COVID-19 pandemic dynamics in South Africa and epidemiological characteristics of three variants of concern (Beta, Delta, and Omicron)

Wan Yang[1*] and Jeffrey Shaman[2]

[1]Department of Epidemiology, [2]Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University, New York, NY, USA

*Correspondence to: wy2202@cumc.columbia.edu

Abstract

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) variants of concern (VOCs) have been key drivers of new coronavirus disease 2019 (COVID-19) pandemic waves. To better understand variant epidemiologic characteristics, here we apply a model-inference system to reconstruct SARS-CoV-2 transmission dynamics in South Africa, a country that has experienced three VOC pandemic waves (i.e. Beta, Delta, and Omicron). We estimate key epidemiologic quantities in each of the nine South African provinces during March 2020 – Feb 2022, while accounting for changing detection rates, infection seasonality, nonpharmaceutical interventions, and vaccination. Model validation shows that estimated underlying infection rates and key parameters (e.g., infection-detection rate and infection-fatality risk) are in line with independent epidemiological data and investigations. In addition, retrospective predictions capture pandemic trajectories beyond the model training period. These detailed, validated model-inference estimates thus enable quantification of both the immune erosion potential and transmissibility of three major SARS-CoV-2 VOCs, i.e., Beta, Delta, and Omicron. These findings help elucidate changing COVID-19 dynamics and inform future public health planning.

Keywords: COVID-19, SARS-CoV-2, variant of concern, immune evasion, transmissibility

INTRODUCTION

Since its emergence in late December 2019, the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has spread globally, causing the coronavirus disease 2019 (COVID-19) pandemic (1). In just two years, SARS-CoV-2 has caused several pandemic waves in quick succession in many places. Many of these repeated pandemic waves have been driven by new variants of concern (VOCs) or interest (VOIs) that erode prior immunity from either infection or vaccination, increase transmissibility, or a combination of both. However, while laboratory and field studies have provided insights into these epidemiological characteristics, quantifying the extent of immune erosion (or evasion) and changes to transmissibility for each VOC remains challenging.

Like many places, by February 2022 South Africa had experienced four distinct pandemic waves caused by the ancestral SARS-CoV-2 and three VOCs (Beta, Delta, and Omicron). However, South Africa is also unique in that the country had the earliest surge for two of the five VOCs identified to date – namely, Beta (2) and Omicron (3). To better understand the COVID-19 dynamics in South Africa and variant epidemiological characteristics, here we utilize a model-inference system similar to one developed for study of SARS-CoV-2 VOCs, including the Beta variant in South Africa (4). We use this system to reconstruct SARS-CoV-2 transmission dynamics in each of the nine provinces of South Africa from the pandemic onset during March 2020 to the end of February 2022 while accounting for multiple factors modulating underlying transmission dynamics. We then rigorously validate the model-inference estimates using independent data and retrospective predictions. The validated estimates quantify the immune erosion potential and transmissibility of three major SARS-CoV-2 variants, i.e., Beta, Delta, and Omicron, in South Africa. Our findings highlight several common characteristics of SARS-CoV-2 VOCs and the need for more proactive planning and preparedness for future VOCs, including development of a universal vaccine that can effectively block SARS-CoV-2 infection as well as prevent severe disease.

RESULTS

Model fit and validation

The model-inference system uses case and death data to reconstruct the transmission dynamics of SARS-CoV-2, while accounting for under-detection of infection, infection seasonality, implemented nonpharmaceutical interventions (NPIs), and vaccination (see Methods). Overall, the model-inference system is able to fit weekly case and death data in each of the nine South African provinces (Fig 1A and Fig S1). We first validated the model-inference estimates using three independent datasets. First, we used serology data. We note that early in the pandemic serology data may reflect underlying infection rates but later, due to waning antibody titers and reinfection, likely underestimate infection. Compared to seroprevalence measures taken at multiple time points in each province, our model estimated cumulative infection rates roughly match corresponding serology measures and trends over time; as expected, model estimates were higher than serology measures taken during later months (Fig 1B). Second, compared to hospital admission data, across the nine provinces, model estimated infection numbers were well correlated with numbers of hospitalizations for all four pandemic waves caused by the ancestral, Beta, Delta, and Omicron variants, respectively ($r > 0.75$, Fig S2 A-D). Third, model-estimated infection numbers were correlated with age-adjusted excess mortality for both the ancestral and Delta wave ($r = 0.86$ and 0.61, respectively; Fig S2 A and C). For the Beta wave, after excluding Western Cape, a province with a very high hospitalization rate but low excess mortality during this wave (Fig S2 B), model-estimated infection numbers

were also correlated with age-adjusted excess mortality for the remaining provinces (r = 0.55; Fig S2 B). For the Omicron wave, like many other places, due to prior infection and/or vaccination (5, 6), mortality rates decoupled from infection rates (Fig S2 D). Overall, comparisons with the three independent datasets indicate our model-inference estimates align with underlying transmission dynamics.

In addition, as a fourth model validation, we generated retrospective predictions of the Delta and Omicron waves at two key time points, i.e. 2 weeks and 1 week, separately, before the observed peak of cases (approximately 3 to 5 weeks before the observed peak of deaths; Fig 2). To accurately predict a pandemic wave caused by a new variant, the model-inference system needs to accurately estimate the background population characteristics (e.g., population susceptibility) before the emergence of the new variant, as well as changes in population susceptibility and transmissibility due to the new variant. This is particularly challenging for South Africa, as the pandemic waves there tended to progress quickly, with cases surging and peaking within 3 to 7 weeks before declining. As a result, often only 1 to 6 weeks of new variant data were available for model-inference before generating the prediction. Despite these challenges, 1-2 weeks before the case peak and 3-5 weeks before the observed death peak, the model was able to accurately predict the remaining trajectories of cases and deaths in most of the nine provinces for both the Delta and Omicron waves (Fig 2 for the four most populous provinces and Fig S3 for the remainder). These accurate model predictions further validate the model-inference estimates.

Pandemic dynamics and key model-inference, using Gauteng province as an example

Next, we use Gauteng, the province with the largest population, as an example to highlight pandemic dynamics in South Africa thus far and develop key model-inference estimates (Fig 3 for Gauteng and Figs S4-S11 for each of the other eight provinces). Despite lower cases per capita than many other countries, infection numbers in South Africa were likely much higher due to under-detection. For Gauteng, the estimated infection-detection rate during the first pandemic wave was 4.59% (95% CI: 2.62 – 9.77%), and increased slightly to 6.18% (95% CI: 3.29 – 11.11%) and 6.27% (95% CI: 3.44 – 12.39%) during the Beta and Delta waves, respectively (Table S1). These estimates are in line with serology data. In particular, a population-level sero-survey in Gauteng found 68.4% seropositivity among those unvaccinated at the end of the Delta wave (7). Combining the reported cases at that time (~6% of the population size) with undercounting of infections in sero-surveys due to sero-reversions and reinfections suggests that the overall detection rate would be less than 10%.

Using our inferred under-detection (Fig 3E), we estimate that 32.83% (95% CI: 15.42 - 57.59%, Table S2) of the population in Gauteng were infected during the first wave, predominantly

during winter when more conducive climate conditions and relaxed public health restrictions existed (see the estimated seasonal and mobility trends, Fig 3A). This high infection rate, while with uncertainty, is in line with serology measures taken in Gauteng at the end of the first wave (ranging from 15% to 27% among 6 sero-surveys during November 2020; Fig 1B) and a study showing 30% sero-positivity among participants enrolled in the Novavax NVX-CoV2373 vaccine phase 2a-b trial in South Africa during August – November 2020 (8).

With the emergence of Beta, another 21.87% (95% CI: 12.16 – 41.13%) of the population in Gauteng – including reinfections – is estimated to have been infected, even though the Beta wave occurred during summer under less conducive climate conditions for transmission (Fig 3A). Consistent with laboratory studies showing low neutralizing ability of convalescent sera against Beta (9, 10), the model-inference system estimates a large increase in population susceptibility with the surge of Beta (Fig 3D). In addition to this immune erosion, an increase in transmissibility is also evident for Beta, after accounting for concurrent NPIs and infection seasonality (Fig 3C). Notably, in contrast to the large fluctuation of the time-varying effective reproduction number over time (R_t, Fig 3B), the transmissibility estimates are more stable and reflect changes in variant-specific properties. Further, consistent with in-depth epidemiological findings (11), the estimated overall infection-fatality risk for Beta was about twice as high as the ancestral SARS-CoV-2 (0.19% [95% CI: 0.10 - 0.33%] vs. 0.09% [95% CI: 0.05 - 0.20%], Fig 3F and Table S3). Nonetheless, these estimates are based on documented COVID-19 deaths and are likely underestimates.

With the introduction of Delta, a third pandemic wave occurred in Gauteng during the 2021 winter. The model-inference system estimates a 49.82% (95% CI: 25.22 – 90.79%) attack rate by Delta, despite the large number of infections during the previous two waves. This large attack rate was possible, due to the high transmissibility of Delta, as reported in multiple studies (12-16), the more conducive winter transmission conditions (Fig 3A), and the immune erosion from Delta relative to both the ancestral and Beta variants (17-19).

Due to these large pandemic waves, prior to the detection of Omicron in Gauteng, estimated cumulative infection numbers surpassed the population size (Fig 4B), indicating the large majority of the population had been infected and some more than once. With the rise of Omicron, the model-inference system estimates a very large increase in population susceptibility (Fig 3D), as well as an increase in transmissibility (Fig 3C); however, unlike previous waves, the Omicron wave progresses much more quickly, peaking 2-3 weeks after initiating marked exponential growth. These estimates suggest that several additional factors may have also contributed to the observed dynamics, including changes to the infection-detection rate (Fig 3E), a summer seasonality increasingly suppressing transmission as the wave

progressed (Fig 3A), as well as a slight change in population mobility suggesting potential behavior changes (Fig 3A). By the end of February 2022, the model-inference system estimates a 44.49% (95% CI: 19.01 – 75.30%) attack rate, with only 4.26% (95% CI: 2.46 – 9.72%) of infections detected as cases, during the Omicron wave in Gauteng. In addition, consistent with the reported 0.3 odds of severe disease compared to Delta infections (6), estimated overall infection-fatality risk during the Omicron wave was about 30% of that during the Delta wave in Gauteng (0.03% [95% CI: 0.02 – 0.06%] vs. 0.11% [95% CI: 0.06 – 0.21%], based on documented COVID-19 deaths; Table S3).

Model inferred epidemiological characteristics across the nine provinces in South Africa

Across all nine provinces in South Africa, the pandemic timing and intensity varied (Fig 4 A-C). In addition to Gauteng, high cumulative infection rates during the first three pandemic waves are also estimated for Western Cape and Northern Cape (Fig 1 C-E, Fig 4B and Table S2). Overall, all nine provinces likely experienced three large pandemic waves prior to the growth of Omicron; estimated average cumulative infections ranged from 60% of the population in Limpopo to 122% in Northern Cape (Fig 4B). Corroboration for these cumulative infection estimates is derived from mortality data. Excess mortality before the Omicron wave was as high as 0.47% of the South African population by the end of November 2021 (20), despite the relatively young population (median age: 27.6 years (21) vs. 38.5 years in the US (22)) and thus lower expected infection-fatality risk (23, 24). Assuming an infection-fatality risk of 0.5% (similar to estimates in (25) for South Africa), these excess deaths would convert to a 94% infection rate.

Combining these model-inference estimates during each wave in each province, we estimate that Beta eroded immunity among 63.4% (95% CI: 45.0 – 77.9%) of individuals with prior ancestral SARS-CoV-2 infection and was 34.3% (95% CI: 20.5 – 48.2%) more transmissible than the ancestral SARS-CoV-2. These estimates for Beta are consistent across the nine provinces (Fig 4D, 1st column), as well as with our previous estimates using national data for South Africa (4). Additional support for the high immune erosion of Beta is evident from recoverees of ancestral SARS-CoV-2 infection who were enrolled in the Novavax NVX-CoV2373 vaccine phase 2a-b trial (8) and found to have a similar likelihood of COVID-19, mostly due to Beta, compared to those seronegative at enrollment.

Estimates for Delta vary across the nine provinces (Fig 4D, 2nd column), given the more diverse population immune landscape among provinces after two pandemic waves. Overall, we estimate that Delta eroded 24.5% (95% CI: 0 – 53.2%) of prior immunity (gained from infection by ancestral SARS-CoV-2 and/or Beta, and/or vaccination) and was 47.5% (95% CI: 28.4 – 69.4%) more transmissible than the ancestral SARS-CoV-2. Consistent with this finding, and in

particular the estimated immune erosion, studies have reported a 27.5% reinfection rate during the Delta pandemic wave in Delhi, India (17) and reduced ability of sera from Beta-infection recoverees to neutralize Delta (18, 19).

For Omicron, estimates also vary by province but still consistently point to its higher transmissibility than all previous variants (Fig 4D, 3rd column). Overall, we estimate that Omicron is 94.0% (95% CI: 73.5 – 121.5%) more transmissible than the ancestral SARS-CoV-2. This estimated transmissibility is higher than Delta and consistent with *in vitro* and/or *ex vivo* studies showing Omicron replicates faster within host than Delta (26, 27). In addition, we estimate that Omicron eroded 54.1% (95% CI: 35.8 – 70.1%) of immunity due to all prior infections and vaccination. Importantly, we note that the estimate for immune erosion is not directly comparable across variants, as it is relative to the combined population immunity accumulated until the rise of each variant. In the case of Beta, it is immunity accumulated from the first wave via infection by the ancestral SARS-CoV-2. In the case of Omicron, it includes immunity from prior infection and refection of any of the previously circulating variants as well as vaccination. Thus, the estimate for Omicron may represent a far broader capacity for immune erosion than was evident for Beta. Supporting the suggestion of broad-spectrum immune erosion of Omicron, studies have reported low neutralization ability of convalescent sera from infections by all previous variants (28, 29), as well as high attack rates among vaccinees in several Omicron outbreaks (30, 31).

DISCUSSION

Using a comprehensive model-inference system, we have reconstructed the pandemic dynamics in each of the nine provinces of South Africa. Uncertainties exist in our findings, due to incomplete and varying detection of SARS-CoV-2 infections and deaths, changing population behavior and public health interventions, and changing circulating variants. To address these uncertainties, we have validated our estimates using three datasets not used by our model-inference system (i.e., serology, hospitalization, and excess mortality data; Fig 1B and Fig S2) as well as retrospective prediction (Fig 2 and Fig S4). In addition, as detailed in the Results, we have showed that estimated underlying infection rates (Fig 1B and Fig S2) and key parameters (e.g., infection-detection rate and infection-fatality risk) are in line with other independent epidemiological data and investigations. The detailed, validated model-inference estimates thus allow quantification of both the immune erosion potential and transmissibility of three major SARS-CoV-2 VOCs, i.e., Beta, Delta, and Omicron.

Specifically, we make three general observations. First, high prior immunity does not preclude new outbreaks, as neither infection nor current vaccination is sterilizing. As shown in South Africa, even with the high infection rate accumulated from preceding waves, new waves can

occur with the emergence or introduction of new variants. Around half of South Africans are estimated to have been infected after the Beta wave, yet the Delta variant caused a third large pandemic wave, followed by a fourth wave with comparable infection rates by Omicron (Fig 4B and Table S2). Second, large numbers of deaths can still occur in later waves with large infection surges, even though prior infection may provide partial protection and to some extent temper disease severity. This is evident from the large Delta wave in South Africa, which resulted in 0.2% excess mortality (vs. 0.08% during the first wave and 0.19% during the Beta wave (20)). Together, the continued transmission and potential severe outcomes highlight the importance of continued preparedness and prompt public health actions as societies learn to live with SARS-CoV-2.

