# THE DAILY COVID-19 EFFECTIVE REPRODUCTIVE NUMBER (R) IN SOUTH AFRICA

NATIONAL INSTITUTE FOR COMMUNICABLE DISEASES
Division of the National Health Laboratory Service



# SUMMARY

## Overview of report

**Disease surveillance is a core function of the National Institute for Communicable Diseases (NICD), a Division of the National Health Laboratory Service (NHLS). This report includes an analysis of data on laboratory-confirmed COVID-19 cases, hospital admissions, and deaths to estimate the effective reproductive number (R) of SARS-CoV-2 over time in South Africa nationally and in provinces where sufficient data are available. The basic reproductive number ($R_0$) is the average number of secondary infections produced by a typical case of an infection in a population where everyone is susceptible, which occurs typically in the first few weeks after introduction of a novel infectious agent into the population. The effective reproduction number (R) is the average number of secondary cases per infectious case in a population composed of both susceptible and non-susceptible hosts (once the infectious agent is circulating). If R>1, the number of new cases per time unit will increase, such as at the start of an epidemic. Where R=1, the number of new cases is stable over time, and where R<1, there will be a decline in the number of new cases per time unit.**

**This report is based on data collected up to 19 July 2021 (week 29 of 2021). The data were adjusted for the delays from illness onset to case report, hospital admission, and death and right censored for 2, 7, and 7 days respectively to account for the time lag between each outcome (test result, admission, or death) and the time of reporting (R estimated up to 15 July). This analysis updates the report released on 9 July 2021. In this report, R is estimated from the data on laboratory-confirmed COVID-19 cases, hospital admissions, and, on a national level, hospital-based deaths. There may be non-overlapping sources of bias for the three data sources, which motivates a comparison of R estimates. R estimates are described for each of the lockdown levels implemented by the South African government – for more information regarding the timing and nature of lockdowns see the South African government website [vi]. Note: COVID-19 is the name of the disease and SARS-CoV-2 is the name of the virus.**

## Highlights

- Civil unrest and disruptions to public transport services in parts of the country during July reduced access to laboratory testing and healthcare facilities. Data series for Gauteng and KwaZulu-Natal provinces have been truncated to include only data recorded prior to 10 July 2021, as estimates beyond this period might provide biased estimates of R. In addition, we have elected not to present national estimates of R (which are strongly driven by trends in Gauteng) in the current report.
- During level 4 lockdown in late June and July 2021, R decreased in six out of seven provinces for which R estimates are available over the full time period.
- R estimates were close to or above 1 in all nine provinces at the end of their respective estimation periods indicating stable or increasing transmission. It is essential that recommended measures to control the spread of COVID-19, including physical distancing, hand hygiene, good ventilation, adherence to venue capacity limits, and wearing of masks, are consistently implemented.

# THE DAILY COVID-19 EFFECTIVE REPRODUCTIVE NUMBER (R) IN SOUTH AFRICA

WEEK **28** 2021

## Methods

*Daily R estimation*

We used data from the first confirmed case in March 2020 until 19 July 2021, based on the national DATCOV dataset on hospitalized cases and in-hospital deaths, and the laboratory-confirmed case line list maintained by the National Institute for Communicable Diseases (NICD). Data from Gauteng and KwaZulu-Natal provinces were truncated on 10 July 2021. The laboratory-confirmed cases data was linked with the national DATCOV dataset to obtain dates of symptom onset. Following data linkage, symptom onset data were available for 6% of laboratory-confirmed cases, while dates of onset were available for 53% of hospitalized cases, and 56% of fatal cases in the DATCOV dataset. 63 cases (0.02%) in the DATCOV database were missing both admission date and date of symptom onset and were excluded from the analyses based on hospital admissions and deaths. The data were adjusted for the delay from symptom onset to reporting of test result / hospital admission and right censored for 2, 7, and 7 days (for cases, hospital admissions, and deaths respectively) to account for reporting delays (last date of estimation for national-level analysis based on lab-confirmed cases: 15 July 2021). The provincial level DATCOV data have different end-dates, so the provincial time series were individually adjusted for right-censoring; national-level analyses were based on the pooled provincial-level time series. Missing dates of symptom onset were imputed using chained equations multiple imputations (50) [i,ii]. A negative-binomial model was fitted to confirmed COVID-19 cases for which the date of symptom onset was available and used to impute the dates of symptom onset for cases with missing information. Separate imputations were done for the case and admissions datasets. The hospital-based deaths data set is a subset of admissions, so the same set of imputations were used. The model predictors for the two imputation procedures were: health sector where sample collection/hospital admission occurred (private or public), age group, month of case report/hospital admission, outcome (for admissions), day of hospital admission (for admissions), and province. The daily R was estimated using the method of Thompson *et al.* (EpiEstim v. 2.2-3) [iii,iv] for each imputed dataset. For the serial interval we used a gamma distribution with mean of 6.6 (s.d. 3.3) and standard deviation of 0.5 (s.d. 0.27) to account for the variability (and uncertainty) of the selected serial interval values. Parameters were estimated by fitting a gamma distribution to data from the PHIRST-C, a community cohort study of COVID-19 transmission (Cheryl Cohen, NICD, unpublished data). We report the medians of the central values and the 2.5$^{th}$-97.5$^{th}$ percentiles of the estimated daily R values obtained from the imputed datasets [i,ii].

