Article

Rapidly shifting immunologic landscape and severity of SARS-CoV-2 in the Omicron era in South Africa

Kaiyuan Sun [1] ✉, Stefano Tempia[2,3,4], Jackie Kleynhans [2,3], Anne von Gottberg [2,5], Meredith L. McMorrow[6], Nicole Wolter [2,5], Jinal N. Bhiman [7,8], Jocelyn Moyes[2,3], Maimuna Carrim[2,5], Neil A. Martinson[9,10], Kathleen Kahn [11], Limakatso Lebina[9], Jacques D. du Toit [11], Thulisa Mkhencele[2], Cécile Viboud [1,12], Cheryl Cohen [2,3,12] ✉ & the PHIRST-C group*

South Africa was among the first countries to detect the SARS-CoV-2 Omicron variant. However, the size of its Omicron BA.1 and BA.2 subvariants (BA.1/2) wave remains poorly understood. We analyzed sequential serum samples collected through a prospective cohort study before, during, and after the Omicron BA.1/2 wave to infer infection rates and monitor changes in the immune histories of participants over time. We found that the Omicron BA.1/2 wave infected more than half of the cohort population, with reinfections and vaccine breakthroughs accounting for > 60% of all infections in both rural and urban sites. After the Omicron BA.1/2 wave, we found few (< 6%) remained naïve to SARS-CoV-2 and the population immunologic landscape is fragmented with diverse infection/immunization histories. Prior infection with the ancestral strain, Beta, and Delta variants provided 13%, 34%, and 51% protection against Omicron BA.1/2 infection, respectively. Hybrid immunity and repeated prior infections reduced the risks of Omicron BA.1/2 infection by 60% and 85% respectively. Our study sheds light on a rapidly shifting landscape of population immunity in the Omicron era and provides context for anticipating the long-term circulation of SARS-CoV-2 in populations no longer naïve to the virus.

One of SARS-CoV-2's most prominent features has been its rapid adaptive evolution throughout the pandemic: every few months, new variants with selective advantages have emerged, displaced resident variants, and reached global dominance. To date, the World Health Organization has classified five SARS-CoV-2 lineages as variants of concern (VOCs) due to their enhanced transmissibility and immune escape properties, including Alpha, Beta, Gamma, Delta and Omicron[1,2]. Following the rise of Omicron (BA.1), new lineages continue to evolve with further mutations, including those on the spike protein

not seen on Omicron BA.1 that evade immune responses (notably BA.2, BA.2.12.1, BA.4, and BA.5[3]. The selective advantage of a new immune-escape variant and subvariants is shaped in part by the host population immunity, first at the location of emergence, and then globally[4].

As of August 1st, 2022, South Africa has experienced five SARS-CoV-2 epidemic waves: the 1st wave was dominated by the ancestral strain carrying the D614G mutation (D614G); the 2nd wave by the Beta VOC (with little impact of the Alpha VOC that was globally dominant at that time[5]); the 3rd wave by the Delta VOC; the 4th wave by Omicron

subvariants BA.1 and BA.2 (BA.1/2 wave); and the 5th wave by Omicron subvariants BA.4 and BA.5 (BA.4/5 wave). In addition to the Beta VOC[6], the Omicron subvariants BA.1, BA.4 and BA.5, are likely to have emerged in South Africa or the surrounding region[7,8]. Detailed studies of the immunologic landscape in South Africa could provide a unique perspective on how immunity contributes to variant success on a population level, near the region of their emergence. The PHIRST-C cohorts have generated detailed prospective data on infection and serology spanning South Africa's first four waves, in carefully sampled populations from randomly selected households[9]. Here we quantify changes in the population immunologic landscape to SARS-CoV-2 over time, and particularly in the aftermath of the Omicron BA.1/2 wave, compare Omicron's epidemiologic properties to those of prior variants, adjusting for changes in prior immunity, and discuss how these factors may interact to determine the fate of new variants.

Results

Serologic specimen collection and SARS-CoV-2 dynamics in the PHIRST-C cohort

As previously described in refs.[9,10], in June 2020, 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, where "C" stands for coronavirus disease

2019 (COVID-19) enrolled a total of 1200 individuals living in 222 randomly selected households. Two sites were selected with 114 households (643 individuals) enrolled in a rural site located in Agincourt, Mpumalanga Province, northeast South Africa and 108 households (557 individuals) in an urban site, located in Jouberton township, Matlosana, North West Province, South Africa. From July 2020 through April 2022, ten sequential serum specimens were collected for each participant (Fig. 1). Blood draws 1-9 were conducted at approximately 2-month intervals, with blood draws 8 and 9 collected one month prior and after the emergence of Omicron BA.1 variant in Southern Africa[8]. Blood draw 10 was conducted at the end of the Omicron BA.1/2 wave, ~4 months after blood draw 9, but prior to the emergence of Omicron BA.4/5 (Fig. 1)[8]. The cohorts included a period of intense follow-up of active infections: ranging from July 16, 2020 to August 28, 2021 for the rural site and from July 27, 2020 to August 28, 2021 for the urban site. During the intense follow-up period, nasal swab samples were collected twice-weekly for SARS-CoV-2 real-time reverse transcription polymerase chain reaction (rRT-PCR), irrespective of symptoms. The intense follow-up period covered the D614G, Beta, and Delta waves in both cohorts.

We used sequential readouts of the Roche Elecsys Anti-SARS-CoV-2 nucleocapsid assay[11] before (blood draws 8, 9) and at the end of (blood draw 10) the Omicron BA.1/2 wave to infer the cumulative rate
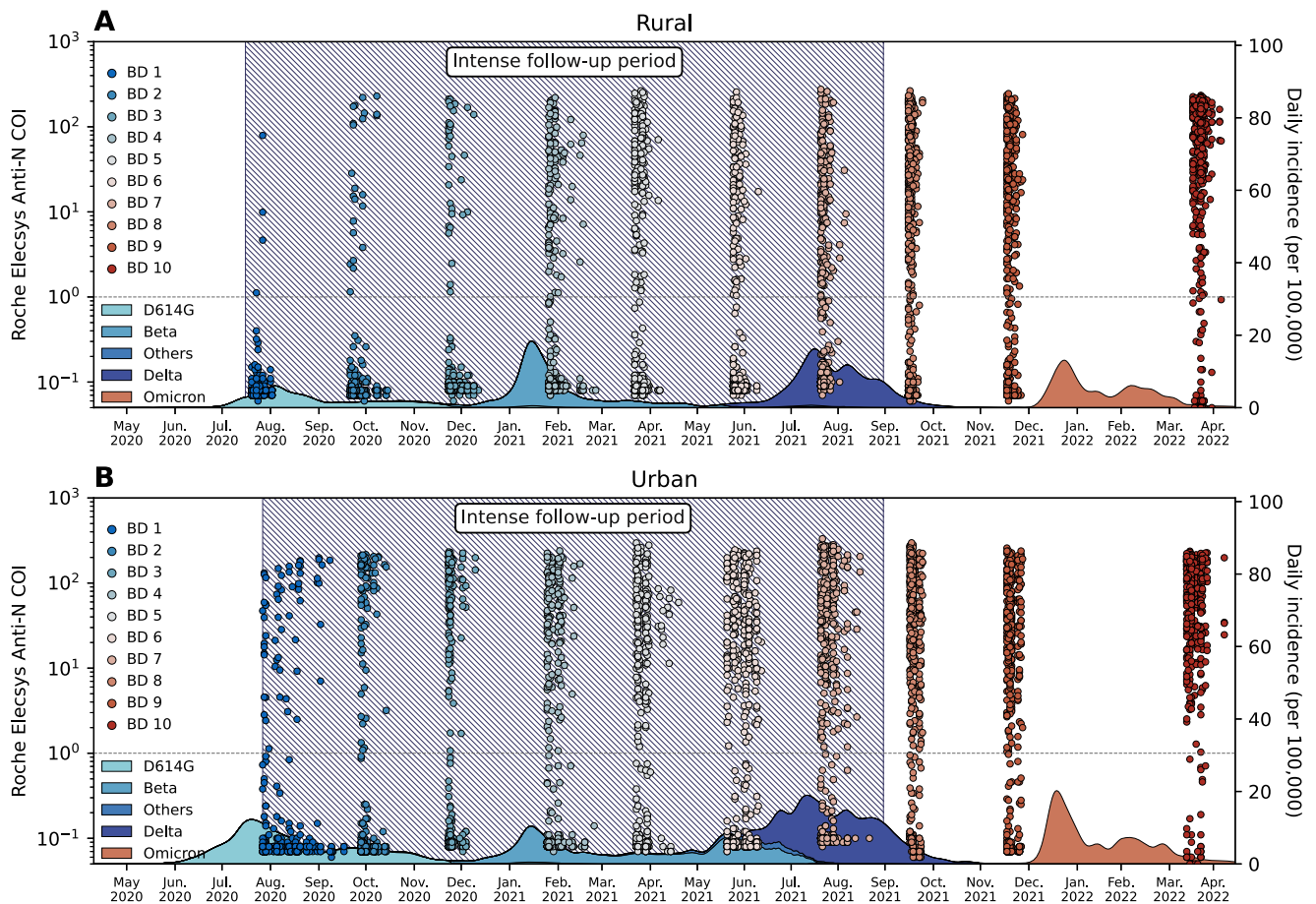
Fig. 1 | PHIRST-C study June 2020 – April 2022, SARS-CoV-2 serology and epidemiologic curve in the two study sites. A Serum samples and epidemiologic curve in the rural site. Dots represent the Roche Elecsys Anti-SARS-CoV-2 nucleocapsid assay cutoff index (COI) at different timepoints of the serum specimen collections; Each dot represents one serum specimen collection, with dot color denoting blood draw collection time, from blue (early) to red (late). The shaded curve at the bottom represents the daily incidence of SARS-CoV-2 cases in routine surveillance data collected from the Ehlanzeni District, Mpumalanga Province.

Colors of the shaded curve represent different variant types. Here, blood draw (BD) 10 was collected at the end of the first Omicron wave. Since in South Africa, Omicron BA.4 and BA.5 only started to rise at April, 2022[8], we assume the Omicron wave prior to BD 10 were BA.1 and BA.2 subvariants. The hatched area represents the period of intense follow-up of the PHIRST-C cohort, when nasal swabs were collected and tested on rRT-PCR at twice-a-week frequency. **B** Same as (**A**) but for the urban site, with shaded curve at the bottom representing routine surveillance data collected from the Dr. Kenneth Kaunda District, North West Province. BD Blood draw.