Third, multiple SARS-CoV-2 VOCs/VOIs have emerged in the two years since pandemic inception. It is challenging to predict the frequency and direction of future viral mutation, in particular, the level of immune erosion, changes in transmissibility, and innate severity. Nonetheless, given high exposure and vaccination in many populations, variants capable of eroding a wide spectrum of prior immunity (i.e., from infection by multiple preexisting variants and vaccination) would have a greater chance of causing new major outbreaks. Indeed, except for the Alpha variant, the other four important VOCs (i.e. Beta, Gamma, Delta, and Omicron) all produced some level of immune erosion. In addition, later VOCs, like Delta and Omicron, appear to have been more genetically distinct from previous variants (32). As a result, they are likely more capable of causing re-infection despite diverse prior exposures and in turn new pandemic waves. Given this pattern, to prepare for future antigenic changes from new variants, development of a universal vaccine that can effectively block SARS-CoV-2 infection in addition to preventing severe disease (e.g. shown in (33)) is urgently needed (34).

The COVID-19 pandemic has caused devastating public health and economic burdens worldwide. Yet SARS-CoV-2 will likely persist in the future. To mitigate its impact, proactive planning and preparedness is paramount.

METHODS

Data sources and processing

We used reported COVID-19 case and mortality data to capture transmission dynamics, weather data to estimate infection seasonality, mobility data to represent concurrent NPIs, and vaccination data to account for changes in population susceptibility due to vaccination in the model-inference system. Provincial level COVID-19 case, mortality, and vaccination data were sourced from the Coronavirus COVID-19 (2019-nCoV) Data Repository for South Africa (COVID19ZA)(35). Hourly surface station temperature and relative humidity came from the Integrated Surface Dataset (ISD) maintained by the National Oceanic and Atmospheric

Administration (NOAA) and are accessible using the "stationaRy" R package (36, 37). We computed specific humidity using temperature and relative humidity per the Clausius-Clapeyron equation (38). We then aggregated these data for all weather stations in each province with measurements since 2000 and calculated the average for each week of the year during 2000-2020.

Mobility data were derived from Google Community Mobility Reports (39); we aggregated all business-related categories (i.e., retail and recreational, transit stations, and workplaces) in all locations in each province to weekly intervals. For vaccination, provincial vaccination data from the COVID19ZA data repository recorded the total number of vaccine doses administered over time; to obtain a breakdown for numbers of partial (1 dose of mRNA vaccine) and full vaccinations (1 dose of Janssen vaccine or 2 doses of mRNA vaccine), separately, we used national vaccination data for South Africa from Our World in Data (40, 41) to apportion the doses each day. In addition, cumulative case data suggested 18,586 new cases on Nov 23, 2021, whereas the South Africa Department of Health reported 868 (42). Thus, for Nov 23, 2021, we used linear interpolation to fill in estimates for each province on that day and then scaled the estimates such that they sum to 868.

Model-inference system

The model-inference system is based on our previous work estimating changes in transmissibility and immune erosion for SARS-CoV-2 VOCs including Alpha, Beta, Gamma, and Delta (4, 43). Below we describe each component.

Epidemic model

The epidemic model follows an SEIRSV (susceptible-exposed-infectious-recovered-susceptible-vaccination) construct per Eqn 1:

$$\begin{cases} \frac{dS}{dt} = \frac{R}{L_t} - \frac{b_t e_t m_t \beta_t IS}{N} - \varepsilon - v_{1,t} - v_{2,t} \\ \frac{dE}{dt} = \frac{b_t e_t m_t \beta_t IS}{N} - \frac{E}{Z_t} + \varepsilon \\ \frac{dI}{dt} = \frac{E}{Z_t} - \frac{I}{D_t} \\ \frac{dR}{dt} = \frac{I}{D_t} - \frac{R}{L_t} + v_{1,t} + v_{2,t} \end{cases} \quad (1)$$

where S, E, I, R are the number of susceptible, exposed (but not yet infectious), infectious, and recovered/immune/deceased individuals; N is the population size; and ε is the number of travel-imported infections. In addition, the model includes the following key components:

297

298 1) Virus-specific properties, including the time-varying variant-specific transmission rate β_t,
299 latency period Z_t, infectious period D_t, and immunity period L_t. Note all parameters are
300 estimated for each week (t) as described below.
301 2) The impact of NPIs. Specifically, we use relative population mobility (see data above) to
302 adjust the transmission rate via the term m_t, as the overall impact of NPIs (e.g., reduction
303 in the time-varying effective reproduction number R_t) has been reported to be highly
304 correlated with population mobility during the COVID-19 pandemic.(44-46) To further
305 account for potential changes in effectiveness, the model additionally includes a
306 parameter, e_t, to scale NPI effectiveness.
307 3) The impact of vaccination, via the terms $v_{1,t}$ and $v_{2,t}$. Specifically, $v_{1,t}$ is the number of
308 individuals successfully immunized after the first dose of vaccine and is computed using
309 vaccination data and vaccine effectiveness (VE) for 1st dose; and $v_{2,t}$ is the additional
310 number of individuals successfully immunized after the second vaccine dose (i.e., excluding
311 those successfully immunized after the first dose). In South Africa, around two-thirds of
312 vaccines administered during our study period were the mRNA BioNTech/Pfizer vaccine
313 and one-third the Janssen vaccine (47). We thus set VE to 20%/85% (partial/full
314 vaccination) for Beta, 35%/75% for Delta, and 10%/35% for Omicron based on reported VE
315 estimates (48-50).
316 4) Infection seasonality, computed using temperature and specific humidity data as described
317 previously (see supplemental material of Yang and Shaman(4)). Briefly, we estimated the
318 relative seasonal trend (b_t) using a model representing the dependency of the survival of
319 respiratory viruses including SARS-CoV-2 to temperature and humidity (51, 52). As shown
320 in Fig 2A, b_t estimates over the year averaged to 1 such that weeks with b_t >1 (e.g. during
321 the winter) are more conducive to SARS-CoV-2 transmission whereas weeks with b_t <1 (e.g.
322 during the summer) have less favorable climate conditions for transmission. The estimated
323 relative seasonal trend, b_t, is used to adjust the relative transmission rate at time t in Eqn 1.
324

325 *Observation model to account for under-detection and delay*
326 Using the model-simulated number of infections occurring each day, we further computed the
327 number of cases and deaths each week to match with the observations, as done in Yang et al
328 (53). Briefly, we include 1) a time-lag from infectiousness to detection (i.e., an infection being
329 diagnosed as a case), drawn from a gamma distribution with a mean of $T_{d,mean}$ days and a
330 standard deviation of $T_{d,sd}$ days, to account for delays in detection (Table S4); 2) an infection-
331 detection rate (r_t), i.e. the fraction of infections (including subclinical or asymptomatic
332 infections) reported as cases, to account for under-detection; 3) a time-lag from infectiousness
333 to death, drawn from a gamma distribution with a mean of 13-15 days and a standard deviation
334 of 10 days; and 4) an infection-fatality risk (IFR_t). To compute the model-simulated number of

new cases each week, we multiplied the model-simulated number of new infections per day by the infection-detection rate, and further distributed these simulated cases in time per the distribution of time-from-infectiousness-to-detection. Similarly, to compute the model-simulated deaths per week and account for delays in time to death, we multiplied the simulated-infections by the IFR and then distributed these simulated deaths in time per the distribution of time-from-infectious-to-death. We then aggregated these daily numbers to weekly totals to match with the weekly case and mortality data for model-inference. For each week, the infection-detection rate (r_t), the infection-fatality risk (IFR_t)., and the two time-to-detection parameters ($T_{d,mean}$ and $T_{d,sd}$) were estimated along with other parameters (see below).

Model inference and parameter estimation

The inference system uses the ensemble adjustment Kalman filter (EAKF (54)), a Bayesian statistical method, to estimate model state variables (i.e., *S*, *E*, *I*, *R* from Eqn 1) and parameters (i.e., β_t, Z_t, D_t, L_t, e_t, from Eqn 1 as well as r_t, IFR_t and other parameters from the observation model). Briefly, the EAKF uses an ensemble of model realizations (n=500 here), each with initial parameters and variables randomly drawn from a *prior* range (see Table S4). After model initialization, the system integrates the model ensemble forward in time for a week (per Eqn 1) to compute the prior distribution for each model state variable and parameter, as well as the model-simulated number of cases and deaths for that week. The system then combines the prior estimates with the observed case and death data for the same week to compute the posterior per Bayes' theorem (54). During this filtering process, the system updates the posterior distribution of all model variables and parameters for each week.

Estimating changes in transmissibility and immune erosion for each variant

As in ref (4), we computed the variant-specific transmissibility (R_{TX}) as the product of the variant-specific transmission rate (β_t) and infectious period (D_t). Note that R_t, the time-varying effective reproduction number, is defined as $R_t = b_t e_t m_t \beta_t D_t S/N = b_t e_t m_t R_{TX} S/N$. To reduce uncertainty, we averaged transmissibility estimates over the period a particular variant of interest was predominant. To find these predominant periods, we first specified the approximate timing of each pandemic wave in each province based on: 1) when available, genomic surveillance data; specifically, the onsets of the Beta wave in Eastern Cape, Western Cape, KwaZulu-Natal, and Northern Cape, were separately based on the initial detection of Beta in these provinces as reported in Tegally et al. (2); the onsets of the Delta wave in each of the nine provinces, separately, were based on genomic sequencing data from the Network for Genomic Surveillance South Africa (NGS-SA)(55); and 2) when genomic data were not available, we used the week with the lowest case number between two waves. The specified calendar periods are listed in Table S5. During later waves, multiple variants could initially co-circulate

before one became predominant. As a result, the estimated transmissibility tended to increase before reaching a plateau (see, e.g., Fig 2C). In addition, in a previous study of the Delta pandemic wave in India (43), we also observed that when many had been infected, transmissibility could decrease a couple months after the peak, likely due to increased reinfections for which onward transmission may be reduced. Thus, to obtain a more variant-specific estimate, we computed the average transmissibility ($\overline{R_{TX}}$) using the weekly R_{TX} estimates over the 8-week period starting the week prior to the maximal R_{tx} during each wave; if no maximum existed (e.g. when a new variant is less transmissible), we simply averaged over the entire wave. We then computed the change in transmissibility due to a given variant relative to the ancestral SARS-CoV-2 as $(\frac{\overline{R_{TX,variant}}-\overline{R_{TX,ancestral}}}{\overline{R_{TX,ancestral}}}) \times 100\%$.

To quantify immune erosion, similar to ref (4), we estimated changes in susceptibility over time and computed the change in immunity as ΔImm = $S_{t+1} - S_t + i_t$, where S_t is the susceptibility at time-t and i_t is the new infections occurring during each week-t. We sum over all ΔImm estimates for a particular location, during each wave, to compute the total change in immunity due to a new variant, $\Sigma\Delta Imm_v$. We then computed the level of immune erosion as the ratio of $\Sigma\Delta Imm_v$ to the model-estimated population immunity prior to the first detection of immune erosion, during each wave. That is, as opposed to having a common reference of prior immunity, here immune erosion for each variant depends on the state of the population immune landscape – i.e., combining all prior exposures and vaccinations – immediately preceding the surge of that variant.

For all provinces, model-inference was initiated the week starting March 15, 2020 and run continuously until the week starting February 27, 2022. To account for model stochasticity, we repeated the model-inference process 100 times for each province, each with 500 model realizations and summarized the results from all 50,000 model estimates.

Model validation using independent data

To compare model estimates with independent observations not assimilated into the model-inference system, we utilized three relevant datasets:

1) Serological survey data measuring the prevalence of SARS-CoV-2 antibodies over time. Multiple serology surveys have been conducted in different provinces of South Africa. The South African COVID-19 Modelling Consortium summarizes the findings from several of these surveys (see Fig 1A of ref (56)). We digitized all data presented in Fig 1A of ref (56) and compared these to corresponding model-estimated cumulative infection rates (computed mid-month for each corresponding month with a seroprevalence measure). Due to unknown survey methodologies and challenges adjusting for sero-reversion and

reinfection, we used these data directly (i.e., without adjustment) for qualitative comparison.

2) COVID-19-related hospitalization data, from COVID19ZA (35)/ We aggregated the total number of COVID-19 hospital admissions during each wave and compared these aggregates to model-estimated cumulative infection rates during the same wave. Of note, these hospitalization data were available from June 6, 2020 onwards and are thus incomplete for the first wave.
3) Age-adjusted excess mortality data from the South African Medical Research Council (SAMRC)(20). Deaths due to COVID-19 (used in the model-inference system) are undercounted. Thus, we also compared model-estimated cumulative infection rates to age-adjusted excess mortality data during each wave. Of note, excess mortality data were available from May 3, 2020 onwards and are thus incomplete for the first wave.

Model validation using retrospective prediction

As a fourth model validation, we generated model predictions at 2 or 1 week before the week of highest cases for the Delta and Omicron waves, separately, and compared the predicted cases and deaths to reported data unknown to the model. Predicting the peak timing, intensity, and epidemic turnaround requires accurate estimation of model state variables and parameters that determine future epidemic trajectories. This is particularly challenging for South Africa as the pandemic waves tended to progress quickly such that cases surged to a peak in only 3 to 7 weeks. Thus, we chose to generate retrospective predictions 2 and 1 weeks before the peak of cases in order to leverage 1 to 6 weeks of new variant data for estimating epidemiological characteristics. Specifically, for each pandemic wave, we ran the model-inference system up to 2 weeks (or 1 week) before the observed peak of cases, halted the inference, and used the population susceptibility and transmissibility of the circulating variant estimated at that time to predict cases and deaths for the remaining weeks. Because the infection detection rate and fatality risk are linked to observations of cases and deaths, changes of these quantities during the prediction period could obscure the underlying infection rate and accuracy of the prediction. Thus, for these two parameters specifically, we used model-inference estimates for corresponding weeks to allow comparison of model-predicted cases and deaths with the data while focusing on testing the accuracy of other key model estimates (e.g., transmissibility of the new variant). As for the model-inference, we repeated each prediction 100 times, each with 500 model realizations and summarized the results from all 50,000 ensemble members.

Data Availability: All data used in this study are publicly available as described in the "Data sources and processing" section.

Code availability: All source code and data necessary for the replication of our results and figures will be made publicly available on Github.

Acknowledgements: This study was supported by the National Institute of Allergy and Infectious Diseases (AI145883 and AI163023), the Centers for Disease Control and Prevention (CK000592), and a gift from the Morris-Singer Foundation.

Author contributions: WY designed the study (main), conducted the model analyses, interpreted results, and wrote the first draft. JS designed the study (supporting), interpreted results, and critically revised the manuscript.

Competing interests: JS and Columbia University disclose partial ownership of SK Analytics. JS discloses consulting for BNI.

Figure Legends:

Fig 1. Pandemic dynamics in South Africa, model-fit and validation using serology data. (A) Pandemic dynamics in each of the nine provinces (see legend); dots depict reported weekly numbers of cases and deaths; lines show model mean estimates (in the same color). (B) For validation, model estimated infection rates are compared to seroprevalence measures over time from multiple sero-surveys summarized in ref (56) (B). Boxplots depict the estimated distribution for each province (middle bar = mean; edges = 50% CrIs) and whiskers (95% CrIs). Red dots show corresponding measurements. Note that reported mortality was high in February 2022 in some provinces with no clear explanation.