Previous versions of this report have included descriptions of trends in daily R values during lockdown levels 5 through 1, and the adjusted level 3 lockdown, as well as the 2021 lockdown levels 1 through 3. The current report focuses on more recent trends in daily R values, starting on 1 May 2021 (for more details regarding the COVID-19 lockdowns in South Africa, please refer to the South African government website [vi]).

## Results

Case counts steadily increased during May and June of 2021. During July, case counts fluctuated, with an overall downward trend following a peak on 29 June.

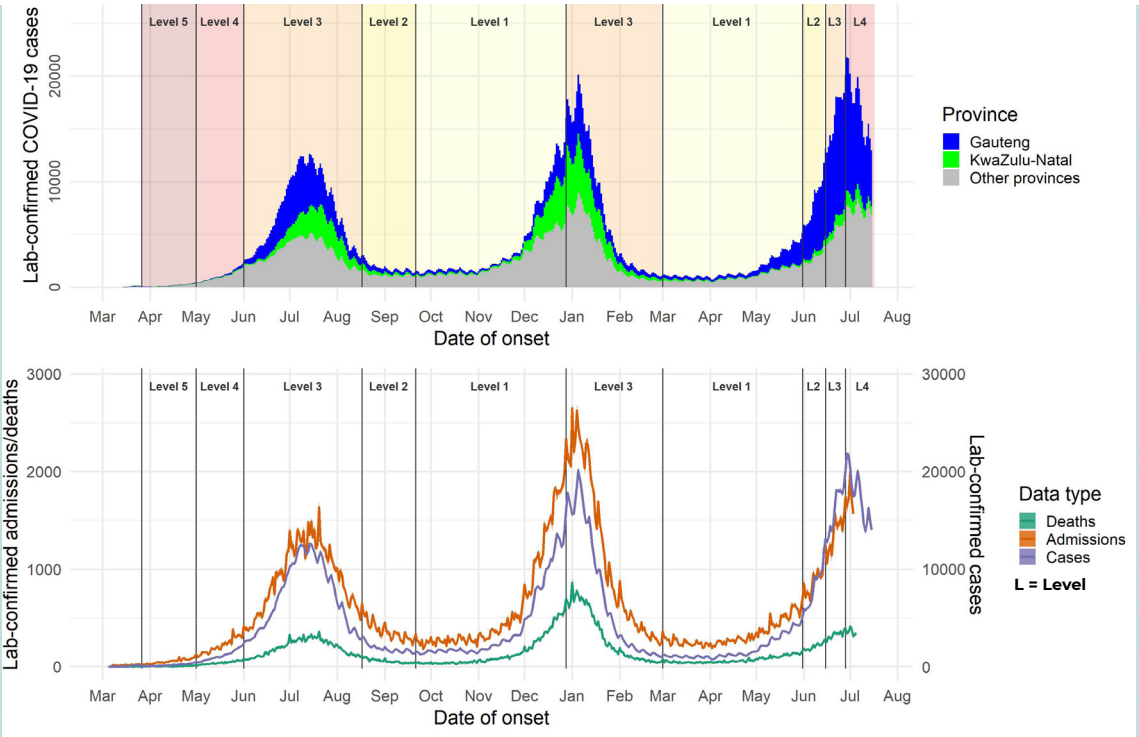


**Figure 1.** Daily number of laboratory-confirmed COVID-19 cases (above) and daily numbers of laboratory-confirmed COVID-19 cases, hospital admissions, and deaths (below), by date of symptom onset and province (missing data imputed; medians and 95% quantiles of imputed time series are shown), South Africa (last date included: 15 July 2021). Daily numbers of laboratory-confirmed COVID-19 cases in the bottom panel are shown on the right-hand y axis. Case counts for Gauteng and KwaZulu-Natal are highlighted in the top panel.

Despite substantial variation in transmission patterns between provinces, trends at the province level were in most cases similar, with R decreasing or remaining steady during level 3 lockdown. R was close to or above 1 in all provinces at the end of their respective estimation periods (Figures 2-8 and Tables 1-2).

