett A, et al. SARS-CoV-2 infection rates of antibody-positive compared with antibody-negative health-care workers in England: a large, multicentre, prospective cohort study (SIREN). *Lancet (London, England)*. 2021;397(10283):1459-1469.

doi:10.1016/S0140-6736(21)00675-9

32. Lumley SF, O'Donnell D, Stoesser NE, et al. Antibody Status and Incidence of SARS-CoV-2 Infection in Health Care Workers. *N Engl J Med*. 2021;384(6):533-540. doi:10.1056/NEJMoa2034545

33. Shinde V, Bhika S, Hoosain Z, et al. Efficacy of NVX-CoV2373 Covid-19 Vaccine against the B.1.351 Variant. *N Engl J Med*. May 2021. doi:10.1056/NEJMOA2103055

34. 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

35. 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

36. Viner RM, Mytton OT, Bonell C, et al. Susceptibility to SARS-CoV-2 Infection Among Children and Adolescents Compared With Adults: A Systematic Review and Meta-analysis. *JAMA Pediatr*. 2021;175(2):143-156. doi:10.1001/jamapediatrics.2020.4573

37. 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

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	167/634 (26)	Reference	Reference
	Urban	239/555 (43)	2.6 (1.6-4.1)	1.5 (1.2-1.9)
Age group (years)	<5	35/152 (23)	0.9 (0.6-1.4)	1.0 (0.7-1.6)
	5-12	83/337 (25)	Reference	Reference
	13-18	74/169 (44)	1.7 (1.3-2.4)	1.7 (1.3-2.4)
	19-39	102/262 (39)	1.5 (1.1-2.0)	1.4 (1.1-2.0)
	40-59	81/166 (49)	1.9 (1.4-2.6)	1.6 (1.1-2.3)
	≥60	31/103 (30)	1.2 (0.8-1.9)	1.1 (0.7-1.7)
Sex	Female	264/714 (37)	1.2 (1.0-1.5)	
	Male	142/475 (30)	Reference	
HIV[a]	Infected	73/173 (42)	1.3 (1.0-1.6)	1.0 (0.8-1.4)
	Uninfected	321/962 (33)	Reference	Reference
CD4+ T cell count/ml[b]	<200	4/14 (29)	0.6 (0.2-1.8)	
	≥200	67/151 (44)	Reference	
HIV viral load copies/ml[b]	≥1000	11/31 (35)	0.8 (0.5-1.5)	
	<1000	61/136 (45)	Reference	
Other underlying illness[c]	Absent	367/1082 (34)	Reference	
	Present	39/107 (36)	1.0 (0.6-1.7)	
BMI[d]	Underweight	24/83 (29)	1.1 (0.7-1.7)	0.9 (0.6-1.5)
	Normal weight	175/637 (27)	Reference	Reference
	Overweight	90/218 (41)	1.5 (1.1-1.9)	1.3 (0.9-1.7)
	Obese	117/249 (47)	1.6 (1.3-2.1)	1.4 (1.0-1.8)
Number of individuals in household	3-5	179/504 (36)	Reference	
	6-10	181/568 (32)	1.1 (0.9-1.4)	
	≥11	46/117 (39)	1.3 (0.8-2.1)	
Crowding (>2 people/sleeping room)	No	230/634 (36)	1.24 (0.9-1.5)	
	Yes	176/555 (32)	Reference	

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 1135 of 1189 (95%) of individuals [b]Among 173 HIV-infected individuals, 165 (95%) had available data on CD4+ T cell count and 167 (97%) had available data on HIV viral load [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 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