Fig 2. Model validation using retrospective prediction. Model-inference was trained on cases and deaths data since March 15, 2020 up to 2 weeks (1st plot in each panel) or 1 week (2nd plot) before the Delta or Omicron wave (see timing on the x-axis); the model was then integrated forward using the estimates made at the time to predict cases (left panel) and deaths (right panel) for the remaining weeks of each wave. Blue lines and surrounding shades show model fitted cases and deaths for weeks before the prediction (line = median, dark blue area = 50% CrIs, and light blue = 80% CrIs). Red lines show model projected median weekly cases and deaths; surrounding shades show 50% (dark red) and 80% (light red) CIs of the prediction. For comparison, reported cases and deaths for each week are shown by the black dots; however, those to the right of the vertical dash lines (showing the start of each prediction) were not used in the model. For clarity, here we show 80% CIs (instead of 95% CIs, which tend to be wider for longer-term projections) and predictions for the four most populous provinces (Gauteng in A and B; KwaZulu-Natal in C and D; Western Cape in E and F; and Eastern Cape in G and H). Predictions for the other five provinces are shown in Fig S3.

Fig 3. Example model-inference estimates for Gauteng. (A) Observed relative mobility, vaccination rate, and estimated disease seasonal trend, compared to case and death rates over time. Key model-inference estimates are shown for the time-varying effective reproduction number R_t (B), transmissibility (C), population susceptibility (D), infection-detection rate (E), and infection-fatality risk (F). Grey shaded areas indicate the approximate circulation period for each variant. In (B) – (F), blue lines and surrounding areas show the estimated mean, 50% (dark) and 95% (light) CrIs; boxes and whiskers show the estimated mean, 50% and 95% CrIs for estimated infection rates. *Note that the transmissibility estimates (in C) have removed the effects of changing population susceptibility, NPIs, and disease seasonality; thus, the trends are more stable than the reproduction number (R_t in B) and reflect changes in variant-specific properties. Also note that infection-fatality risk estimates were based on reported COVID-19 deaths and may not reflect true values due to likely under-reporting of COVID-19 deaths.*

Fig 4. Model-inferred epidemiological properties for different variants across SA provinces. Heatmaps show (A) Estimated mean infection rates by week (x-axis) and province (y-axis), (B) Estimated mean *cumulative* infection numbers relative to the population size in each province, and (C) Estimated population susceptibility (to the circulating variant) by week and province. (D) Boxplots in the top row show the estimated distribution of changes in transmissibility for Beta, Delta, and Omicron, relative to the Ancestral SARS-CoV-2, for each province (middle bar = median; edges = 50% CIs; and whiskers =95% CIs); boxplots in the bottom row show, for each variant, the estimated distribution of immune erosion to all adaptive immunity gained from infection and vaccination prior to that variant. Red lines show the mean across all provinces.

References:

1. Koelle K, Martin MA, Antia R, Lopman B, & Dean NE (2022) The changing epidemiology of SARS-CoV-2. *Science* 375(6585):1116-1121.
2. Tegally H*, et al.* (2021) Detection of a SARS-CoV-2 variant of concern in South Africa. *Nature* 592(7854):438-443.
3. Viana R*, et al.* (2022) Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa. *Nature*.
4. Yang W & Shaman J (2021) Development of a model-inference system for estimating epidemiological characteristics of SARS-CoV-2 variants of concern. *Nature Communications* 12:5573.
5. Nyberg T*, et al.* (Comparative analysis of the risks of hospitalisation and death associated with SARS-CoV-2 omicron (B.1.1.529) and delta (B.1.617.2) variants in England: a cohort study. *The Lancet*.
6. Wolter N*, et al.* (2022) Early assessment of the clinical severity of the SARS-CoV-2 omicron variant in South Africa: a data linkage study. *The Lancet* 399(10323):437-446.

7. Madhi SA, *et al.* (2022) South African Population Immunity and Severe Covid-19 with Omicron Variant. *medRxiv*:2021.2012.2020.21268096.
8. Shinde V, *et al.* (2021) Efficacy of NVX-CoV2373 Covid-19 Vaccine against the B.1.351 Variant. *N Engl J Med* 384(20):1899-1909.
9. Garcia-Beltran WF, *et al.* (2021) Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity. *Cell* 184(9):2372-2383 e2379.
10. Wall EC, *et al.* (2021) Neutralising antibody activity against SARS-CoV-2 VOCs B.1.617.2 and B.1.351 by BNT162b2 vaccination. *Lancet*.
11. Abu-Raddad LJ, *et al.* (2021) Severity, Criticality, and Fatality of the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Beta Variant. *Clinical Infectious Diseases*.
12. Public Health England (2021) SARS-CoV-2 variants of concern and variants under investigation in England. Technical briefing 14.
13. Allen H, *et al.* (2021) Household transmission of COVID-19 cases associated with SARS-CoV-2 delta variant (B.1.617.2): national case-control study. *The Lancet regional health. Europe*:100252.
14. Challen R, *et al.* (2021) Early epidemiological signatures of novel SARS-CoV-2 variants: establishment of B.1.617.2 in England. *medRxiv*:2021.2006.2005.21258365.
15. Earnest R, *et al.* (2021) Comparative transmissibility of SARS-CoV-2 variants Delta and Alpha in New England, USA. *medRxiv*.
16. Vohringer HS, *et al.* (2021) Genomic reconstruction of the SARS-CoV-2 epidemic in England. *Nature*.
17. Dhar MS, *et al.* (2021) Genomic characterization and epidemiology of an emerging SARS-CoV-2 variant in Delhi, India. *Science*:eabj9932.
18. Liu C, *et al.* (2021) Reduced neutralization of SARS-CoV-2 B.1.617 by vaccine and convalescent serum. *Cell* 184(16):4220-+.
19. de Oliveira T & Lessells R (2021) Update on Delta and other variants in South Africa and other world.
20. The South African Medical Research Council (SAMRC) (2021) Report on Weekly Deaths in South Africa.
21. Anonymous (South Africa Population (live).
22. United States Census Bureau (2020) Census Bureau Releases 2020 Demographic Analysis Estimates.
23. Levin AT, *et al.* (2020) Assessing the age specificity of infection fatality rates for COVID-19: systematic review, meta-analysis, and public policy implications. *Eur J Epidemiol* 35(12):1123-1138.
24. O'Driscoll M, *et al.* (2021) Age-specific mortality and immunity patterns of SARS-CoV-2. *Nature* 590(7844):140-145.
25. Anonymous (Variation in the COVID-19 infection-fatality ratio by age, time, and geography during the pre-vaccine era: a systematic analysis. *The Lancet*.
26. Garcia-Beltran WF, *et al.* (2022) mRNA-based COVID-19 vaccine boosters induce neutralizing immunity against SARS-CoV-2 Omicron variant. *Cell* 185(3):457-466.e454.
27. Hui KPY, *et al.* (2022) SARS-CoV-2 Omicron variant replication in human bronchus and lung ex vivo. *Nature*.

28. Rössler A, Riepler L, Bante D, von Laer D, & Kimpel J (2022) SARS-CoV-2 Omicron Variant Neutralization in Serum from Vaccinated and Convalescent Persons. *New England Journal of Medicine* 386(7):698-700.
29. Cele S*, et al.* (2022) Omicron extensively but incompletely escapes Pfizer BNT162b2 neutralization. *Nature* 602(7898):654-656.
30. Brandal LT*, et al.* (2021) Outbreak caused by the SARS-CoV-2 Omicron variant in Norway, November to December 2021. *Eurosurveillance* 26(50):2101147.
31. Helmsdal G*, et al.* (2022) Omicron outbreak at a private gathering in the Faroe Islands, infecting 21 of 33 triple-vaccinated healthcare workers. *Clinical Infectious Diseases*.
32. van der Straten K*, et al.* (2022) Mapping the antigenic diversification of SARS-CoV-2. *medRxiv*:2022.2001.2003.21268582.
33. Mao T*, et al.* (2022) Unadjuvanted intranasal spike vaccine booster elicits robust protective mucosal immunity against sarbecoviruses. *bioRxiv*:2022.2001.2024.477597.
34. Morens DM, Taubenberger JK, & Fauci AS (2021) Universal Coronavirus Vaccines — An Urgent Need. *New England Journal of Medicine* 386(4):297-299.
35. Data Science for Social Impact Research Group at University of Pretoria (2021) Coronavirus COVID-19 (2019-nCoV) Data Repository for South Africa.
36. Iannone R (2020) Package 'stationaRy'.
37. Iannone R (2020) stationaRy.
38. Wallace J & Hobbs P (2006) *Atmospheric Science: An Introductory survey* (Academic Press, New York) 2nd Edition Ed p 504.
39. Google Inc. (2020) Community Mobility Reports.
40. Anonymous (2020) Data on COVID-19 (coronavirus) vaccinations by Our World in Data.
41. Mathieu E*, et al.* (2021) A global database of COVID-19 vaccinations. *Nat Hum Behav* 5(7):947-953.
42. Department of Health Republic of South Africa (2021) Update on Covid-19 (Tuesday 23 November 2021).
43. Yang W & Shaman J (2021) COVID-19 pandemic dynamics in India, the SARS-CoV-2 Delta variant, and implications for vaccination. *medRxiv*:2021.2006.2021.21259268.
44. Yang W, Shaff J, & Shaman J (2021) Effectiveness of non-pharmaceutical interventions to contain COVID-19: a case study of the 2020 spring pandemic wave in New York City. *J R Soc Interface* 18(175):20200822.
45. Lasry A*, et al.* (2020) Timing of Community Mitigation and Changes in Reported COVID-19 and Community Mobility - Four U.S. Metropolitan Areas, February 26-April 1, 2020. *MMWR. Morbidity and mortality weekly report* 69(15):451-457.
46. Kraemer MUG*, et al.* (2020) The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* 368(6490):493-497.
47. Department of Health Republic of South Africa (2021) Latest Vaccine Statistics.
48. Abu-Raddad LJ, Chemaitelly H, & Butt AA (2021) Effectiveness of the BNT162b2 Covid-19 Vaccine against the B.1.1.7 and B.1.351 Variants. *New Engl J Med*.
49. Bernal JL*, et al.* (2021) Effectiveness of Covid-19 Vaccines against the B.1.617.2 (Delta) Variant. *New England Journal of Medicine* 385(7):585-594.
50. Andrews N*, et al.* (2021) Effectiveness of COVID-19 vaccines against the Omicron (B.1.1.529) variant of concern. *medRxiv*:2021.2012.2014.21267615.

51. Biryukov J*, et al.* (2020) Increasing Temperature and Relative Humidity Accelerates Inactivation of SARS-CoV-2 on Surfaces. *mSphere* 5(4):e00441-00420.
52. Morris DH*, et al.* (2021) Mechanistic theory predicts the effects of temperature and humidity on inactivation of SARS-CoV-2 and other enveloped viruses. *Elife* 10.
53. Yang W*, et al.* (2021) Estimating the infection-fatality risk of SARS-CoV-2 in New York City during the spring 2020 pandemic wave: a model-based analysis. *The Lancet. Infectious diseases* 21(2):203-212.
54. Anderson JL (2001) An ensemble adjustment Kalman filter for data assimilation. *Mon. Weather Rev.* 129(12):2884-2903.
55. The National Institute for Communicable Diseases (NICD) of the National Health Laboratory (NHLS) on behalf of the Network for Genomics Surveillance in South Africa (NGS-SA) (2021) Network for Genomic Surveillance South Africa (NGS-SA) SARS-CoV-2 Sequencing Update 19 August 2021.
56. The South African COVID-19 Modelling Consortium (2021) COVID-19 modelling update: Considerations for a potential fourth wave (17 Nov 2021).

Fig 1. Pandemic dynamics in South Africa, model-fit and validation using serology data. (A) Pandemic dynamics in each of the nine provinces (see legend); dots depict reported weekly numbers of cases and deaths; lines show model mean estimates (in the same color). (B) For validation, model estimated infection rates are compared to seroprevalence measures over time from multiple sero-surveys summarized in ref (56) (B). Boxplots depict the estimated distribution for each province (middle bar = mean; edges = 50% CrIs) and whiskers (95% CrIs). Red dots show corresponding measurements. Note that reported mortality was high in February 2022 in some provinces with no clear explanation.

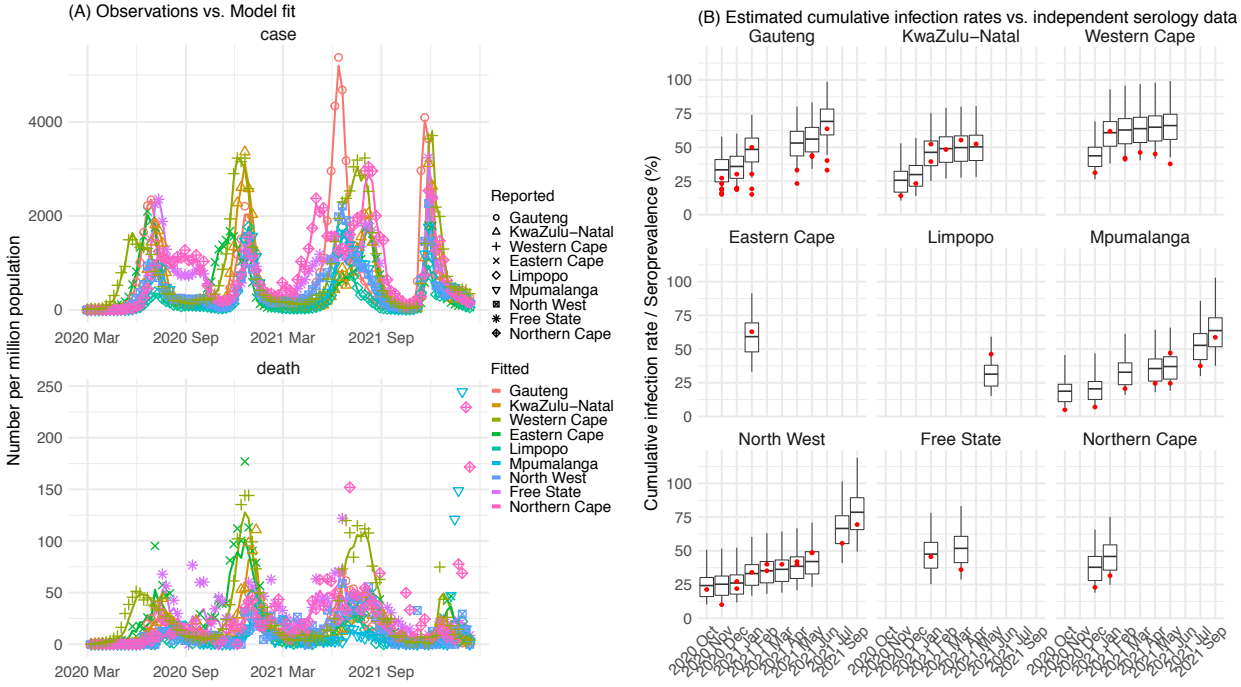


Fig 2. Model validation using retrospective prediction. Model-inference was trained on cases and deaths data since March 15, 2020 up to 2 weeks (1st plot in each panel) or 1 week (2nd plot) before the Delta or Omicron wave (see timing on the x-axis); the model was then integrated forward using the estimates made at the time to predict cases (left panel) and deaths (right panel) for the remaining weeks of each wave. Blue lines and surrounding shades show model fitted cases and deaths for weeks before the prediction (line = median, dark blue area = 50% CrIs, and light blue = 80% CrIs). Red lines show model projected median weekly cases and deaths; surrounding shades show 50% (dark red) and 80% (light red) CIs of the prediction. For comparison, reported cases and deaths for each week are shown by the black dots; however, those to the right of the vertical dash lines (showing the start of each prediction) were not used in the model. For clarity, here we show 80% CIs (instead of 95% CIs, which tend to be wider for longer-term projections) and predictions for the four most populous provinces (Gauteng in A and B; KwaZulu-Natal in C and D; Western Cape in E and F; and Eastern Cape in G and H). Predictions for the other five provinces are shown in Fig S3.