**Table 1.** Daily R estimates by province (excluding Gauteng and KwaZulu-Natal) for R based on lab-confirmed cases and hospital admissions. Each cell contains median values with 95% confidence intervals.

| | Cases | Cases | Admissions |
|---|---|---|---|
| | 15 July 2021 | 11 July 2021 | 11 July 2021 |
| Western Cape | 1.06 (1.04,1.09) | 1.14 (1.10,1.17) | 1.19 (1.13,1.26) |
| Eastern Cape | 1.10 (1.07,1.13) | 1.11 (1.07,1.15) | 1.12 (1.03,1.23) |
| Free State | 0.97 (0.94,1.01) | 1.03 (0.99,1.06) | 1.12 (1.02,1.22) |
| Northern Cape | 1.05 (1.00,1.10) | 1.02 (0.98,1.07) | 1.24 (1.08,1.43) |
| North West | 0.95 (0.93,0.97) | 1.05 (1.03,1.08) | 1.13 (1.06,1.22) |
| Mpumalanga | 1.11 (1.08,1.15) | 1.24 (1.18,1.29) | 1.25 (1.14,1.35) |
| Limpopo | 1.02 (1.00,1.05) | 1.23 (1.18,1.29) | 1.15 (1.06,1.23) |

**Table 1.** Daily R estimates in Gauteng and KwaZulu-Natal for R based on lab-confirmed cases and hospital admissions. Data were truncated on 10 July 2021 because of concerns about data validity after this date following widespread unrest. Each cell contains median values with 95% confidence intervals.

| | Cases | Cases | Admissions |
|---|---|---|---|
| | 8 July 2021 | 3 July 2021 | 3 July 2021 |
| Gauteng | 1.02 (1.01,1.04) | 1.20 (1.17,1.23) | 1.17 (1.13,1.20) |
| KwaZulu-Natal | 1.35 (1.29,1.42) | 1.48 (1.41,1.57) | 1.41 (1.30,1.53) |

# THE DAILY COVID-19 EFFECTIVE REPRODUCTIVE NUMBER (R) IN SOUTH AFRICA



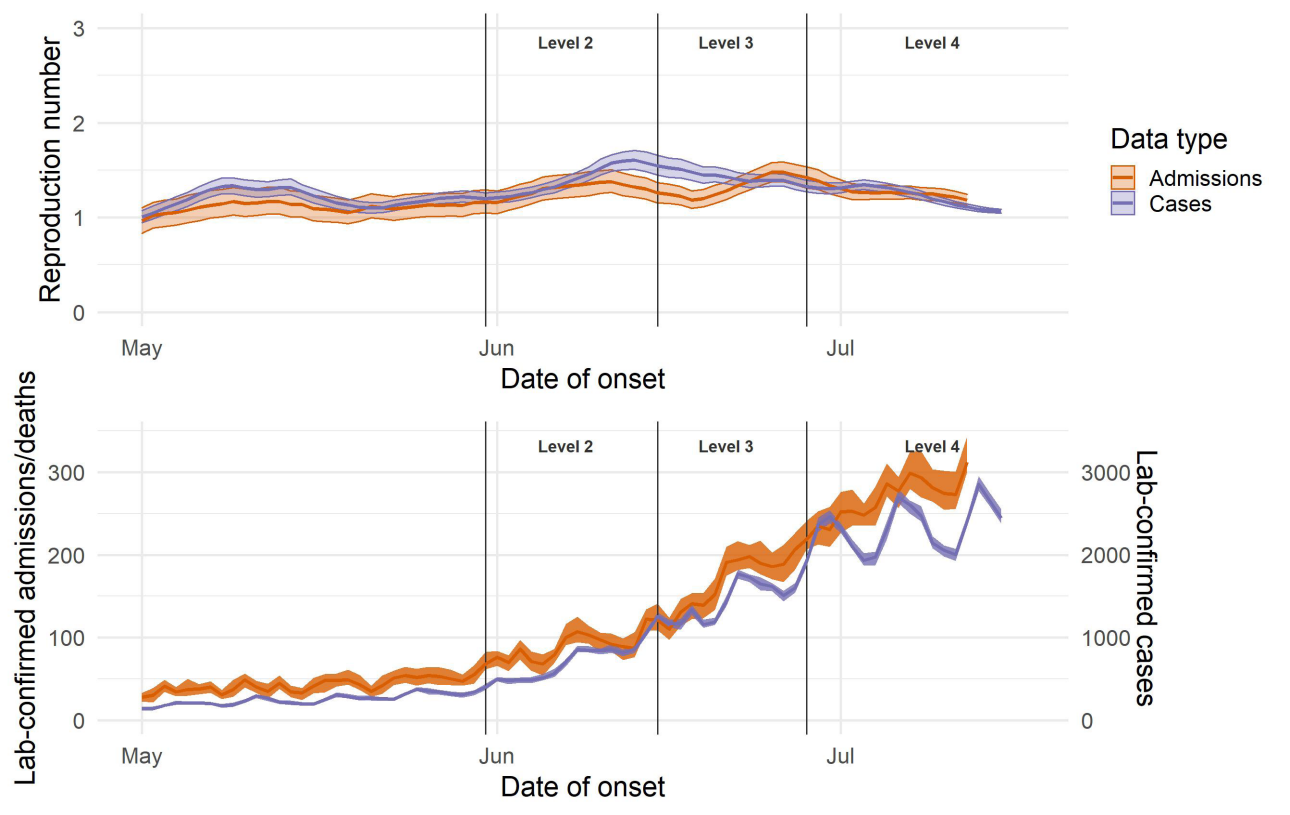


**Figure 2.** Upper panel: Estimated daily reproduction number (R), with 95% confidence intervals, Western Cape (last date included in the estimation: 15 July 2021). Lower panel: estimated number of laboratory-confirmed COVID-19 cases and hospital admissions by onset date with missing data imputed. The medians and 95% ranges for the imputed datasets are shown. Daily numbers of laboratory-confirmed COVID-19 cases in the bottom panel are shown on the right-hand y axis.