Table 1 | Omicron BA.1/2 infection attack rate (AR) in rural and urban cohorts of South Africa based on 905 participants with complete serologic information (75% of the total 1200 participants)

Characteristics	Rural (478)	Urban (427)
	n/N (AR%)	n/N (AR%)
Age group, in years		
0–4	42/64 (65.6)	19/37 (51.4)
5–12	129/185 (69.7)	76/113 (67.3)
13–18	39/59 (66.1)	46/67 (68.7)
19–39	44/64 (68.8)	50/77 (64.9)
40–59	37/71 (52.1)	43/91 (47.3)
60+	18/35 (51.4)	14/42 (33.3)
Sex		
Male	116/171 (67.8)	108/188 (57.4)
Female	193/307 (62.9)	140/239 (58.6)
Household size (no. individuals)		
3–5	129/190 (67.9)	121/212 (57.1)
6–8	91/156 (58.3)	84/148 (56.8)
9–12	71/99 (71.7)	31/50 (62.0)
13+	18/33 (54.5)	12/17 (70.6)
HIV status		
Negative	257/392 (65.6)	201/345 (58.3)
PLWH[*]	39/63 (61.9)	41/71 (57.7)
Unknown	13/23 (56.5)	6/11 (58.2)
SARS-CoV-2 immunity status		
Naïve	117/144 (81.3)	71/94 (75.5)
Janssen Ad26.COV2.S (1dose)	9/11 (81.8)	2/2 (100)
Pfizer–BioNTech BNT162b2 (1st dose)	5/10 (50.0)	2/3 (66.7)
Pfizer–BioNTech BNT162b2 (2nd dose)	3/5 (60.0)	6/11 (54.5)
Prior infection (D614G)	20/27 (74.1)	47/67 (70.2)
Prior infection (Beta)	50/69 (72.5)	56/92 (60.9)
Prior infection (Delta)	81/149 (54.4)	33/61 (54.1)
Hybrid immunity[†]	19/41 (46.3)	15/47 (31.9)
Repeated infections	4/19 (21.1)	10/37 (27.0)
Rest[‡]	1/3 (33.3)	6/13 (46.2)

[*]PLWH People living with HIV. [†]Hybrid immunity: one episode of prior infection and vaccination. [‡]Rest: this category consists of repeat infections plus vaccination, vaccination with unknow vaccine types, and prior infection with other less frequent lineages including Alpha and C.1.2 variants.

of infections and re-infections during the Omicron BA.1/2 wave among the cohort populations. The inference of primary and repeat infections with Omicron was based on the dynamics of sequential serologic assay readouts, after calibrating the method on the Delta epidemic wave, for which we had both serology and PCR data. The analysis was performed on a subset of 905 of 1200 cohort participants with complete serum specimens collected during the relevant Delta and Omicron periods. Details of the serologic inference and calibration are described in Methods Section 3. We ascertained infections predating the Omicron BA.1/2 wave based on both variant-specific PCR tests and serology, as detailed in a prior study[10], covering the period through blood draw 8, or mid-September, 2021. All COVID-19 vaccinations were recorded, although the vaccine coverage remained low in this population, with < 20% individuals receiving 1 or more doses prior to the emergence of the Omicron variant at both sites. We estimated Omicron BA.1/2's infection attack rate in each stratum as the total number of Omicron BA.1/2 infections within the stratum divided by the total stratum population. In Table 1, we summarize the site-specific infection attack

rate during the Omicron BA.1/2 epidemic wave, stratified by age, sex, household size, HIV infection status, and prior SARS-CoV-2 exposures. Here prior SARS-CoV-2 exposures include prior SARS-CoV-2 infection(s), SARS-CoV-2 vaccination, or both.

Increasing population immunity and changing immunologic landscape over four epidemic waves among the PHIRST-C cohort population

Infection attack rates of the Omicron BA.1/2 subvariants were substantially higher than those of previously circulating variants in both study sites. In the rural site, infection rates for pre-Omicron variants were 8.2% (95% CI 5.7–11.0%) for D614G, 20.7% (95% CI 17.1–24.3%) for Beta and 38.7% (95% CI 34.3–43.1%) for Delta, with higher attack rates in each successive wave (Fig. 2A, left panel). In the urban site, infection attack rates were more similar among pre-Omicron variants: 25.8% (95% CI 21.6–30.0%) for D614G, 32.1% (95% CI 27.7–36.5%) for Beta, 23.2% (95% CI 19.2–27.2%) for Delta (Fig. 2B, left panel). Differences in infection attack rates between the urban site and the rural site reflect the spatial heterogeneity of SARS-CoV-2 circulation in South Africa. Attack rates for the Omicron BA.1/2 subvariants reached 64.6% (95% CI 60.4–68.9%) in the rural site (Fig. 2A, left panel) and 58.1% (95% CI 53.4–62.8%) in the urban site (Fig. 2B, left panel).

Prior to the emergence of Omicron, in September 2021, the cumulative infection rate (including reinfections), was 67.6% (95% CI 61.4–73.8%) in the rural site and 81.0% (95% CI 73.8– 88.3%) in the urban site (Fig. 2A, B, right panels), with primary infections accounting for the majority of all infections in both sites (Fig. S5, left panels). In contrast, reinfections were predominant with Omicron BA.1/2, with primary infections representing only 37.9% (95% CI 32.5–43.3%) and 28.6% (95% CI 23.0–34.3%) of all Omicron BA.1/2 infections in the rural and urban site, respectively (Fig. S5, right panels). As a result, cohort participants had experienced an average of 1.32 (95%CI 1.25–1.40) and 1.39 (95% CI 1.31–1.48) SARS-CoV-2 infection episodes in the rural and the urban site, respectively, approximately 2 years into the pandemic (Fig. 2A, B).

Longitudinal rRT-PCR and serologic follow up at the individual level, combined with household-level random sampling scheme, provide a unique opportunity to track the history of SARS-CoV-2 immunizing events in the cohort populations, including vaccination (partially or fully) and infection. Fig. 2C, D are Sankey flow diagrams tracking the transition of different SARS-CoV-2 exposure histories after each of the four epidemic waves. The first three epidemic waves were dominated by primary exposures to D614G, Beta, and Delta variants. In addition, less than 20% of the population was vaccinated by either the Janssen Ad26.COV2.S or the Pfizer–BioNTech BNT162b2 vaccines prior to the emergence of Omicron. At the end of the third wave, only 30.1% and 22.0% of the population remained naïve to SARS-CoV-2 in the rural and the urban site, respectively. Most of the population had experienced a single SARS-CoV-2 exposure prior to Omicron's emergence (55.9% in the rural and 54.8% in the urban site), while a minority had two or more exposures (14.0% and 23.2% in the rural and urban site respectively, Fig. S5). This observation is in line with the finding of a durable immune protection conferred by prior infection and vaccination in the pre-Omicron era[10]. In contrast, the large contribution of re-infections during the Omicron wave shifted the population immune landscape towards a dominance of repeat exposures. 52.1% of individuals in rural site and 60.4% in the urban site had experienced more than one exposure, whether prior infection or vaccination, after the fourth epidemic wave. The high proportion of reinfections observed during the Omicron wave is in line with a high level of population immunity predating Omicron BA.1/2's arrival, combined with this variant's immune evasion properties (Fig. S5). As a result, the Omicron BA.1/2 wave left behind a heterogeneous immunological landscape, with population subgroups characterized by distinct exposure histories, and no exposure category accounting for more than 25% of the population (Fig. 2C, D).
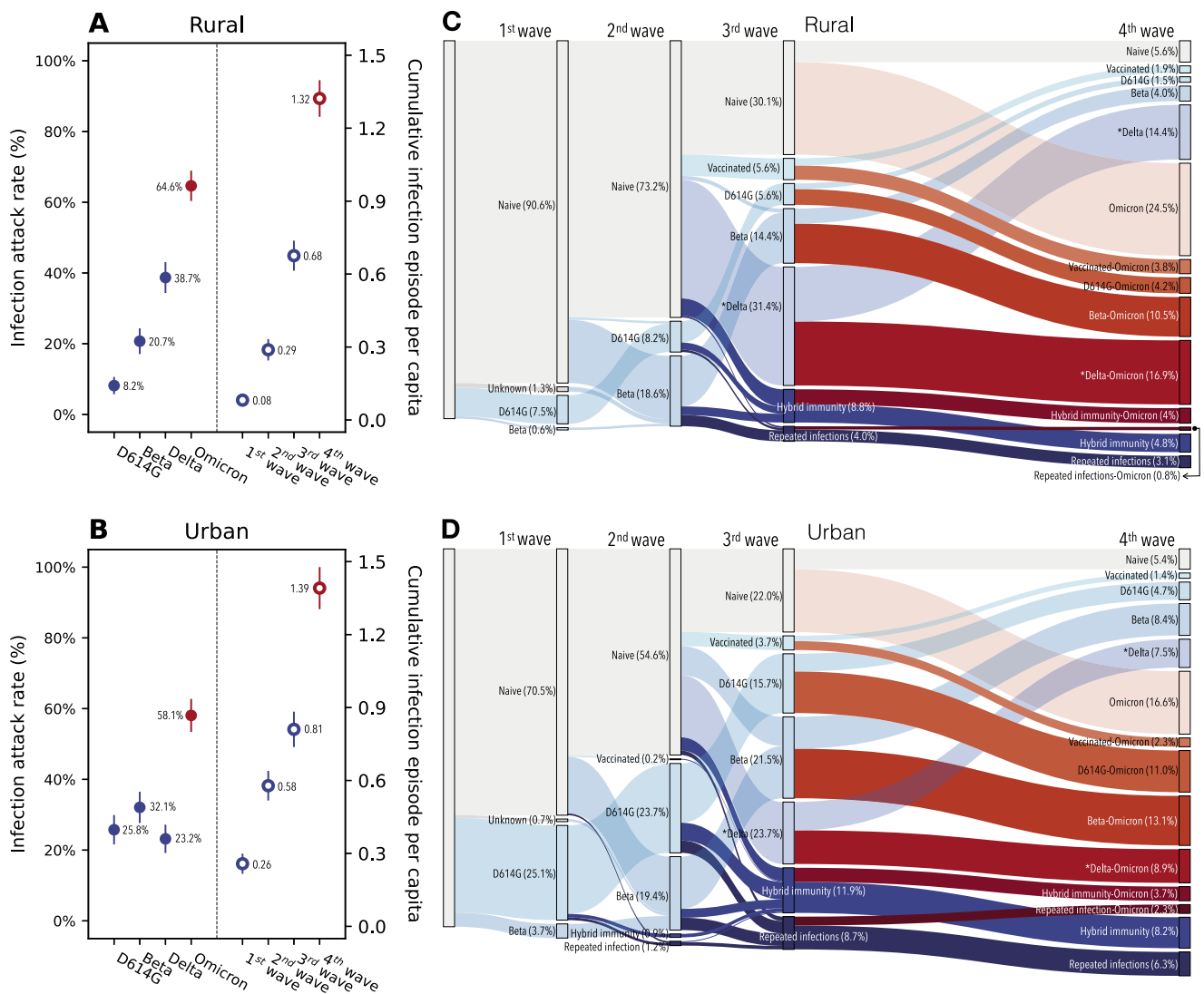
Fig. 2 | SARS-CoV-2 infection attack rates and shifts in immunologic landscape.
A Infection attack rates in the rural site by variant type (left) and the cumulative number of infection episodes per capita after each epidemic wave (right), based on n = 905 participants. Dots and lines represent mean and 95% confidence intervals. The end of the 1st wave is marked by blood draw 2, 2nd wave is marked by blood draw 5, 3rd wave is marked by blood draw 8, 4th wave is marked by blood draw 10. **B** Same as (**A**) but for the urban site. **C** Sankey diagram demonstrating the distribution of different type of immunologic exposures (including vaccination and infection) in the population of the rural site after each epidemic wave and the transition of immunologic exposures in-between waves. In the Sankey diagram, rectangular nodes of the same color represent proportion of population of a given immunologic state: gray color represents SARS-CoV-2 immunologic naïve