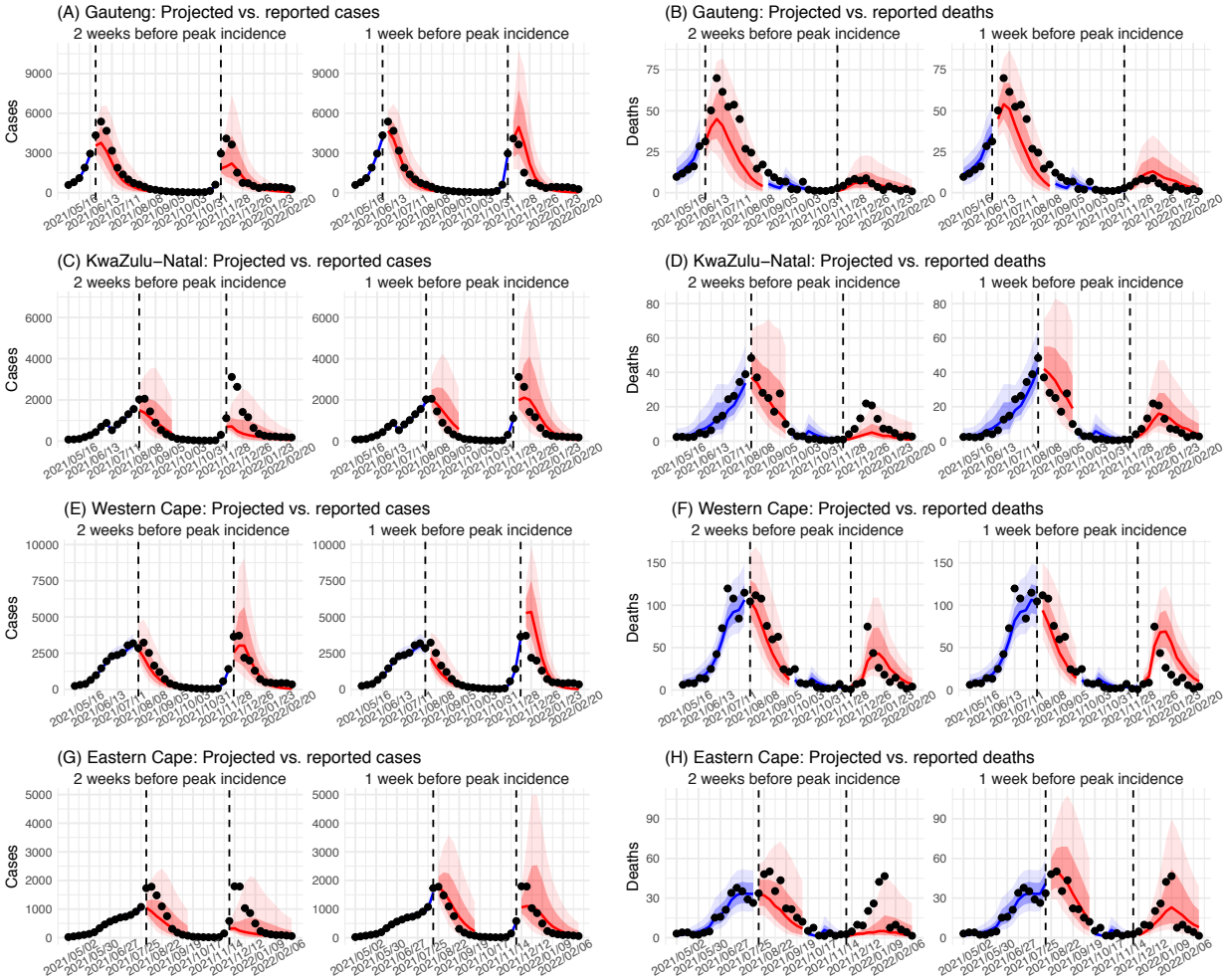


Fig 3. Example model-inference estimates for Gauteng. (A) Observed relative mobility, vaccination rate, and estimated disease seasonal trend, compared to case and death rates over time. Key model-inference estimates are shown for the time-varying effective reproduction number R_t (B), transmissibility (C), population susceptibility (D), infection-detection rate (E), and infection-fatality risk (F). Grey shaded areas indicate the approximate circulation period for each variant. In (B) – (F), blue lines and surrounding areas show the estimated mean, 50% (dark) and 95% (light) CrIs; boxes and whiskers show the estimated mean, 50% and 95% CrIs for estimated infection rates. *Note that the transmissibility estimates (in C) have removed the effects of changing population susceptibility, NPIs, and disease seasonality; thus, the trends are more stable than the reproduction number (R_t in B) and reflect changes in variant-specific properties. Also note that infection-fatality risk estimates were based on reported COVID-19 deaths and may not reflect true values due to likely under-reporting of COVID-19 deaths.*

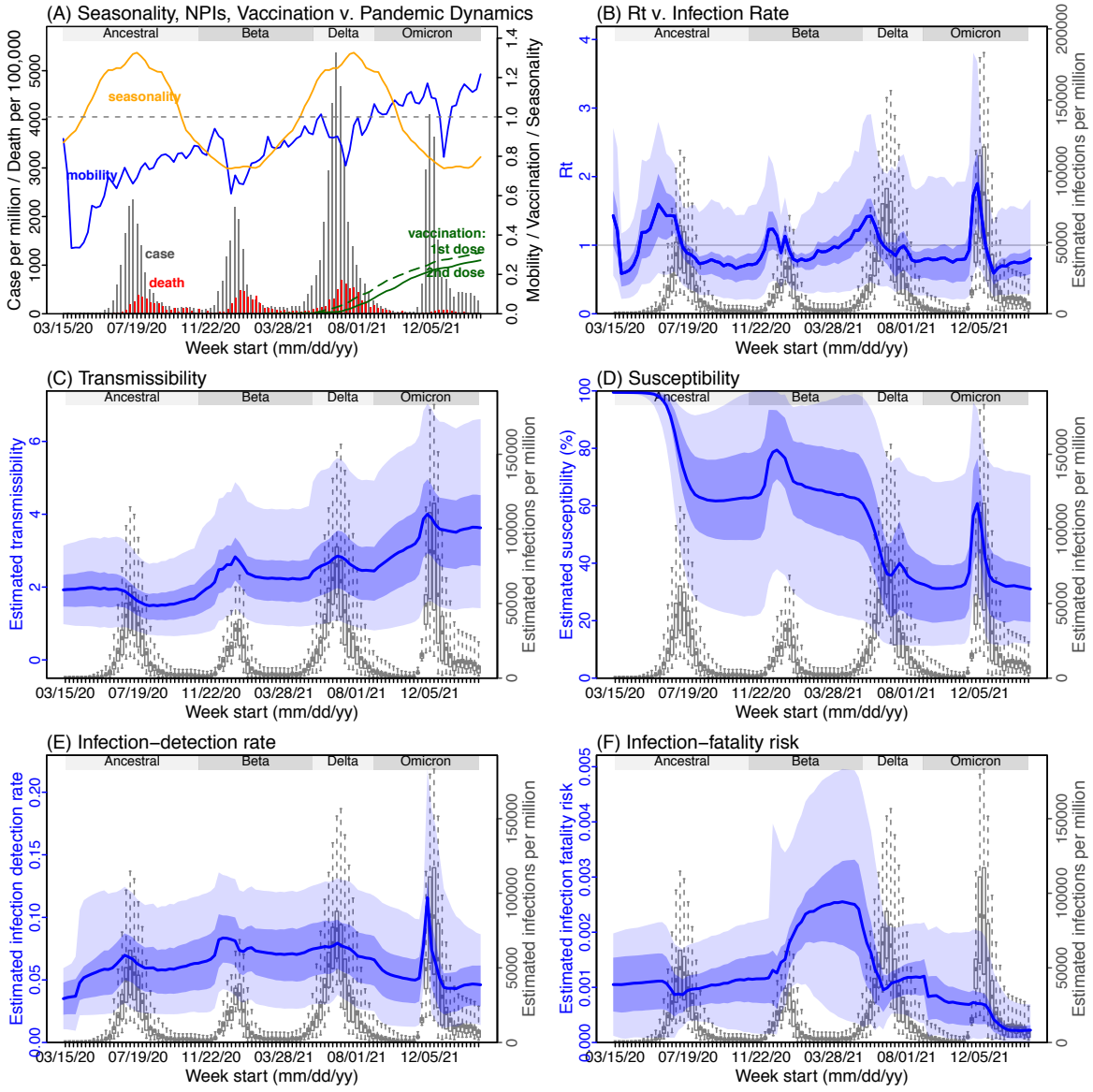
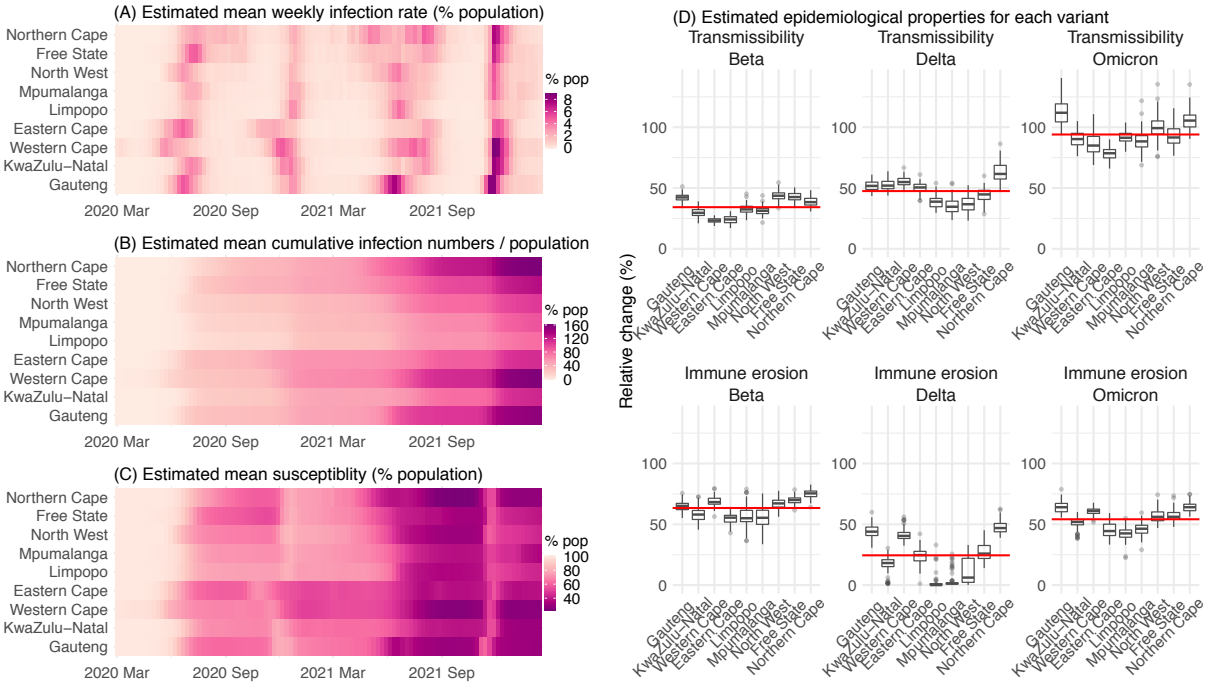


Fig 4. Model-inferred epidemiological properties for different variants across SA provinces. Heatmaps show (A) Estimated mean infection rates by week (x-axis) and province (y-axis), (B) Estimated mean *cumulative* infection numbers relative to the population size in each province, and (C) Estimated population susceptibility (to the circulating variant) by week and province. (D) Boxplots in the top row show the estimated distribution of changes in transmissibility for Beta, Delta, and Omicron, relative to the Ancestral SARS-CoV-2, for each province (middle bar = median; edges = 50% CIs; and whiskers =95% CIs); boxplots in the bottom row show, for each variant, the estimated distribution of immune erosion to all adaptive immunity gained from infection and vaccination prior to that variant. Red lines show the mean across all provinces.



Supplemental Figures and Tables

Fig S1. Model-fit to case and death data in each province. Dots show reported SARS-CoV-2 cases and deaths by week. Blue lines and surrounding area show model estimated median, 50% (darker blue) and 95% (lighter blue) credible intervals. Note that reported mortality was high in February 2022 in some provinces with no clear explanation.

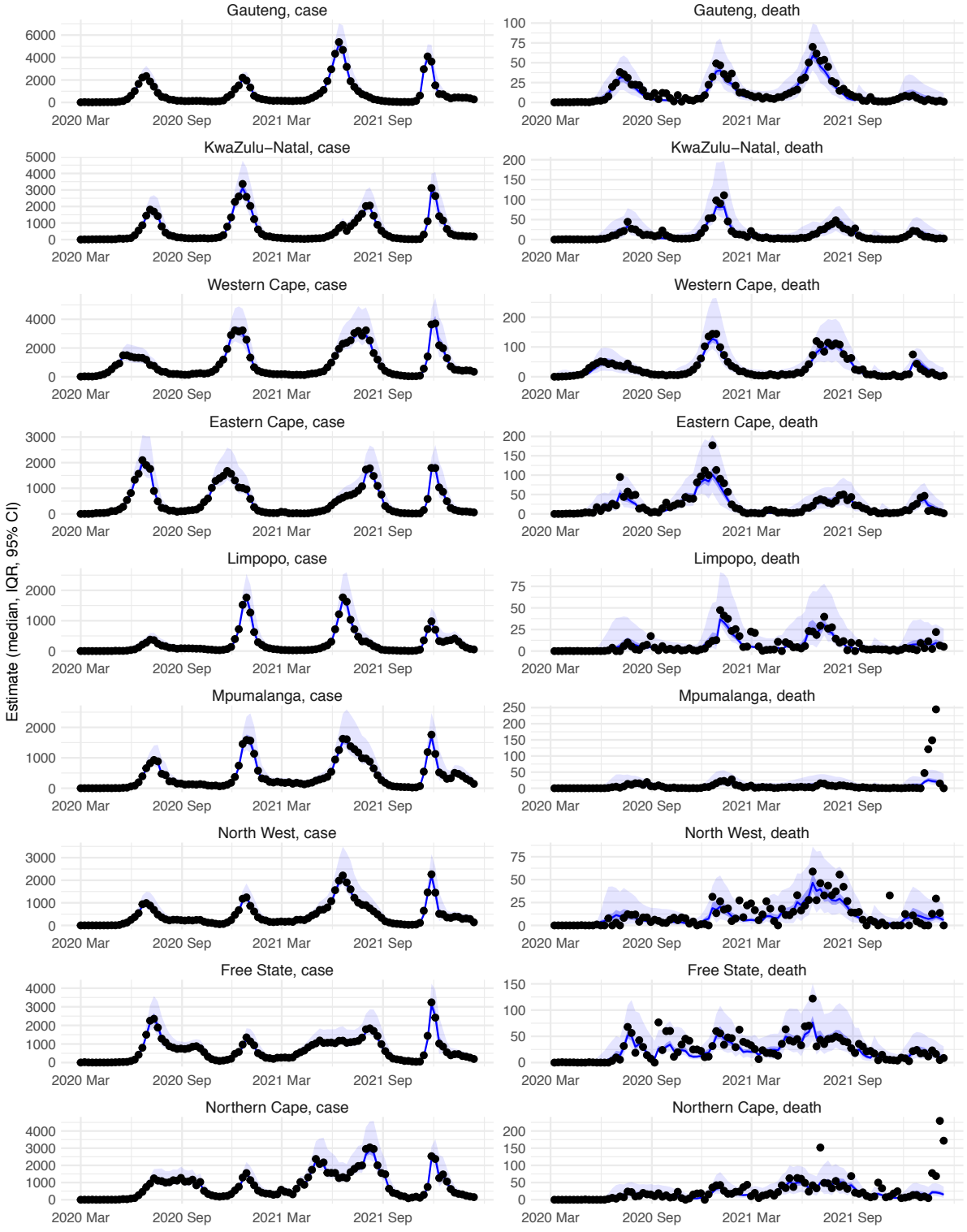


Fig S2. Model validation using hospitalization and excess mortality data. Model estimated infection rates are compared to COVID-related hospitalizations (left panel) and excess mortality (right panel) during the Ancestral (A), Beta (B), Delta (C), and Omicron (D) waves. Boxplots show the estimated distribution for each province (middle bar = mean; edges = 50% CrIs) and whiskers (=95% CrIs). Red dots show corresponding measurements. Correlation (*r*) between model estimated cumulative infection rate and cumulative hospitalization or age-adjusted excess mortality for each wave is shown in each plot. *Note that hospitalization data begin from 6/6/20 and excess mortality data begin from 5/3/20 and thus are incomplete for the ancestral wave.*