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	3/25 (12)	1.7 (0.3-8.1)	1.6 (0.3-8.0)
	5-12	4/53 (8)	Reference	Reference
	13-18	7/44 (16)	2.3 (0.6-8.5)	2.5 (0.7-9.6)
	19-39	11/72 (15)	2.2 (0.7-7.4)	2.5 (0.7-8.5)
	40-59	11/36 (31)	5.4 (1.6-18.7)	5.6 (1.6-19.9)
	≥60	7/24 (29)	5.0 (1.3-19.4)	5.0 (1.3-19.3)
Sex	Female	31/170 (18)	1.3 (0.6-2.8)	
	Male	12/84 (14)	Reference	
HIV	Infected	9/44 (20)	1.3 (0.6-3.1)	
	Uninfected	32/199 (16)	Reference	
Other underlying illness[b]	Absent	37/232 (16)	Reference	
	Present	6/22 (27)	2.0 (0.7-5.4)	
BMI[c]	Underweight	0/14 (0)	0.2 (0.0-3.3)	
	Normal weight	18/120 (15)	Reference	
	Overweight	8/51 (16)	1.1 (0.4-2.6)	
	Obese	17/69 (25)	1.8 (0.9-3.8)	
Duration of viral RNA shedding (days)	≤4	6/56 (11)	Reference	
	>4	37/198 (19)	1.9 (0.8-4.8)	
Minimum Ct value	≤30	38/202 (19)	2.2 (0.8-5.8)	1.5 (0.5-4.4)
	>30	5/52 (10)	Reference	Reference
Number of rRT-PCR targets positive	1	2/18 (11)	Reference	
	>1	39/231 (17)	1.6 (0.4-7.4)	
PCR or seropositive before episode	No	42/228 (18)	Reference	Reference
	Yes	1/187 (6)	0.3 (0.1-2.0)	0.3 (0.4-2.7)
Epidemic wave	1	7/69 (10)	Reference	
	2	36/185 (19)	2.1 (0.9-5.1)	
Variant	Wild type	7/65 (11)	Reference	Reference
	Beta	36/188 (19)	2.0 (0.8-4.7)	2.0 (0.8-4.8)
	Alpha	0/1 (0)	Not estimated	Not estimated

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

[a]≥1 symptom vs no symptom reported, analysis restricted to 254 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

Table 3: Factors associated with duration of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) viral RNA shedding 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	12.6±7.6 (4-31)	0.9 (0.5-1.4)	1.0 (0.6-1.6)
	5-12	11.2±8.6 (4-46)	Reference	Reference
	13-18	12.1±9.8 (4-60)	0.9 (0.6-1.3)	0.8 (0.5-1.2)
	19-39	12.1±8.0 (4-40)	0.9 (0.6-1.3)	0.9 (0.6-1.3)
	40-59	12.9±8.4 (4-43)	0.8 (0.5-1.2)	0.7 (0.4-1.0)
	≥60	16.6±11.2 (4-52)	0.6 (0.3-0.9)	0.5 (0.3-0.9)
Sex	Female	13.1±7.2 (4-39)	0.8 (0.6-1.0)	
	Male	11.1.3±9.6 (4-60)	Reference	
HIV	Infected	13.8±10.5 (4-52)	0.8 (0.6-1.1)	
	Uninfected	12.3±8.5 (4-60)	Reference	
	Unknown	10.0±8.5 (4-33)	Not included	
Other underlying illness	Absent	12.6±8.8 (4-60)	Reference	
	Present	10.8±7.9 (4-34)	1.3 (0.8-2.0)	
BMI[b]	Underweight	13.0±14.2 (4-60)	0.7 (0.4-1.2)	
	Normal weight	11.9±7.9 (4-52)	Reference	
	Overweight	13.2±9.4 (4-46)	0.8 (0.6-1.1)	
	Obese	12.9±8.3 (4-43)	0.9 (0.6-1.2)	
Symptoms	Absent	12.4±8.8 (4-60)	Reference	
	Present	13.0±8.3 (4-35)	0.9 (0.7-1.4)	
Minimum Ct value	≤30	14.0±8.8 (4-60)	0.3 (0.2-0.4)	0.4 (0.3-0.5)
	>30	6.5±5.4 (4-28)	Reference	Reference
Number of rRT-PCR targets positive	1	4.3±1.3 (4-10)	Reference	Reference
	≥2	13.1±8.8 (4-60)	0.1 (0.1-0.2)	0.3 (0.2-0.6)
PCR or seropositive before episode	No	13.1±9.1 (4-60)	Reference	
	Yes	6.3±4.9 (4-24)	3.1 (1.9-5.1)	
	Unknown	10.4±4.2 (4-18)		
Epidemic wave	First	10.9±7.3 (4-43)	Reference	
	Second	13.1±9.2 (4-60)	0.7 (0.6-0.9)	
Variant	Wild type	10.4±7.1 (4-43)	Reference	Reference
	Beta	13.3±9.3 (4-60)	0.7 (0.5-0.9)	0.7 (0.5-0.9)
	Alpha	4±0 (4-4)	5.4 (0.7-39.7)	Not estimated

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. [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