individuals; blue shades represent non-Omicron exposures; red shades represent Omicron exposures; darker colors represent repeat exposures while transparent shading represents primary exposures. Each column of nodes represented the distribution of immunologic state within the cohort population post a given epidemic wave. The vertical height of a node is proportional to the fraction of the population with the specific immunity. The band connecting nodes between waves represent the fraction of population (proportion to band width) transitioning from one immunologic state to another due to the impact of the epidemic wave of interest. **D** same as (**C**) but for the urban site. *In additional to Delta, here also includes other less frequent lineages including other lineages including Alpha and C.1.2 variants.

Risk factors, increased infectivity and immune evasion of the Omicron BA.1/2 variant, relative to Delta

Next, we fitted a chain-binomial household transmission model to the inferred serologic infections to contrast the characteristics of the Delta and Omicron variants, including the role of age, sex, household size, and prior exposure history (see Methods Section 4 for details). We further differentiated the risk of transmission of primary infections and vaccine breakthroughs or reinfections, as well as the risk of acquiring infection within the household and from the community (stratified by age, sex and cohort site). We found that the Omicron BA.1/2 variant was more than twice as likely to transmit as the Delta variant (odds ratio Omicron vs. Delta: 2.36, 95% CI 2.12–2.63), after

controlling for other risk factors (Fig. 3). Vaccine breakthroughs and reinfections were 41% less likely to transmit than primary infections (odds ratio 0.59, 95% CI 0.46–0.73), suggesting a transmission reduction effect of prior immunity, in line with other findings[12].

Prior work has demonstrated Omicron BA. 1/2's significant immune escape properties[13–17]. In our data (Fig. 3), the most recent prior infection with the Delta variant still conferred significant residual protection against Omicron (odds ratio 0.49, 95% CI 0.40–0.61), but protection declined with earlier infections: the odds ratio of prior Beta infection (2nd wave) was 0.66 (95% CI 0.52–0.84), while a prior D614G infection (1st wave) did not offer significant protection (0.87 (95% CI 0.63–1.21)). In contrast, prior Beta and D614G infections conferred a
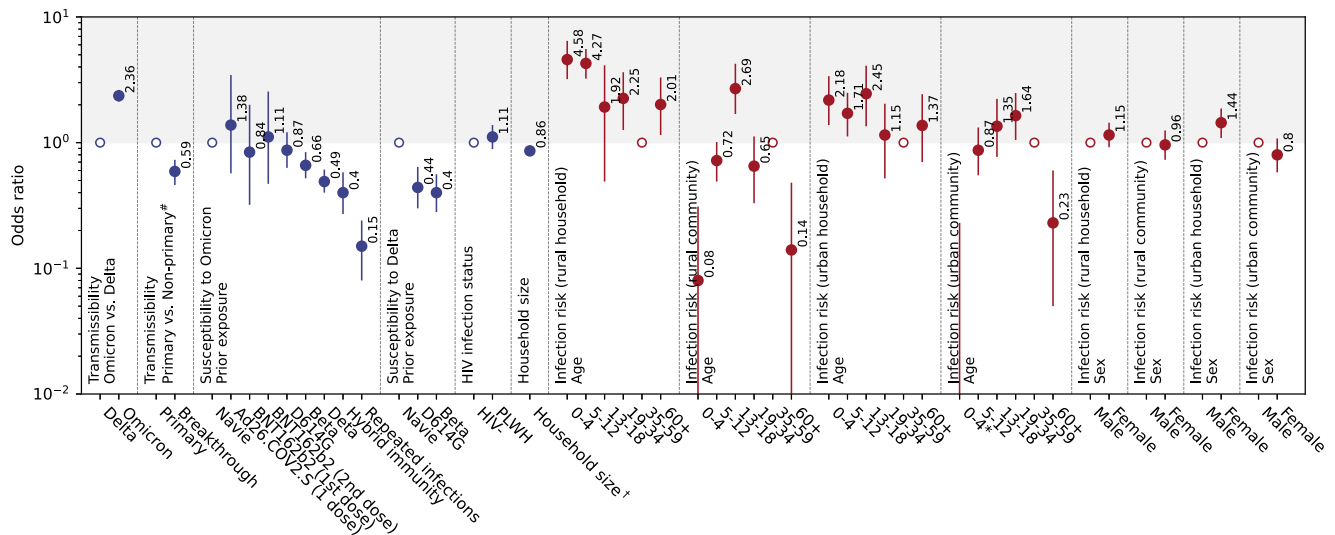
Fig. 3 | Risk factors associated with SARS-CoV-2 Omicron BA.1/2 and Delta infection. Odds ratios (adjusted after controlling for other risk factors, see Methods Section 4 for details) were estimated by a chain-binomial model fitted to the infection outcome of $n = 905$ participants, where the Omicron BA.1/2 and Delta infections was inferred by the serologic approach. Empty circles are reference classes. Solid dots and lines represent maximum likelihood estimate and 95% confidence intervals. Abbreviation: PLWH: persons living with HIV. Category "Unknown" for "HIV infection status" and category "Rest" in "Prior exposure" (Table 1) were included in the model but omitted here due to small sample size in the strata. *0-4 age group have odds ratio point estimate less than 0.01, thus not shown in the figure. #Non-primary infections represent repeat/breakthrough infections. †Household size denotes the number of household members within a household and is analyzed as a continuous variable.

higher level of protection against Delta (odds ratios: 0.40 (0.28−0.56) and 0.44 (0.30−0.64), respectively) than against Omicron. Neither Janssen Ad26.COV2.S (1 dose) nor Pfizer−BioNTech BNT162b2 vaccination (1 or 2 doses) showed significant protection against Omicron (Fig. 3), despite >90% being administered within 5 months of Omicron's emergence. It is also worth noting that < 5% of participants (43/905, Table 1) had received a complete vaccine schedule before Omicron, resulting in limited statistical power and wide confidence intervals on the effects of vaccination. However, hybrid immunity, defined as a combination of one prior infection (with any variant) with one or two doses of vaccines, conferred significant protection (odds ratio 0.40, 95% CI 0.27−0.58). Of note, repeat infections provided the strongest protection against the Omicron BA.1/2 variant, reducing the risk of infection by 85% (odds ratio 0.15, 95% CI 0.08−0.24). To explore how immune waning may also affect the risk of infection, we performed a sensitivity analysis by including an exponential waning term on the effects of prior exposure (Detailed in Methods Section 4). Waning did not significantly improve model fit, likely because it conflated with the period when specific variants circulated. While addition of a waning term decreased some of the variant-specific effects, the trends remained the same, with the earlier variant generating less protection against Omicron than Delta (Fig. S7). The model with waning also indicates a protection "half-life" of 229 days (95% CI 150−389 days), in line with the time scale of antibody waning described in other studies[18].

In terms of behavioral and demographic factors, we found that the risk of transmission to household contacts (per-contact risk) decreased significantly with the number of household members (odds ratio 0.86, 95% CI 0.84−0.89, Fig. 3). Interestingly, we found that the risk of acquiring infection within the household or from the community varied substantially by age, sex and study site. Pre-school children aged 0-4 years had much higher risk of acquiring infection within the household relative to adults 35−49 years (odds ratio, urban site: 2.18, 95%CI 1.38−3.38; odds ratio, rural site: 4.58, 95% CI 3.20−6.45) but much lower risk of acquiring infection from the community (odds ratio, urban site: < 0.001, 95% CI 0.00−0.23; odds ratio, rural site: 0.08, 95% CI 0.01−0.31, Fig. 3). Similarly, individuals 60 years and older had

very low risk of acquiring infection from the community (odds ratio, urban site: 0.23, 95% CI 0.05−0.60; odds ratio, rural site: 0.14, 95% CI 0.01−0.48). In the urban cohort, age group 19−34 years had the highest risk of acquiring infection from the community as compared to acquiring infection from the household (odds ratio: 1.64 with 95% CI 1.05−2.48), and so did age group 13−18 years in the rural site (odds ratio: 2.69 with 95% CI 1.69−4.24). Females had a significantly higher risk of infection within the household relative to males in the urban site (odds ratio female vs. male: 1.44, 95% CI 1.09−1.87) but there was no difference in the rural site.