# THE DAILY COVID-19 EFFECTIVE REPRODUCTIVE NUMBER (R) IN SOUTH AFRICA



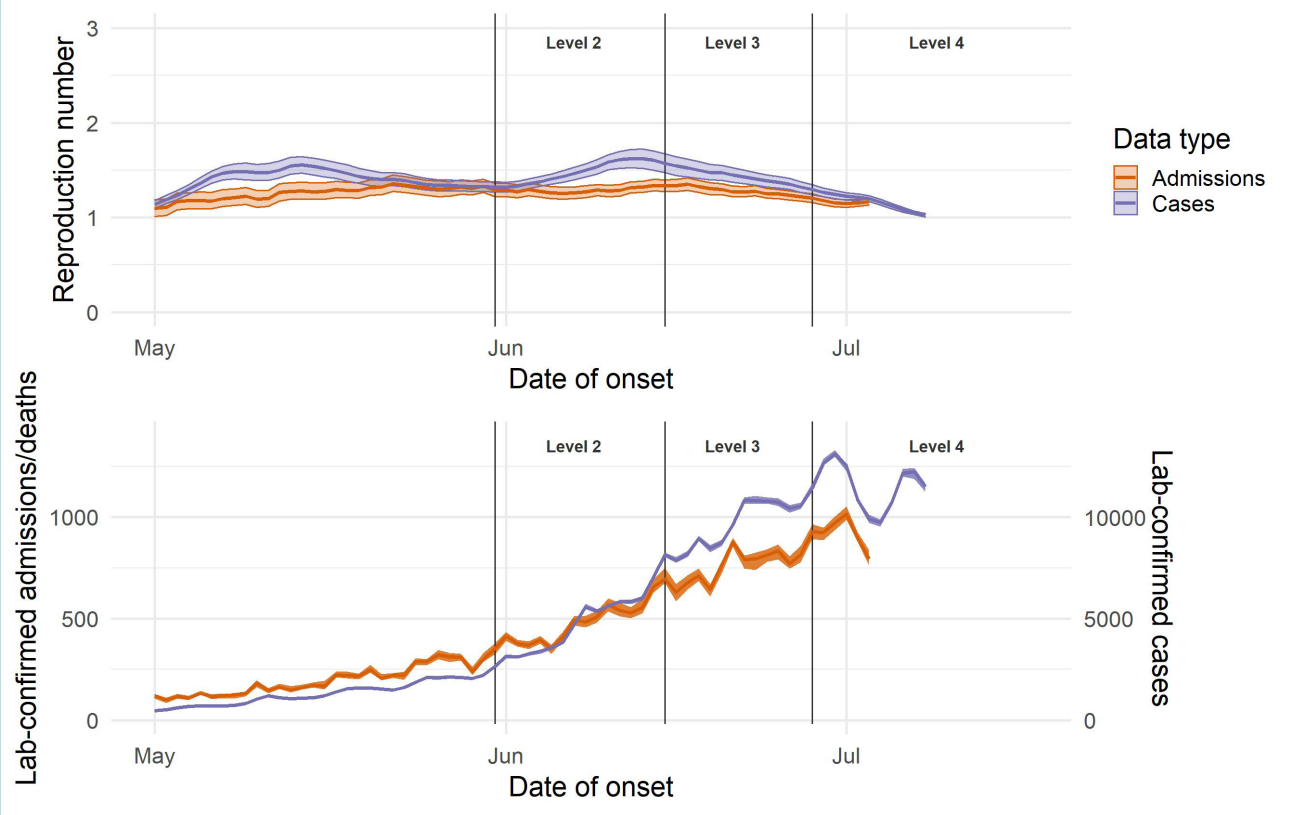


**Figure 3.** Upper panel: Estimated daily reproduction number (R), with 95% confidence intervals Gauteng (last date included in the estimation: 10 July 2021). Lower panel: estimated number of laboratory-confirmed COVID-19 cases and hospital admissions by onset date with missing data imputed. The medians and 95% ranges for the imputed datasets are shown. Daily numbers of laboratory-confirmed COVID-19 cases in the bottom panel are shown on the right-hand y axis.