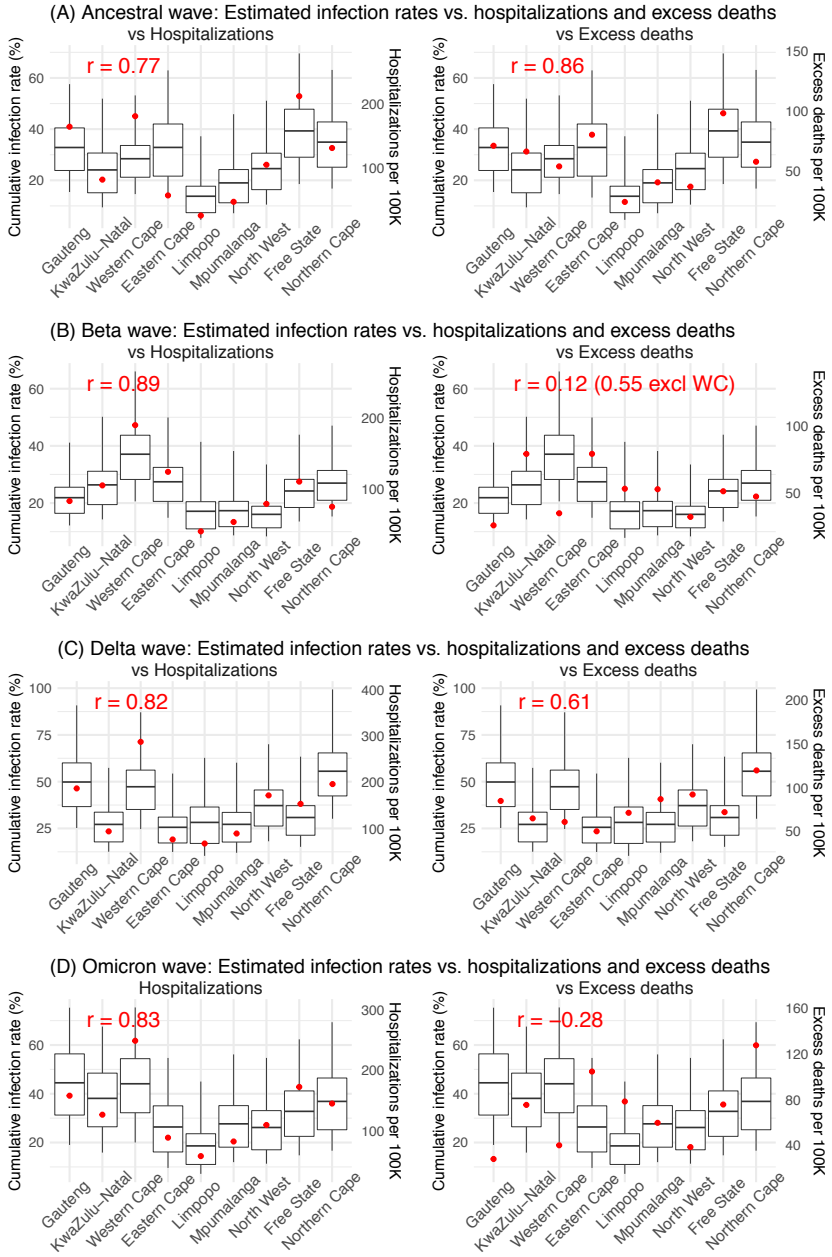


Fig S3. Model validation using retrospective prediction, for the remaining 5 provinces. Model-inference was trained on cases and deaths data since March 15, 2020 up to 2 weeks (1st plot in each panel) or 1 week (2nd plot) before the Delta or Omicron wave (see timing on the x-axis); the model was then integrated forward using the estimates made at the time to predict cases (left panel) and deaths (right panel) for the remaining weeks of each wave. Blue lines and surrounding shades show model fitted cases and deaths for weeks before the prediction (line = median, dark blue area = 50% CrIs, and light blue = 80% CrIs). Red lines show model projected median weekly cases and deaths; surrounding shades show 50% (dark red) and 80% (light red) CIs of the prediction. For comparison, reported cases and deaths for each week are shown by the black dots; however, those to the right of the vertical dash lines (showing the start of each prediction) were not used in the model. For clarity, here we show 80% CIs (instead of 95% CIs, which tend to be wider for longer-term projections) and predictions for the five least populous provinces (Limpopo in A and B; Mpumalanga in C and D; North West in E and F; Free State in G and H; and Northern Cape in I and J). Predictions for the other 4 provinces are shown in Fig 2.

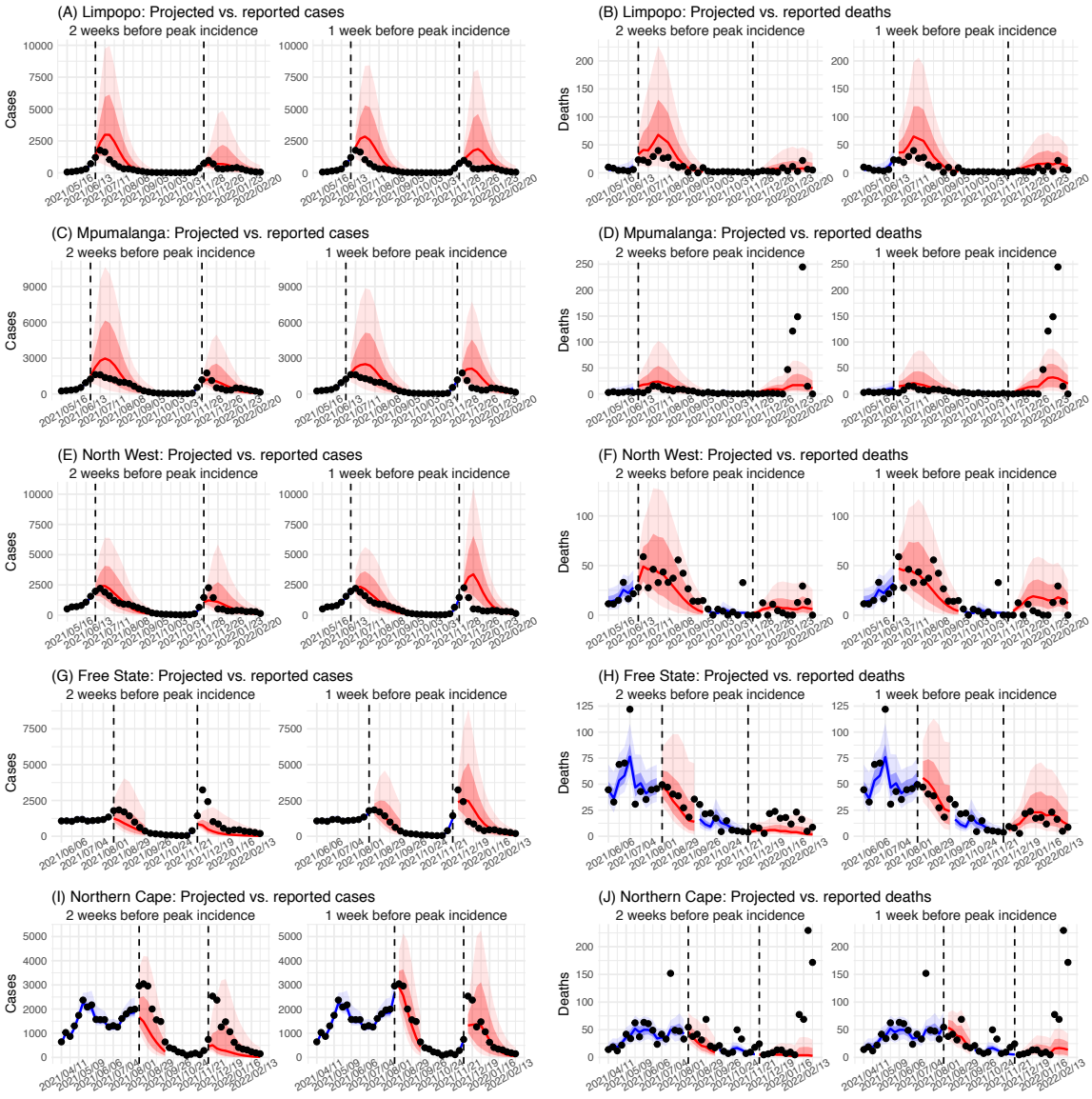


Fig S4. Model inference estimates for *KwaZulu-Natal*. (A) Observed relative mobility, vaccination rate, and estimated disease seasonal trend, compared to case and death rates over time. Key model-inference estimates are shown for the time-varying effective reproduction number R_t (B), transmissibility (C), population susceptibility (D), infection-detection rate (E), and infection-fatality risk (F). Grey shaded areas indicate the approximate circulation period for each variant. In (B) – (F), blue lines and surrounding areas show the estimated mean, 50% (dark) and 95% (light) CrIs; boxes and whiskers show the estimated mean, 50% and 95% CrIs for estimated infection rates. *Note that the transmissibility estimates (in C) have removed the effects of changing population susceptibility, NPIs, and disease seasonality; thus, the trends are more stable than the reproduction number (R_t in B) and reflect changes in variant-specific properties. Also note that infection-fatality risk estimates were based on reported COVID-19 deaths and may not reflect true values due to likely under-reporting of COVID-19 deaths.*

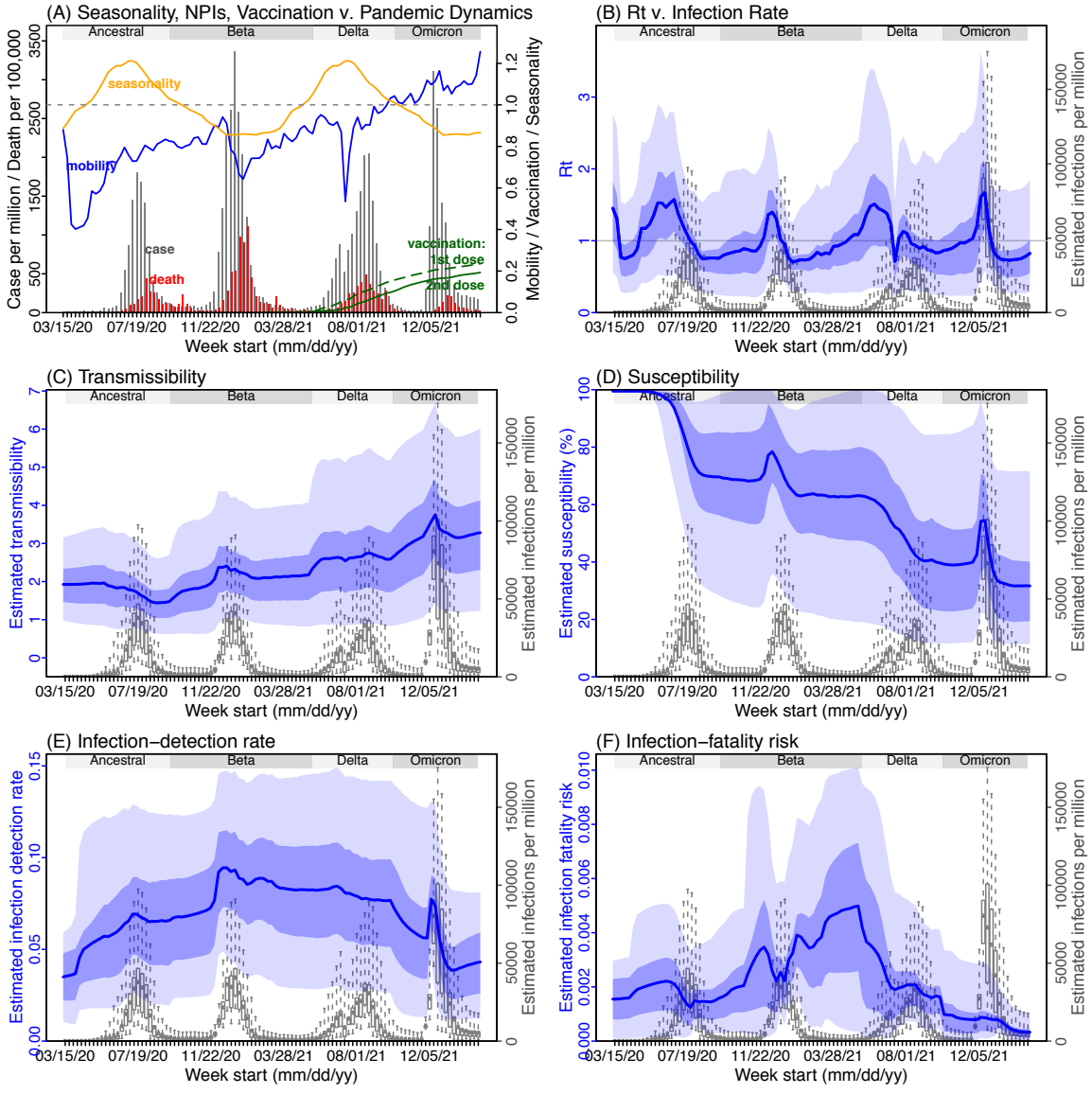


Fig S5. Model inference estimates for *Western Cape*. (A) Observed relative mobility, vaccination rate, and estimated disease seasonal trend, compared to case and death rates over time. Key model-inference estimates are shown for the time-varying effective reproduction number R_t (B), transmissibility (C), population susceptibility (D), infection-detection rate (E), and infection-fatality risk (F). Grey shaded areas indicate the approximate circulation period for each variant. In (B) – (F), blue lines and surrounding areas show the estimated mean, 50% (dark) and 95% (light) CrIs; boxes and whiskers show the estimated mean, 50% and 95% CrIs for estimated infection rates. *Note that the transmissibility estimates (in C) have removed the effects of changing population susceptibility, NPIs, and disease seasonality; thus, the trends are more stable than the reproduction number (R_t in B) and reflect changes in variant-specific properties. Also note that infection-fatality risk estimates were based on reported COVID-19 deaths and may not reflect true values due to likely under-reporting of COVID-19 deaths.*