Comparing the disease severity of Omicron with that of earlier variants

Estimating the severity of Omicron BA.1/2 remains difficult due to profound changes in case reporting and the impact of prior exposures on clinical presentation, relative to prior pandemic waves. We compared our infection attack rate estimates with surveillance data to evaluate the extent of under-reporting[19]. The cumulative SARS-CoV-2 incidence rate reported by the surveillance system for the Omicron wave was 0.54 per 100 individuals in the health district of the rural site (Ehlanzeni District) and 0.76 per 100 in the urban site district (Dr. Kenneth Kaunda District). The infection ascertainment rate was estimated at 0.84% in the rural site district and 1.31% in the urban site district, considerably lower than the ascertainment rates of prior waves[10,11], indicating that the surveillance system captured only a very small fraction of all Omicron BA.1/2 infections.

In Fig. 4A, we estimate the infection fatality ratio (IFR) of each epidemic wave in the urban site of the study, which is more representative of South Africa's urbanized population. We used the in-hospital death rate reported to the COVID-19 National Hospital Surveillance[20] at the district level (as numerator, Fig. 4B) and the age-specific infection rates estimated in the PHIRST-C urban cohort (as denominator, Fig. 4C). We estimate that the IFR was 0.043% (95% CI 0.040−0.047%) during the Omicron BA.1/2 wave, significantly lower than in the prior three waves (0.15% (95% CI 0.13−0.17%) during the 1st wave dominated by D614G, 0.36% (95% CI 0.30−0.46%) during the 2nd wave dominated by Beta,
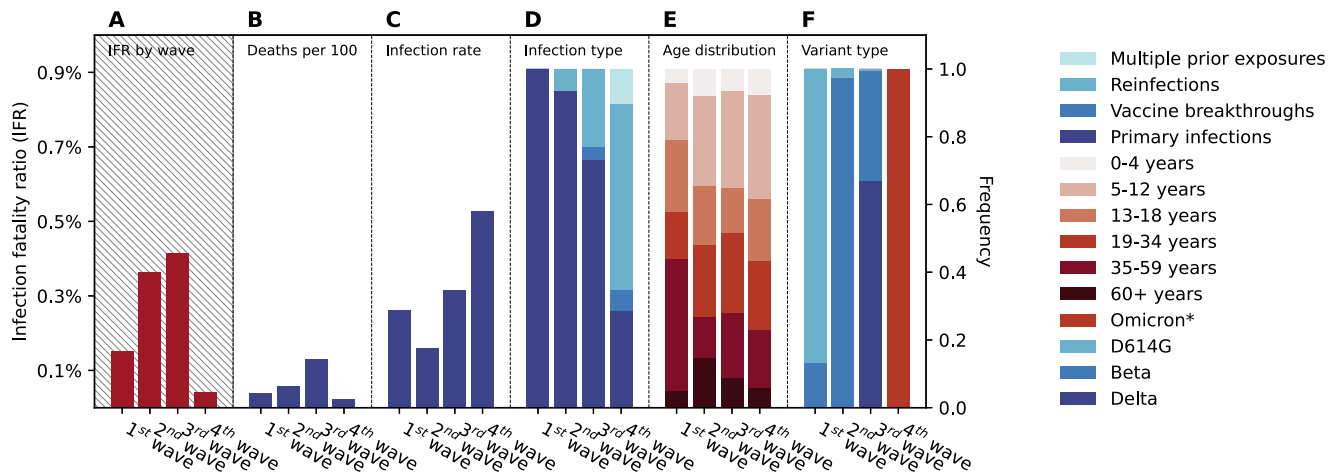
Fig. 4 | The infection fatality ratios and the factors associated with SARS-CoV-2 disease severity for different epidemic waves in the urban site's district. A The estimated infection fatality ratio for each epidemic wave. **B** The mortality burden of each epidemic wave measured by the cumulative rate of in-hospital deaths per 100 individuals. **C** The infection attack rate of each epidemic wave in the North West based on the PHIRST-C urban cohort, assuming that the urban cohort population is representative of the population of the North West Province. **D** The wave-specific distribution of infection types based on prior exposure histories, including primary infection, vaccine breakthroughs (1 or 2 doses of vaccines), reinfections (infection after one prior infection), and multiple prior exposures (infection with two or more prior infections or a mixture of prior infection and vaccination). **E** The wave-specific age distribution of infections. **F** The wave-specific distribution of variant type among infections. **B**–**F** Share the same axis on the right. *For the 4th wave, we could not confirm variant type by variant-specific rRT-PCR or sequencing, however, judging from the timing of emergence and dominance of Omicron in South Africa in late November 2021, we assumed here that all infections during the 4th wave were due to Omicron BA.1/2 variants.

0.41% (95% CI 0.37–0.47%) during the 3rd wave dominated by Delta).

In Fig. 4D–F, we deconstruct the infection profile of each epidemic wave to highlight the role of key factors affecting severity. Prior immunity has an important role in shaping IFR, with the fraction of primary infections decreasing with each new epidemic wave, from 100% during the 1st wave to 93%, and 73% and 28% during the Beta (2nd) and Delta (3rd) and Omicron BA.1/2 (4th) waves (Fig. 4D). Notable changes in the age patterns of infections in different waves have also affected the IFR (Fig. 4E). In particular, the second wave had the highest proportion of infections in individuals 60 years and older (15%), followed by the 3rd wave (9%), 4th wave (6%), and 1st wave (5%). Finally, each epidemic wave was dominated by a different variant (Fig. 4F), and variations in the intrinsic severity of each variant would also impact the observed variations in IFRs across different epidemic waves.

Discussion

Our findings indicate that the Omicron BA.1/2 wave had significantly higher attack rates than any of the previously circulating variants in South Africa. The Omicron BA.1/2 wave infected 58% and 65% of the population in our urban and rural cohorts respectively, despite high levels of pre-existing population immunity. Reinfections and vaccine breakthrough infections accounted for the majority of Omicron BA.1/2 infections, likely attributable to a minority of population remaining naïve prior to the Omicron BA.1/2 wave and to Omicron BA.1/2's immune-escape properties[16,21,22]. Additionally, a household transmission model estimated that participants were more than twice as likely to get infected during the Omicron BA.1/2 wave than the Delta wave, after adjustment for prior immunity and other factors. These findings support that the fitness advantage of Omicron BA.1/2 over Delta was not solely due to immune escape but also higher intrinsic transmissibility[23]. However, relaxing nonpharmaceutical interventions during the Omicron wave likely also contributed to the higher infection risk during the Omicron BA.1/2 wave (vs. Delta wave), as prior to the 4th epidemic wave (October 1, 2021), the national COVID-19 measures were tuned down to Alert Level 1, the lowest level in South Africa's COVID-19 alert system[24]. It is worth noting that our serologic

estimates of Omicron infection attack rates and the large contribution of reinfections and vaccine breakthroughs are in close agreement with earlier model projections for South Africa[10] in the urban site, the infection attack rate was estimated at 58.1% by serology vs. 44–81% by model projections, while the proportion of reinfections was 71% by serology vs. 49–72% by model projections. These projections were based on a dynamic transmission model calibrated to the PHIRST-C urban site population and solely relied on epidemiological evidence of Omicron's growth advantage and immune protection reduction[10]. The concordance between model projections and post-projection serologic surveys highlights how mathematical modelling can synthetize routine surveillance data and detailed cohort information to anticipate epidemic size and other dynamical features of public health relevance, months before serologic surveys become available.

A unique feature of our cohort populations includes intense monitoring since the first SARS-CoV-2 epidemic wave, with nasal swab collection at a twice weekly frequency, combined with frequent serum specimen collection[10]. Based on variant-typed respiratory specimens during the intense follow-up period[9,10], serologic testing, as well as information on vaccine administration, we had the unusual opportunity to track each individual participant's SARS-CoV-2 antigen exposure history in chronological order. Random sampling of households ensured that the immune exposure distribution within these cohorts reflected the immunologic landscape of the broader population at the study sites. After the third epidemic wave dominated by Delta, and prior to Omicron's emergence, more than 60% of the population at both sites had been infected by and/or vaccinated against SARS-CoV-2 at least once, with limited occurrence of reinfections (Fig. 2, Fig. S5). While there are different hypotheses addressing the evolutionary origins of Omicron BA.1's emergence and this variant's unusual number of mutations[25], the rising level of SARS-CoV-2 exposure in South Africa in 2020–2021 would be expected to promote the fitness advantage of immune escape variants over earlier variants. Globally, especially in high income countries, a similar shift in the immunologic landscape appears to have occurred, mainly via increased vaccine coverage, allowing Omicron to out-compete other variants and rapidly reach global dominance within 3 months of initial detection in Southern Africa[5]. Our analysis suggests that even after controlling for prior

exposure, the transmissibility of SARS-CoV-2 during the Omicron wave is substantially higher than that of the Delta wave (Fig. 3). This could be due to Omicron's enhanced infectiousness over Delta (higher intrinsic transmissibility due to biological properties), further relaxation of the non-pharmaceutical interventions during the Omicron wave (Fig. S8), or a combination of both.

After the Omicron BA.1/2 epidemic wave in South Africa, only 5.7% and 5.4% of the cohort population remained naïve to SARS-CoV-2 in the rural and urban sites respectively, and past immune histories in the rest of the population were highly diverse. Remarkably, the immunologic landscape of the population has become highly fragmented, with no single SARS-CoV-2 exposure category representing more than 25% of the population (Fig. 2). For instance, among 379 individuals (across both study sites) who experienced a single SARS-CoV-2 exposure, 49.6% (188/379) were infected by the Omicron variant and the rest were exposed to pre-Omicron variants or vaccines. These two population groups primed by different antigens will likely have different antibody responses as the sera of individuals primed by pre-Omicron variants or vaccines poorly neutralize the Omicron BA.1/2 variant and vice versa[26,27]. Among individuals that were primed by different types of pre-Omicron variants or vaccines, differences in antibody responses are also expected, though they would be less prominent than the differences between individuals primed by the Omicron vs. pre-Omicron variants[28].