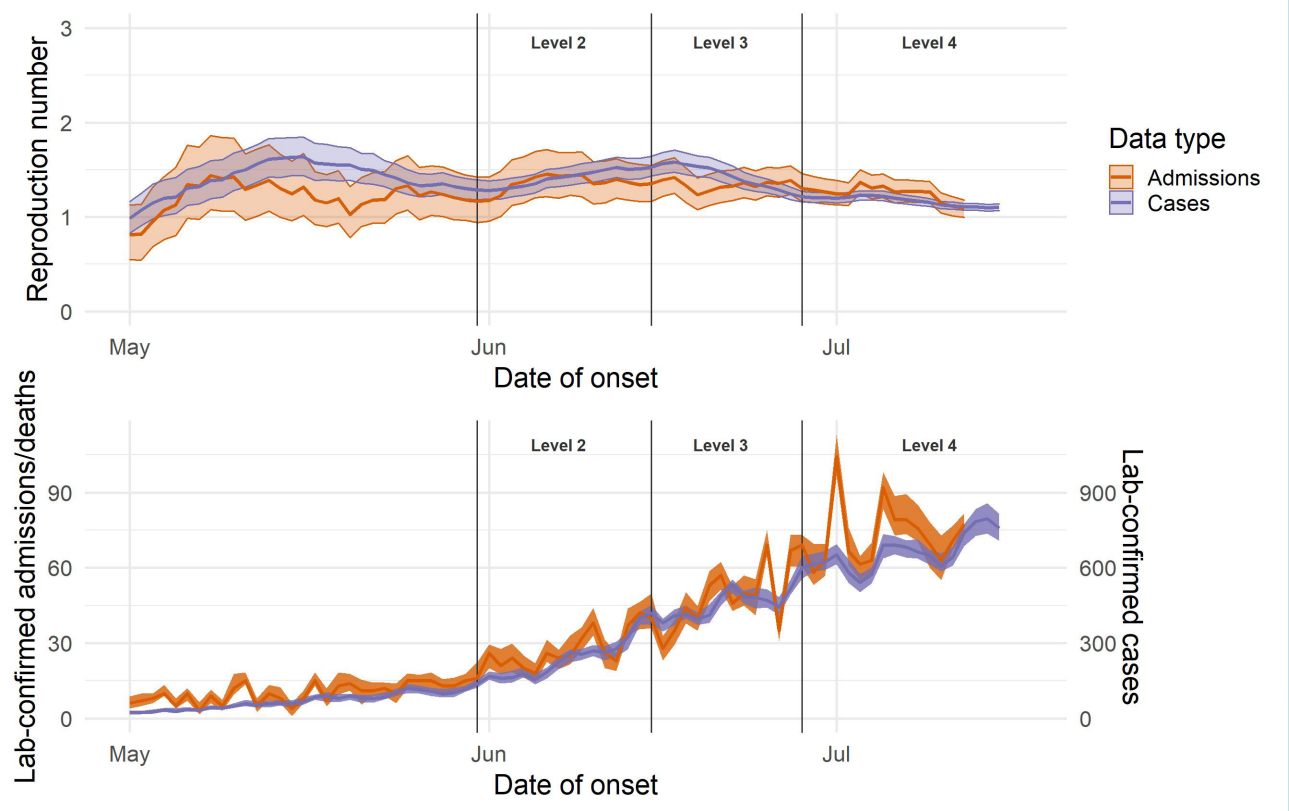


**Figure 4.** Upper panel: Estimated daily reproduction number (R), with 95% confidence intervals, Eastern Cape (last date included in the estimation: 15 July 2021). Lower panel: estimated number of laboratory-confirmed COVID-19 cases and hospital admissions by onset date with missing data imputed. The medians and 95% ranges for the imputed datasets are shown. Daily numbers of laboratory-confirmed COVID-19 cases in the bottom panel are shown on the right-hand y axis.