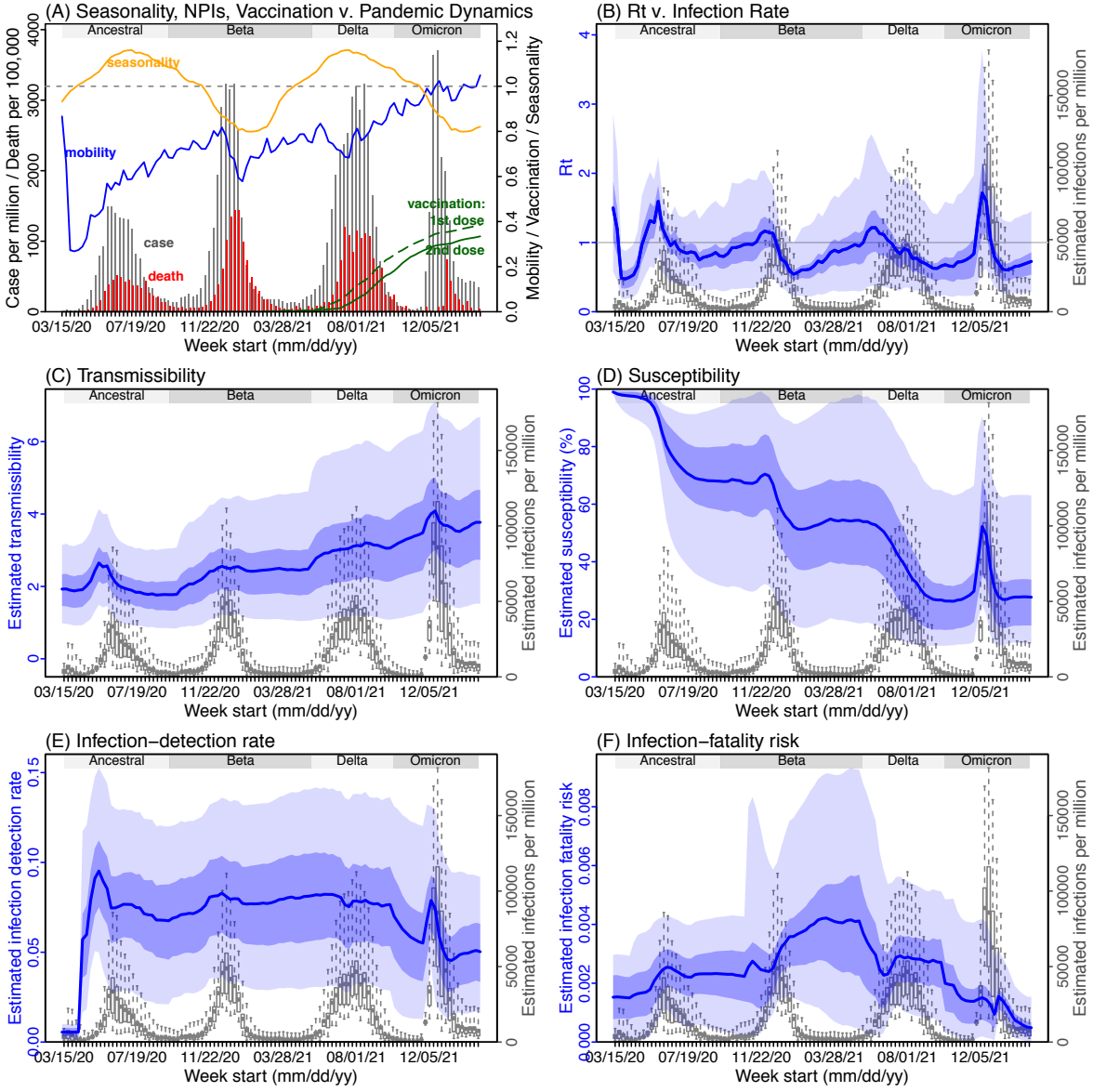


Fig S6. Model inference estimates for *Eastern Cape*. (A) Observed relative mobility, vaccination rate, and estimated disease seasonal trend, compared to case and death rates over time. Key model-inference estimates are shown for the time-varying effective reproduction number R_t (B), transmissibility (C), population susceptibility (D), infection-detection rate (E), and infection-fatality risk (F). Grey shaded areas indicate the approximate circulation period for each variant. In (B) – (F), blue lines and surrounding areas show the estimated mean, 50% (dark) and 95% (light) CrIs; boxes and whiskers show the estimated mean, 50% and 95% CrIs for estimated infection rates. *Note that the transmissibility estimates (in C) have removed the effects of changing population susceptibility, NPIs, and disease seasonality; thus, the trends are more stable than the reproduction number (R_t in B) and reflect changes in variant-specific properties. Also note that infection-fatality risk estimates were based on reported COVID-19 deaths and may not reflect true values due to likely under-reporting of COVID-19 deaths.*

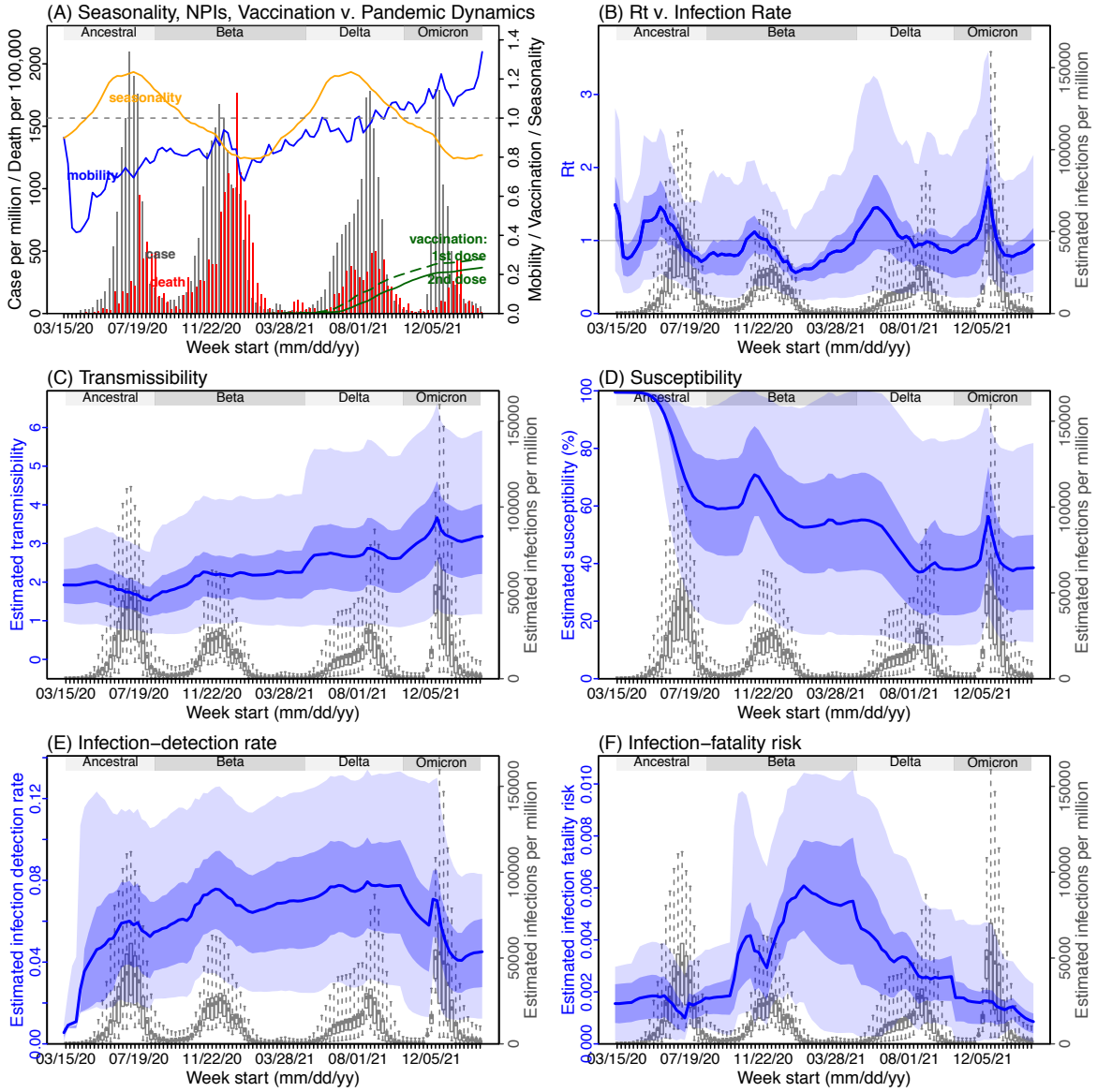


Fig S7. Model inference estimates for *Limpopo*. (A) Observed relative mobility, vaccination rate, and estimated disease seasonal trend, compared to case and death rates over time. Key model-inference estimates are shown for the time-varying effective reproduction number R_t (B), transmissibility (C), population susceptibility (D), infection-detection rate (E), and infection-fatality risk (F). Grey shaded areas indicate the approximate circulation period for each variant. In (B) – (F), blue lines and surrounding areas show the estimated mean, 50% (dark) and 95% (light) CrIs; boxes and whiskers show the estimated mean, 50% and 95% CrIs for estimated infection rates. *Note that the transmissibility estimates (in C) have removed the effects of changing population susceptibility, NPIs, and disease seasonality; thus, the trends are more stable than the reproduction number (R_t in B) and reflect changes in variant-specific properties. Also note that infection-fatality risk estimates were based on reported COVID-19 deaths and may not reflect true values due to likely under-reporting of COVID-19 deaths.*

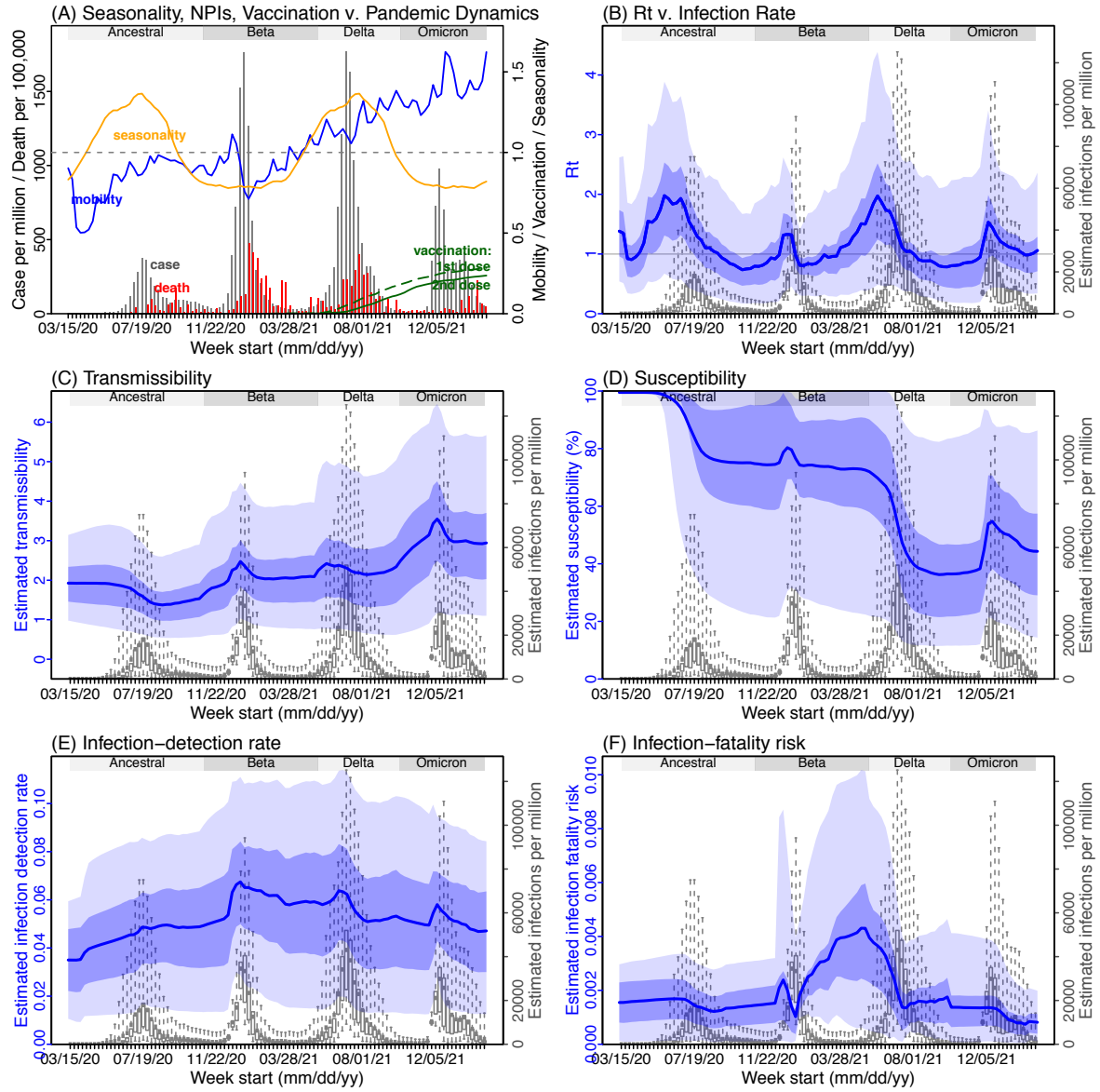


Fig S8. Model inference estimates for *Mpumalanga*. (A) Observed relative mobility, vaccination rate, and estimated disease seasonal trend, compared to case and death rates over time. Key model-inference estimates are shown for the time-varying effective reproduction number R_t (B), transmissibility (C), population susceptibility (D), infection-detection rate (E), and infection-fatality risk (F). Grey shaded areas indicate the approximate circulation period for each variant. In (B) – (F), blue lines and surrounding areas show the estimated mean, 50% (dark) and 95% (light) CrIs; boxes and whiskers show the estimated mean, 50% and 95% CrIs for estimated infection rates. *Note that the transmissibility estimates (in C) have removed the effects of changing population susceptibility, NPIs, and disease seasonality; thus, the trends are more stable than the reproduction number (R_t in B) and reflect changes in variant-specific properties. Also note that infection-fatality risk estimates were based on reported COVID-19 deaths and may not reflect true values due to likely under-reporting of COVID-19 deaths.*