After the 4th wave, among individuals who experienced two or more SARS-CoV-2 exposures, most had experienced Omicron BA.1/2 reinfections or vaccine breakthroughs with Omicron (80.3%, Fig. 2C, D). As SARS-CoV-2 moves into the endemic phase, multiple antigen exposures will become the norm, and these complex exposure histories will define population immune pressure to select new variants. We can see some of this play out already in the rapid succession of BA.1/2 and BA.4/5 waves, with in vitro experimental and computational predictive approaches informing which mutations are likely to become successful. Deep mutational scanning experiments of the SARS-CoV-2 receptor binding domain suggest that further mutations of Omicron BA.1/2 at sites 452 and 486 could lead to further immune evasion against antibodies isolated from individuals who had pre-Omicron and Omicron BA.1 breakthrough infections[21,28,29]. Our study suggests that after the Omicron BA.1/2 wave, the population was dominated by Omicron BA.1/2 convalescent individuals (Omicron BA.1/2 breakthroughs/reinfections in particular). Hence, the strongest immune pressure would arise from these individuals. Strikingly, mutations L452R and F486V have appeared in Omicron BA.4/5 lineages and these variants have now caused a fifth epidemic wave in South Africa. Interestingly, repeat infections by two or more pre-Omicron variants accounted for only a small fraction of the population after the third wave in our data, but these individuals retained great protection against Omicron BA.1/2.

As a large fraction of the population in South Africa (and globally) have experienced two or more antigenically distinct exposures involving Omicron lineages, prior variants, and/or vaccination episodes, the potential impact of immune imprinting needs to be closely monitored. The effects of imprinting have already been observed, where, for instance, individuals primed by spike-based vaccines have a more spike-focused immune response, compared to individuals primed by pre-Omicron infections (in which case, priming also includes non-spike proteins)[30,31]. It will be important to evaluate how imprinting by Omicron and pre-Omicron spike antigens differentially affects the immune response. Several studies have demonstrated that the antibody response after Omicron BA.1 breakthrough infection is dominated by recall memory B-cells against conserved epitopes shared between pre-Omicron strains and Omicron BA.1/2. Over reliance on immune memory may impact the breadth and strength of antibody and B-cell immunity as new antigenically drifted variants arise[15,21,32,33]. With each new variant wave, and vaccine reformulation anticipated for later in 2022, imprinting could play a role in shaping future disease trajectories. Ensuring the continuation of long-term cohort studies is particularly key to answering this question and guiding future vaccine updates.

Comparing our serology-based infection estimates with surveillance data indicates that only the "tip of the iceberg" of all SARS-CoV-2 infections (less than 1%) were captured by routine surveillance during the Omicron BA.1/2 epidemic wave in South Africa. On the other hand, weekly case surveillance accurately captured the speed of Omicron's spread, and mathematical models calibrated to these data generated early projections of epidemic size that are comparable to serologic estimates[10]. Despite a particularly high rate of infection with the Omicron BA.1/2 wave, there were fewer in-hospital deaths reported during this wave than in all prior waves in South Africa, suggesting a marked reduction in the overall infection fatality ratio (Fig. 4). The lower estimated disease severity of Omicron BA.1/2 needs to be interpreted in the context of changes in infection demographics, circulating variant, and immunologic landscape across waves[34,35]. Firstly, the intrinsic severity of Omicron BA.1/2 could be lower than that of prior variants, especially when compared to Delta, as suggested by observational studies[36,37]. In addition, in vitro studies support Omicron's limited ability to infect the lung[38], which could be a potential mechanism for Omicron BA.1/2's observed attenuated severity. Secondly, the level of prior immunity was much higher during the Omicron BA.1/2 epidemic wave, as shown in our data. If prior immunity confers significant protection against SARS-CoV-2 deaths, as many studies show, then the high level of prior immunity at the start of the Omicron BA.1/2 wave would also contribute to the perceived attenuation of disease severity. It is also worth noting that "survival bias" could further underestimate the crude IFR, whereby frail individuals may have succumbed to earlier waves and left a healthier population at the start of the Omicron wave[39–41]. Finally, the age profile of infections could also influence the overall IFR, as the risk of SARS-CoV-2 death increases exponentially with age, with most COVID-19 deaths concentrated in the population 60 years and older[42]. The proportion of infections among individuals 60 years and older was 6.9% during the Omicron wave, intermediate between prior waves. Future cohort analyses of Omicron BA.1/2 disease severity need to carefully adjust for age, prior infection and vaccination, and potential observational biases.

Our study has several limitations. First, Omicron BA.1/2 infections were ascertained by serology without confirmation by rRT-PCR, and infections may have been misclassified due to imperfect sensitivity and specificity of the serologic approach. We trialed the approach during the Delta wave, by comparing serology with rRT-PCR results. We optimized sensitivity and specificity of the serologic approach based on the Youden's J statistics (Fig. S3). Among 237 PCR-confirmed primary Delta infections, only 11 were missed by our serologic approach (Fig. S9A). Of the 11 mis-classified Delta primary infections, 7 were non-responders measured by the Roche Elecsys Anti-SARS-CoV-2 nucleocapsid assay across all three blood draws, 2 were weak responders (highest readout < 10) and 2 had mis-classified trajectory. Thus, the serologic approach may have missed a small fraction of the primarily infected individuals who mounted a very weak antibody response. On the other hand, among 51 PCR confirmed Delta reinfections, 10 were missed by the serologic approach (Fig. S9B). This suggests the serologic approach was less sensitive in detecting reinfections and we may be underestimating reinfections during the Omicron wave. Refinement of the serologic analysis and/or improved serologic assays more sensitive to waning[43] could further reduce potential misclassification. Second, the lack of virologic samples hindered differentiation between the Omicron BA.1 and BA.2 lineages, which may have different epidemiological properties in terms of transmissibility, immune evasion, and disease severity. Both the urban and rural sites had low vaccination rates during the study period. As a result, our assessment of the impact of vaccination

remains uncertain due to limited sample size (Table 1), with wide confidence intervals estimated by the chain-binomial model (Fig. 3). Lastly, our study was concentrated on a relatively confined geographic scale and findings may not be representative of the broader population in South Africa.

In conclusion, our study suggests the massive scale of the Omicron BA.1/2 wave of infections in South Africa, orders of magnitude larger than the observations based on surveillance data. Reinfections and vaccine breakthrough infections dominated this wave due to combined effects of Omicron BA.1/2's immune evasion properties and a high proportion of the population with pre-existing immunity. The overall disease severity of Omicron BA.1/2 was much lower than that of pre-Omicron waves, reflecting the contribution of attenuated intrinsic severity of this variant, but also changes in the age patterns of infections, and the impact of prior immunity. The remarkably diverse population immunologic landscape left by successive waves of SARS-CoV-2 will affect the epidemiological success of future variants. Our study sites are uniquely located in a region which has witnessed the emergence of key SARS-CoV-2 variants including Beta, Omicron BA.1, BA.4, and BA.5. As we reach the endemic phase of SARS-CoV-2, longitudinal studies will continue to be important to monitor shifts in population immunologic landscape and provide important context for understanding the fitness advantage and adaptive evolution of current and future variants.

Methods
Cohort design and the timing of serum sample collection for the Omicron BA.1/2 wave
The study builds on prior work on the PHIRST-C cohorts (PHIRST-C is an acronym for Prospective Household study of SARS-CoV-2, Influenza and Respiratory Syncytial virus community burden, Transmission dynamics and viral interaction in South Africa, PHIRST-C, where "C" stands for COVID-19). 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 at:
- https://clinicaltrials.gov/ct2/show/NCT02519803
- https://clinicaltrials.gov/ct2/show/NCT05277298

Informed consent was obtained from all adult participants (aged ≥ 18 years), assent from children aged 7 to 17 years, and consent from a parent or guardian for children younger than 18 years before data collection. Participants receive grocery store vouchers of ZAR50 (USD 3) per visit to compensate for time required for specimen collection and interview. Details on these South African cohorts have been previously described elsewhere[9,10]. To briefly summarize, the PHIRST-C cohorts consist of 114 households (638 participants) in a rural site located in Agincourt, a rural community in Mpumalanga Province, and 108 households (557 participants) in an urban site located in Jouberton township, Matlosana, North West Province. These cohorts were subject to an intense period of follow-up, from July 2020 to August 2021, during which nasal swab specimens were collected twice weekly and a total of 7 blood draws (BDs) were collected at enrollment and then approximately every two months. We used these data to reconstruct the prior exposure history of each cohort individual before Omicron arose, and to calibrate a serology-based approach to infer infections during the Omicron period. After the intense follow-up period concluded, three additional BDs were collected at both study sites (Fig. 1), with BD 8 collected in mid-September 2021 (the end of Delta wave), BD 9 in mid-November 2021 (at the time Omicron BA1/2's emergence in Southern Africa, but prior to the rise in reported cases), and BD 10 in late March 2022 (at the end of South Africa's Omicron BA.1/2 wave). We

used these data to infer the Omicron infection status of each cohort participant.

Laboratory methods
The laboratory methods of the intense follow-up period have been described in detail in prior study by Cohen et al.[9]. Serum specimens collected at blood draws 8, 9, and 10 follow the same protocol as those for prior blood draws detailed in ref.[9]: briefly, serum specimens were collected using venous blood, centrifuged into serum separator tubes, refrigerated immediately and transported to the NICD laboratories. According to manufacturer instructions, aliquots of prespecified volume were tested for the presence of SARS-CoV-2 antibodies by the Roche Elecsys Anti-SARS-CoV-2 nucleocapsid (N) assay[44].

Inference of SARS-CoV-2 Omicron infections among PHIRS-C participants based on the antibody trajectory estimated from pre- and post-Omicron BA.1/2 (4th) wave blood draws
Overview. As the PHIRST-C intense follow-up period ended in August 2021, individuals were no longer tested by rRT-PCR at twice-weekly frequency, thus Omicron BA.1/2 infections in the cohort population can only be ascertained through serology. We borrowed a paired sera approach from prior influenza studies, where serum samples are collected before and after the epidemic wave, with the rise in antibody titers between the pre- and post-wave sera used as a marker of influenza infection. This approach has been the gold standard to ascertain influenza infection attack rate, as prior exposures were common in the population[45]. To identify SARS-CoV-2 primary infections and reinfections during the Omicron BA.1/2 wave at both sites, we relied on the serial serologic results of BD 8, 9, (pre-Omicron BA.1/2 wave) and BD 10 (post-Omicron BA.1/2 wave) by the Roche Elecsys Anti-SARS-CoV-2 nucleocapsid (N) assay (refer to as Roche anti-N here after). To capture both primary infections and reinfections occurring in-between BD 8, 9, and 10, rather than relying on a boost of pre-post season paired sera as the sole marker of infection[45], we finely categorize the serial BDs patterns based on the sequential seroconversion and boosting/waning patterns of the Roche anti-N assay readouts. To trial the approach, we relied on a study period surrounding the Delta wave where we had both periodic serology and twice-weekly Rt-PCR testing on all individuals. We matched BDs 8, 9, and 10 (4th wave dominated by Omicron) with BDs 5, 6 and 8 (3rd wave dominated by Delta) based on the similar timing of serum specimen collection with respect to the epidemic curves of these waves (Fig. 1). We identified the serial serologic patterns most strongly associated with rRT-PCR confirmed SARS-CoV-2 infections during the Delta wave. We then generalized the serial serologic patterns to BDs 8, 9, and 10 to infer Omicron BA.1/2 infections during the 4th epidemic wave.