# THE DAILY COVID-19 EFFECTIVE REPRODUCTIVE NUMBER (R) IN SOUTH AFRICA



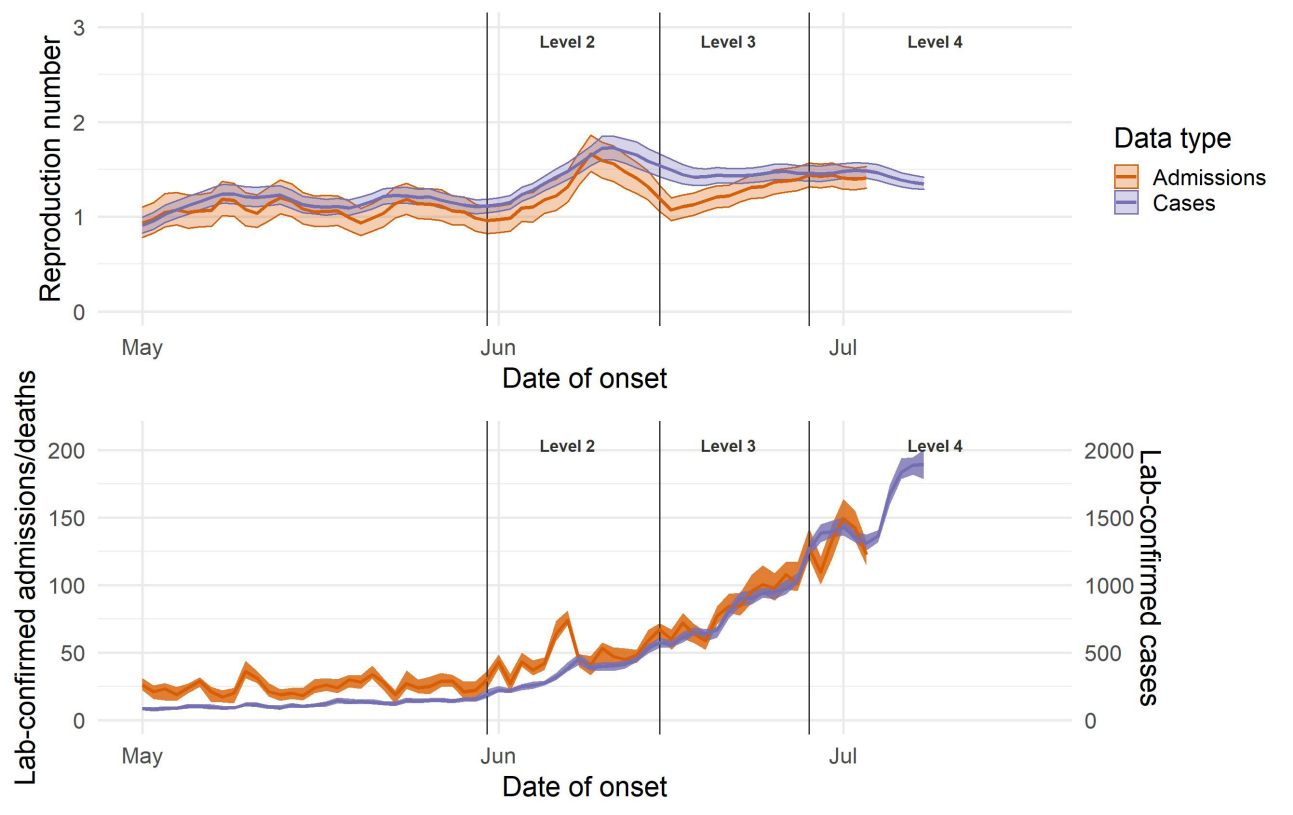


**Figure 5.** Upper panel: Estimated daily reproduction number (R), with 95% confidence intervals, KwaZulu-Natal (last date included in the estimation: 10 July 2021). Lower panel: estimated number of laboratory-confirmed COVID-19 cases and hospital admissions by onset date with missing data imputed. The medians and 95% ranges for the imputed datasets are shown. Daily numbers of laboratory-confirmed COVID-19 cases in the bottom panel are shown on the right-hand y axis.