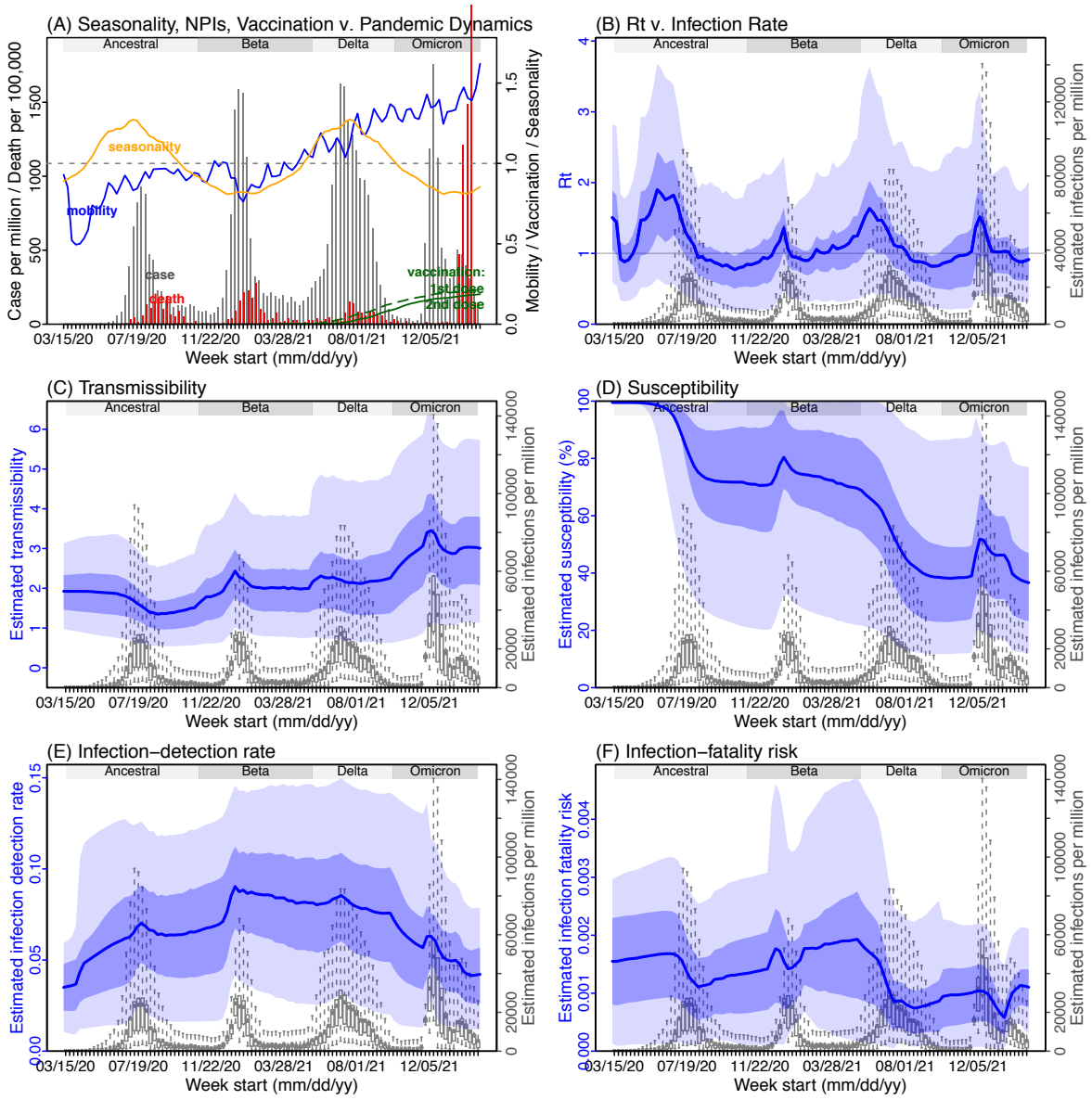
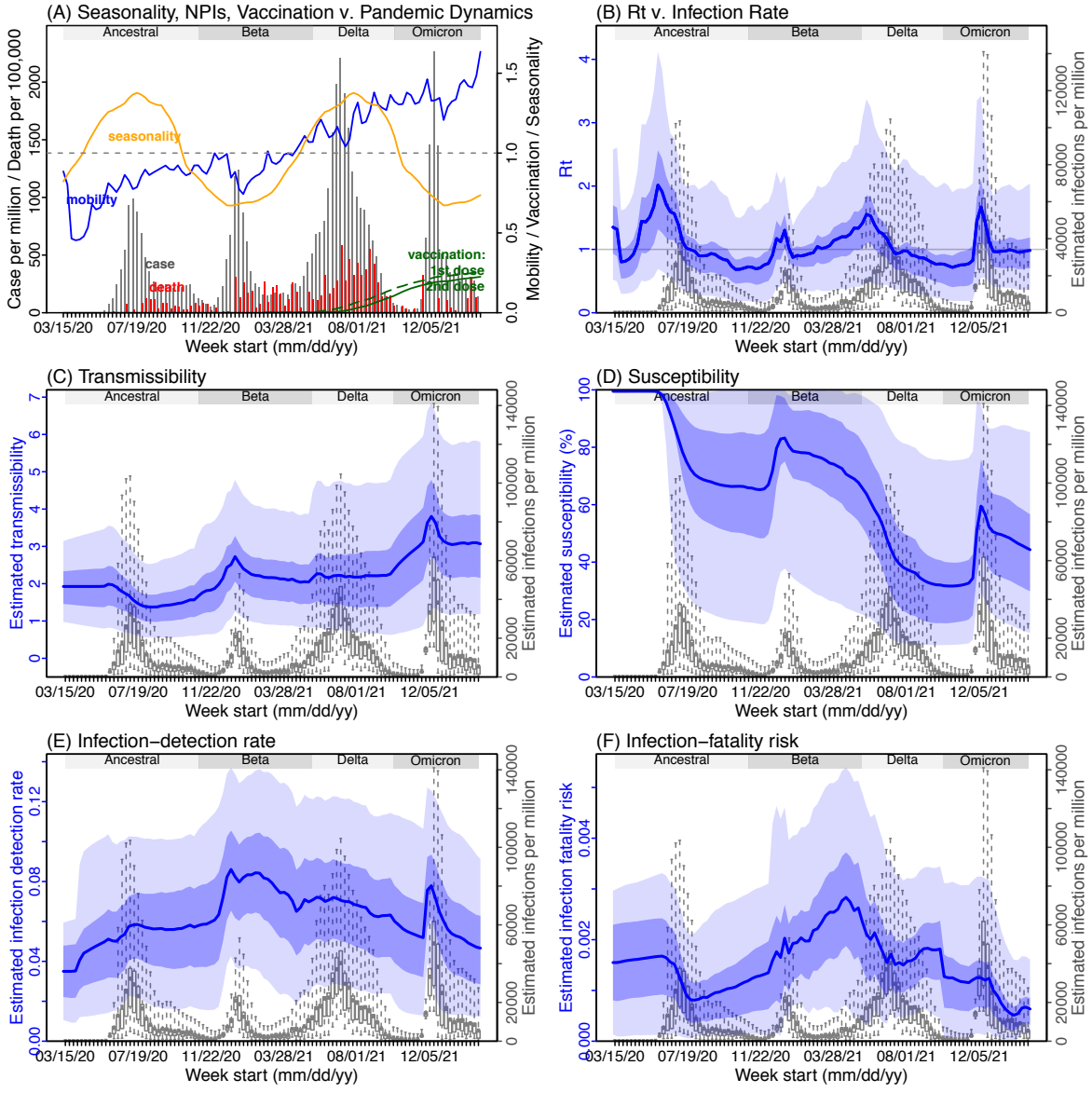
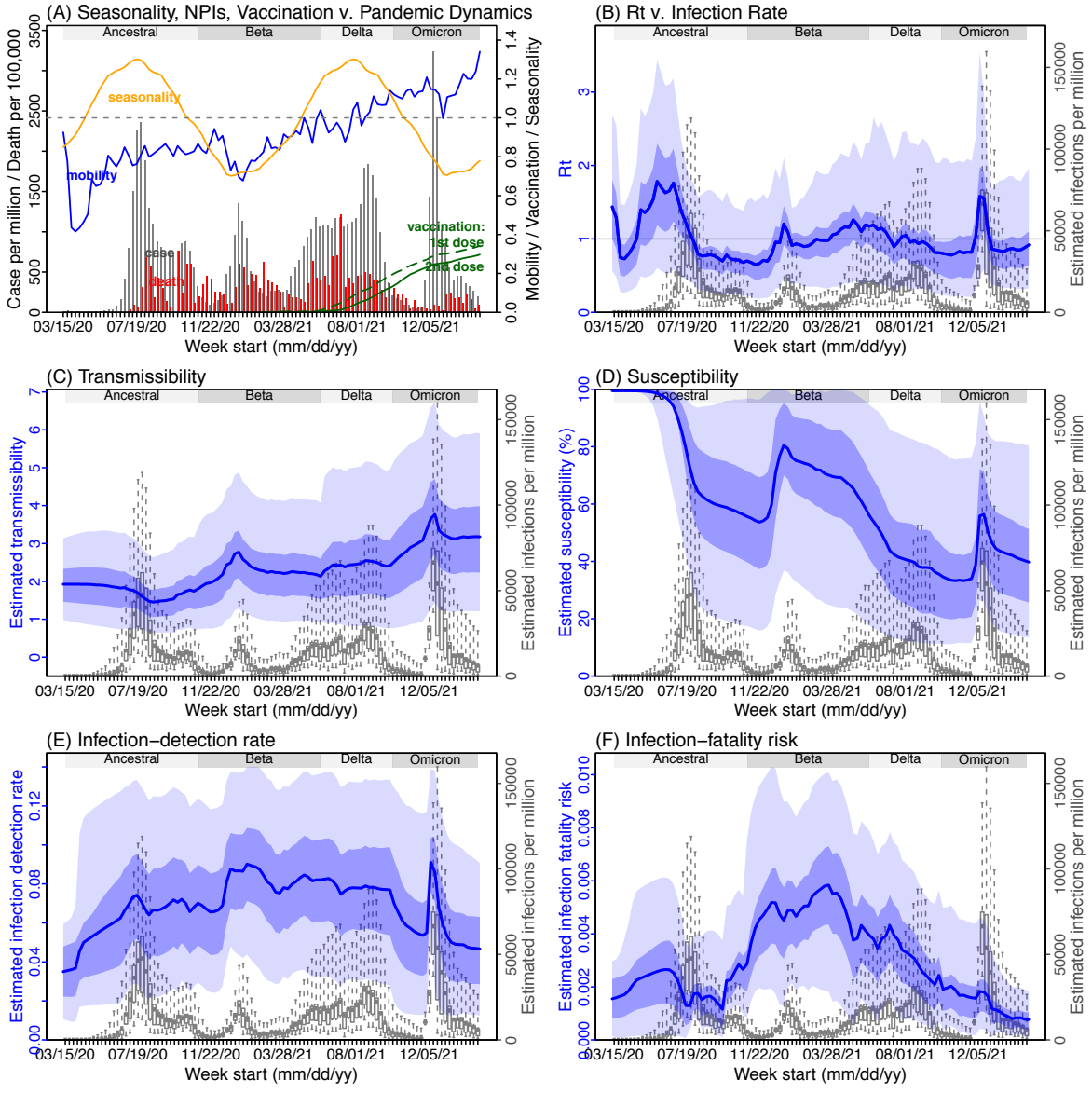


Fig S9. Model inference estimates for *North West*. (A) Observed relative mobility, vaccination rate, and estimated disease seasonal trend, compared to case and death rates over time. Key model-inference estimates are shown for the time-varying effective reproduction number R_t (B), transmissibility (C), population susceptibility (D), infection-detection rate (E), and infection-fatality risk (F). Grey shaded areas indicate the approximate circulation period for each variant. In (B) – (F), blue lines and surrounding areas show the estimated mean, 50% (dark) and 95% (light) CrIs; boxes and whiskers show the estimated mean, 50% and 95% CrIs for estimated infection rates. *Note that the transmissibility estimates (in C) have removed the effects of changing population susceptibility, NPIs, and disease seasonality; thus, the trends are more stable than the reproduction number (R_t in B) and reflect changes in variant-specific properties. Also note that infection-fatality risk estimates were based on reported COVID-19 deaths and may not reflect true values due to likely under-reporting of COVID-19 deaths.*



788 **Fig S10. Model inference estimates for *Free State*.** (A) Observed relative mobility, vaccination
789 rate, and estimated disease seasonal trend, compared to case and death rates over time. Key
790 model-inference estimates are shown for the time-varying effective reproduction number R_t
791 (B), transmissibility (C), population susceptibility (D), infection-detection rate (E), and infection-
792 fatality risk (F). Grey shaded areas indicate the approximate circulation period for each variant.
793 In (B) – (F), blue lines and surrounding areas show the estimated mean, 50% (dark) and 95%
794 (light) CrIs; boxes and whiskers show the estimated mean, 50% and 95% CrIs for estimated
795 infection rates. *Note that the transmissibility estimates (in C) have removed the effects of*
796 *changing population susceptibility, NPIs, and disease seasonality; thus, the trends are more*
797 *stable than the reproduction number (R_t in B) and reflect changes in variant-specific properties.*
798 *Also note that infection-fatality risk estimates were based on reported COVID-19 deaths and*
799 *may not reflect true values due to likely under-reporting of COVID-19 deaths.*



800

Fig S11. Model inference estimates for *Northern Cape*. (A) Observed relative mobility, vaccination rate, and estimated disease seasonal trend, compared to case and death rates over time. Key model-inference estimates are shown for the time-varying effective reproduction number R_t (B), transmissibility (C), population susceptibility (D), infection-detection rate (E), and infection-fatality risk (F). Grey shaded areas indicate the approximate circulation period for each variant. In (B) – (F), blue lines and surrounding areas show the estimated mean, 50% (dark) and 95% (light) CrIs; boxes and whiskers show the estimated mean, 50% and 95% CrIs for estimated infection rates. *Note that the transmissibility estimates (in C) have removed the effects of changing population susceptibility, NPIs, and disease seasonality; thus, the trends are more stable than the reproduction number (R_t in B) and reflect changes in variant-specific properties. Also note that infection-fatality risk estimates were based on reported COVID-19 deaths and may not reflect true values due to likely under-reporting of COVID-19 deaths.*

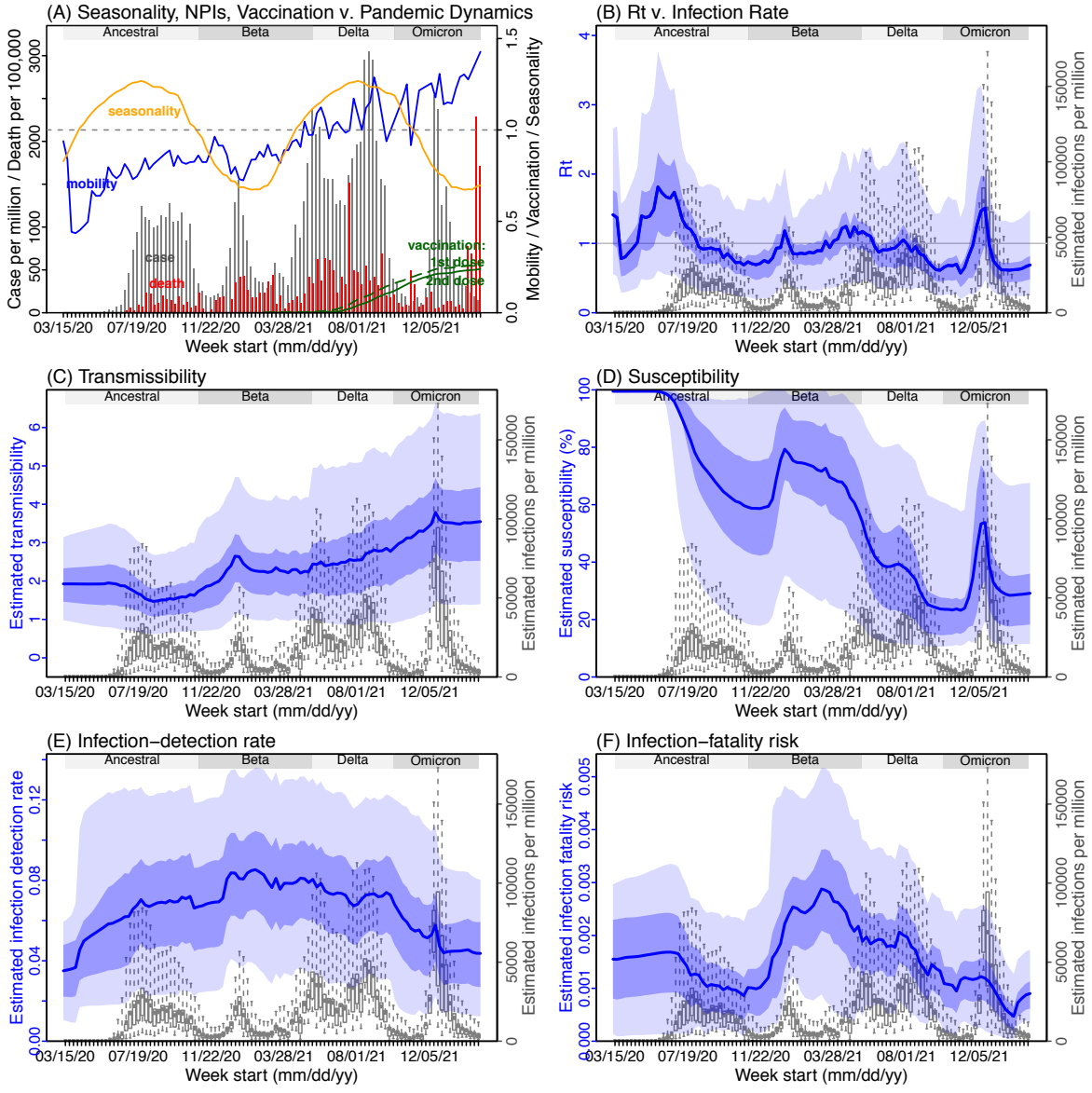


Table S1. Model estimated infection-detection rate during each wave. Numbers show the estimated percentage of infections (including asymptomatic and subclinical infections) documented as cases (mean and 95% CI in parentheses).