Categorization of serial serologic patterns during the 3rd and 4th epidemic waves in South Africa
The Roche anti-N is a commercial assay that was calibrated to detect recent and prior SARS-CoV-2 infections, based on the level of serum antibody against the SARS-CoV-2 N protein. The assay cutoff index (COI) above or equal to 1 marks seropositivity, while a COI below 1 is deemed seronegative[44]. We assessed each participant's serial serologic trajectory from pre- and post-wave serum specimens, measured by Roche anti-N COIs. We used the assay seroconversion (anti-N COIs going from below 1 to above 1) as evidence of primary infection. We also used further rises in COI from a seropositive baseline (COI above 1) as a marker of reinfection, as a new exposure would be expected to generate anamnestic boosting of anti-N antibody levels above prior levels. We inferred SARS-CoV-2 primary infections and reinfections during the 3rd epidemic wave based on participants' serial serologic trajectory from pre- and post-wave serum specimen (measured in Roche anti-N COIs). We selected two BDs prior to the 3rd (Delta) epidemic wave (BD 5 and 6) and one BD at the end of the 3rd epidemic

wave (BD 8) (Fig. 1). We did not consider BD 7, because this offered an extra sampling time point that we would not have for the Omicron wave. We aimed to infer the occurrence of primary/repeat infections that occurred between May 1, 2021 (roughly the mid-point between BD 5 and BD 6) and BD 8, covering the majority of the 3rd epidemic wave. It's worth noting that we chose the mid-point between BD 5 and BD 6 as the start date of 3rd wave inference to mirror the timing of the Omicron emergence in South Africa during the 4th wave, which roughly falls in-between BD 8 and 9, in early November 2021.

As the majority of the 3rd wave infections occurred in-between BD 6 and BD 8 (Fig. 1), changes in the COIs from BD 6 to BD 8 are likely most informative of SARS-CoV-2 infections between the two blood draws. We categorized the sequential serologic patterns as follows, based on the Roche anti-N assay:

1. Seroconversion from BD 6 (seronegative, COI < 1) to BD 8 (ser-opositive, COI ≥ 1). This was likely due to a SARS-CoV-2 primary infection occurring between BD 6 and BD 8. However, sero-negativity at BD 6 could also correspond to a prior infection that had sero-reverted by BD 6. In this case, seroconversion from BD 6 to BD 8 would suggest a reinfection between BD 6 and BD8. If sero-reversion had occurred by BD6, then the individual should be seropositive at the earlier blood draw (BD5).

2. If a participant was already seropositive at BD 6, we hypothesized that a reinfection occurring between BD 6 and BD 8 would induce anamnestic "boosting" of the anti-N antibody level which could lead to an increase in BD 8's COI when compared to BD 6's. Evidence of reinfection is particularly strong if the COI of BD 6 was lower than the COI of BD 5, establishing a waning baseline prior to further exposure. In this case, serial COI is expected to follow a "V-shape" trajectory from BD5-8, as suggested by a prior study also using serial blood samples to detect reinfections[46]. Establishing a waning baseline is especially informative for the Roche anti-N assay. A prior study suggests that a small fraction of individuals can show a gradual increase in Roche anti-N COI up to 4 months after SARS-CoV-2 infection[47], likely due to this assay using a dual-antigen antibody detection method tuned for high avidity antibodies[48]. Thus "double boosting" at BD 5, BD 6 then BD 8 may not necessarily suggest reinfection, and a boosting threshold may be needed to differentiate strong boosting (most likely due to reinfection) from weak boosting (due to the nature of the Roche-N longitudinal kinetics or measurement error). We will formally test these serial serologic patterns on their predictability of SARS-CoV-2 infections through comparing them with rRT-PCR confirmed infections during the Delta wave (detailed below then in Methods Section 3.3).

Building on the above logic, we start by categorizing the serial Roche anti-N COIs of BD 5, 6, and 8 into eight crude categories A, B, C, D, E, F, G, and H based on seroconversion and boosting/waning of the serial BDs' COIs, specifically:

- Category A: Seronegative at each BD 5, 6, and 8. This pattern is consistent with no infection before and during Delta wave.
- Category B: Seropositive at BD 5 followed by decreasing COIs (waning) from BD 5 to BD 6 and further decline from BD 6 to BD 8. This suggests a prior infection before the Delta wave, and waning during Delta wave.
- Category C: Seropositive at BD 5 followed by a rise in COI from BD 5 to BD 6 (boosting) then a decline in COI from BD 6 to BD 8. This could be evidence of re-infection between BD 5 and BD 6; we further refine this logic to distinguish reinfection "boosting" from long-term COI increase due to infection prior to BD 5 or measurement noise (see below).
- Category D: Seronegative at BD 5 followed by seroconversion at BD 6 then declining COI from BD 6 to BD 8. This is evidence of a Delta infection between BD 5 and BD 6.

- Category E: Seropositive at BD 5 followed by rise in COI from BD 5 to BD 6 then further rise of COI from BD 6 to BD 8. This could be evidence of reinfection between BD 5 and BD 6 or reinfection between BD 6 and BD 8; we further refine this logic to distinguish reinfection "boosting" from long-term COI increase due to infection prior to BD 5 or measurement noise (see below).
- Category F: Seronegative at BD 5 followed by seroconversion at BD 6 and then boosting COI from BD 6 to BD 8. This could be evidence of primary infection between BD 5 and BD 6 and/or reinfection between BD 6 and BD 8; we further refine this logic to distinguish reinfection "boosting" from long-term COI increase due to infection prior to BD 6 or measurement noise (see below).
- Category G: Seropositive at BD 5 followed by waning COI from BD 5 to BD 6 and boosting COI from BD 6 to BD 8. This could be evidence of reinfection between BD 5 and BD 8; we further refine this logic to distinguish reinfection "boosting" from measurement noise (see below).
- Category H: Seronegative at BD 5 and BD 6 followed by ser-oconversion in-between BD6 and BD 8. This is evidence of primary infection between BD 6 and BD 8.

Figure S1A–H visualize the serologic trajectory of BD5, 6, and 8's Roche anti-N COIs for individuals in Categories A through H, respectively. The quantitative criteria for the crude serial serologic pattern categorizations are listed in Table S1 under "Categorization (crude)".

A further rise of Roche anti-N COIs among seropositive individuals could be a marker for SARS-CoV-2 reinfections. To identify the degree of boosting most concordant with reinfection, we refine the crude categories listed above into finer categories by introducing a hyper-parameter of boosting threshold $\gamma > 1$: in a seropositive individual, a further increase of COI above the threshold γ may indicate a stronger signal of anamnestic boosting in antibody level due to reinfection, while boosting below γ is consistent with the slow long-term rise of antibody level from prior infection measured by the Roche anti-N assay[47,48] and/or measurement noise (see next section for calibrating γ for the optimized sensitivity/specificity). This refined characterization of serial serologic patterns is applied to crude categories C, E, F, and G (categories with further boosting from seropositive baseline), specifically:

- Category C is further divided into two sub-categories C_0 and C_1, where C_1 requires $BD_6/BD_5 > \gamma$ and C_0 requires $BD_6/BD_5 \leq \gamma$.
- Category E is further divided into three sub-categories E_0, E_1, and E_2, where E_2 requires $BD_8/BD_6 > \gamma$, E_1 requires $BD_8/BD_6 \leq \gamma$ & $BD_6/BD_5 > \gamma$, and E_0 requires $BD_8/BD_6 \leq \gamma$ & $BD_6/BD_5 \leq \gamma$.
- Category F is further divided into to two sub-categories F_0 and F_1, where F_1 requires $BD_8/BD_6 > \gamma$ and F_0 requires $BD_8/BD_6 \leq \gamma$.
- Category G is further divided into two sub-categories G_0 and G_1, where G_1 requires $BD_8/BD_6 > \gamma$ and G_0 requires $BD_8/BD_6 \leq \gamma$.

The quantitative criteria for the refined subcategorization described above are also listed in Table S1 under the "Categorization (refined)" of the 3rd epidemic wave. Similarly, we categorized serologic patterns in BDs 8, 9, 10 (with BD 8 matching BD 5, BD 9 matching BD 6, and BD 8 matching BD 10) based on quantitative criteria listed in Table S2. Fig. S2A–H visualize, from Category A through H respectively, the serologic trajectory of BD8, 9, and 10's Roche anti-N COIs.

Calibrating γ and selecting serial serologic patterns associated with SARS-CoV-2 primary infections and reinfections among PHIRST-C cohorts during the 3rd epidemic wave in South Africa (intense follow-up period)

We optimized the γ value as well as serial serologic patterns that best differentiated between individuals with SARS-CoV-2 primary infections and reinfections from non-infected individuals. This analysis was based on infections occurring between the mid-point

of BDs 5 and 6 and BD 8, corresponding to the 3rd wave dominated by Delta, in the period where rRT-PCR was available. We scanned through γ values from 1 to 3 using a step size of 0.1. Under each γ value, we iterated through each of the 13 refined serial serology subcategories (A, B, C_0, C_1, D, E_0, E_1, E_2, F_0, F_1, G_0, G_1, H, detailed in Table S1). We calculated the net contribution to the Youden's J statistics if the pattern of interest was considered as a marker of infection (primary or reinfections) during the time window of interest (Youden's J = sensitivity + specificity − 1, the higher the value of Youden's J, the better the performance of the classification balancing both sensitivity and specificity). We then selected all serial serologic patterns with net-positive contribution to the Youden's J statistics as markers for infection, while the rest of the 13 serial serologic patterns were markers of non-infection. This gave the best Youden's J statistics under a given γ value. We found that at $\gamma = 1.4$, we arrived at the highest Youden's J statistics of 0.820 and good sensitivity (0.804) for reinfections (Fig. S3) with patterns E_2, F_1, G_1, H indicating SARS-CoV-2 infections while the rest (A, B, C_0, C_1, D, E_0, E_1, F_0, G_0) indicating absence of infections (Fig. S4 & Table S1). We generalize this categorization of infections (E_2, F_1, G_1, H) vs non-infections (A, B, C_0, C_1, D, E_0, E_1, F_0, G_0) to BDs 8, 9, 10 to infer infections during the 4th (Omicron BA.1/2) wave.