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 (95% CI)	Multivariable aOR (95% CI)
Characteristics of the index case				
Age group (years)	<5	6/31 (19)	2.9 (0.3-14.3)	
	5-12	7/49 (14)	Reference	
	13-18	14/73 (19	1.1 (0.3-5.0)	
	19-39	22/165 (13)	1.0 (0.2-3.8)	
	40-59	13/56 (23)	2.6 (0.5-12.5)	
	≥60	4/37 (11)	0.7 (0.1-5.1)	
Sex	Female	49/275 (18)	1.6 (0.7-3.4)	
	Male	17/136 (13)	Reference	
HIV status	Infected	14/83 (17)	0.6 (0.2-1.6)	
	Uninfected	50/310 (16)	Reference	
BMI[b]	Underweight	9/27 (33)	3.8 (1.0-15.2)	
	Normal weight	21/181 (12)	Reference	
	Overweight	9/87 (10)	1.1 (0.4-3.3)	
	Obese	27/116 (23)	4.2 (1.5-11.4)	
Symptoms	Absent	61/373 (16)	Reference	
	Present	8/67 (12)	0.7 (0.2-2.0)	
Duration of viral RNA shedding (days)	≤4	5/96 (5)	Reference	
	>4	64/344 (19)	4.7 (1.5-15.0)	
Minimum Ct value	≤30	64/354 (18)	4.7 (1.4-16.3)	5.8 (1.8-19.1)
	>30	5/86 (6)	Reference	Reference
Number of rRT-PCR targets positive	1	2/41 (5)	Reference	
	>1	67/388 (17)	4.8 (0.8-29.0)	
Epidemic wave	1	17/162 (10)	Reference	
	2	52/278 (19)	2.6 (1.1-5.9)	
Variant	Wild type	15/154 (10)	Reference	Reference
	Beta	54/281 (19)	2.8 (1.2-6.3)	3.7 (1.6-8.4)
	Alpha	0/5 (0)	Not estimated	Not estimated
Site	Rural	27/239 (11)	Reference	Reference
	Urban	39/172 (23)	2.4 (1.1-5.2)	3.1 (1.5-6.2)
Characteristics of the household contact				
Age group (years)	<5	7/43 (16)	1.1 (0.4-3.2)	
	5-12	20/129 (16)	Reference	
	13-18	11/58 (19)	1.4 (0.5-3.7)	
	19-39	11/88 (13)	0.7 (0.3-1.8)	
	40-59	8/52 (15)	1.1 (0.4-3.1)	
	≥60	9/41 (22)	1.6 (0.5-4.6)	
Sex	Female	43/248 (17)	1.4 (0.7-2.6)	
	Male	23/163 (14)	Reference	
HIV status	Infected	11/52 (21)	1.4 (0.6-3.4)	
	Uninfected	55/342 (16)	Reference	
Other underlying illness	Absent	59/366 (16)	Reference	
	Present	7/45 (16)	0.9 (0.3-2.4)	

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, 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. [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,j]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 - rRT-PCR missing or negative and serology negative, dark blue - rRT-PCR missing or negative and serology positive, purple rRT-PCR positive and serology positive, green – rRT-PCR negative and serology positive. Bottom panel: Percent of rRT-PCR-positive samples typed as wild type (WT) or variant of concern (VOC) by follow-up visit (10 visit moving average used for smoothing). All VOC typed as Beta except for a single Alpha on the final follow-up visit at the urban site.