Province	Ancestral wave	Beta wave	Delta wave	Omicron wave
Gauteng	4.59 (2.62, 9.77)	6.18 (3.29, 11.11)	6.27 (3.44, 12.39)	4.16 (2.46, 9.72)
KwaZulu-Natal	4.33 (2.01, 11.02)	7.4 (3.89, 13.67)	5.69 (2.69, 12.34)	3.25 (1.84, 7.81)
Western Cape	5.62 (3, 10.93)	7.1 (3.99, 12.78)	6.83 (3.71, 13.08)	4.26 (2.49, 9.37)
Eastern Cape	3.79 (1.98, 9.39)	6.1 (3.35, 11.27)	5.58 (2.63, 11.52)	2.91 (1.4, 7.99)
Limpopo	2.13 (0.79, 6.46)	4.57 (1.89, 10.01)	3.4 (1.53, 9.3)	2.9 (1.2, 7.55)
Mpumalanga	3.42 (1.42, 9.1)	6.28 (2.85, 12.51)	5.71 (2.58, 12.96)	3.13 (1.54, 7.24)
North West	3.37 (1.62, 7.88)	5.79 (2.77, 11.14)	5.26 (2.8, 10.8)	3.73 (1.78, 8.62)
Free State	5.02 (2.83, 10.63)	6.69 (3.69, 11.97)	6.5 (3.16, 13.23)	4.03 (2.12, 8.95)
Northern Cape	4.96 (2.75, 10.34)	6.49 (3.72, 11.44)	6.69 (3.74, 12.32)	3.71 (1.97, 8.21)

Table S2. Model estimated attack rate during each wave. Numbers show estimated cumulative infection numbers, expressed as percentage of population size (mean and 95% CI in parentheses).

Province	Ancestral wave	Beta wave	Delta wave	Omicron wave
Gauteng	32.83 (15.42, 57.59)	21.87 (12.16, 41.13)	49.82 (25.22, 90.79)	44.49 (19.01, 75.3)
KwaZulu-Natal	24.06 (9.45, 51.91)	26.36 (14.28, 50.18)	27.15 (12.52, 57.39)	38.11 (15.87, 67.56)
Western Cape	28.44 (14.61, 53.17)	37.09 (20.61, 66.04)	47.29 (24.68, 87.1)	44.1 (20.02, 75.4)
Eastern Cape	32.85 (13.27, 62.95)	27.44 (14.86, 49.95)	25.59 (12.4, 54.34)	26.38 (9.59, 54.69)
Limpopo	13.78 (4.55, 37.21)	17.12 (7.82, 41.41)	28.22 (10.33, 62.74)	18.62 (7.15, 45.01)
Mpumalanga	18.99 (7.14, 45.83)	17.33 (8.7, 38.21)	27.18 (11.97, 60.14)	27.67 (11.96, 56.13)
North West	24.57 (10.51, 51.09)	16.04 (8.34, 33.49)	37.21 (18.13, 70.02)	26.17 (11.33, 54.71)
Free State	39.31 (18.54, 69.57)	24.23 (13.54, 43.92)	30.85 (15.16, 63.38)	32.79 (14.76, 62.32)
Northern Cape	34.92 (16.77, 63.13)	26.98 (15.3, 47.09)	55.59 (30.18, 99.32)	36.87 (16.65, 69.34)

Table S3. Model estimated infection-fatality risk during each wave. Numbers are percentages (%; mean and 95% CI in parentheses). Note that these estimates were based on reported COVID-19 deaths and may be biased due to likely under-reporting of COVID-19 deaths.

Province	Ancestral wave	Beta wave	Delta wave	Omicron wave
Gauteng	0.09 (0.05, 0.2)	0.19 (0.1, 0.33)	0.11 (0.06, 0.21)	0.03 (0.02, 0.06)
KwaZulu-Natal	0.09 (0.04, 0.24)	0.27 (0.14, 0.49)	0.14 (0.06, 0.29)	0.03 (0.02, 0.08)
Western Cape	0.21 (0.11, 0.41)	0.3 (0.17, 0.54)	0.25 (0.14, 0.48)	0.06 (0.04, 0.14)
Eastern Cape	0.11 (0.06, 0.27)	0.5 (0.27, 0.91)	0.2 (0.1, 0.42)	0.08 (0.04, 0.22)
Limpopo	0.06 (0.02, 0.17)	0.18 (0.08, 0.4)	0.1 (0.04, 0.27)	0.05 (0.02, 0.12)
Mpumalanga	0.07 (0.03, 0.18)	0.1 (0.05, 0.2)	0.04 (0.02, 0.1)	0.21 (0.11, 0.5)
North West	0.05 (0.02, 0.11)	0.21 (0.1, 0.4)	0.16 (0.08, 0.32)	0.05 (0.03, 0.12)
Free State	0.13 (0.08, 0.28)	0.42 (0.23, 0.75)	0.26 (0.13, 0.52)	0.09 (0.05, 0.2)
Northern Cape	0.06 (0.03, 0.13)	0.21 (0.12, 0.37)	0.17 (0.1, 0.32)	0.22 (0.12, 0.48)

Table S4. Prior ranges for the parameters used in the model-inference system. All initial values are drawn from uniform distributions using Latin Hypercube Sampling.

Parameter/ variable	Symbol	Prior range	Source/rationale
Initial exposed	$E(t=0)$	1 – 500 times of reported cases during the Week of March 15, 2020 for Western Cape and Eastern Cape; 1 – 10 times of reported cases during the Week of March 15, 2020, for other provinces	Low infection-detection rate in first weeks; earlier and higher case numbers reported in Western Cape and Eastern Cape than other provinces.
Initial infectious	$I(t=0)$	Same as for $E(t=0)$	
Initial susceptible	$S(t=0)$	99 – 100% of the population	Almost everyone is susceptible initially
Population size	N	N/A	Based on population data from COVID19ZA (main text ref 24)
Variant-specific transmission rate	β	For all provinces, starting from U[0.4, 0.7] at time 0 and allowed to increase over time using space re-probing(1) with values drawn from U[0.5, 0.9] during the Beta wave, U[0.7, 1.25] during the Delta wave, and U[0.7, 1.3] during the Omicron wave.	For the initial range at model initialization, based on R_0 estimates of around 1.5-4 for SARS-CoV-2.(2-4) For the Beta, Delta and Omicron variants, we use large bounds for space re-probing (SR)(1) to explore the parameter state space and enable estimation of changes in transmissibility due to the new variants. Note that SR is only applied to 3-10% of the ensemble members and β can migrate outside either the initial range or the SR ranges during EAKF update.
Scaling of effectiveness of NPI	e	[0.5, 1.5], for all provinces	Around 1, with a large bound to be flexible.

Latency period	Z	[2, 5] days, for all provinces	Incubation period: 5.2 days (95% CI: 4.1, 7)(2); latency period is likely shorter than the incubation period
Infectious period	D	[2, 5] days, for all provinces	Time from symptom onset to hospitalization: 3.8 days (95% CI: 0, 12.0) in China,(5) plus 1-2 days viral shedding before symptom onset. We did not distinguish symptomatic/asymptomatic infections.
Immunity period	L	[730, 1095] days, for all provinces	Assuming immunity lasts for 2-3 years
Mean of time from viral shedding to diagnosis	T_m	[5, 8] days, for all provinces	From a few days to a week from symptom onset to diagnosis/reporting,(5) plus 1-2 days of viral shedding (being infectious) before symptom onset.
Standard deviation (SD) of time from viral shedding to diagnosis	T_{sd}	[1, 3] days, for all provinces	To allow variation in time to diagnosis/reporting
Infection-detection rate	r	Starting from U[0.001, 0.01] at time 0 for Western Cape and Eastern Cape as these two provinces had earlier and higher case numbers during March – April 2020 than other provinces, suggesting lower detection rate at the time; for the rest starting from U[0.01, 0.06]. For all provinces,	Large uncertainties; therefore, in general we use large prior bounds and large bounds for space re-probing (SR). Note that SR is only applied to 3-10% of the ensemble members and r can migrate

	allowed *r* to increase over time using space re-probing (1) with values drawn from uniform distributions with ranges between roughly 0.01 to 0.12.	outside either the initial range or the SR ranges during EAKF update.
Infection fatality risk (IFR)	For Gauteng: starting from [0.0001, 0.002] at time 0 and allowed to change over time using space re-probing(1) with values drawn from U[0.0001, 0.005] during 12/13/2020 – 5/15/21 (due to Beta), U[0.0001, 0.002] during the Delta wave, and U[0.00001, 0.00075] starting 9/1/21 (Omicron wave). For KwaZulu-Natal: starting from U[0.0001, 0.003] at time 0 and allowed to change over time using space re-probing (1) with values drawn from U[0.0001, 0.005] during 4/19/20 –10/31/20 (ancestral wave), U[0.0001, 0.01] during 11/1/20 – 5/15/21 (Beta wave), U[0.0001, 0.002] during the Delta wave, and U[0.00001, 0.00075] starting 10/1/21 (Omicron wave). For Western Cape: starting from U[0.00001, 0.003] at time 0 and allowed to change over time using space re-probing (1) with values drawn from U[0.00001, 0.0004] during 4/19/20 – 10/31/20 (ancestral wave), U[0.00001, 0.01] during 11/1/20 – 5/15/21 (Beta wave), U[0.00001, 0.005] during 5/16/21 – 9/30/21 (Delta wave) and U[0.00001, 0.002] starting 10/1/21 (Omicron wave). For Eastern Cape: starting from U[0.0001, 0.003] at time 0 and allowed to change over time using space re-probing(1) with values drawn from U[0.0001, 0.004] during 4/19/20 – 9/30/20 (Ancestral wave), U[0.0001, 0.01] during 10/1/20 – 40/30/21 (Beta wave), [0.0001, 0.005] during the Delta wave, and U[0.00001, 0.002] or starting 10/16/21 (Omicron wave).	Based on previous estimates(6) but extend to have wider ranges. Note that SR is only applied to 3-10% of the ensemble members and IFR can migrate outside either the initial range or the SR ranges during EAKF update. Western Cape had earlier and higher case numbers during March – April 2020 than other provinces, suggesting lower detection rate at the time. Initial mortality rate in Gauteng was relatively low because initial infections occurred mainly among middle-aged, returning holiday makers.(7) Earlier spread of Beta in Eastern Cape, KwaZulu-Natal, and Northern Cape, higher numbers of deaths per capita reported. Free State reported higher number of deaths per capita.

	For Limpopo and Mpumalanga: starting from U[0.0001, 0.003] at time 0 and allowed to change over time using space re-probing (1) with values drawn from U[0.0001, 0.01] during the Beta wave, U[0.0001, 0.005] during the Delta wave, U[0.00001, .002] for the Omicron wave. For Free State: starting from U[0.0001, 0.003] at time 0 and allowed to change over time using space re-probing (1) with values drawn from U[0.0001, 0.006] during 3/16/20 – 10/31/20, U[0.0001, 0.01] during the Beta wave, U[0.0001, 0.008] during the Delta wave, and U[0.00001, 0.002] starting 10/1/21 (Omicron wave). For North West and Northern Cape: starting from U[0.0001, 0.003] at time 0 and allowed to change over time using space re-probing (1) with values drawn from U[0.0001, 0.005] during the Beta wave, U[0.0001, 0.003] during the Delta wave, and U[0.00001, 0.0015] starting 10/1/21 (Omicron wave).

References including in Table S4:

1. Yang W & Shaman J (2014) A simple modification for improving inference of non-linear dynamical systems. *arXiv*:1403.6804.
2. Li Q*, et al.* (2020) Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. *New Engl J Med*.
3. Wu JT, Leung K, & Leung GM (2020) Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet*.
4. Li R*, et al.* (2020) Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science* 368(6490):489-493.
5. Zhang J*, et al.* (2020) Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside Hubei province, China: a descriptive and modelling study. *The Lancet. Infectious diseases*.
6. Verity R*, et al.* (2020) Estimates of the severity of coronavirus disease 2019: a model-based analysis. *The Lancet. Infectious diseases*.
7. Giandhari J*, et al.* (2021) Early transmission of SARS-CoV-2 in South Africa: An epidemiological and phylogenetic report. *Int J Infect Dis* 103:234-241.

Table S5. Approximate epidemic timing (mm/dd/yy) for each wave in each province, used in the study. Note 3/5/22 is the last date of the study period.

Province	Variant	Start date	End date
Gauteng	Ancestral	3/15/20	10/31/20
Gauteng	Beta	11/1/20	5/15/21
Gauteng	Delta	5/16/21	8/31/21
Gauteng	Omicron	9/1/21	3/5/22
KwaZulu-Natal	Ancestral	3/15/20	9/15/20
KwaZulu-Natal	Beta	9/16/20	5/15/21
KwaZulu-Natal	Delta	5/16/21	9/30/21
KwaZulu-Natal	Omicron	10/1/21	3/5/22
Western Cape	Ancestral	3/15/20	9/15/20
Western Cape	Beta	9/16/20	5/15/21
Western Cape	Delta	5/16/21	9/30/21
Western Cape	Omicron	10/1/21	3/5/22
Eastern Cape	Ancestral	3/15/20	8/15/20
Eastern Cape	Beta	8/16/20	4/30/21
Eastern Cape	Delta	5/1/21	10/15/21
Eastern Cape	Omicron	10/16/21	3/5/22
Limpopo	Ancestral	3/15/20	10/31/20
Limpopo	Beta	11/1/20	5/15/21
Limpopo	Delta	5/16/21	9/30/21
Limpopo	Omicron	10/1/21	3/5/22
Mpumalanga	Ancestral	3/15/20	10/31/20
Mpumalanga	Beta	11/1/20	5/15/21
Mpumalanga	Delta	5/16/21	9/30/21
Mpumalanga	Omicron	10/1/21	3/5/22
North West	Ancestral	3/15/20	10/31/20
North West	Beta	11/1/20	5/15/21
North West	Delta	5/16/21	9/30/21
North West	Omicron	10/1/21	3/5/22
Free State	Ancestral	3/15/20	10/31/20
Free State	Beta	11/1/20	5/31/21
Free State	Delta	6/1/21	9/30/21
Free State	Omicron	10/1/21	3/5/22
Northern Cape	Ancestral	3/15/20	10/31/20
Northern Cape	Beta	11/1/20	5/15/21
Northern Cape	Delta	5/16/21	9/30/21
Northern Cape	Omicron	10/1/21	3/5/22