# THE DAILY COVID-19 EFFECTIVE REPRODUCTIVE NUMBER (R) IN SOUTH AFRICA



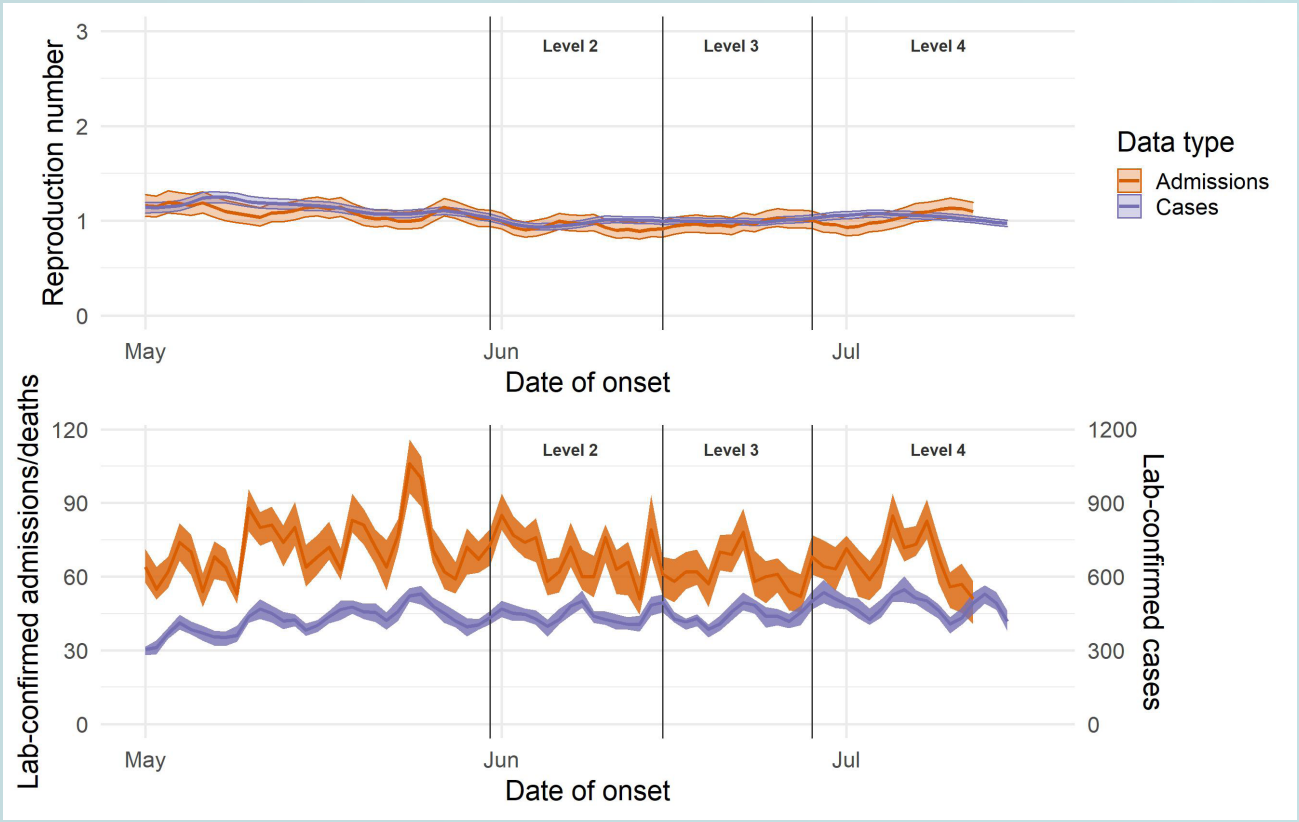


**Figure 6.** Upper panel: Estimated daily reproduction number (R), with 95% confidence intervals, Free State (last date included in the estimation: 15 July 2021). Lower panel: estimated number of laboratory-confirmed COVID-19 cases and hospital admissions by onset date with missing data imputed. The medians and 95% ranges for the imputed datasets are shown. Daily numbers of laboratory-confirmed COVID-19 cases in the bottom panel are shown on the right-hand y axis.


**Figure 7.** Upper panel: Estimated daily reproduction number (R), with 95% confidence intervals, Northern Cape (last date included in the estimation: 15 July 2021). Lower panel: estimated number of laboratory-confirmed COVID-19 cases and hospital admissions by onset date with missing data imputed. The medians and 95% ranges for the imputed datasets are shown. Daily numbers of laboratory-confirmed COVID-19 cases in the bottom panel are shown on the right-hand y axis.

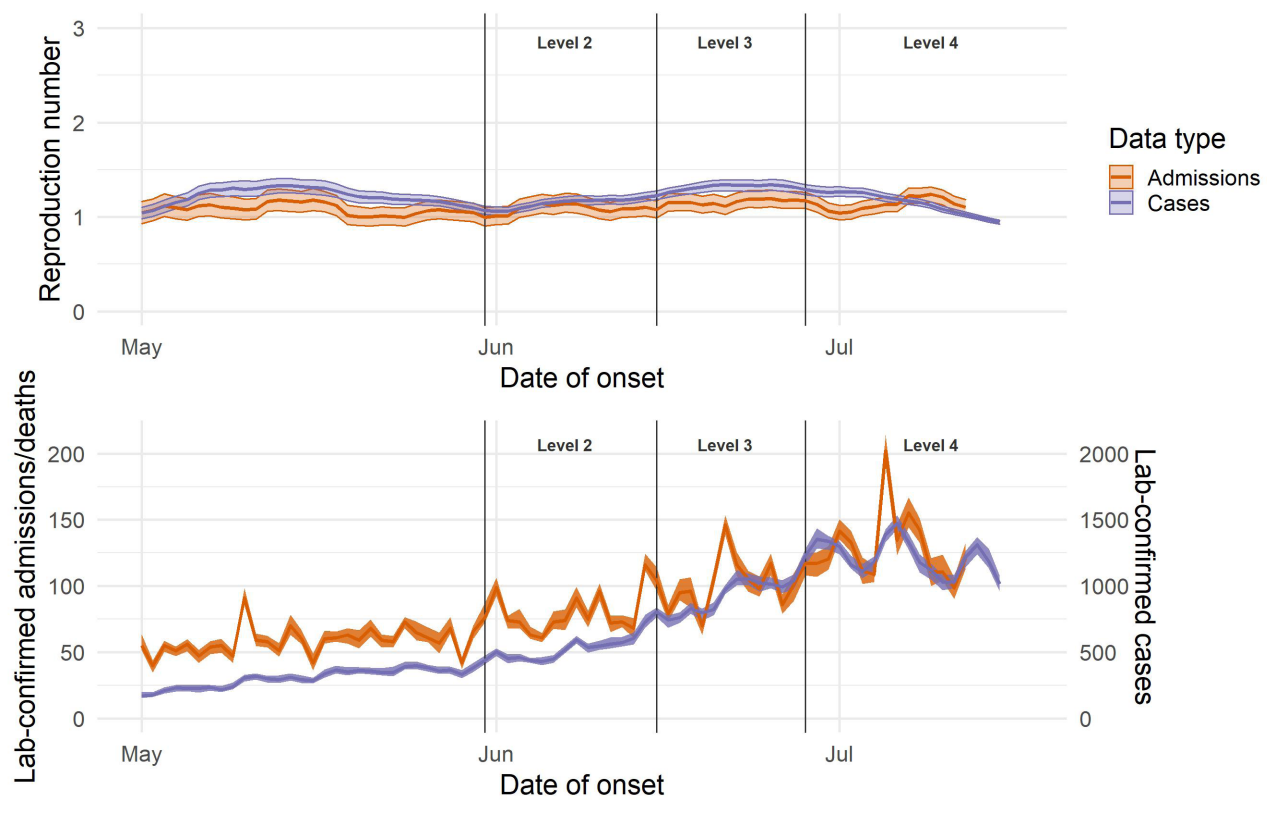


**Figure 8.** Upper panel: Estimated daily reproduction number (R), with 95% confidence intervals, North West (last date included in the estimation: 15 July 2021). Lower panel: estimated number of laboratory-confirmed COVID-19 cases and hospital admissions by onset date with missing data imputed. The medians and 95% ranges for the imputed datasets are shown. Daily numbers of laboratory-confirmed COVID-19 cases in the bottom panel are shown on the right-hand y axis.