Chain-binomial SARS-CoV-2 household transmission model

Here we studied predictors and risk factors of SARS-CoV-2 infection during the Delta and the Omicron waves, in particular the contribution of prior immunity, demographic and medical characteristics, other infections in the household, and the risk of infection from the community. We used a chain-binomial household transmission model for SARS-CoV-2, as an extension of prior household models developed to study influenza transmission[49,50]. We jointly fitted the model to both the 3rd (Delta dominated) and 4th waves (Omicron BA.1/2), using serology-inferred infections as the outcome, as described in Section 3. Our model considers both community-acquired infections as well as multigenerational transmission within the household.

We denote P_c^j as the risk of individual j acquiring infection from outside the household (community); to model the risk of transmission between household contacts, we denote P_{hh}^{ij} as the risk of an infected household contact i infecting household contact j in household h. We can express P_c^j, P_h^{ij} as:

$$P_c^j = P_c \exp\left(\sum_k \alpha_k a_k + \sum_l \theta_m c_m\right); P_h^{ij} = P_{hh} \exp\left(\sum_k \alpha_k a_k + \sum_l \beta_l b_l\right)$$

(1)

where:
- P_c denotes the baseline SARS-CoV-2 community infection risk.
- P_{hh} denotes the baseline risk of SARS-CoV-2 transmission between an infected and another uninfected household contact.
- a_k denotes risk factors k shared by community-acquired infection and household transmission, including, variant-specific transmissibility, variant-specific susceptibility due to prior exposure history, and HIV infection status (see Fig. 3).
- b_l denotes household-specific risk factor l that could potentially influence household transmission, including transmissibility of primary vs breakthrough infections, household size, site-specific age & sex (see Fig. 3).
- c_m denotes community-specific risk factor m that could potentially influence the acquisition of infection from the community, including site-specific age & sex (see Fig. 3).

Let h denote a household, i an individual, and i_-^h an individual i who remained infection-free during the Delta/Omicron waves in household h, while i_+^h is an individual i who was infected in household h (as determined by serology). We can then express the probability of household member i escaping infection from all infected household contacts as:

$$e_{hh}^i = \prod_{\{j \neq i\}} (1 - P_h^{ij})$$

(2)

where $\{j \neq i\}$ represents all infected household contacts. We can also express the probability of household member i escaping infections from the community as:

$$e_c^i = \left(1 - P_c^i\right)$$

(3)

Thus, within household h, the likelihood of a household contact i being non-infected is given by:

$$l_-^i = e_c^i e_{hh}^i$$

(4)

And the likelihood of an individual i being infected is given by:

$$l_+^i = 1 - l_-^i$$

(5)

For household h, the loglikelihood of observing the infection status of all household contacts is given by:

$$\log\left(l^h\right) = \sum_{\{l_+^h\}} \log\left(l_+^i\right) + \sum_{\{l_-^h\}} \log\left(l_-^i\right)$$

(6)

The overall likelihood of the observations across all households is given by:

$$\log(L) = \sum_h \log\left(l^h\right)$$

(7)

We used maximum likelihood method to optimize the function of log(L), and obtain point estimates of $\{\alpha_k\}, \{\beta_l\}, \{\theta_m\}$. The full list of the risk factors $\{\alpha_k\}, \{\beta_l\}$, and $\{\theta_m\}$ can be found in Table S3. We performed a model selection analysis on the risk factors by incorporating the risk factors in a stepwise fashion and calculate the loglikelihood and the Akaike information criterion (AIC) for each model. The results of the model selection analysis are reported in Table S3, with "Model 0" representing null model having no risk factors included and "Model 9" representing the full model used in the main text. The estimates of "Model 9" are presented in Fig. 3.

As a sensitivity analysis for the Delta wave infection, in addition to fitting "Model 9" to Delta wave infections inferred solely by the serologic approach, we consider another scenario where Delta wave infections were either inferred by the serologic approach or confirmed by the rRT-PCR. Among 237 PCR confirmed primary Delta infections, only 11 were missed by our serologic approach (Fig. S9A). Of the 11 misclassified Delta primary infections, 7 were non-responders measured by the Roche Elecsys Anti-SARS-CoV-2 nucleocapsid assay across all three blood draws, 2 were weak responders (highest readout < 10) and 2 had mis-classified trajectory. Thus, the serologic approach may have missed a small fraction of the primarily infected individuals who mounted a very weak antibody response. On the other hand, among 51 PCR confirmed Delta reinfections, 10 were missed by the serologic approach (Fig. S9B). In the sensitivity analysis, we combined the 21 rRT-PCR confirmed Delta infections with the serologic inferred Delta infections and fitted it to "Model 9". The results of this sensitivity analysis were presented in Fig. S6.

Lastly, we consider a model explicitly incorporating the waning of immunity, indexed "Model 10". On top of "Model 9", which explicitly

consider the effect of waning. Specifically, if we denote $\{\alpha_{im}\}$ ($\{\alpha_{im}\} \subset \{\alpha_k\}$) as relative odds of infection associated with immunity due to prior exposure with respect to naïve individuals, for each $\alpha_{im} \in \{\alpha_{im}\}$, we can express α_{im} as $\alpha_{im} = \alpha_{im}^{t=0} \times \exp(-\ln 2 \frac{t}{\tau})$, where t denotes the time since prior infection/immunization, τ denotes the half-life of the relative odds, $\alpha_{im}^{t=0}$ denotes the relative odds of infection right after infection/immunization. We assume a constant waning rate for all types of prior exposures; thus "Model 10" introduce only one extra parameter τ over "Model 9". The estimates of "Model 10" were presented in Fig. S7.

For all models, we minimized the negative of the $\log(L)$, using the optimize.minimize function of python's scipy package[51] (python version 3.8.11, scipy version 1.7.1) with the "Broyden–Fletcher–Goldfarb–Shanno" algorithm[52]. The corresponding 95% confidence intervals were calculated based on Wilks statistic (likelihood ratio of 1.92).

Reporting summary
Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability
Aggregate data to reproduce the figures are available at https://doi.org/10.5281/zenodo.7260083. Individual-level data cannot be publicly shared because of ethical restrictions and the potential for identifying included individuals. Accessing individual participant data and a data dictionary defining each field in the dataset would require provision of protocol and ethics approval for the proposed use. To request individual participant data access, please submit a proposal to C.C. who will respond within 1 month of request. Upon approval, data can be made available through a data sharing agreement.

Code availability
Code to reproduce the figures, and the chain-binomial transmission model are available at https://doi.org/10.5281/zenodo.7260083.

References
1. Tracking SARS-CoV-2 variants, (available at https://www.who.int/activities/tracking-SARS-CoV-2-variants).
2. CoVariants, (available at https://covariants.org/).
3. Obermeyer, F. et al. Analysis of 6.4 million SARS-CoV-2 genomes identifies mutations associated with fitness. *Science* **376**, 1327–1332 (2022).
4. Telenti, A., Hodcroft, E. B., Robertson, D. L. The Evolution and Biology of SARS-CoV-2 Variants. *Cold Spring Harb. Perspect. Med.* **12** (2022), https://doi.org/10.1101/cshperspect.a041390.
5. the Nextstrain team, Genomic epidemiology of novel coronavirus - Global subsampling. *Nextstrain* https://nextstrain.org/ncov/gisaid/global (2022).
6. Tegally, H. et al. Emergence of a SARS-CoV-2 variant of concern with mutations in spike glycoprotein. *Nature* https://doi.org/10.1038/s41586-021-03402-9 (2021).
7. Viana, R. et al. Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa. *Nature* https://doi.org/10.1038/d41586-021-03832-5 (2022).
8. Tegally, H., Moir, M., Everatt, J., Giovanetti, M. & Scheepers, C. Emergence of SARS-CoV-2 Omicron lineages BA.4 and BA.5 in South Africa. *Nat. Med.* **28**, 1785–1790 (2022).
9. Cohen, C., Kleynhans, J., von Gottberg, A., McMorrow, M. L. & Wolter, N. SARS-CoV-2 incidence, transmission, and reinfection in a rural and an urban setting: results of the PHIRST-C cohort study, South Africa, 2020-21. *Lancet Infect. Dis.* **22**, 821–834 (2022).
10. Sun, K. et al. SARS-CoV-2 transmission, persistence of immunity, and estimates of Omicron's impact in South African population cohorts. *Sci. Transl. Med.* eabo7081 (2022).
11. Kleynhans, J. 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.* **27**, 3020–3029 (2021).
12. Prunas, O. et al. Vaccination with BNT162b2 reduces transmission of SARS-CoV-2 to household contacts in Israel. *Science* **375**, 1151–1154 (2022).
13. Altarawneh, H. N. et al. Effects of previous infection and vaccination on symptomatic omicron infections. *N. Engl. J. Med.* **387**, 21–34 (2022).
14. Andrews, N., Stowe, J., Kirsebom, F., Toffa, S. & Rickeard, T. Covid-19 vaccine effectiveness against the Omicron (B.1.1.529) Variant. *N. Engl. J. Med.* **386**, 1532–1546 (2022).
15. Richardson, S. I. et al. SARS-CoV-2 Omicron triggers cross-reactive neutralization and Fc effector functions in previously vaccinated, but not unvaccinated, individuals. *Cell Host Microbe* **30**, 880-886.e4–886.e4 (2022).
16. Cele, S., Jackson, L., Khoury, D. S., Khan, K. & Moyo-Gwete, T. Omicron extensively but incompletely escapes Pfizer BNT162b2 neutralization. *Nature* https://doi.org/10.1038/s41586-021-04387-1 (2021).
17. Garcia-Beltran, W. F. et al. Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity. *Cell* **184**, 2372–2383.e9 (2021).
18. Cohen, K. W. et al. Longitudinal analysis shows durable and broad immune memory after SARS-CoV-2 infection with persisting antibody responses and memory B and T cells. *Cell Rep. Med* **2**, 100354 (2021).
19. NICD, Weekly Epidemiological Brief. https://www.nicd.ac.za/diseases-a-z-index/disease-index-covid-19/surveillance-reports/weekly-epidemiological-brief/. (2020)
20. The National Institute For Communicable Diseases Of South Africa, "NICD National COVID-19 Hospital Surveillance" (The National Institute For Communicable Diseases Of South Africa, 2022), https://www.nicd.ac.za/wp-content/uploads/2022/02/NICD-COVID-19-Daily-Sentinel-Hospital-Surveillance-report-National-20220203.pdf.
21. Cao, Y. et al. Omicron escapes the majority of existing SARS-CoV-2 neutralizing antibodies. *Nature* **602**, 657–663 (2022).
22. McCallum, M. et al. Structural basis of SARS-CoV-2 Omicron immune evasion and receptor engagement. *Science* **375**, 864–868 (2022).
23. Pearson, C. A. B. et al. Bounding the levels of transmissibility & immune evasion of the Omicron variant in South Africa. *bioRxiv* https://doi.org/10.1101/2021.12.19.21268038 (2021).
24. About alert system. https://www.gov.za/covid-19/about/about-alert-system.
25. Kupferschmidt, K. Where did "weird" Omicron come from? *Science* **374**, 1179–1179 (2021).
26. Rössler, A., Knabl, L., von Laer, D. & Kimpel, J. Neutralization Profile after Recovery from SARS-CoV-2 Omicron Infection. *N. Engl. J. Med.* **386**, 1764–1766 (2022).
27. Mykytyn, A.Z. et al. Antigenic cartography of SARS-CoV-2 reveals that Omicron BA.1 and BA.2 are antigenically distinct. *Sci. Immunol.* eabq4450 (2022).
28. Allison J.Greaney, Tyler N. Starr, Jesse Bloom, Escape calculator for SARS-CoV-2 RBD. *Escape calculator for SARS-CoV-2 RBD*. https://jbloomlab.github.io/SARS2_RBD_Ab_escape_maps/escape-calc/.
29. Greaney, A. J., Starr, T. N. & Bloom, J. D. An antibody-escape estimator for mutations to the SARS-CoV-2 receptor-binding domain. *Virus Evol.* **8**, veac021 (2022).