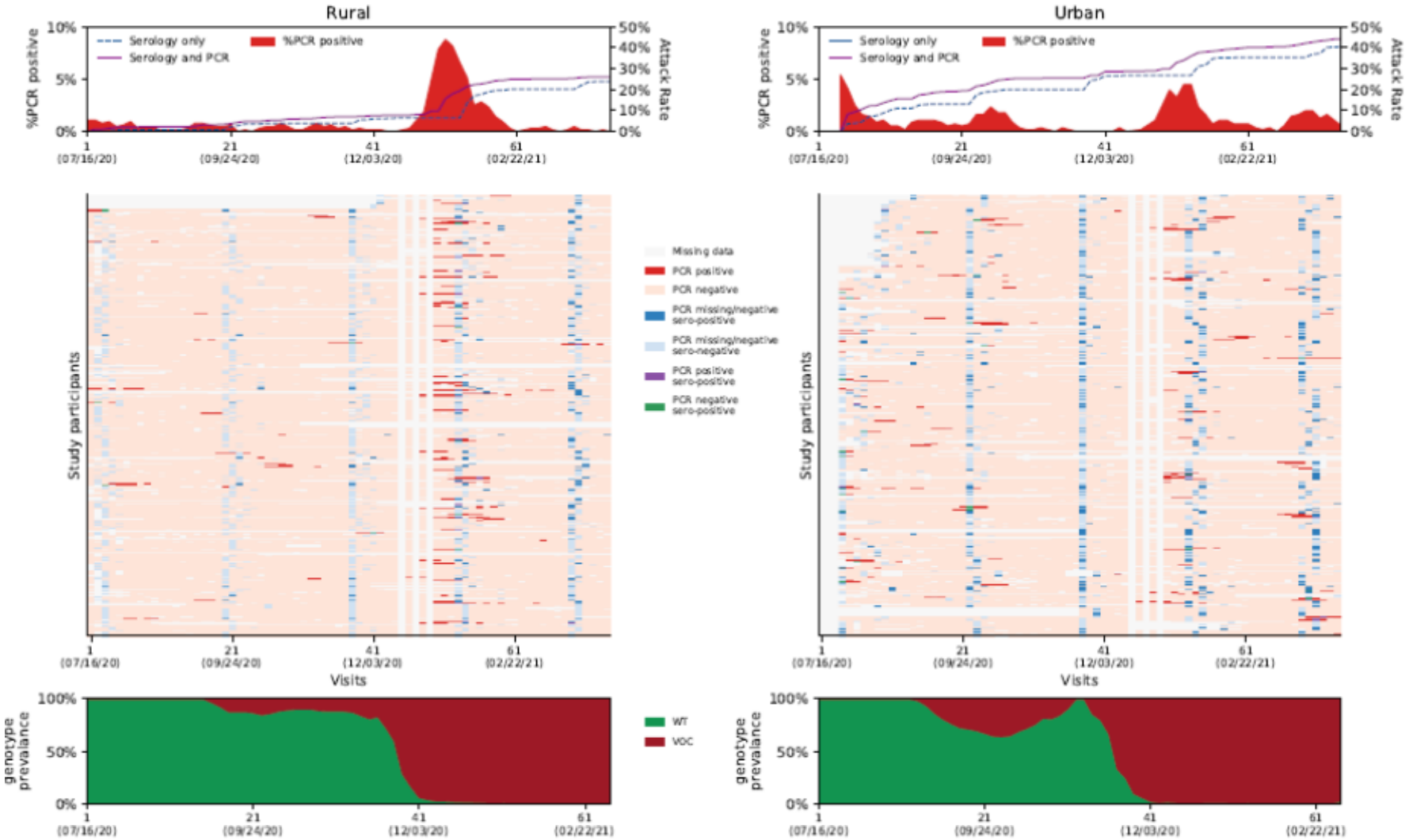
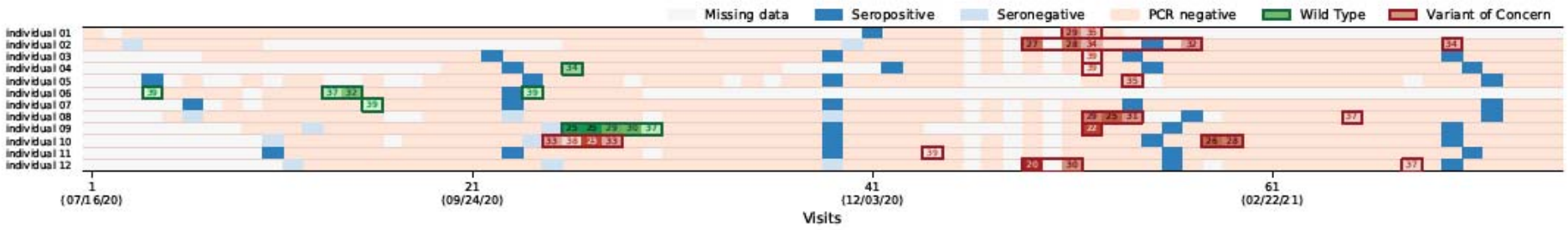

Rural
Urban
Serology only
Serology and PCR
%PCR positive
%PCR positive
Attack Rate
10%
5%
0%
50%
40%
30%
20%
10%
0%
1 (07/16/20)
21 (09/24/20)
41 (12/03/20)
61 (02/22/21)
Study participants
Visits
Missing data
PCR positive
PCR negative
PCR missing/negative sero-positive
PCR missing/negative sero-negative
PCR positive sero-positive
PCR negative sero-positive
genotype prevalance
100%
50%
0%
WT
VOC

Figure 2: Timing of results of serology and real-time reverse transcription polymerase chain reaction (rRT-PCR)-of 12 individuals with definite 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 and outlined green if infection ise due to wild type (non-variant) SARS-CoV-2 and dark red if infection is due to a variant of concern (all Beta variant in this figure). 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. Possible reinfections, individuals 1, 2, 6, 7, 8, 12; probable reinfections individuals 3, 4, 5, 10, 11; confirmed reinfections individuals 9