## Limitations

The main limitation of this analysis is that the ascertainment rate of COVID-19 cases and deaths, along with the proportion of cases which are admitted to hospital, may change over time, potentially affecting R estimation. These effects are likely driven in part by changes in the criteria for testing and hospital admission, as well as by shifting care seeking behavior during the epidemic. The increased use of antigen tests through time, which have lower sensitivity than the more-commonly used PCR tests, may result in lower ascertainment rates, particularly during the second wave. In addition, capturing of cases confirmed on antigen detection tests may not be complete in the NMC line list data and may vary by province and over time potentially leading to biased estimates. Because use of antigen detection tests has been increasing through time, this could result in underestimation of current R values. Due to concerns about the impact of civil unrest leading to reductions in testing rates in KwaZulu-Natal and Gauteng provinces, we have elected to censor data included for R estimates from these provinces on 10 July prior to these disruptions.

Along with the ascertainment rate, the delay between symptom onset and reporting of case/admission/death may change over time, which would affect the accuracy of the adjustment for right-censoring the end of the time series. In addition, the relatively low numbers of deaths recorded between waves results in high levels of uncertainty and large fluctuation in R estimates based on daily deaths. We do not present death-based R estimates in province-level analyses due to instability of the estimates, which results from small numbers. Furthermore, a number of factors may have altered mortality outcomes over time, including the introduction of dexamethasone treatment in mid-June, the use of oxygen administration via high flow nasal cannula, changes in quality of healthcare provided if health systems are overwhelmed, and potential differences in severity between earlier circulating viruses, the 501Y.V2 / B.1.351 (“beta”) variant that dominated the second wave, and the B.1.617.2 (“delta”) variant which has been prominent during the third wave. Combined, these factors may lead to perturbations in the time series data that are unrelated to transmission. Comparing R estimates from the 3 data sources may help in assessing the severity of some of these biases, as indicated by inconsistent results across analyses of the three data sources. In addition to limitations in the ability of the available time series data to reflect underlying transmission, the serial interval distribution is estimated based on a relatively small dataset from an on-going study.

This report was jointly prepared by the National Institute for Communicable Disease (NICD) and the DSI-NRF Centre of Excellence in Epidemiological Modelling and Analysis (SACEMA). Inquiries should be referred to Prof Cheryl Cohen (cherylc@nicd.ac.za).

## References

i van Buuren, S. (2018) Flexible Imputation of Missing Data, Second Edition. Chapman and Hall/CRC: New York. DOI: 10.1201/9780429492259. https://stefvanbuuren.name/fimd

ii Kleinke, K, & Reinecke, J. (2015) Multiple imputation of overdispersed multilevel count data. In: Uwe Engel (Ed.), Survey Measurements. Techniques, Data Quality and Sources of Error (pp. 209–226). Frankfurt A. M.: Campus/The University of Chicago Press. http://press.uchicago.edu/ucp/books/book/distributed/S/bo22196267.html

iii Thompson, RN, JE Stockwin, RD van Gaalen, JA Polonsky, ZN Kamvar, PA Demarsh, E Dahlqwist, S Li, E Miguel, T Jombart, J Lessler, S Cauchemez, and A Cori. (2019) Improved inference of time-varying reproduction numbers during infectious disease outbreaks. Epidemics 29: 100356. DOI: 10.1016/j.epidem.2019.100356

iv Cori, A. (2020) EpiEstim: Estimate time varying reproduction numbers from epidemic curves. R package version 2.2-3. https://github.com/mrc-ide/EpiEstim

v Pitzer, VE, Chitwood, M, Havumaki, J, Menzies, NA, Perniciaro, M, Warren, JL, Weinberger, DM, and T Cohen. (2020) The impact of changes in diagnostic testing practices on estimates of COVID-19 transmission in the United States. medRxiv https://doi.org/10.1101/2020.04.20.20073338

vi Ali, ST, L Wang, EHY Lau, XK Xu, Z Du, Y Wu, GM Leung, and BJ Cowling. (2020) Serial interval of SARS-CoV-2 was shortened over time by nonpharmaceutical interventions. Science eabc9004. DOI: 10.1126/science.abc9004

vii Government of South Africa. (2021). Regulations and guidelines – coronavirus COVID-19. Retrieved from https://www.gov.za/covid-19/resources/regulations-and-guidelines-coronavirus-covid-19#