30. Reynolds, C.J., Pade, C., Gibbons, J. M., Otter, A. D., Lin, K.-M. Immune boosting by B.1.1.529 (Omicron) depends on previous SARS-CoV-2 exposure. *Science*, eabq1841 (2022).

31. Blom, K. et al. Immune responses after omicron infection in triple-vaccinated health-care workers with and without previous SARS-CoV-2 infection. *Lancet Infect. Dis.* **22**, 943–945 (2022).

32. Park, Y.-J. et al. Imprinted antibody responses against SARS-CoV-2 Omicron sublineages. *bioRxiv* https://doi.org/10.1101/2022.05.08.491108 (2022). 2022.05.08.491108.

33. Quandt, J. et al. Omicron BA.1 breakthrough infection drives cross-variant neutralization and memory B cell formation against conserved epitopes. *Sci. Immunol.* eabq2427 (2022)

34. Sigal, A. Milder disease with Omicron: Is it the virus or the pre-existing immunity. *Nat. Rev. Immunol.* **22**, 69–71 (2022).

35. Sigal, A., Milo, R. & Jassat, W. Estimating disease severity of Omicron and Delta SARS-CoV-2 infections. *Nat. Rev. Immunol.* **22**, 267–269 (2022).

36. Wolter, N. et al. Early assessment of the clinical severity of the SARS-CoV-2 omicron variant in South Africa: a data linkage study. *Lancet* **399**, 437–446 (2022).

37. Lewnard, J. A. et al. Clinical outcomes associated with SARS-CoV-2 Omicron (B.1.1.529) variant and BA.1/BA.1.1 or BA.2 subvariant infection in southern California. *Nat. Med.* **28**, 1933–1943 (2022).

38. Meng, B., Abdullahi, A., Ferreira, I. A. T. M., Goonawardane, N. & Saito, A. Altered TMPRSS2 usage by SARS-CoV-2 Omicron impacts infectivity and fusogenicity. *Nature* **603**, 706–714 (2022).

39. Report on Weekly Deaths in South Africa. *South African Medical Research Council*. https://www.samrc.ac.za/reports/report-weekly-deaths-south-africa.(2022)

40. Wang, H., Paulson, K. R., Pease, S. A., Watson, S. & Comfort, H. Estimating excess mortality due to the COVID-19 pandemic: a systematic analysis of COVID-19-related mortality, 2020–21. *Lancet* **399**, 1513–1536 (2022).

41. Bradshaw, D., Dorrington, R., Laubscher, R., Groenewald, P., Moultrie, T. COVID-19 and all-cause mortality in South Africa – the hidden deaths in the first four waves. *S. Afr. J. Sci.* **118** (2022), https://doi.org/10.17159/sajs.2022/13300.

42. O'Driscoll, M. et al. Age-specific mortality and immunity patterns of SARS-CoV-2. *Nature* **590**, 140–145 (2021).

43. Stone, M. et al. Evaluation of commercially available high-throughput SARS-CoV-2 serologic assays for serosurveillance and related applications. *Emerg. Infect. Dis.* **28**, 672–683 (2022).

44. Elecsys® Anti-SARS-CoV-2. https://diagnostics.roche.com/us/en/products/params/elecsys-anti-sars-cov-2.html.

45. Cauchemez, S. et al. Influenza infection rates, measurement errors and the interpretation of paired serology. *PLoS Pathog.* **8**, e1003061 (2012).

46. Prete, C. A. Jr et al. Reinfection by the SARS-CoV-2 Gamma variant in blood donors in Manaus, Brazil. *BMC Infect. Dis.* **22**, 127 (2022).

47. Favresse, J. et al. Long-term kinetics of anti-SARS-CoV-2 antibodies in a cohort of 197 hospitalized and non-hospitalized COVID-19 patients. *Clin. Chem. Lab. Med* **59**, e179–e183 (2021).

48. Nakagama, Y. et al. Detecting waning serological response with commercial immunoassays: 18-month longitudinal follow-up of Anti-SARS-CoV-2 Nucleocapsid Antibodies. *Microbiol. Spectr.* e0098622 (2022).

49. Longini, I. M. Jr, Koopman, J. S., Monto, A. S. & Fox, J. P. Estimating household and community transmission parameters for influenza. *Am. J. Epidemiol.* **115**, 736–751 (1982).

50. Gordon, A. et al. Influenza transmission dynamics in urban households, Managua, Nicaragua, 2012-2014. *Emerg. Infect. Dis.* **24**, 1882–1888 (2018).

51. Virtanen, P. et al. SciPy 1.0 Contributors, SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).

52. Nocedal, J., Wright, S. J. *Numerical Optimization* (Springer New York).

Author contributions

K.S., S.T., J.K., Av.G., M.L.M., N.W., J.M., N.A.M., K.K., L.L., C.V., C.C. designed the experiments. C.C., J.K., and S.T. accessed and verified the underlying data. S.T., J.K., Av.G., M.L.M., N.W., J.N.B., J.M., M.C., N.A.M., K.K., L.L., Jd.T., T.M., C.C. collected the data and performed laboratory experiments. K.S., S.T., J.K., Av.G., M.L.M., N.W., J.N.B., J.M., M.C., N.A.M., K.K., L.L., Jd.T., T.M., C.V., and C.C. analyzed the data and interpreted the results. K.S., S.T., J.K., Av.G., C.V., and C.C. drafted the manuscript. All authors critically reviewed the article. All authors had access to all the data reported in the study.

Competing interests

C.C. has received grant support from Sanofi Pasteur, Advanced Vaccine Initiative, and payment of travel costs from Parexel. Av.G. has received grant support from Sanofi Pasteur, Pfizer related to pneumococcal vaccine, C.D.C. and the Bill & Melinda Gates Foundation. N.W. reports grants from Sanofi Pasteur and the Bill & Melinda Gates Foundation. N.A.M. 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. J.M. has received grant support from Sanofi Pasteur. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-022-35652-0.

Correspondence and requests for materials should be addressed to Kaiyuan Sun or Cheryl Cohen.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

[1]Division of International Epidemiology and Population Studies, Fogarty International Center, National Institutes of Health, Bethesda, MD, USA. [2]Centre for Respiratory Diseases and Meningitis, National Institute for Communicable Diseases of the National Health Laboratory Service, Johannesburg, South Africa. [3]School of Public Health, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa. [4]Influenza Division, Centers for Disease Control and Prevention, Atlanta, GA, USA. [5]School of Pathology, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa. [6]COVID-19 Response, Centers for Disease Control and Prevention, Atlanta, GA, USA. [7]Centre for HIV and STIs, National Institute for Communicable Diseases of the National Health Laboratory Service, Johannesburg, South Africa. [8]SAMRC Antibody Immunity Unit, University of the Witwatersrand, Johannesburg, South Africa. [9]Perinatal HIV Research Unit, University of the Witwatersrand, Johannesburg, South Africa. [10]Johns Hopkins University Center for TB Research, Baltimore, MD, USA. [11]MRC/Wits Rural Public Health and Health Transitions Research Unit (Agincourt), School of Public Health, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa. [12]These authors jointly supervised this work: Cécile Viboud, Cheryl Cohen. ✉e-mail: kaiyuan.sun@nih.gov; cherylc@nicd.ac.za

the PHIRST-C group

Jinal N. Bhiman [7,8], Amelia Buys[2], Maimuna Carrim[2,5], Cheryl Cohen [2,3,12]✉, Linda de Gouveia[2], Mignon du Plessis[2,5], Jacques D. du Toit [11], Francesc Xavier Gómez-Olivé[11], Kathleen Kahn [11], Kgaugelo Patricia Kgasago[9], Jackie Kleynhans [2,3], Retshidisitswe Kotane[2], Limakatso Lebina[9], Neil A. Martinson[9,10], Meredith L. McMorrow[6], Tumelo Moloantoa[9], Jocelyn Moyes[2,3], Stefano Tempia[2,3,4], Stephen Tollman[11], Anne von Gottberg [2,5], Floidy Wafawanaka[11] & Nicole Wolter [2,5]